Exploring new concepts in biomolecular recognition and ligand discovery

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

Doctor of Philosophy

Na Youn Cho

Keble College, University of Oxford, Hilary 2019



Abstract

Aptamers are short nucleic acids that can bind with high affinity to a variety of biological targets through shape complementary. However, aptamers composed of natural bases are susceptible to nuclease degradation and have a limited range of intermolecular interactions with their targets. Here we introduce novel modifications at one of the four bases in SELEX libraries to enhance the interaction capabilities of the DNA and thereby broaden the target spectrum. Initially, azidoethoxy and propargylamino groups are introduced to C2′ position of thymidine or C5 position of deoxyuridine with benefits of further conjugation to novel functional groups – such as hydrophobic amino acid side chains to supplement protein-like interaction or halogen substituted modifications to add halogen bonding interaction. By providing the additional interactions that are not available in a canonical nucleic acid aptamer, next-generation aptamers with high target binding affinity and kinetic are anticipated. For their uses in SELEX to discover next-generation aptamer, the acceptance of chemically functionalised triphosphates by DNA polymerase is firstly examined. Base-modified nucleotides are better substrates for DNA polymerases than sugar-modified nucleotides due to the steric gate of the enzyme. Especially, 5-propargylamino-iodotriazole-dUTP (5-ITZ-dUTP), a monomer which can potentially allow aptamers to form halogen bonding interaction is used in SELEX. Promisingly, a G-quadruplex streptavidin aptamer containing nine 5-ITZ-dUTPs which has a high affinity to the target is discovered. This aptamer shows 6.5-fold and 9-fold higher binding affinities than the appropriate control sequences which contain non-halogen bearing nucleotides (5-propargylamino-prototriazole-dU or canonical dT). Further studies to demonstrate halogen bonding in aptamer-target interaction are required to confirm this as the first example of halogen bonding in aptamer-protein interactions.

Declaration of authorship

I, Na Youn Cho declare that this thesis and all 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

Foremost, I would like to express my sincere gratitude to my supervisor Professor Tom Brown for offering me a DPhil position and the continuous support of my DPhil study and research. I am also thankful for his immense knowledge, patience, endless ideas and encouragement which have made the research possible. Many thanks to Dr. Afaf El-Sagheer for all the guidance, encouragement and supervision during my DPhil course and also thanks for synthesising many oligonucleotides for my project. I am also grateful to ATDBio Ltd. for funding my study.

I would like express my appreciation to Professor Paul Beer for his advice and expertise during collaboration work for halogen bonding project. I also would like to thank Dr Asha Brown for helping me with chemical synthesis during halogen bonding project and her time to proofread this thesis.

Particularly special thanks are due to Dr Arun Shivalingam for always being so helpful, answering all the questions and for his contribution in the aptamer project and all the effort he put in correcting my thesis. I want to also thank Pui and Pheobe. We had a wonderful time together and always stayed side by side in tears and laughter. Thank you to all my friends for supporting me throughout my DPhil. In particular, Mei, Anna, Carolyn, Pietro, Jack C, Sara, Yssy, Ashley, Sarah, Jack H, Piotr, Cam, Tanishta, Sven and Yutong for providing such an enjoyable research environment. I could not have gone through it without all of you.

Last, but not least, I would like to thank my mum, dad and sister, Sally and my boyfriend, Jinsung for their endless love, support, great patience and for always believing me.

Abbreviations

A/dA/dATP	Adenine/deoxyadenosine/deoxyadenosine triphosphate
Å	Å ngström (10^{-10} m)
aq	Aqueous
bp	Base pair
B/W	Binding and Wash
C/dC/dCTP	Cytosine/deoxycytidine/ deoxycytidine triphosphate
℃	Degree Celsius
Calc	Calculated
cm	Centimetre (s)
COSY	^{1}H-^{1}H correlation spectroscopy
Cq	Quantitation cycle
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DIEA	*N*,*N*-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	4,4'-Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
eq	Equivalent
ESI	Electrospray ionisation

Et_3N	Triethylamine
EtOAc	Ethyl acetate
EtOH	Ethanol
FAM	6′-Carboxyfluorescein
FC	Flow cell
Fmoc	Fluorenylmethyloxycarbonyl
g	Gram(s)
h	Hour(s)
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBTU	O-(Benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate
HEG	Hexaethylene glycol
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HMBC	Heteronuclear Multiple Bond correlation
HOBt	1-Hydroxybenzotriazole
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear single quantum correlation
Hz	Hertz
J	Coupling constant
k_a	Association rate constant
k_d	Dissociation rate constant
K_D	Equilibrium dissociation (binding) constant
L	Litre(s)
LRMS	Low Resolution Mass Spectrometry
M	Molar concentration

MeCN	Acetonitrile
MeOH	Methanol
mg	Milligram(s)
min	Minute(s)
mL	Millilitre(s)
mol	Mole(s)
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass/charge
NBu_3	Tributylamine
NGS	Next generation sequencing
NHS	*N*-Hydroxysuccinimide
NMM	*N*-Methylmorpholine
NMR	Nuclear magnetic resonance
nt	Nucleotide(s)
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
pH	Potential of hydrogen (negative logarithm of the positive hydrogen ion concentration)
ppm	parts per million (NMR)
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate

Q-PCR	quantitative polymerase chain reaction
R_f	Retention factor
RNA	Ribonucleic acid
RP-HPLC	Reversed-phase high performance liquid chromatography
RT	Room temperature
RU	Response unit
SAXS	Small-angle X-ray scattering
SELEX	Systematic evolution of ligands by exponential enrichment
SOMAmer	Solw off-rate modified aptamer
ssDNA	Single-stranded DNA
T/dT/dTTP/TTP	Thymine/deoxythymidine/deoxythymidine triphosphate
TBE	Tri-Borate-EDTA
TEA	Triethylamine
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid/trifluoroacetyl
TLC	Thin layer chromatography
T_m	Melting temperature
TOF	Time-of-flight
U	Unit(s)
U/dU/dUTP	Uracil/deoxyuridine/deoxyuridine triphosphate
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
V	Volt(s)
W	Watt(s)

XB	Halogen bonding
δ	Chemical shift in parts per million
ε	Extinction coefficient
Δ	Delta

Contents

Abstract ..... **I**

Declaration of authorship ..... **II**

Acknowledgements ..... **III**

Abbreviations ..... **IV**

Contents ..... **IX**

Chapter 1 ..... **1**

1 Introduction **2**

1.1 Nucleic acids 2

1.1.1 Primary structure of nucleic acids 2

1.1.2 Secondary structure of nucleic acids 5

1.1.3 Thermal stability of DNA duplexes 8

1.2 Aptamers 10

1.2.1 Comparison with antibodies 15

1.2.2 Aptamer selection (SELEX) 16

1.2.3 Improvement in SELEX techniques 18

1.3 Chemical modifications in aptamer generation 26

1.3.1 Sugar modifications 27

1.3.2 Nucleobase modifications 30

1.3.3 Backbone modifications 33

1.4 Aptamer sequencing 36

1.4.1 Sanger sequencing 36

1.4.2 Next-generation sequencing (NGS) 36

1.5 Synthesis of modified oligonucleotides 40

1.5.1 Solid phase synthesis of oligonucleotides 40

1.5.2 Post-synthesis chemical labelling reactions 42

1.5.3 Synthesis of nucleoside triphosphates 46

1.5.4 Enzymatic synthesis of oligonucleotides 49

1.6 Aptamer validation 59

1.6.1 Surface plasmon resonance (SPR) 60

1.6.2 Electrophoretic mobility shift assay (EMSA) 63

1.7 Objectives of the research 65

Chapter 2 ..... **66**

2 Synthesis of C2′ sugar modified thymidine triphosphates and their enzymatic reactions **67**

2.1 Background 67

2.2 Primer extension reaction using a 2′-AmE-T modified template 72

2.3 Preparation of 2′-AzE-TTP 85

2.4 Primer extension reactions using 2′-AzE-TTP 86

2.5 PCR using 2′-AzE-TTP 93

2.6 Primer extension reactions using 2′-ethynyl-UTP 96

2.7 Conclusions 99

Chapter 3 ..... **101**

3 Synthesis of C5 base modified dUTP and their enzymatic reactions . **102**

3.1 Background 102

3.2 Preparation of modified triphosphates derived from 5-PA-dUTP 107

3.2.1 Synthesis of modified triphosphates bearing hydrophobic substituents 111

3.2.2 Synthesis of modified triphosphates bearing halogen bonding substituents 114

3.3 Primer extension reactions using modified dUTP bearing hydrophobic modifications 116

3.4 PCR reactions using modified dUTP bearing hydrophobic modifications 126

3.5 Primer extension reactions using modified dUTP bearing halogen bonding substituents 134

3.6 PCR reactions using modified dUTP bearing halogen bonding substituents 141

3.7 Conclusions 155

Chapter 4 ..... **157**

4 DNA aptamer for HER2 extracellular epitopes **158**

4.1 Background 158

4.2 Target selection 161

4.3 ssDNA library design 162

4.3.1 ssDNA generation method 163

4.3.2 ssDNA library and primers 170

4.4 Amplification of DNA library and optimisation 172

4.4.1 PCR amplification with natural dNTPs 172

4.4.2 Size separation for ssDNA 174

4.4.3 Quantification of ssDNA by fluorescence measurement 175

4.5 Adaptation of the selection method 176

4.5.1 Peptide immobilisation and quantification ... 176
4.5.2 Selection ... 185
4.5.3 Melting temperatures from rounds 1–7 of SELEX ... 190
4.6 Binding of aptamer candidates ... 192
4.7 Conclusions ... 195
Chapter 5 ... 197
5 Selection of halogen bearing DNA aptamer against streptavidin ... 198
5.1 Background ... 198
5.2 Design of the 5-ITZ-dU library ... 203
5.3 PCR amplification of 5-ITZ-dU library and ssDNA generation ... 205
5.4 Selection of a modified DNA aptamer for streptavidin ... 207
5.4.1 Q-PCR quantification ... 210
5.5 Identification of aptamer candidates ... 213
5.5.1 Base composition ... 213
5.5.2 Unique sequence enrichment ... 215
5.5.3 G4 enrichment ... 217
5.5.4 NMer cluster ... 219
5.6 Validation of aptamer candidates ... 220
5.6.1 Synthesis of aptamer candidates ... 222
5.6.2 SPR binding analysis ... 238
5.6.3 EMSA analysis ... 246
5.7 Conclusions ... 250
Chapter 6 ... 252
6 Conclusion and future work ... 253
Chapter 7 ... 258
7 Experimental ... 259
7.1 General methods ... 259
7.2 A general method for oligonucleotide synthesis, purification and analysis ... 260
7.2.1 Oligonucleotides and triphosphates quantification ... 261
7.2.2 Denaturing polyacrylamide gel electrophoresis (PAGE) ... 262
7.3 General buffers ... 263
7.3.1 RP-HPLC buffer ... 263
7.3.2 Anion exchange buffer ... 263
7.3.3 Gel electrophoresis buffer ... 263

7.4 Chemical Synthesis 264
7.5 Primer extension experiment protocol 307
7.5.1 Primer extension results analysis by mass spectrometry 307
7.6 Polymerase chain reaction (PCR) protocol 308
7.6.1 Melting temperature (T_m) experiments protocol 309
7.7 Single-stranded DNA generation 310
7.7.1 Lambda(λ) exonuclease digestion protocol 310
7.7.2 Magnetic separation with streptavidin-coated beads 310
7.7.3 Size separation method 311
7.8 SELEX techniques 312
7.8.1 FAM-labelled ssDNA quantification 312
7.8.2 Fluorescamine assay for peptide quantification 312
7.8.3 Peptide labelling with biotin 312
7.8.4 Non-templated control PCR test 313
7.8.5 PCR amplification and purification for each selection cycle 313
7.8.6 Q-PCR quantification 314
7.8.7 Enzymatic synthesis of aptamer sequences for target binding validation 315
7.8.8 SPR binding analysis 316
7.8.9 Electrophoretic Mobility Shift Assay 320

References 322

Appendix 355

Chapter 1

Introduction

1 Introduction

1.1 Nucleic acids

1.1.1 Primary structure of nucleic acids

Nucleic acids are small biological molecules which carry genetic information in living systems. They are comprised of monomers called nucleotides, which made of three subunits (Figure 1.1): a pentose sugar (2′-deoxy-D-ribose for DNA and D-ribose for RNA), a nitrogen-containing heterocyclic base (nucleobase) and a phosphate linker.

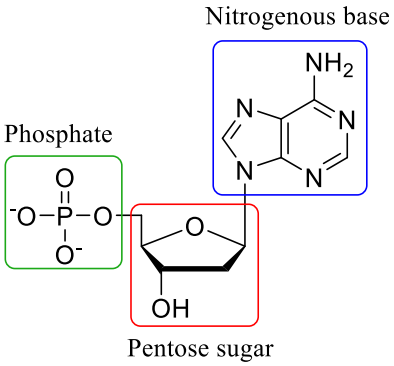


Figure 1.1. Structure of nucleotide. The pentose here depicted is 2′-deoxyribose with adenosine as a base.

The nucleobases are grouped in two families, pyrimidines (cytosine (C) and thymine (T) or uracil (U)) and purines (adenine (A) and guanine (G)) (Figure 1.2). DNA has four bases, adenine (A), guanine (G), cytosine (C) and thymine (T) whereas RNA has uracil (U) instead of thymine (T).

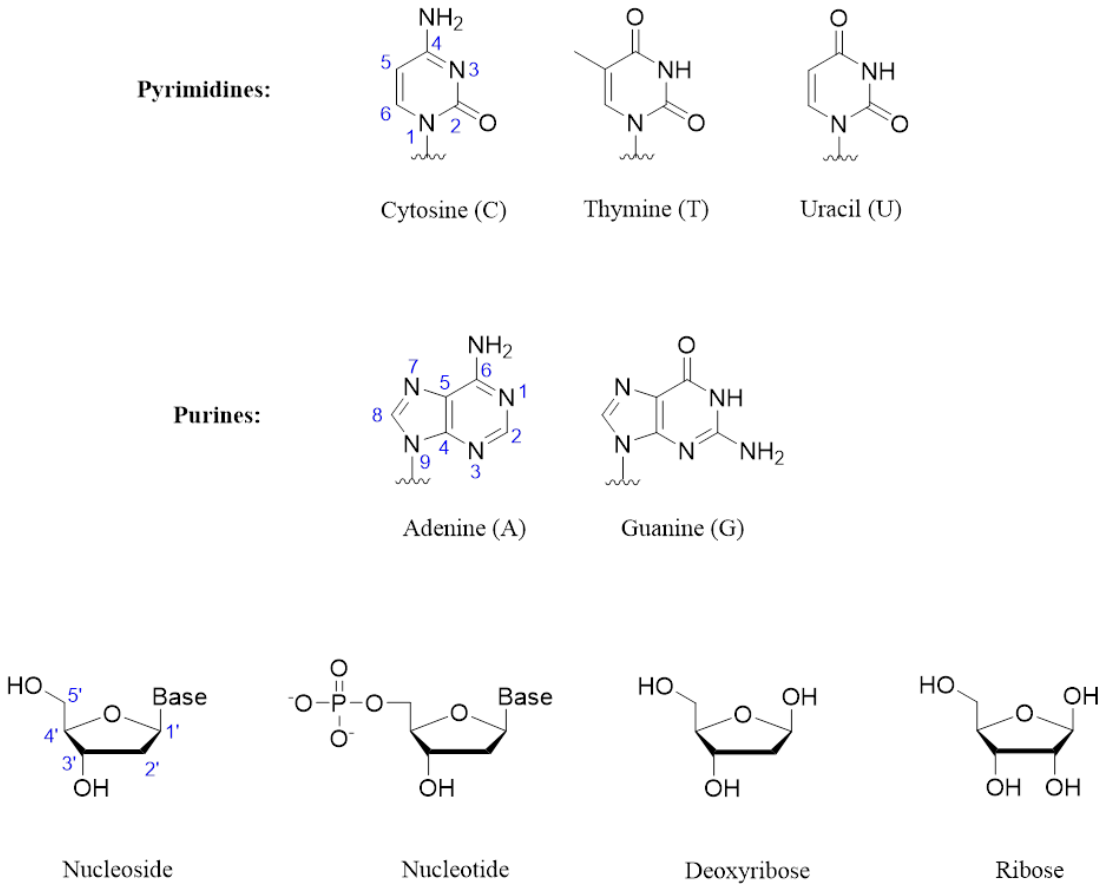


Figure 1.2. The five nitrogenous bases most commonly present in DNA and RNA. Thymine is present in DNA only, while uracil is present only in RNA.

The heterocyclic base is attached to the pentose sugar via an *N*-glycosidic bond which links between the C1′ position of the ribose with the N1 position of pyrimidines or N9 position of purines. The nucleobase can be located on the same side of the sugar ring (β-anomer) or the opposite side (α-anomer) to the 5′-hydroxyl group (Figure 1.3). The β-configuration is present in all naturally occurring oligonucleotides. The rotation around the glycosidic bond allows the nucleoside to have *anti*- and *syn*-conformations (Figure 1.3). Nucleosides usually exhibit *anti*-conformation due to the restriction of bond rotation and steric interference.

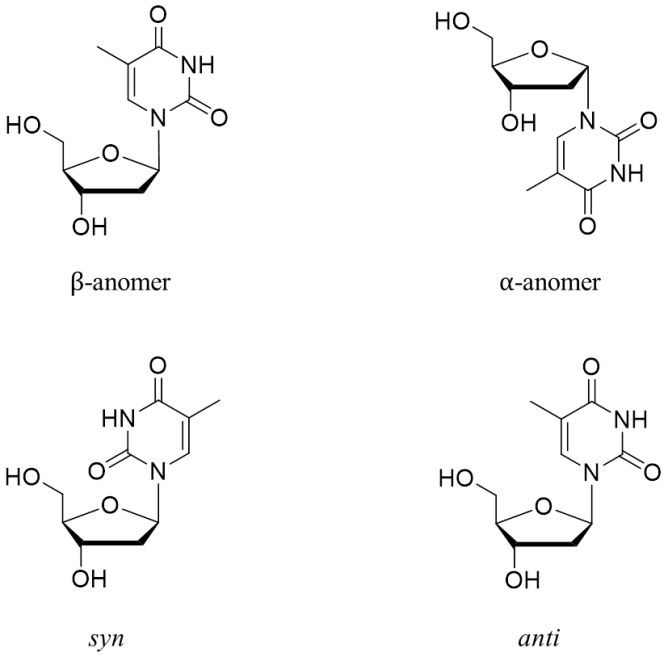


Figure 1.3. Conformations and configurations of nucleosides.

The furanose rings are twisted out of the plane so that electronic and steric interactions between their substituents can be reduced. This is referred to as sugar puckering. DNA adopts mainly the S-type configuration where the C2′ on the same side of the ring (endo) as the C5′, while RNA adopts mostly the N-type configuration with C3′-endo ribose puckering as a consequence of the 2′-OH (Figure 1.4).

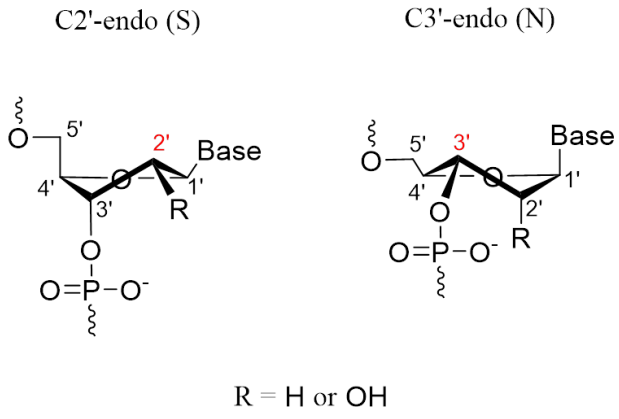


Figure 1.4. C2′-endo (S) and C3′-endo (N) pucker conformations of deoxyribose (R = H) or ribose (R = OH).

1.1.2 Secondary structure of nucleic acids

In an oligonucleotide (oligomer), nucleotides are covalently linked through the 5′-phosphate group of one nucleotide to the 3′-hydroxyl group of the next nucleotide forming a phosphodiester bond. At physiological pH (approximately pH 7.4), all phosphodiester groups are deprotonated, and nucleic acids are polyanionic molecules. Conventionally, the sequence of nucleic acids is written from the 5′ to 3′ direction as oligonucleotides have directionality with a 5′-hydroxyl group on one end of a strand and a 3′-hydroxyl group on the other end (Figure 1.5 **A**).

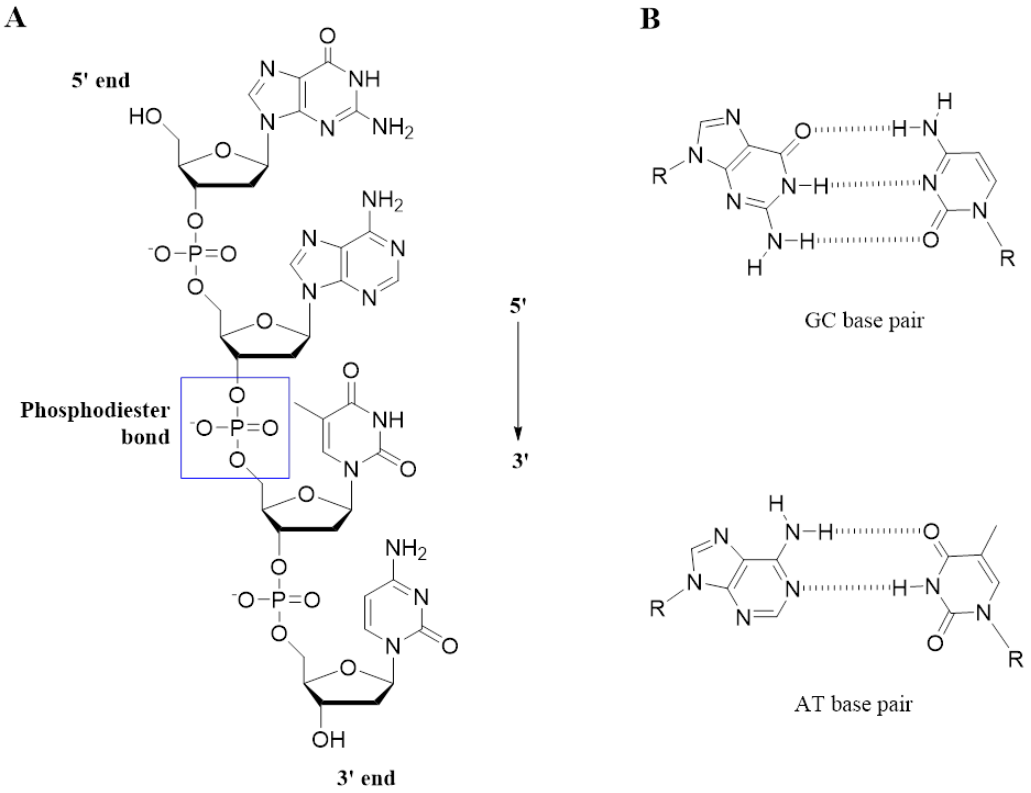


Figure 1.5. **A**) Chemical structure of oligonucleotide dGATC with phosphodiester group highlighted. **B**) Watson-Crick base pairs (GC and AT pairs).

In 1953, James Watson and Francis Crick discovered that DNA could form a double-stranded (ds) structure which has two helical chains each coiled round the same axis.[1]

Crick and Watson established their model based on several observations. First, an equal ratio between purine and pyrimidine bases (Chargaff's rule), *i.e.* A/T and G/C = 1 or the sum of purines is equal to the sum of pyrimidines (Figure 1.5 **B**). Second, Wilkins and Franklin observed a helical structure for DNA using X-ray diffraction.[2–4] Third, Jerry Donohue pointed out that Watson was using the wrong tautomeric forms of the nitrogen bases and corrected the base pairing rule which leads to hydrogen bonds between A-T and G-C.[5] Watson and Crick's experimental results confirmed that the polydeoxyribonucleotide chains were held together by hydrogen bonds.

The heterocyclic base pairs are projected towards the centre of the helix, and the negatively charged phosphate chains form the external hydrophilic shield. The two strands of DNA are complementary and antiparallel. The planar hydrophobic bases are parallel to one another (base stacking), and the overlapping π-orbitals of neighbouring base pairs stabilise DNA duplex (Figure 1.6 **A**). Two continuous bases are 3.4 Å apart, and there are 10.5 base pairs per one 360° turn of the helix (Figure 1.6 **B**). Adenine (A) pairs with thymine (T) (or uracil in RNA) via two hydrogen bonds; guanine (G) hydrogen bonds with cytosine (C) by three hydrogen bonds (Figure 1.5 **B**). The higher the G-C/A-T ratio the more is stable dsDNA since the G-C base pair has an additional hydrogen bond compared to the A-T base pair. Both hydrogen bonding and base stacking play important roles in stabilising the DNA duplex.

The two DNA strands that form the duplex can adopt different conformations – A-, B-, or Z-forms.[6] Both B-form and A-form are right-handed helices, with the most biologically prevalent conformation for DNA being B-form (Figure 1.6 **A**).[7] B-DNA is characterised by the presence of a wide major groove (width 11.6 Å, depth 8.5 Å) and a narrow minor groove (width 6 Å, depth 8.2 Å).[8] The major groove is more accessible to

biological macromolecules interacting with DNA,[9] whereas small molecules and few proteins can bind in the minor groove due to steric hindrance (Figure 1.6 **B**).[10–12] The transition from B-DNA to A-DNA occurs when the relative humidity of the sample is lowered and also RNA more readily adopts the A-form.[13,14] The bases in A-form duplexes are located away from the central axis, resulting in a ribbon-like helical structure with a wider core and larger diameter (Figure 1.6 **A**). Moreover, A-form helices have deeper major grooves and shallower minor grooves compared to B-form (Figure 1.6 **A**).[15]

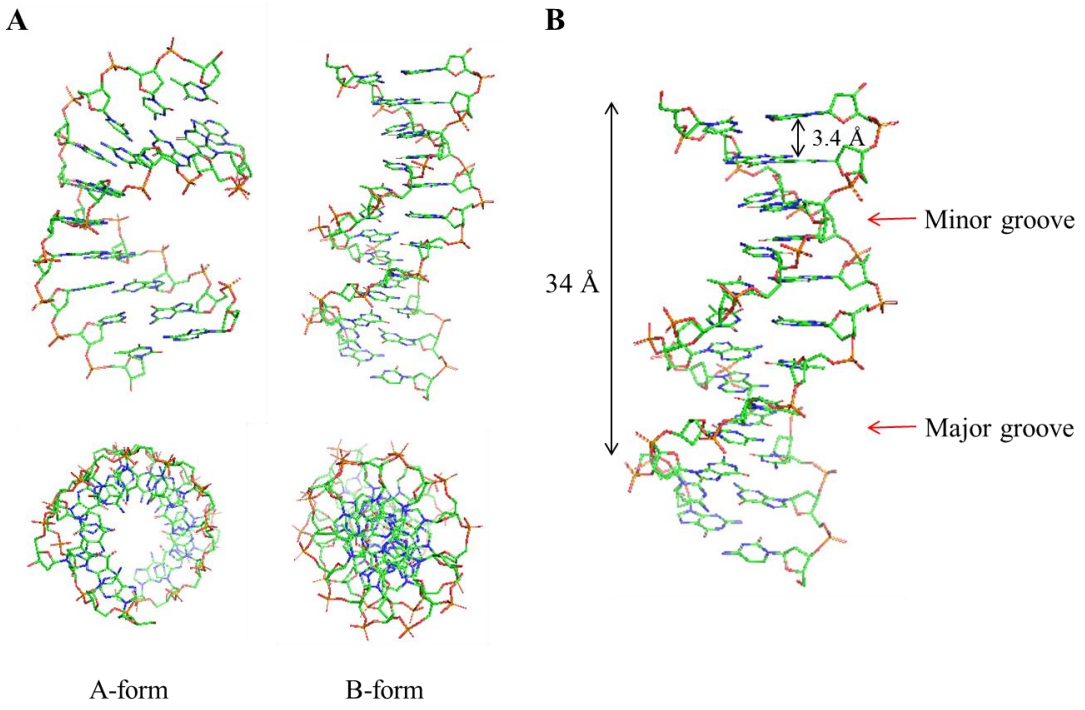


Figure 1.6. Structures of A- and B-form DNA duplexes (PDB code 440D[16] and 1BNA[17], respectively) **A**) Side and top views of A- and B-form DNA. **B**) Structural parameters of B-form DNA. The structural presentations were generated using PyMOL program.[371]

1.1.3 Thermal stability of DNA duplexes

The melting temperature (T_m) of a nucleic acid duplex is the point at which half of the DNA duplex is dissociated into single-stranded (ss) DNA (Figure 1.7). T_m is dependent on the length of DNA duplex, base stacking and the number of hydrogen bonds, *i.e.* longer or GC-rich oligonucleotides have higher T_m than shorter or AT-rich oligonucleotides (*i.e.* they exhibit higher duplex stability). The stability of hybridised oligonucleotides is also affected by the concentration of monovalent (e.g. sodium and potassium), divalent (e.g. magnesium) and polyvalent cations. Divalent cations bring a larger impact on T_m than monovalent cations.

T_m is an important parameter to consider when performing enzymatic reaction such as polymerase chain reaction (PCR), quantitative PCR (Q-PCR) or primer extension reaction, where thermostable polymerases are used, which are active at a higher temperature. The temperature that primers bind to the template should be low enough but not so low as to enable non-specific (mismatch containing) duplex formation. Forward and reverse primers in PCR should be designed to have a similar T_m value.

In order to measure thermal stability, the intrinsic properties of nucleic acids are monitored. Nucleotide bases of DNA are absorbers of ultraviolet (UV) light. Due to strong π-stacking interaction, the absorbance of duplex DNA is lower than single strands (Figure 1.7 **A**) and therefore, the change in absorbance can be used to measure T_m. However, at very low concentrations of oligonucleotides (at nanomolar (nM) concentrations or lower), a more sensitive method is required. Fluorescence melting (Figure 1.7 **B**) using DNA intercalators such as SYBR Green (Figure 1.8 **A**) or EvaGreen (Figure 1.8 **B**) can be used to measure the T_m of DNA. The majority of intercalators are positively charged, aromatic and planar. The cationic intercalator is initially attracted electrostatically to the negatively charged DNA. The intercalator

displaces a mono- or divalent cation that always surrounds DNA and form a weak electrostatic interaction. Then, it is finally substituted into the hydrophobic environment found between the two neighbouring nucleobase pairs via π–π, hydrophobic, polar and electrostatic interactions. These interactions stabilise the DNA-intercalator complex.[18–20] Intercalation also changes the torsion angles of the phosphodiester backbone, resulting in the unwinding and lengthening of the DNA helix by approximately 3.4 Å per intercalation event, while maintaining the overall B-form of DNA.[21,22]

As DNA duplexes are stabilised in the presence of intercalators, this method of monitoring duplex stability only provides an approximate T_m, depending on the intercalators used in the reaction.[23] These intercalators have been widely used to monitor the progression of PCR, with the higher amounts of duplex product generated during amplification giving a stronger signal as more intercalators are bound.

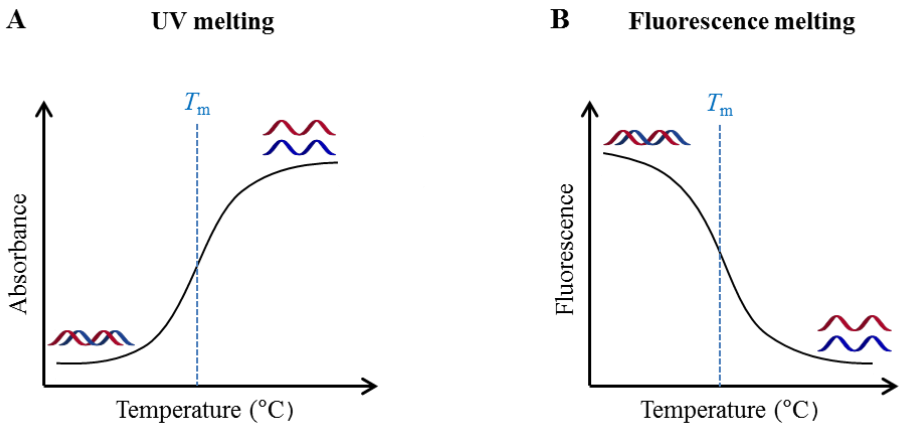


Figure 1.7. Schematic of melting curves for UV (**A**) and fluorescence melting (**B**) using a DNA intercalator. T_m = melting temperature.

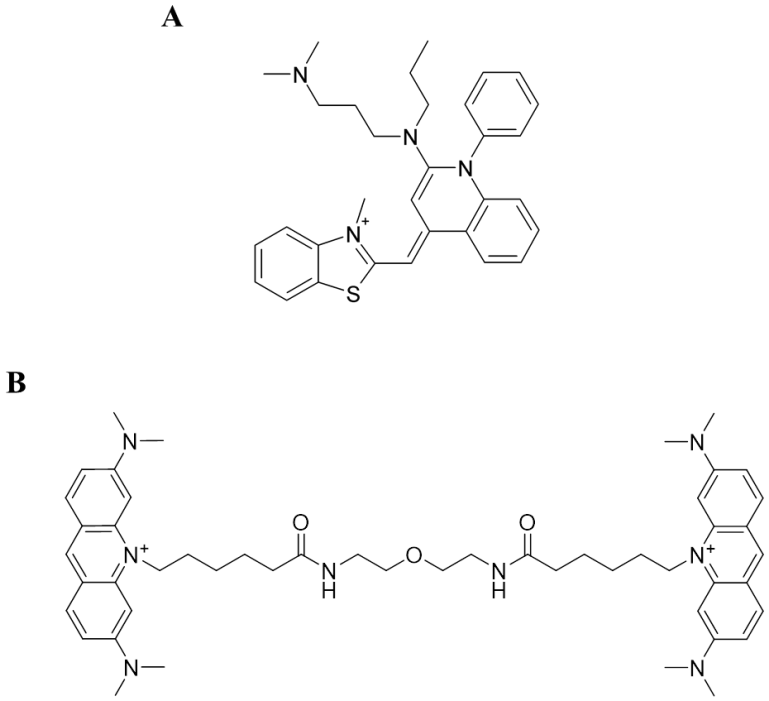


Figure 1.8. SYBR Green (**A**) and EvaGreen (**B**) intercalator.

1.2 Aptamers

Aptamers, a promising class of targeted therapeutics and diagnostic agents, are short, single-stranded oligonucleotides (DNA or RNA), which bind with high affinity and specificity to their targets (e.g. small molecules, peptides, viruses, bacteria, proteins, cells and even whole tumours in the aminal models).[24–28] Aptamers can fold into various well-defined structures (e.g. stem, loop, bulge, pseudoknot, hairpin and G-quadruplex) for target recognition by shape complementarity (Figure 1.9 **A-D**). Hydrophobic and electrostatic interactions – such as hydrogen bonding, van der Waals forces and base stacking are also essential for aptamer binding affinity and specificity.[29] As a consequence, they generally have a very high affinity (typically nanomolar dissociation constants) and high specificity (>1,000-fold) for their targets, almost equivalent to that of antibodies.

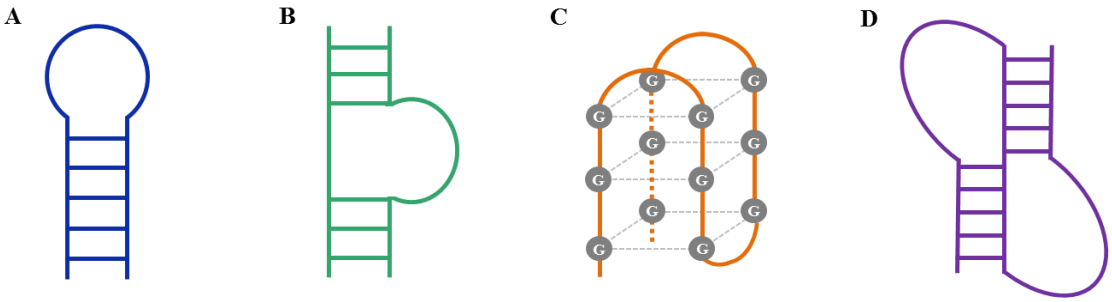


Figure 1.9. Examples of the secondary structures of aptamers. **A**) Hairpin (stem-loop) structure. **B**) Bulge loop structure. **C**) G-quadruplex structure. **D**) Pseudoknot structure.

Other than Watson-Crick duplex regions, G-quadruplexes are the tertiary structures most frequently found in DNA aptamers.[30] Oligonucleotides containing guanine repeats which are separated by other nucleobases tend to form G-quadruplex structures under certain conditions. The basic unit of a G-quadruplex is the guanine tetrad (G-tetrad), which is composed of a square planar array of four guanine bases stabilised by Hoogsteen hydrogen bonding (Figure 1.10).[31] The structures are formed by stacking between at least two G-tetrads which are held together by loops. The stabilisation of G-tetrads is partly via π-π stacking interactions and the entire structure is further stabilised by binding of monovalent cations (Na^+ or K^+).[32,33] This cation coordinates within the quadruplex core and interacts electrostatically with the O6 atoms of guanine base (Figure 1.10) and the anionic phosphodiester backbone.[34] Intensive studies on G-quadruplexes are currently ongoing due to their potential formation in single-stranded telomeres and oncogene promoter regions.[35–38]

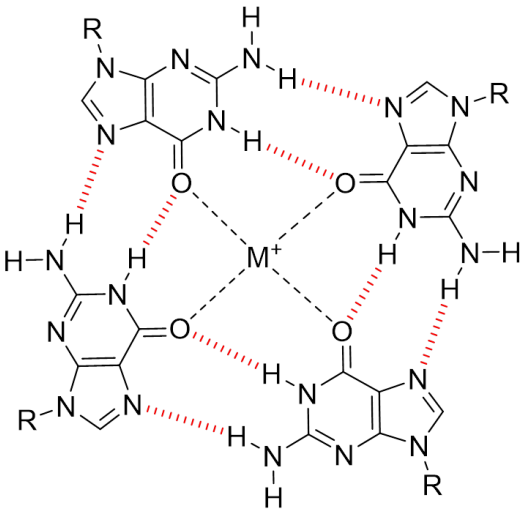


Figure 1.10. Structure of G-quadruplex which arises from the interaction of four guanines into a cyclic arrangement stabilised by Hoogsten hydrogen bonding (N1–O6 and N2–N7). The structure is further stabilised by monovalent cation (M^+) binding.

G-quadruplex architecture is topologically polymorphic and therefore a wide range of different topologies are known to exist. The combination of the number of stacked G-tetrads, the polarity of the strands and the location and length of the loops give many possible structural arrangements of G-quadruplex structures.[31] G-rich strands can form intra- or inter-molecular folds. When four or more G-tetrads are present in one strand, intra-molecular folding can occur. Inter-molecular folding can arise from two or four strands giving rise to parallel or antiparallel structures depending on the orientation of the strands (Figure 1.11).[39] G-quadruplex structures can be predicted by analysing the sequences which match a quadruplex folding rule which is G_n-$X_{1\text{-}7}$-G_n-$X_{1\text{-}7}$-G_n-$X_{1\text{-}7}$-G_n (G = guanine, X = any base, $3< n <5$).[40,41] G-quadruplexes are known to be thermally stable compared to dsDNA and their thermodynamic stability depends on the underlying oligonucleotide sequences such as the number of G-tetrad resides in the structure and length and nucleobase composition of the loops formed by adenine, cytosine and thymine or uracil bases.[42,43] Higher number of G-tetrads extend the π-π

stacking and lead to higher thermal stability.[44] In addition, the parallel strand conformation of G-quadruplexes has been found to be thermodynamically more favourable due to its relatively short loops compared to the antiparallel conformation which must have longer loops (Figure 1.11).[43–45] G-quadruplex DNA usually exhibits high thermodynamic stability and its unfolding kinetics are much slower than the canonical hairpin loop DNA and RNA structures.[46]

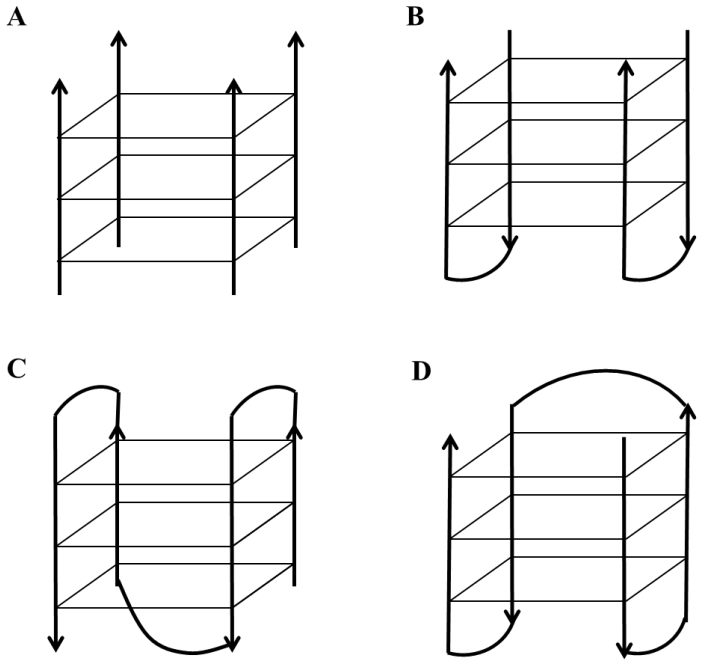


Figure 1.11. G-quadruplex molecularities. **A**) Tetramolecular, **B**) Biomolecular and **C** and **D**) intramolecular with different loop topologies.

G-quadruplexes have been commonly found in DNA aptamers for a wide range of protein targets, possibly partly due to the high charge density in G-quadruplexes which have twice the negative charge per unit length compared to dsDNA.[47] This high electrostatic potential of G-quadruplexes contributes to binding to positively charged target proteins in particular.[48] Nagatoshi *et al*. suggested a mechanism of G-quadruplex stabilisation by target binding.[49] The loop regions of a 15 mer G-quadruplex thrombin-

binding DNA aptamer were site-specifically mutated and the results revealed that the specific interactions between the nucleobases in loops and amino-acid residues of thrombin stabilise the G-quadruplex aptamer upon binding.[49] Loop regions of the G-quadruplex play an important role in thermodynamic stabilisation of G-quadruplexes via π stacking between the G-tetrads and loop nucleobases.[45,50,51] Through this, specific interactions of G-quadruplexes with their targets can be made.[52]

Interestingly, the protonated C-rich sequence can also fold into three-dimensional tetraplex structures called i-motifs (Figure 1.12 **A**) in slightly acidic or even neutral pH environment, but i-motifs are predominant in the acidic pH region.[53,54] Structure and the stability of i-motif are related to the number of cytosines in repeating C-tracts or length of loops (Figure 1.12 **B**).[55,56] In contrast to G-quadruplex, i-motifs has a limitation due to its low stability *in vivo* application where exhibiting neutral pH.[55] However, i-motif structures have been recently used in analytical and biomedical applications by monitoring its structural changes (i-motif folding/unfolding) upon the pH changes inside cellular system.[57–59]

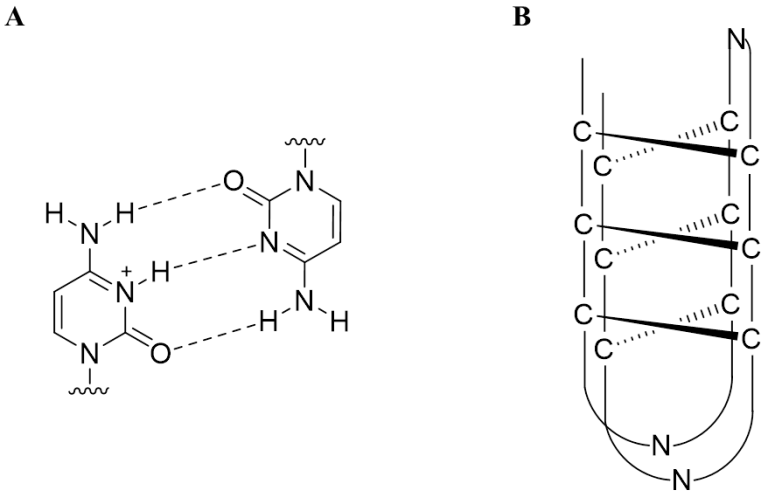


Figure 1.12. **A**) Protonated-cytosine:cytosine base pair. **B**) Schematic representation of an i-motif structure, where N denotes any DNA heterocyclic base.

1.2.1 Comparison with antibodies

Importantly, aptamers offer several advantages over antibodies. Firstly, from a synthetic perspective, aptamers can be cost-effectively produced on a large scale and site-specifically modified by chemical methods. In addition, they are intrinsically stable to heat and, unlike antibodies, can undergo multiple heat denaturation steps and still refold. Secondly, for therapeutic applications, they penetrate tissue faster due to their smaller size (8–25 kDa aptamers vs ~150 kDa of antibodies) and generate lower toxicity and immunogenicity as they are not normally recognised by the host immune system.[24,25]

From a clinical perspective, an anti-VEGF aptamer, (Pegaptanib or Macugen) has been approved for the treatment of neovascular (wet) age-related macular degeneration disease (AMD).[60,61] In fact, numerous aptamers are now at preclinical or clinical trial stages for coagulation therapy,[62,63] gene therapy,[64–66] immunotherapy,[67,68] and cancer therapy.[69,70] Many aptamers have been discovered to bind selectively to a diverse range of extracellular targets including vascular endothelial growth factor (VEGF),[71–73] basic fibroblastic growth factor (bFGF),[74] and platelet-derived growth factor (PDGF).[75–77] Another interesting target is coagulation factor IXa, where aptamers prevent arterial thrombosis by providing a rapid and tightly regulated system for controlling the coagulation cascade mechanism.[62,78,79] Aptamers also bind to many oncogenic reactors. Upon binding, the ligand-receptor interaction is blocked and therefore, cancer cell survival, growth and proliferation are controlled.[29,80] The clinically proven aptamers for cancer treatment are AS1411 and NOX-A12. AS1411 binds to the external domain of nucleolin with high affinity,[70,81] and NOX-A12 binds to chemokine C-X-C motif ligand 12 (CXCL12; also known as stromal cell-derived factor-1).[82,83]

Most of the published aptamers were evolved against extracellular targets, due to the importance of these targets in tumour biology and the ease of access to these proteins.

Intracellular targets are more challenging than extracellular targets because highly polar aptamers pass through the lipid bilayer of the cell membrane with great difficulty.[84] Many studies are still ongoing to discover aptamers against intracellular targets including transcription factors, enzymes and coenzymes, as well as signal transduction molecules.[85–87] A major limitation is an alteration of aptamer structure when delivering intracellularly. Nevertheless, efficient delivery of aptamers to intracellular targets, while maintaining function, has been achieved by different groups.[88–90]

1.2.2 Aptamer selection (SELEX)

Aptamers are normally derived from a random library of 10^{13}–10^{16} ssDNA or ssRNA molecules through an *in vitro* selection technique called SELEX (**S**ystematic **E**volution of **L**igands by **EX**ponential enrichment). The technique was developed independently by Szostak *et al*. and Gold *et al*. in the 1990s and is depicted in Figure 1.13.[91,92]

A ssDNA library, typically 80–100 nucleotides in length containing a random region (the diversity of which is determined by the equation 4^n where n = the size of the randomised region) flanked by two constant primer binding sites, it allowed to bind to the desired target. Oligonucleotides that bind to the target are retained, while non-binding sequences are removed, typically by washing if the target is immobilised. The recovered sequences are then amplified by PCR to enrich and bias the library towards sequences that bind to the target. The process is repeated multiple times (typically 8–20 rounds) until the DNA sequences are sufficiently enriched for target binding. This process can be monitored by Q-PCR. Often, the binding conditions are made more stringent as SELEX proceeds through each round, increasing the selection pressure. For example, selection pressure can be increased by reducing the concentration of target

molecules presented during the binding step during successive cycles. Additional selection steps using counter-targets (a similar molecule to the target) are performed to improve the ability of aptamers to discriminate between structurally similar molecules.[29,93] In addition, selection pressure can be increased by increasing the number and duration of washing steps to give a slower (lower) dissociation rate constant (k_d) so that aptamers with lower equilibrium binding constant (K_D) can be obtained.[93] The early SELEX cycles have low selection pressure to ensure the recovery of rare sequences. When the desired sequences are enriched after 2–3 cycles of selection, higher stringency can be applied to discover aptamers with high binding affinity and selectivity.[94]

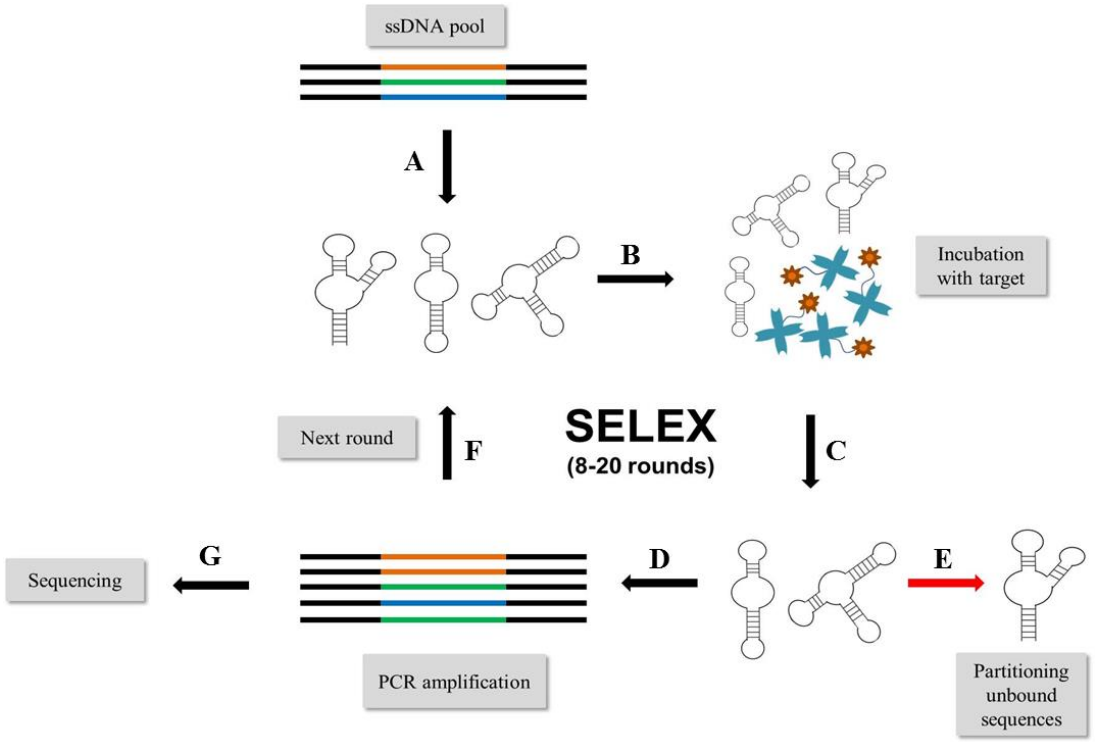


Figure 1.13. Overview of a typical SELEX procedure. **A**) Initial ssDNA library preparation. **B**) ssDNA and target binding. **C**) Capture of ssDNA-target complexes. **D**) PCR amplification. **E**) Removal of unbound sequences. **F**) ssDNA generation for the next round of selection. **G**) Sequencing.

1.2.3 Improvement in SELEX techniques

While SELEX has been a successful method of aptamer generation for over two decades, there is still room for improvement.

1.2.3.1 Bias in PCR amplification

The presence of 10^{13}–10^{16} ssDNA containing undefined random regions can produce problems during PCR amplification. Typically, short dsDNA products produced by primer-primer hybridisation, mis-priming or longer ss/dsDNA products are found. For example, primers can bind to complementary nucleotides in the random region.[95,96] Many techniques to optimise PCR conditions to reduce by-product formation have been evaluated including optimising the primer design, the initial concentration of the library, the number of PCR cycles, the extension time and temperature, total length of the heterogeneous DNA library and the concentration of DNA polymerases.[95,97] However, repetitive amplification of the library inevitably leads to a degree of by-product formation. PCR amplification of random DNA libraries must be carefully optimised for efficient aptamer selection. For example, droplet digital PCR (ddPCR) or emulsion PCR (ePCR) have been recently introduced to reduce by-product formation in PCR. This technique is successful because one copy of ssDNA is compartmentalised in a droplet in a water-in-oil emulsion and the sequences in the droplet are amplified during PCR.[98–100] These droplets can then be broken up and the sequences pooled. These methods have been shown to maintain library diversity.[98] Compared to solution PCR, ddPCR requires additional steps, including droplet generation and amplicon extraction by organic solvent from the droplets. Therefore, ddPCR is expensive and laborious.[99]

1.2.3.2 ssDNA generation methods

One of the crucial steps in SELEX is the generation of ssDNA libraries from dsDNA products after PCR amplification. Various methods have been reported for the generation of ssDNA including asymmetric PCR, lambda exonuclease digestion of one strand, streptavidin-biotin bead separation using a magnet and size separation by denaturing polyacrylamide gel electrophoresis (denaturing PAGE).[101]

The most widely used method in SELEX is magnetic separation using streptavidin-coated beads (Figure 1.14 **A**). A biotinylated primer is used during PCR amplification to produce dsDNA with one biotinylated strand. This is then captured by the streptavidin-coated magnetic beads. Sodium hydroxide (alkaline) treatment or heating the beads to 95 ℃ eventually releases the non-biotinylated strand into the supernatant.[102] However, Paul *et al*. reported that non-covalent binding between streptavidin and the beads could be disrupted by alkaline treatment. This dissociation of biotinylated strands leads to re-annealing of them to the desired complementary strand forming dsDNA.[103]

An alternative method is to carry out an asymmetric PCR where one strand of the template DNA is amplified more efficiently than the other by using an unequal ratio of forward and reverse primers in the PCR reaction. Hence, ssDNA is produced during each cycle when the primer at lower concentration becomes exhausted.[104] However, asymmetric PCR generates not only ssDNA but also dsDNA products which may limit the diversity of enriched ssDNA pools.

Lambda (λ) exonuclease digestion is also frequently used to generate ssDNA during SELEX. One of the dsDNA strands is modified through the use of a 5′-phosphorylated primer during PCR. This phosphorylated strand is selectively digested by lambda exonuclease (Figure 1.14 **B**).[105–107] However, incomplete digestion of the

phosphorylated strand or partial digestion of the unphosphorylated complementary strands may also lead to low yield in ssDNA generation.

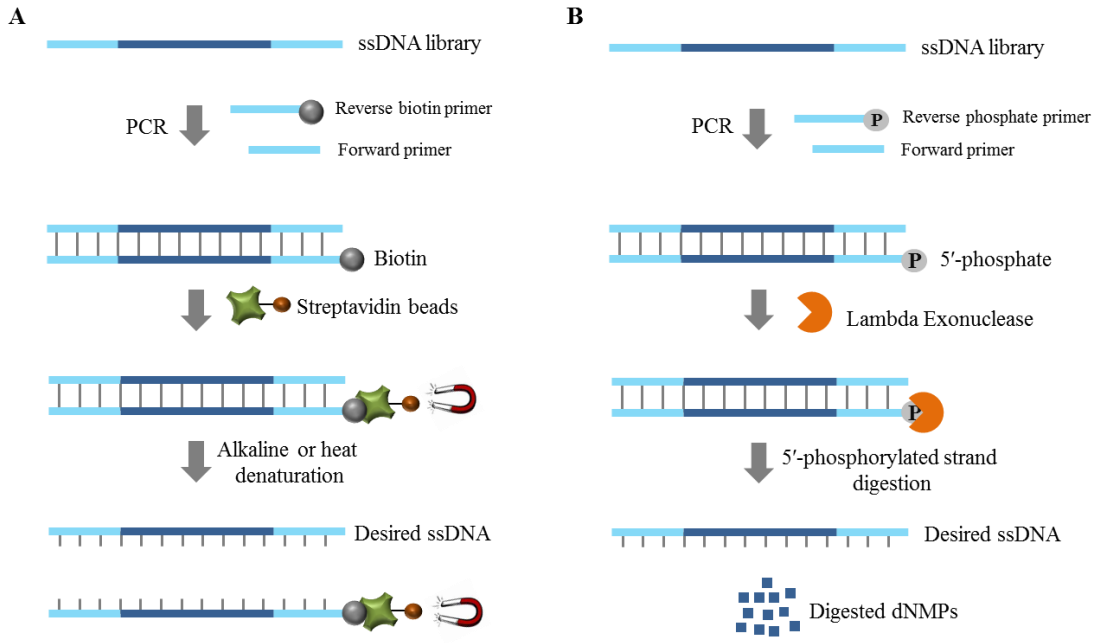


Figure 1.14. ssDNA generation methods using streptavidin beads separation (**A**) and lambda exonuclease digestion (**B**).

Another method to separate the DNA strands is size separation through the use of primers containing chemical modifications (e.g. 18-carbon ethylene glycol spacer (HEG) with an extension of 20 nucleotides length of a string of adenosines (PolyA tail) (Figure 1.15 **A**) or structural modification (GC-rich stem-loop) (Figure 1.15 **B**).[108,109] After PCR, the two strands have different lengths and the desired stand can be extracted from a denaturing PAGE gel. However, size separation derived from a primer with a GC-rich stem-loop structure gave a significantly lower recovery rate of ssDNA compared to size separation derived from primer with a chemical terminator (*i.e.* HEG).[101] This is because GC-rich stem-loop primers produced not only the desired unequal length product strands but also unwanted equal length strands of the products with a larger

molecular weight (MW), as the stem-loop structure can be partly disrupted at high temperatures.

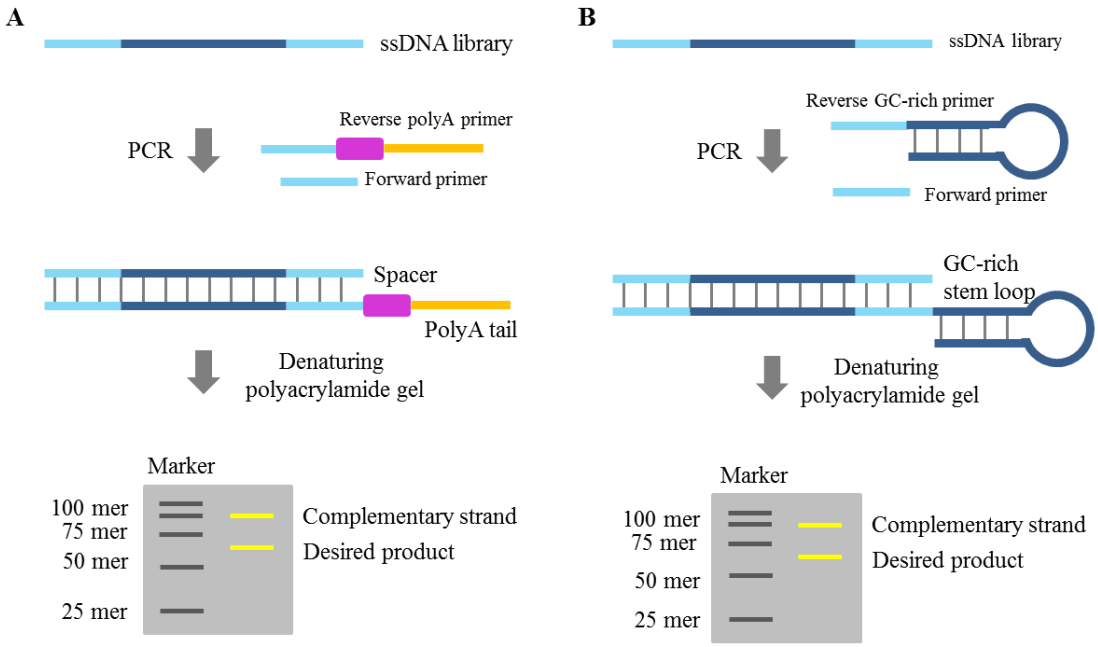


Figure 1.15. ssDNA generation by size separation method using a spacer (e.g. HEG) and the polyA tail (**A**) or using GC-rich stem-loop (**B**).

The type of ssDNA generation method does not affect the success of aptamer selection, implying that the quality of the ssDNA and the yield of the ssDNA are the most important points.[110]

1.2.3.3 Modified SELEX techniques

SELEX can be laborious and time-consuming (up to several weeks).[111] Therefore, numerous alternative methods have been developed to address these issues.

In general, the target molecules are attached to a solid support to separate target-bound oligonucleotide sequences from unbound sequences. The original SELEX used nitrocellulose filters to partition target-bound oligonucleotides from unbound

oligonucleotides. However, aptamers found to be not only bind to their target but also bind to the nitrocellulose membrane, leading to enrichment of non-specific oligonucleotides.[92] This is a common problem with immobilisation of the target on a matrix. To address this, in 1992, Ellington *et al*. demonstrated that 'negative SELEX' rounds, where the library is incubated with purification support (in their case agarose), can help eliminate non-specifically bound sequences.[112]

As an example method for partitioning target bound and unbound sequences, target molecules were attached to magnetic beads and partitioned from unbound sequences using a magnet.[113] The benefits of this approach were speed, ease of handling, use of very small amounts of the target during the aptamer selection, efficient partitioning between bound and unbound complexes and compatibility with stringent washing steps.[114] In 2005, Stoltenburg *et al*. developed 'FluMag-SELEX' which is based on magnetic bead immobilisation but also uses fluorophores (e.g. fluorescein) in ssDNA allowing easy quantification in each round of the SELEX cycle.[114] By using this approach, aptamers against streptavidin,[114,115] polychlorinated biphenyls,[116] and ibuprofen[117] were generated.

In light of these results, a SELEX variant, called capture-SELEX was developed by Stoltenburg in 2012.[118] During capture-SELEX, the DNA library is immobilised instead of the target by annealing DNA library to a bead-bound capture oligo (*i.e*. docking sequence). This selection method enables the selection of aptamers against small organic molecules (e.g. pharmaceuticals) which cannot be immobilised on a solid support. Capture-SELEX was used to discover aptamers against kanamycin A, sulfacarbamide, sulfamethoxazole and solatol hydrochloride.[118]

A major imitation of the conventional SELEX, with very few exceptions, is that additional analytical methods (e.g. SPR, spectroscopy, gel shift assays) are required to evaluate the binding characteristics of the pools or the selected aptamer candidates. Especially, label-free surface plasmon resonance (SPR) instruments (see section 1.6.1) are commonly used to study the binding properties of aptamers. In 2003, Khati *et al*. discovered 2′-F modified aptamers binding to the gp120 of the R5 strain, HIV-1_{Ba-L}. They used a SPR sensor chip to immobilise their target and monitored the binding of aptamers in real-time, which can be alternative selection method to conventional SELEX.[119] In 2016, Dausse *et al*. also applied real time SPR based SELEX. By using this method, they discovered RNA aptamers binds to a *XBP1* pre-mRNA derivative via highly stable loop–loop complexes.[120] As real time SPR-SELEX allows the selection and evaluation simultaneously; it is expected to speed up the entire aptamer selection process.

In 2004, capillary electrophoresis SELEX (CE-SELEX) was developed, where target-bound sequences were separated from unbound oligonucleotides according to their differences in electrophoretic mobility. The advantage of CE-SELEX is the high speed of aptamer selection compared to conventional SELEX methods. The high resolving power of CE minimises the number of required selection cycles from 8–15 in conventional SELEX to only 1–4 for CE-SELEX.[121] Using CE-SELEX, aptamers were discovered against human IgE,[122] human immunodeficiency virus reverse transcriptase (HIV-1 RT),[123] neuropeptide Y,[124] ricin toxin,[125] protein kinase C-δ,[126] and anthrax protective antigen.[127]

Another variation of CE-SELEX called ‘Non-Equilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM)’ has also been described. It is a unique non-SELEX

technique to select an aptamer without an amplification step and is based on the principles of non-equilibrium capillary electrophoresis.[128] When an oligonucleotide library and its target are incubated in solution and reach the equilibrium, the complexes are injected into a capillary and separated by an electric field. The unbound oligonucleotide and targets migrate in a single electrophoretic area whereas oligonucleotide-target complexes dissociate slowly by creating a non-equilibrium complex of oligonucleotide and target. By using this technique, selection time was dramatically reduced (1 hour) in contrast to several weeks required in conventional SELEX,[129] and also the binding kinetics, association and dissociation rate constants (k_a and k_d) can be calculated.

However, the limitation of CE-SELEX and NECEEM is that only small injection volume is allowed (5–50 nL) and therefore, the number of aptamer sequences in each injection is limited (~$10^{12–13}$) and the samples are in very low concentration.[122,123]

All previously described SELEX methods use the most common form of a purified target protein. However, the full-length membrane proteins expressed in prokaryotic or lower eukaryotic systems may lack or have different post-translational modifications. Hence, it might be possible that they cannot fold into the correct structure that is formed under physiological conditions. Consequently, aptamers generated using a purified recombinant protein may not be able to selectively recognise and bind to their targets in the native conformation.[130–133] To this end, whole live cells can be employed as targets for aptamers in 'cell-SELEX'. The prior knowledge of the specific cell surface marker is unnecessary. The most important advantage of this method is that protein targets are in their natural state and therefore, it is a way to discover unknown cell surface markers which can more precisely differentiate a diseased cell from a normal cell. Cell-SELEX

was firstly discovered by Gold in 2003.[134] Since its discovery, many different types of cancer cells including human hepatocarcinoma,[135] mouse tumour endothelial,[136], metastatic liver cancer,[137] glioblastoma multiform,[138] HER2 positive breast cancer,[139] small cell lung cancer,[140] lung adenocarcinoma,[141] prostate cancer,[142] and colorectal cancer[143] were used as targets in cell-SELEX.

The main limitation of cell-CELEX is that the presence of dead cells during selection which may lead to non-specific binding of oligonucleotides. In order to resolve this issue, Raddatz *et al*. applied fluorescence-activated cell-sorting to cell-CELEX. When aptamers were incubated with target cells containing both vital and dead Ramos Burkitt's lymphoma B cells, only aptamers that bound to calcein-AM-stained vital cells were collected by FACS.[144] However, the complexity of target cell surface components means that additional counter selections are compulsory steps to improve the specificity of aptamers. This makes the entire process tedious and costly.[130] In addition, the cell surface is very complex and therefore, it is difficult to successfully identify the target of the aptamer generated by cell-SELEX. Some aptamers not only bind to the target cell surface but also internalise into the cells making the identification of target more problematic.[130]

1.3 Chemical modifications in aptamer generation

There are several reasons why chemical modifications are introduced into nucleic acid aptamers:

- to increase chemical diversity. Aptamers naturally contain only a negatively charged phosphate backbone, sugar, and the four hydrophobic heterocyclic bases.

- to enhance nuclease resistance – for therapeutic applications, aptamer must be stable to endonucleases which are present in cells and rapidly digest unmodified DNA and RNA.

- to increase cellular uptake by conjugating with lipophilic compounds (cholesterol and tocopherol), cell-penetrating peptides (penetratin) or polyamines (spermine). For specific receptor binding, small-molecule-based targeting moieties such as the carbohydrate *N*-acetyl galactosamine (GalNAc, Figure 1.16) can be conjugated to aptamers.[145,146]

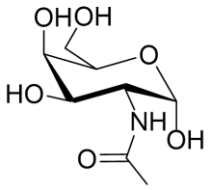


Figure 1.16. Chemical structure of carbohydrate *N*-acetyl galactosamine (GalNAc)

- to increase structural stability and rigidity for higher binding affinity. Conformational flexibility of aptamers results in a reduction in entropy upon binding to the target which is thermodynamically unfavourable, meaning that aptamers sometimes bind weakly to their target.[147]

Post-SELEX modification of aptamers is possible, but it is not ideal as the aptamers are not selected with the modification already present. It requires trial and error to find aptamers that retain good target binding characteristics. Therefore, SELEX is often performed using modified oligonucleotides. In order to modify aptamers during SELEX, there are several criteria:[148]

1. The base pairing interactions should not be disrupted by the modification.
2. DNA and/or RNA polymerase should tolerate the modified nucleotides (*i.e.* the polymerase should be able to read-through the modified template as well as incorporate the modified nucleotide).
3. The modification should not introduce sequence bias into enzymatic ssDNA or ssRNA synthesis.

1.3.1 Sugar modifications

For therapeutic applications, a major limitation of natural nucleic acid-based aptamers is their poor stability in biological media. They have short half-lives as a result of endo- and exonuclease enzymatic activities present in biological fluids and the hostile local environment (e.g. acidic or basic media, metal ions).[149,150] To overcome the former limitation, the 2′-position of the ribose sugar is often modified with fluoro (-F), amino (-NH_2), O-alkyl (e.g. OMe) or thiol (-SH) groups (Figure 1.17).[149,151,152] They are often introduced post-SELEX and retain good binding affinities (K_D values in the low nM down to the pM range) and also confer strong nuclease resistance. Yet, very few chemically modified nucleotides (e.g. 2′-F- and 2′-NH_2-pyrimidines) are tolerated as substrates by the polymerase enzymes used in SELEX.[71,153–157]

For the 2′-NH_2-ribose sugar, the amino group has a pK_a in the range of 9–10 and is therefore protonated at physiological pH. This cationic group dramatically reduces the susceptibility of oligonucleotides to nucleases, possibly by displacing the metal ion cofactors at the nuclease active sites of nucleases enzymes.[158] However, RNA aptamers with 2′-NH_2-pyrimidines adopt a B-like conformation (as in DNA) which destabilises intramolecular helices, giving rise to aptamers with thermodynamically less stable tertiary structures and weaker target-binding affinities.[152] In contrast, 2′-F-pyrimidine and 2′-OMe-pyrimidine both favour an RNA-like A-helix and impart duplex and tertiary structure stabilisation. 2′-F-pyrimidine modified RNA aptamers often have high binding affinity towards their targets.[156,159,160] Unlike 2′-F- and 2′-NH_2-pyrimidine-modified aptamers, 2′-OMe-RNA pyrimidine nucleotides are commonly introduced by trial and error post-SELEX modification.[73] This is because the bulky 2′-OMe-modification is poorly accepted by wild-type T7 RNA polymerase especially when compared to the smaller 2′-F- and 2′-NH_2-modifications.[151] Post-SELEX modification of aptamers with 2′-OMe-pyrimidine nucleotides results in a reduced target affinity compared to the natural aptamers. Therefore, the incorporation of 2′-modified nucleotides using a mutant T7 RNA polymerase has been studied and 2′-F-, 2′-NH_2-, 2′-OMe- and 2′-N_3-NTP have been successfully accepted as substrates.[161–164] Nevertheless, the introduction of bulky substituents at the 2′-position to produce stable aptamers remains challenging.[165,166]

Other sugar modifications have also been introduced into the RNA pool during the selection process for aptamer generation. In 2005, Matsuda's group successfully discovered an RNA aptamer by using the 4′-thio sugar modification in uridine and cytidine (Figure 1.17).[167] The resulting 4′-thio-modified RNA aptamer bound to human α-thrombin with high affinity (K_D = 4.7 nM) and had high RNase A stability compared

to wild-type RNA (50-fold increase). The 4′-thio-modification was accepted by DNA polymerases and therefore, the direct use of this sugar analogue in SELEX was possible.[168,169]

Locked-nucleic acids (LNAs) bear a methylene ether bridge between the 2′-oxygen atom and the 4′-carbon atom, 'locking' the ribose sugar moiety (Figure 1.17). Therefore, LNAs adopt fixed C3′-endo conformation mimicking RNA structure. In earlier studies, this modification was used only post-SELEX.[170,171] More recently, Wengel's group showed LNA nucleoside triphosphates are also compatible with enzymatic synthesis.[172,173] In 2015, an LNA aptamer for the human CD73 target containing LNA nucleotides in the primer and the random region was selected.[174]

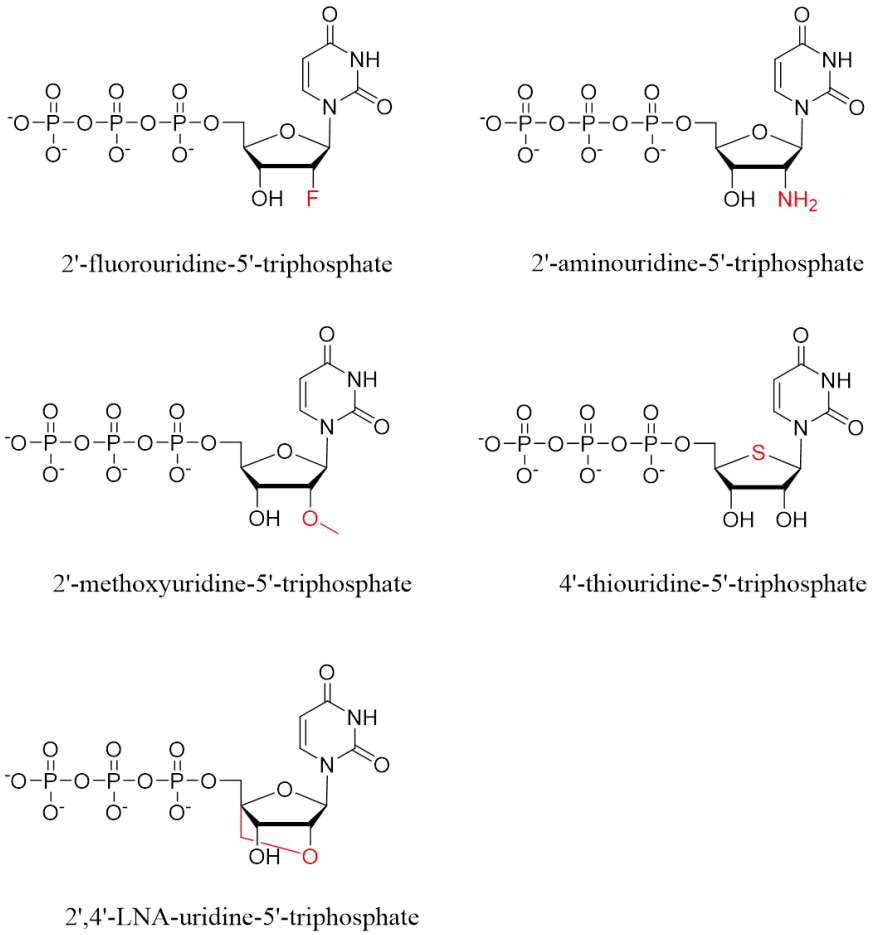


Figure 1.17. Chemical structures of sugar modified triphosphates used in aptamer generation.

1.3.2 Nucleobase modifications

In order to improve aptamer selection during SELEX, another major goal has been to expand the structural diversity of aptamers. Additional functional groups attached to the nucleoside can enhance target recognition by creating new interactions or altering nucleic acid secondary structures. In this regard, modifications or substitutions at the C5 position of pyrimidines or N7 position of 7-deaza-purines (Figure 1.2) have been widely investigated and shown to be tolerated by DNA polymerases as they are accommodated well in the major groove and do not give rise to steric clashes.[175,176] Therefore, nucleotides modified in these positions have been shown to be good substrates for polymerases.[177–180] The thermophilic B family of DNA polymerases with no 3′-exonuclease activity such as Vent exo-, Deep Vent exo-, Pfu and KOD Dash (KOD XL) are superior in tolerating sugar/base modification during SELEX.[181–183] However, there is not one specific polymerase which can accept the entire spectrum of non-natural substrates.

Functional groups present at the C5 of pyrimidine nucleobases can have a variety of properties including hydrophobic, acidic, basic and aromatic. In 2010, Vaught *et al*. synthesised C5-modified carboxamide dUTP derivatives to attach a range of hydrophobic groups. These modified dUTPs were tested with various DNA polymerases and eventually, modified DNA aptamers for the tumour necrosis factor receptor superfamily member 9 (TNFRSF9), with a K_D value of 0.5–9 nM, were selected.[179] Interestingly, this finding resulted in the development of novel Slow Off-rate Modified Aptamers or 'SOMAmers'. SOMAmers contain hydrophobic/aromatic side chains at the C5 position of dUTP and generally possess very low dissociation rates (Figure 1.18).[184–186]

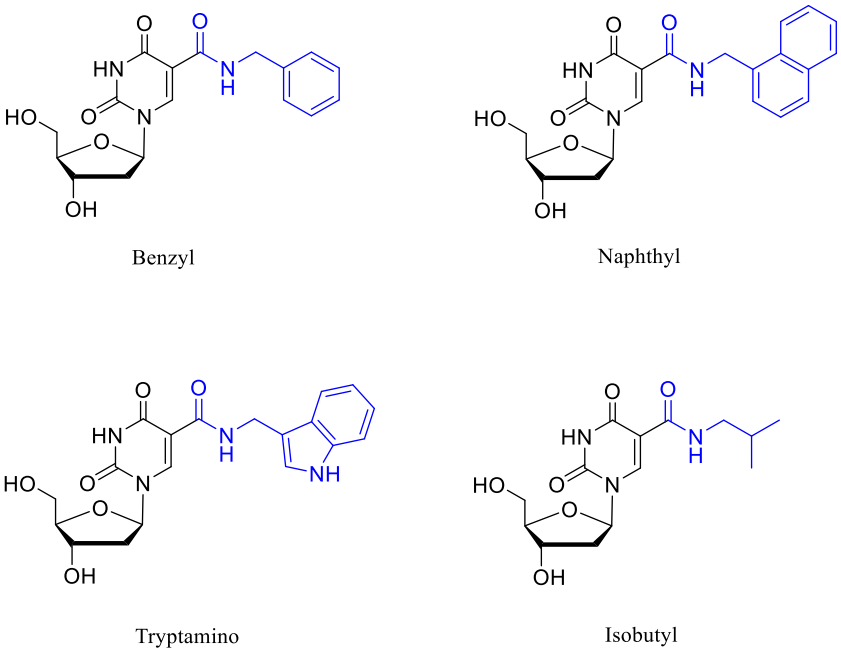


Figure 1.18. Modifications in SOMAmers. The modifications are introduced at the C5 position of dUTP. 5-benzylaminocarbonyl-dU, 5-naphthylmethylaminocarbonyl-dU, 5-tryptaminomethylcarbonyl-dU and 5-isobutylaminocarbonyl-dU.

Janjic's group studied SOMAmers with amino-acid-like modifications on both dCTP and dUTP to expand the physiochemical properties and diversity of their aptamers. Their rationale design is to mimic amino acid side chains that are overrepresented in complementarity determining regions (CDRs) of antibodies such as tryptophan and tyrosine.[185] By adding a second modification to achieve specific pairwise combinations, there was an improvement in binding affinity and specificity of the resultant aptamers, compared to those with only a single modification.[187] X-ray crystal structure analysis of SOMAmers showed that the modifications participate in secondary structure folding and present hydrophobic binding surfaces to their targets, unlike traditional protein binding aptamers which mainly exhibit polar interactions.[185]

However, if the modifications are too bulky, the modified nucleotides are not accepted by the polymerases. To overcome this limitation, Tolle *et al.* introduced a new

technique whereby the chemical modification is introduced via a click reaction after each enzymatic replication using a mixture of dATP, dGTP, dCTP and 5-ethynyl-dUTP (Figure 1.19). A broad range of functional groups can be attached using this click reaction as 5-ethynyl-dUTP is well tolerated by polymerases.[93,188]

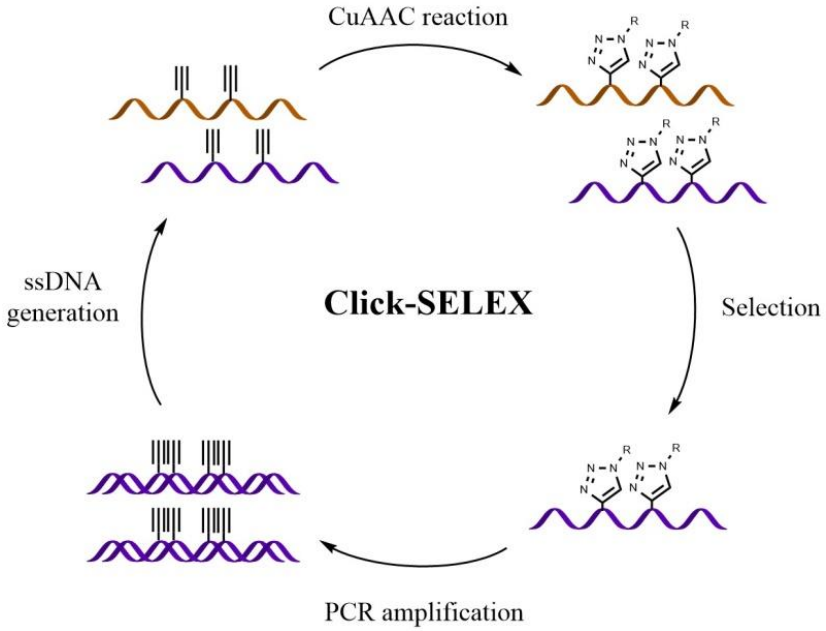


Figure 1.19. Schematic figure of click-SELEX.[188]

Recently, researchers tried to expand the genetic alphabet beyond the natural Watson-Crick base pairs to develop nucleic acids with additional functionality. Using a 'similar shape complementarity' approach, Hirao's group introduced an artificial third base pair (the **Ds**:**Px** pair) that is functional during DNA replication (Figure 1.20 **A**). It was proved that this base pairing was retained during 100 PCR cycles,[189] and this promising result enabled the selection of aptamers to VEGF-165 and INF-γ with very high affinity. The drawback was that the position of the unnatural base **Ds** had to be predetermined in the library to facilitate cloning and sequencing.[190,191] Thereby, unnatural bases cannot be used in the randomised positions during SELEX.

In a similar manner, the pyrimidine analogue **Z** and complementary purine analogue **P** pair, another non-cognate base-pair present rearranged hydrogen bonding but still having a Watson-Crick-like geometry (Figure 1.20 **B**).[192] This unique base pair was used in SELEX experiments using an expanded ACGTZP sequence space for the generation of aptamers that can selectively recognize HepG2 liver cancer cells with higher affinity compared to natural DNA aptamers.[193]

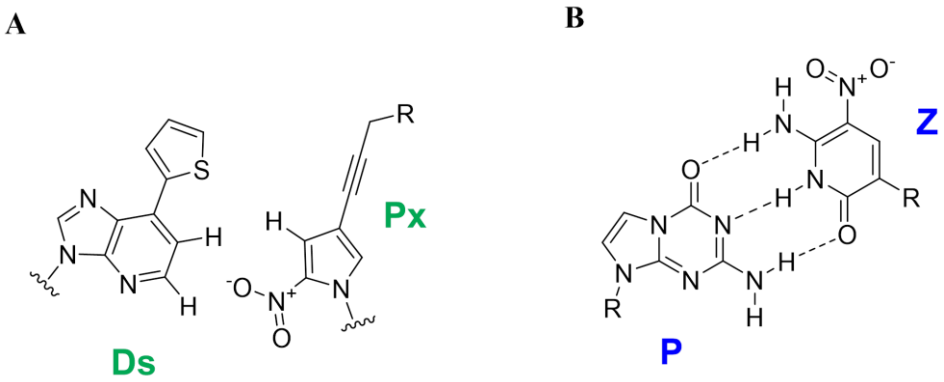


Figure 1.20. Non-cognate third base pair. **A**) Ds:Px base pairing[189] and **B**) P:Z base pairing[192].

1.3.3 Backbone modifications

Several phosphate backbone modifications have been investigated in order to improve aptamer stability *in vivo*. These modifications typically maintaining DNA and RNA secondary structure and also, provide nuclease resistance and greater cell uptake.[194,195]

The non-bridging oxygen atoms of the phosphodiester backbones of DNA and RNA aptamers can be substituted with one or two sulfur atoms to make phosphorothioate (PS) or phosphorodiothioate (PDS) linkages, respectively (Figure 1.21).[196–198] Phosphorothioate is the most common backbone modification used in the general area of oligonucleotide therapeutics.[199–202]

The PS linkage introduces chirality to internucleotide linkages and there are two stereoisomers (S_p and R_p, Figure 1.21). The individual PS stereoisomers possess different biophysical and biological properties; the S_p isomer is highly resistant to nuclease-mediated cleavage whereas the R_p isomer has better duplex stabilising (T_m) and RNase H activating profiles.[203,204] It is common to synthesise PS modified oligonucleotides as a diastereomeric mixture, but they also can be synthesised in a stereocontrolled manner.[205] Kang *et al.* have discovered a transforming growth factor-β1 (TGF-β1) binding DNA aptamer (K_D value of 90 nM) that was selected using a DNA library containing R_p phosphorothioate linkages at the 5′-side of adenosines and cytosines.[206] They also found an RNA aptamer from an RNA library containing R_p phosphorothioate linkages on the 5′-side of adenosines. This aptamer binds to the Venezuelan equine encephalitis virus capsid protein with K_D of 7 nM.[207] Another study performed by King *et al.* showed that PS modifications at the 5′-side of adenosines was used to select a high binding affinity DNA aptamer to human NF-kappa B proteins RelA(p65) and p50 (K_D values of 4.8 nM and 0.5 nM, respectively).[199] For post-SELEX modifications, PS groups, were substituted into the different internucleotide sites to study the thermodynamic stability of the phosphodiester linkage using thrombin binding aptamers.[147,208] Abeydeera *et al.* reported that the substitution of the PDS internucleotide linkage at a single position in thrombin binding RNA aptamers improves binding affinity by approximately 1,000-fold.[197]

PS groups can be placed at the 5'- or 3'- end of the oligonucleotide. The 3'-side of the internal pyrimidines known to be the most efficient in improving exonuclease resistance.[209] However, PS has been reported to be "stickier" towards proteins than the normal phosphodiester backbone, and this may cause non-specific binding in aptamer development.[210]

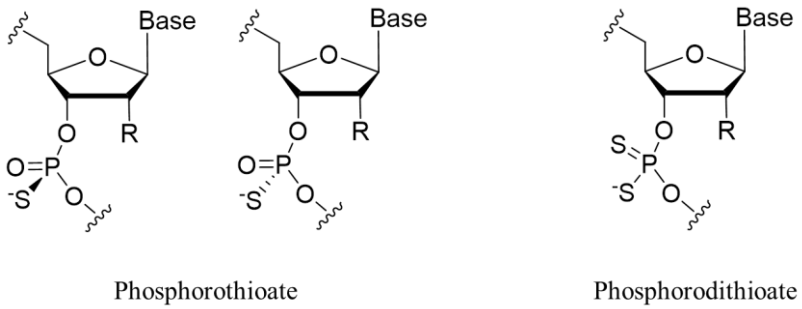


Figure 1.21. Structures of phosphorothioate (PS) (R isomer on the left, S isomer on the right) and phosphorodithioate (PDS) (where R = H or OH).

In 2013, Varizhuk *et al*. used triazole linkages for replacement of the phosphodiester linkage in their DNA aptamer generation for thrombin. They showed that the aptamer containing triazole linkages had increased stability to nuclease hydrolysis (Figure 1.22 **B**).[211] In addition, boranophosphate modifications were introduced on the α-phosphorus of GTP and UTP and they were used to select modified aptamers that bind to ATP.[212]

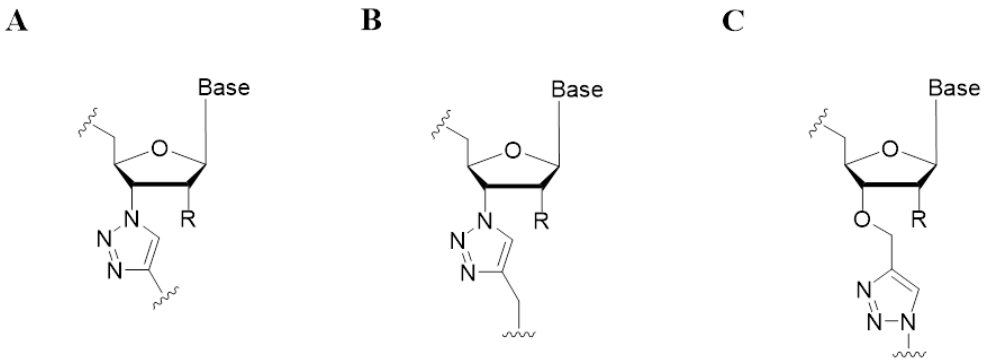


Figure 1.22. **A**, **B** and **C** represent three different examples of triazole linkages (where R = H or OH)[213] and **B** was used for thombin binding aptamer.[211]

1.4 Aptamer sequencing

Aptamer candidates are enriched after several rounds of selection and identified by sequencing of the DNA library. In particular, the frequencies of sequences in the last few cycles of SELEX must be carefully analysed.

1.4.1 Sanger sequencing

Aptamers candidates were traditionally analysed by cloning and sequencing the library from the final SELEX round. This library is cloned into a plasmid and transformed into competent bacteria. After amplification of the bacteria, the plasmids are purified and sequenced by the chain terminator method. This method was developed by Nobel laureate Frederic Sanger and is referred to as Sanger sequencing.[214] Usually only 30–100 clones are analysed due to the relatively low throughput of the sequencing method even in its automated form. Therefore, in order to identify aptamers using Sanger sequencing, it is essential to have a highly enriched population of sequences that bind to the target.

1.4.2 Next-generation sequencing (NGS)

Next-generation sequencing (NGS) was developed with the aim of rapidly decoding the whole genome. As a result, up to 100 million sequences can be sequenced in one experiment. NGS can be used for any type of nucleic acid; it has been used for PCR product sequencing, biomarker discovery,[215–217] and nucleic acid aptamer selection/identification.[218–220] This technique provides higher coverage of the sequences in a SELEX library in a relatively shorter time compared to Sanger sequencing,[218] and therefore, candidate sequences from all rounds of SELEX cycles can be analysed. The first rounds of SELEX may show enrichment of the library, which can be monitored

through to the final round. This allows the evolution of the aptamer to be tracked in response to varying selection conditions. A list of most frequently found aptamer in each round can be used to determine the best candidates, and from these sub-structures sequence motifs may be identified. NGS can also track the distribution of nucleotide bases at all positions in the random region of the library and thereby, the critical positions of certain bases are identified. The success of aptamer identification and optimisation has been improved by NGS.[221–223]

The most popular NGS platform was developed by Illumina. The basic principle behind their system is that a fluorescently labelled deoxyribonucleotide triphosphate with a reversible terminator is incorporated into a DNA template by a DNA polymerase during sequential cycles of DNA synthesis. During each cycle, the nucleotides are identified by fluorophore excitation when each new nucleotide is incorporated.[218,221–223]

The Illumina methods consist of four steps.

1. Library preparation (Figure 1.23 **A**)

   The nucleic acid sample is converted to dsDNA which is then ligated to 5′ and 3′ adapters.

2. Cluster generation (Figure 1.23 **B**)

   The library is then loaded into a flow cell and each fragment is hybridised to surface-bound oligonucleotides that are complementary to the adapters. Then, they are amplified into clonal clusters through bridge amplification. Sequencing starts when the cluster generation is finished.

3. Sequencing (Figure 1.23 **C**)

   Now, all four dNTPs which are bound with reversible terminator are present during each sequencing cycle and each nucleotide is fluorescently labelled with a different colour. These four nucleotides then undergo natural competition to be incorporated into the complementary copy of the immobilised DNA template. After each synthesis, the flow cell is imaged after each nucleotide addition and the nature of the emission from each cluster is recorded. The cycle is repeated "n" times until "n" bases are read through. The advantage of Illumina sequencing is that multiple strands sequencing is possible in parallel on a short timescale.[224,225]

4. Data analysis

   Using the imaged data, sequences are aligned to a reference genome. Many other analysis methods are available.

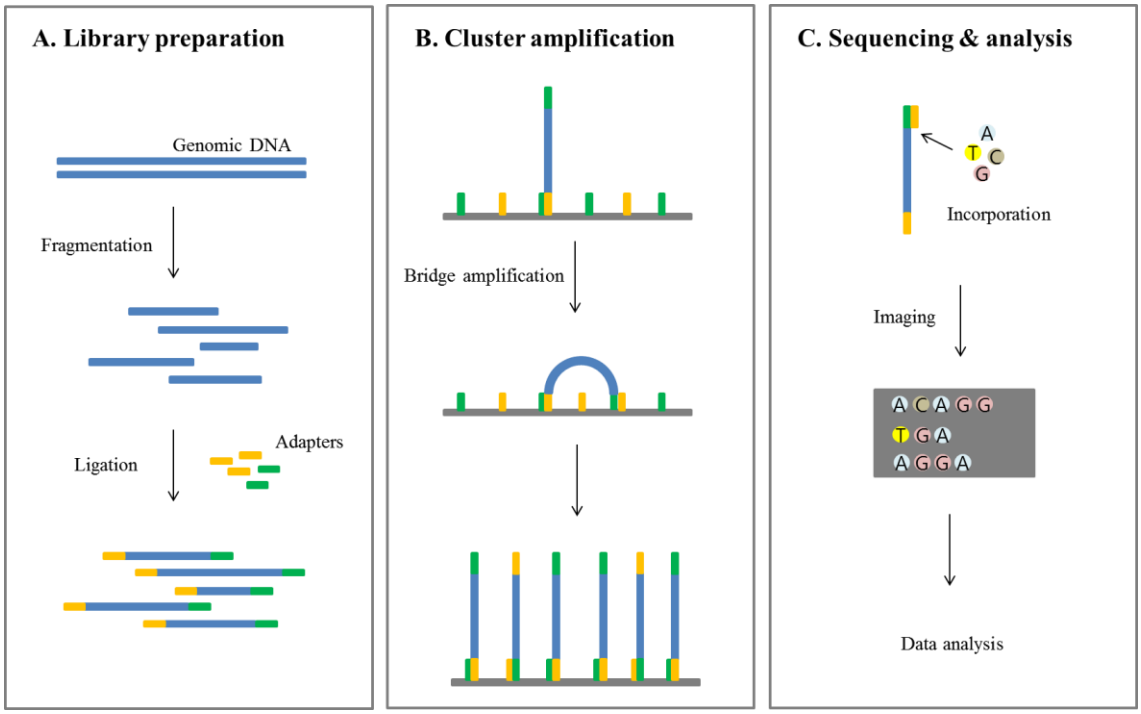


Figure 1.23. Illumina's next-generation sequencing overview.[226] The figure was modified from https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf.

1.5 Synthesis of modified oligonucleotides

In order to perform a binding assay on aptamer candidates, the aptamers must be resynthesised. Modified nucleic acids can be made by automated solid phase oligonucleotide synthesis using phosphoramidite monomers or enzymatic synthesis using modified triphosphates as substrates. In some cases, a post-synthesis labelling step is required to attach modifications.

1.5.1 Solid phase synthesis of oligonucleotides

Automated solid-phase oligonucleotide synthesis using phosphoramidite monomers developed by Caruthers and Beaucage is the method of choice for the synthesis of natural or modified oligonucleotides.[227,228] At the beginning of the synthesis, the 5′-DMT protecting group (DMT = 4,4′-dimethoxytrityl) of the support bound 3′-nucleoside is removed by treatment with trichloroacetic acid (TCA). The next nucleotide in the desired sequence is then coupled as a phosphoramidite monomer, using an activator (e.g. tetrazole) (Figure 1.24, step 1). Then, any unreacted nucleoside is removed from the synthetic cycle by acetylation using acetic anhydride, *N*-methylimidazole and pyridine (Figure 1.24, step 2). The oxidation of the phosphite-triester (P^{III}) formed during the phosphoramidite coupling reaction is converted to a stable (P^{V}) form using iodine in the presence of water and pyridine (Figure 1.24, step 3). This step is required before the next detritylation step due to the acid liability of P^{III}. The cyclic is then repeated until the correct oligonucleotide sequence is assembled. When the synthesis is complete, the 5′-DMT group of the last nucleotide is removed using the same detritylating agent (Figure 1.24, step 4). The oligonucleotides are then cleaved from the solid support and all protecting groups are removed, typically using concentrated ammonia in a single step (Figure 1.24). The advantages of solid phase

synthesis are (i) large excess of solution phase reagents are used to drive reactions to completion, (ii) preparation of oligonucleotides is fast due to the high concentrations and molar excess of the reagents, (iii) excess reagents are easily removed and (iv) modified nucleotides can be introduced at specified positions using modified phosphoramidites monomers. However, high chemical stability of any modified phosphoramidites under the solid phase synthesis conditions (acidic, alkaline and oxidative) is crucial.

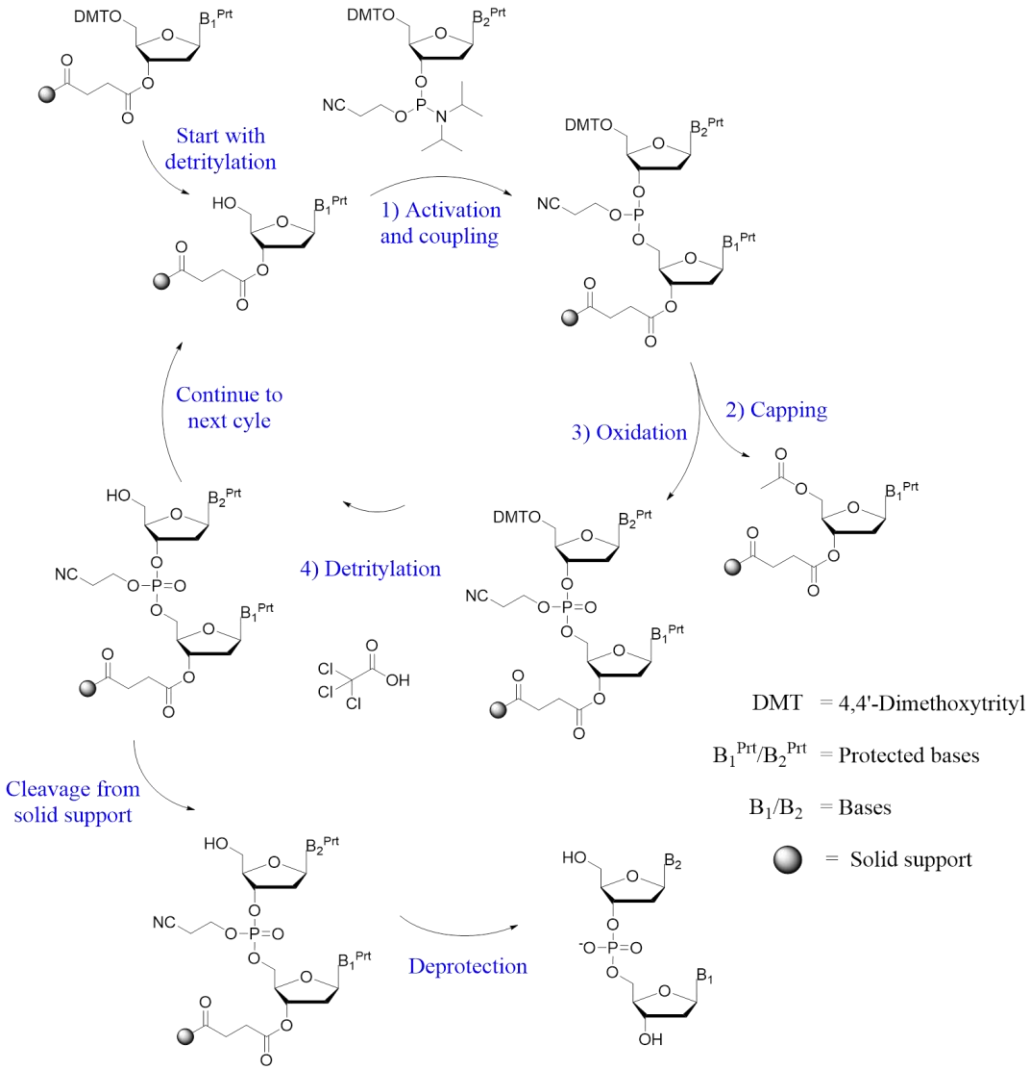


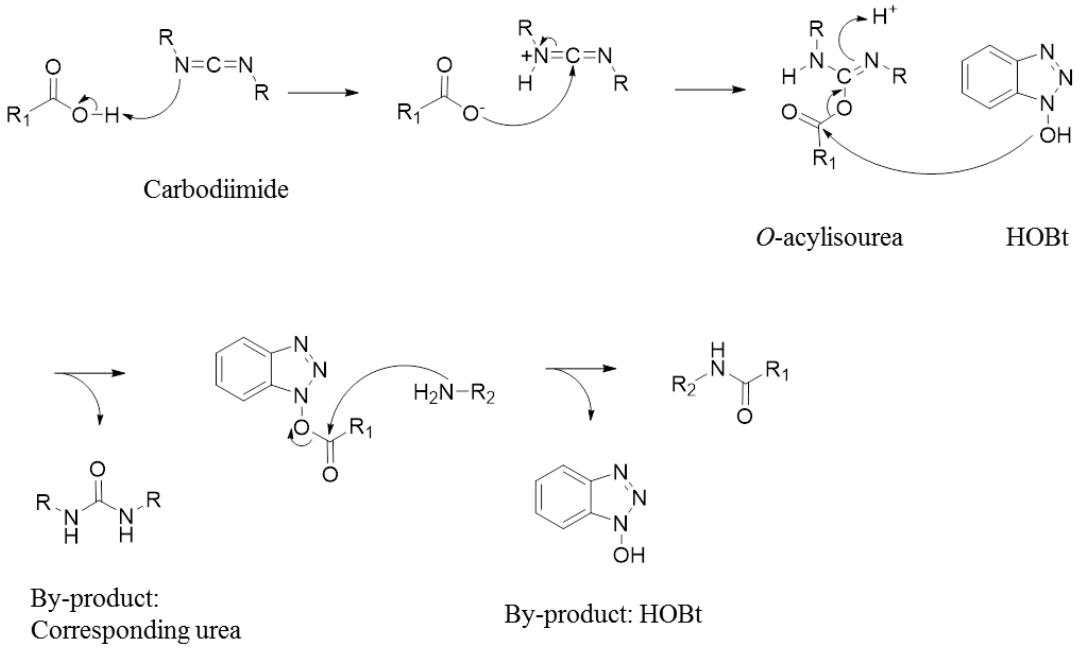
Figure 1.24. The synthesis cycle in automated solid-phase phosphoramidite oligonucleotide synthesis.

1.5.2 Post-synthesis chemical labelling reactions

Modifications can be introduced to oligonucleotides on the solid support in the form of phosphoramidite monomers. However, the synthesis of phosphoramidite monomers containing modifications is sometimes time-consuming and some modifications are not compatible with reagents used in oligonucleotide synthesis or deprotection conditions or both. Alternatively, there are various methods for attaching functional groups to an oligonucleotide in solution.[229] The most common is through amide coupling, where an amino group on the oligonucleotide is reacted with an activated carboxylic acid derivative of the modification of interest. The limitation of this method is that the nucleophilicity of amino groups is pH dependent, with greater nucleophilicity at higher pH, but activated carboxylic acids are decomposed rapidly at higher pH. Therefore, a compromise between the two (pH between 7.2 and 9.0) is needed.

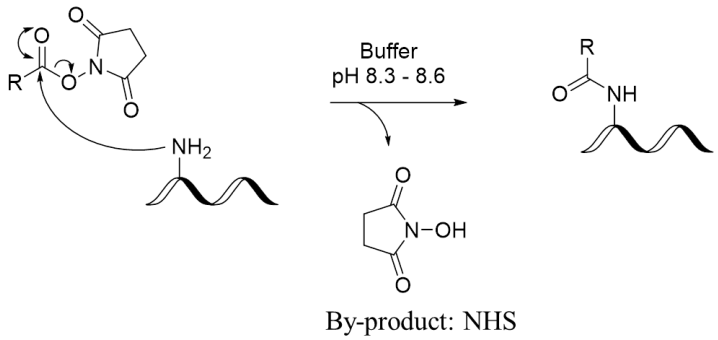
Carbodiimides react with carboxylic acids to give *O*-acylisourea, which are reactive towards primary amines, resulting in amide bond formation (Scheme 1.1). The most commonly used carbodiimide for oligonucleotide labelling reaction is 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) (Scheme 1.1 **A**) as it is soluble in water. Other carbodiimides, dicyclohexylcarbodiimide (DCC) (Scheme 1.1 **B**) and *N*,*N*′-diisopropylcarbodiimide (DIC) (Scheme 1.1 **C**) can be used, but they are insoluble in water and therefore, organic solvents are required for their reactions. In order to form an *O*-acylisourea, acidic pH (3.5–4.5) is desirable, but amide formation between the *O*-acylisourea requires higher pH to suppress protonation of the amine. In general, pH 5.0–6.5 is used in the presence of *N*-hydroxysuccinimide (NHS) as a catalyst.[230] In order to reduce epimerisation in peptide synthesis, 1-hydroxy-1*H*-benzotriazole (HOBt) was introduced by Koenig and Geiger, allowing the reaction to proceed via a HOBt ester

(Scheme 1.1).[231] HOBt also accelerates the coupling process and therefore the formation of unreactive N-acylurea can be reduced.[232,233]



Scheme 1.1. Mechanism of activation by 1-hydroxy-1*H*-benzotriazole (HOBt) when used in conjunction with carbodiimides. Examples of carbodiimides are shown in a box. **A**) 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC), **B**) dicyclohexylcarbodiimide (DCC) and **C**) *N*,*N*′-diisoproptlcarbodiimide (DIC).

The NHS ester of the required modification can be synthesised and used to react with the DNA strand bearing a primary amino group (Scheme 1.2). An optimal pH between 8.2 and 8.6 and normally a large excess of the NHS ester is used to make the reaction go to completion.



Scheme 1.2. General methods for the DNA amino-NHS ester coupling reaction.

In addition, several ‘‘salts’’ are often associated as coupling reagents based on 1*H*-benzotriazoles, including uronium/aminium (e.g. HBTU, HATU) (Figure 1.25) and phosphonium (e.g. BOP, PyBOP) (Figure 1.26) salts in the presence of a tertiary base, *N*,*N*-diisopropylethylamine (DIEA) or *N*-methylmorpholine (NMM) (Scheme 1.3).[234] Uronium/aminium salt reagents react with carboxylic acids to form OAt/OBt active esters, which then react with amines. However, they also react with the amine to form a guanidinium by-product (Scheme 1.4) and therefore, the order of addition and timing is important to avoid by-product formation. Phosphonium salts have the advantage of not yielding guanidinium by-products during the reaction of the coupling reagent with amines.[232]

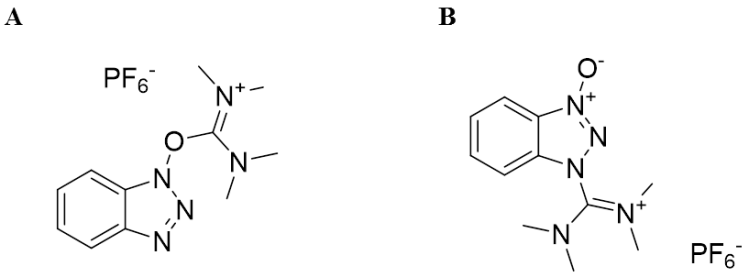


Figure 1.25. Uronium (**A**) and aminium (**B**) isomers of HBTU.

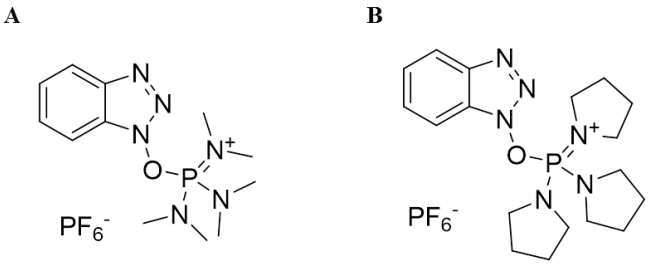
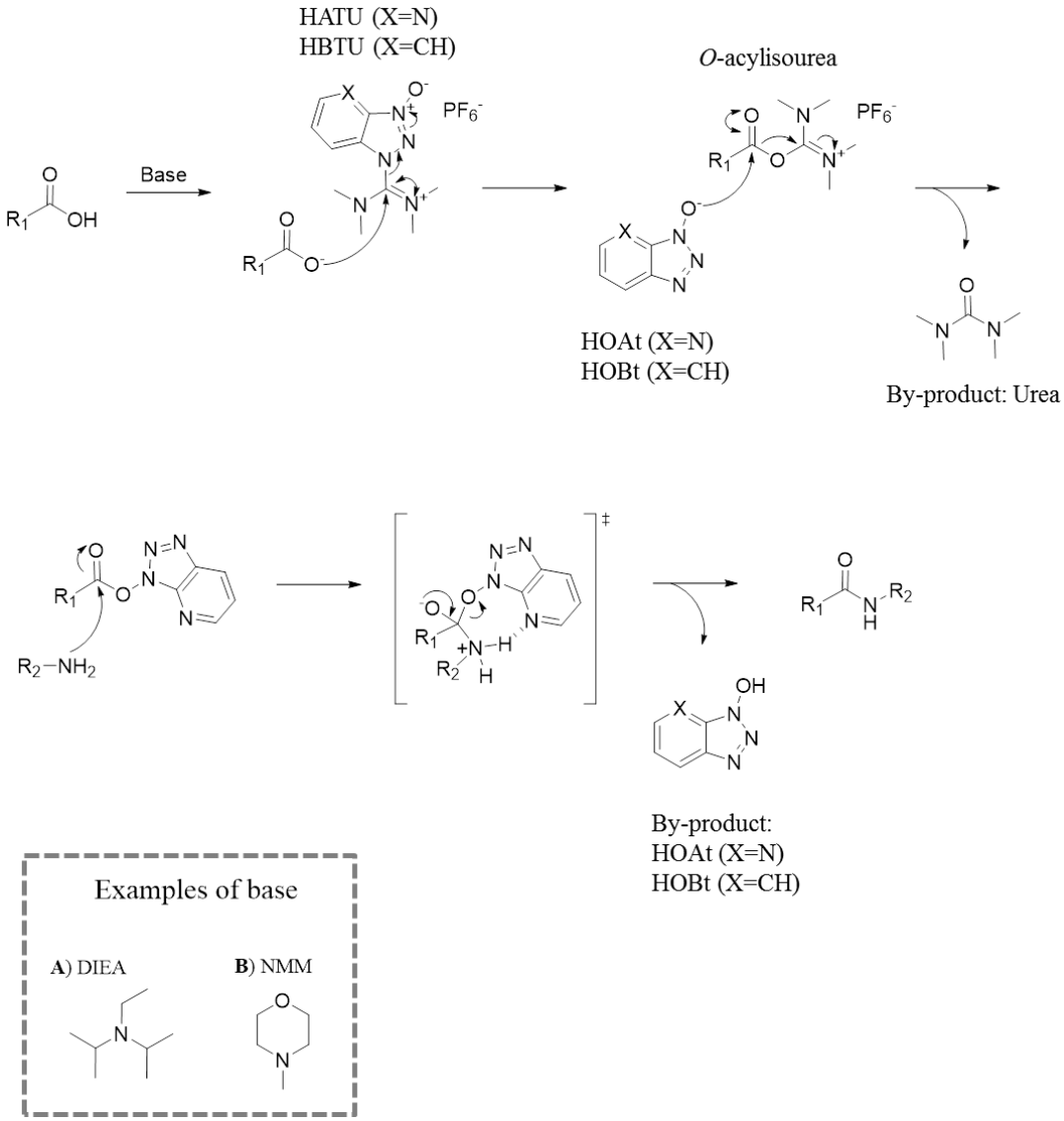
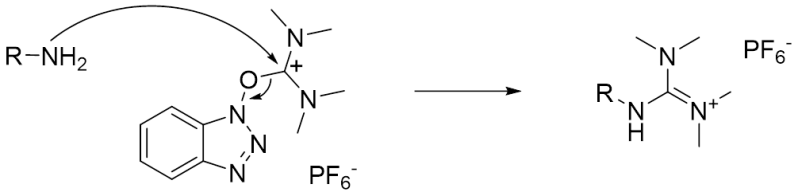


Figure 1.26. Examples of phosphonium type coupling reagents; BOP (**A**) and PyBOP (**B**).



Scheme 1.3. Activation mechanism using aminium type reagents (HATU or HBTU). Examples of base are shown in a box. **A**) *N*,*N*-diisopropylethylamine (DIEA) and **B**) *N*-methylmorpholine (NMM).



Scheme 1.4. Guanidinium formation with uronium/aminium type coupling reagents.

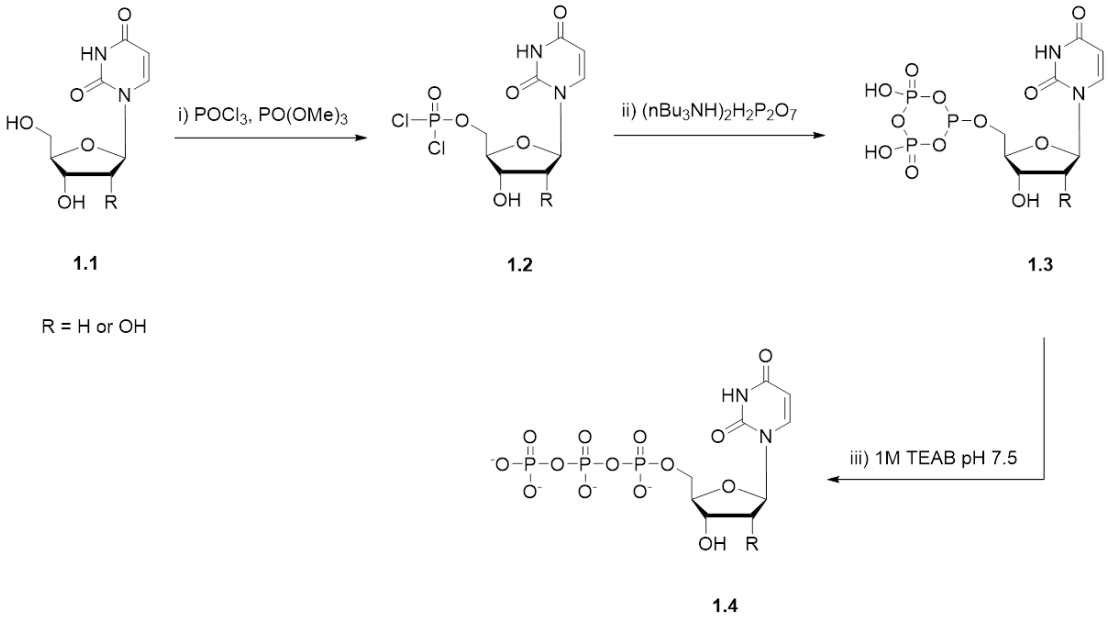
1.5.3 Synthesis of nucleoside triphosphates

Nucleoside triphosphates can be easily modified chemically to generate a versatile toolbox for functionalised nucleic acids synthesis. Although there are several synthesis routes to make modified triphosphates, synthesis of triphosphates from nucleoside precursors in high yield is still challenging.[177,235]

Two of the most popular methods for nucleoside 5′-triphosphate synthesis are the Yoshikawa and Ludwig-Eckstein approaches. The Yoshikawa method starts with unprotected 5′-OH nucleoside precursor **1.1** and uses phosphorus oxychloride ($POCl_3$) as a phosphorylating agent to yield the highly reactive phosphorodichlorate intermediate **1.2** (Scheme 1.5).[236] Subsequently, the intermediate reacts with pyrophosphate to generate the cyclic triphosphate **1.3** which is then hydrolysed to produce the 5′-triphosphate **1.4**. The use of $POCl_3$ and pyrophosphate in excess results in the formation of the triphosphate as the major product.[237]

The merits of Yoshikawa method are that the procedure is a one-pot reaction with short reaction times. Protecting groups are unnecessary if trialkyl phosphate solvents are used, which can help direct selectivity for the 5′-regioisomer. However, the strong electrophilic phosphorylating reagent, $POCl_3$ can add to the 3′-OH (for deoxyribose) and both 3′- and 2′-OH (for ribose) of the sugar moieties and thereby, generates 3′-

monophosphate, 2′-monophosphate and 3′, 5′-diphosphate by-products **1.5**–**1.9** (Figure 1.27).[238]



Scheme 1.5. Yoshikawa method for uridine 5′-triphosphate synthesis.

This method can give very low yields, and consequently, modifications to the original procedure have been explored. To increase the phosphorylation reaction rate, a Proton Sponge can be added during the first step of the reaction.[239] Reduction of reaction temperature from 0 to -15 ℃ and increased reaction time in the second step was shown to improve the yield of triphosphate over monophosphate.[240]

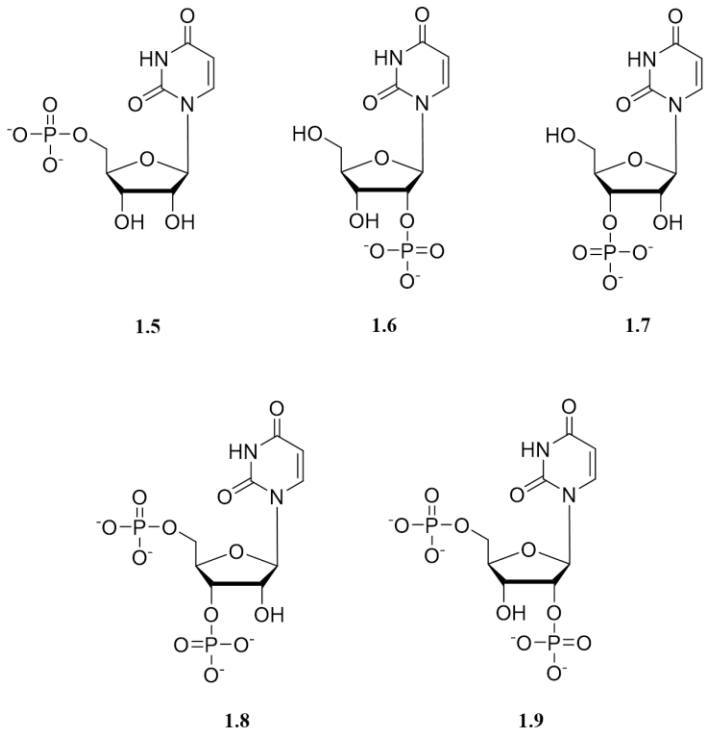
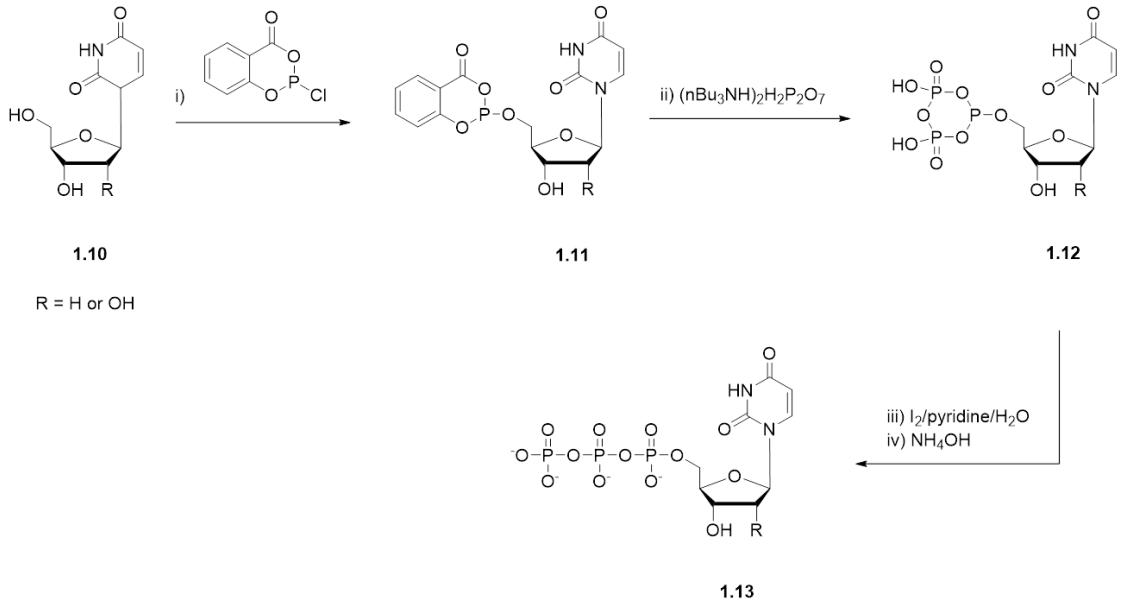


Figure 1.27. By-products which can be formed during triphosphate synthesis.

Alternatively, the Ludwig-Eckstein method uses salicyl phosphorochlorite which actively and selectively reacts with the 5′-hydroxyl of the nucleoside precursor **1.10** to generate the intermediate **1.11** (Scheme 1.6).[241] Salicylic acid is then displaced by pyrophosphate to form the cyclic triphosphate **1.12** which in turn is hydrolysed by iodine solution (to oxidise the P^{III} to P^{V}) to produce the 5′-triphosphate **1.13**. This method has also been widely used in base-modified and sugar-modified triphosphates synthesis.[242,243] The advantage of using the Ludwig-Eckstein method is that bulky salicyl phosphorochlorite helps to improve selectivity for the primary 5′-OH groups over the secondary 2′- and 3′-OH groups. Consequently, fewer by-products (regioisomers, 3′-monophosphate, 2′-monophosphate and 3′, 5′-diphosphates, **1.5**–**1.9** in Figure 1.27) are produced than when using Yoshikawa methodology.[244,245]



Scheme 1.6. Ludwig-Eckstein method for uridine 5′-triphosphate synthesis.

1.5.4 Enzymatic synthesis of oligonucleotides

DNA is assembled from nucleoside triphosphates by DNA polymerases and then transcribed to ribonucleic acid by RNA polymerases in living organisms. One DNA strand acts as a template to produce a new strand DNA or RNA that is complementary to the original strand. Enzymatic synthesis using polymerases is an alternative to chemical synthesis. The advantages of enzymatic synthesis of the oligonucleotide are that long oligonucleotides containing high levels of modifications can be synthesised. However, this is only achievable if the modified nucleotides are accepted as substrates by the polymerase. Site-specific modification and a large scale synthesis are not possible using enzymatic synthesis.

1.5.4.1 DNA polymerases

All DNA polymerases catalyse nucleotide addition to a growing primer strand and excise incorrectly incorporated dNTPs via their intrinsic 3′–5′ exonuclease activity.[246–252] There are two metal ions binding sites (Figure 1.28 **A**). Metal ion in the A site lowers the pK_a of the terminal 3′-OH of the primer and brings the primer and the α-phosphorus atom (P_α) of the incoming dNTP closer. Then, a nucleophilic attack on the P_α of the incoming dNTP is facilitated, leading to the formation of a phosphodiester bond. Metal ion in the B site, however, interacts with the α, β and γ non-bridging phosphate oxygens of the incoming dNTP, helping to neutralise the developing negative charge as the ternary complex approaches the transition state in the nucleotidyl transfer reaction and assists in the departure of the pyrophosphate group.[246,247] The crystal structures of DNA polymerases with metal ions and an incoming dNTP bound in their active sites have been widely studied by several groups.[247,253,254] Xia *et al*. used a family B DNA polymerase (RB69 DNA polymerase) to study the specific structural features of the catalytic metal-ion coordination geometry during the incorporation of dNTP in ternary complexes.[247,253] Out of four mutations made adjacent to the nucleotide binding pocket, the triple mutant was used to capture a ternary complex. This mutant was used for ease of crystallisation and proved to exhibit the same kinetic characteristics for incorporation of complementary dNMPs as the wild-type RB69 DNA polymerase. Mn^{2+} and dUpNPP, a non-hydrolysable dNTP analogue was used to capture the closed pre-insertion ternary complex. The structures revealed that Mn^{2+} present in the B site adopted an octahedral geometry (Figure 1.28 **B**). Mn^{2+} is coordinated by the triphosphate tail of the incoming dUpNpp along with the carboxylate side chains of Aps_{411} (D411) and Asp_{623} (D623) residues and the carbonyl backbone oxygen of Leu_{412} (L412) (Figure 1.28 **B**). Mn^{2+} in the A site adopted a distorted octahedral geometry (Figure 1.28 **B**). It is coordinated

with the 3′-OH of the extending primer (*pt*O3′) along with two carboxylate side chains of Asp_{411} (D411) and Asp_{623} (D623) residues and the oxygen of the P_α of the dUpNpp. The distance between the 3′-OH of the extending primer and the P_α of the dUpNpp is close enough to form a phosphodiester bond and this distance is made closer by Mn^{2+} in the A site when the reaction approaches the transition state (Figure 1.28 **C**). This relocation of the 3′-OH group creates ideal octahedron geometry during the transition state (Figure 1.28 **C** and **D**). This helps in the formation of a covalent bond between them. Mg^{2+} showed similar results when it was substituted for Mn^{2+} meaning that Mg^{2+} and Mn^{2+} share similar coordination geometries in the RB69 pol ternary complexes.[247]

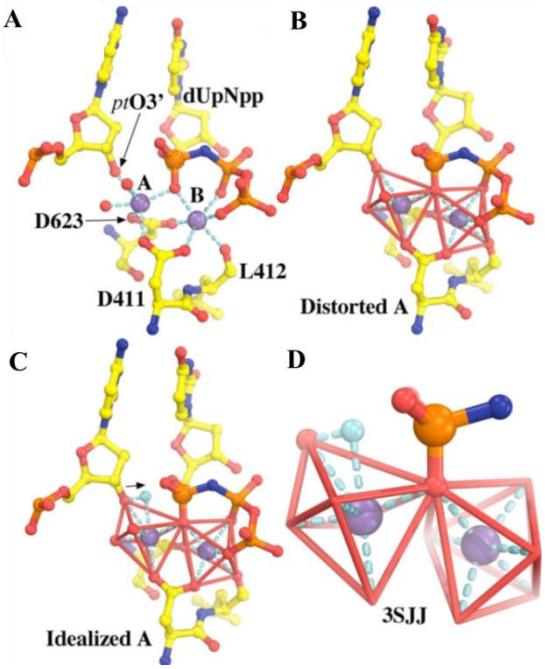


Figure 1.28 The structural representatation of the triple mutant (tm) RB69 DNA polymerase ternary complex (PDB code 3SJJ).[246] **A**) The active site when the nucleotide, dUpNpp is bound in the presence of Mn^{2+} at two metal binding sites. **B**) Mn^{2+} in site B shows perfect octahedral geometry whereas a distorted octahedral geometry is observed with Mn^{2+} in site A. **C**) The possible relocation of the 3′-OH group to make an ideal octahedron geometry during the transition state. **D**) Zoom-in of the two Mn^{2+} ions, the 3′-OH group and the α-phosphate of dUpNpp in a complex form. Picture reuse was permitted by the Journal of Biological Chemistry (see appendix 8.1).

DNA polymerases appear to share a common overall structure. They generally have the shape of a right hand in which the finger, palm and thumb sub-domains form the DNA-binding crevice. The role of the palm domain is to catalyse the transfer reaction of phosphate groups whereas the finger domain interacts with the incoming nucleoside triphosphate and bases on the DNA template. The thumb domain is responsible for the processivity, translocation and positioning of the DNA duplex (Figure 1.29). The structure and sequences of fingers and thumb domains are different among polymerases, but the palm domain has a common structural core which contains the conserved two aspartate residues. However, DNA polymerase β has no architectural relationship with other polymerases. Polymerases can be grouped into more than five different families based on their amino acid sequences and crystal structures.[255] DNA polymerase family A includes the large fragment (Klenow fragments) of *Escherichia coli* and a *Bacillus* DNA polymerase I, *Thermus aquaticus* DNA polymerase, and the T7 RNA and DNA polymerases. All eukaryotic replicating DNA polymerases from phage T4 and RB69 are grouped into DNA polymerase family B.

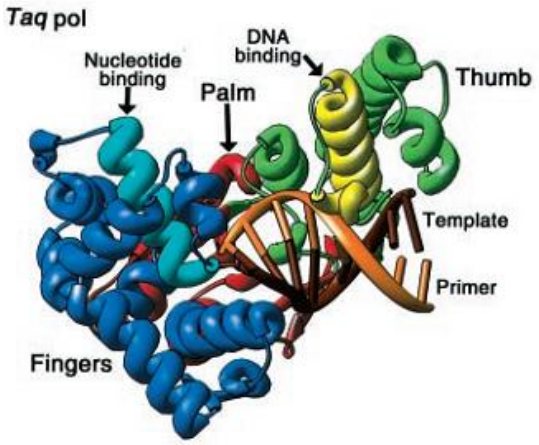


Figure 1.29. The overall structure of Taq polymerase with DNA template and primer.[256] Picture reuse was permitted by the Elsevier (see appendix 8.2)

It has been established that DNA polymerases of family B are more efficient enzymes in incorporating modified nucleotides than DNA polymerases of family A.[257,258] The reason for the preferential acceptance of modified substrates by family B DNA polymerases has not yet been elucidated. Hottin *et al*. hypothesised that the dsDNA in the KOD DNA polymerase adopts a B-form DNA conformation showing the ideal puckering for B-DNA (C2′-endo). In contrast, family A DNA polymerases were observed to have A-form DNA near the insertion site. Compare to A-form DNA, B-form DNA is more elongated with a wide major groove and therefore, DNA duplex conformation in family B DNA polymerases might favour the acceptance of modified substrates.[176]

In order to tolerate unnatural nucleotides more efficiently during enzymatic synthesis, DNA polymerases have been engineered.[259] KOD DNA polymerase is a genetically engineered family B polymerase, originally isolated from the hyperthermophilic archaeon, *Pyrococcus kodakaraensis*.[260] The mutation of a "unique loop" in the exonuclease domain of KOD DNA polymerase alters 3′–5′ exonuclease activity.[261] The kinetic and enzymatic capability of engineered KOD DNA polymerase was assessed using various modified thymidines as building blocks and the successive incorporation of bridged nucleotides with a 2′-ONHCH$_2$CH$_2$-4′ linkage was achieved.[262] Also, variants of Vent DNA polymerase, originating from the hyperthermophilic archaeon, *Thermococcus litoralis* was studied and used for synthesis and enzymatic incorporation of modified triphosphates.[263–265] 9°N DNA polymerase was also evolved to reduce 3′–5′ exonuclease activity enabling modified nucleotides to be accepted as substrates during enzymatic reactions.[266]

Fa *et al.* made mutations to residues proximal to the triphosphate binding site of Stoffel fragment (Sf) DNA polymerase. The engineered enzyme increased rates for the

incorporation of each 2′-OMe modified triphosphate (~10^4-fold compared to wild-type enzyme) with high fidelity.[267] Ong *et al.* engineered the active site of Taq DNA polymerase to incorporate both dNTPs and rNTPs or 2′-modified NTPs. This mutated DNA polymerase successfully incorporates 2′-N_3-dATP and 2′-F-dUTP.[268] Likewise, engineered DNA polymerase enables mixed RNA–DNA PCR amplification using a single enzyme demonstrating DNA polymerase, RNA polymerase and reverse transcriptase activity.[268] The Klenow fragment DNA polymerase I, bearing the I709F mutation also more readily incorporates ribonucleotides than its wild-type.[269,270] DNA polymerases such as SequiTherm, Tht, AmpliTaq FS exo-, Pfu exo-, Vent exo-, and deep Vent exo- as well as the human DNA polymerases α and γ, are also capable of accepting 2′-F-NTP.[271,272]

1.5.4.2 RNA polymerases

Most prokaryotic and eukaryotic DNA-dependent RNA polymerases consist of multi subunits to translocate DNA template and generate a complementary RNA. This complicated mechanism is called transcription.[273,274] Transcription initiates when RNA polymerase recognises and binds to a specific DNA promoter sequence (Figure 1.30 **A**).[275] Then, the dsDNA unwinds near the gene that is being transcribed. One of the two unwound DNA strands is used as a template during transcription and the RNA strand is then synthesised using incoming ribonucleoside triphosphates which are complementary to the bases in the DNA template strand. During this initial step, RNA polymerase repeatedly synthesises and releases short RNA fragments (8–10 nucleotides) and then, the enzyme undergoes conformational changes to form the more stable elongation complex and continues to its elongation step. During the elongation step, the enzyme adds a matching RNA nucleotide in the 3′ to 5′ direction (Figure 1.30 **B**).

Polymerisation lasts until the enzyme reaches the end of the template or encounters a terminator sequence (Figure 1.30 **C**).[276]

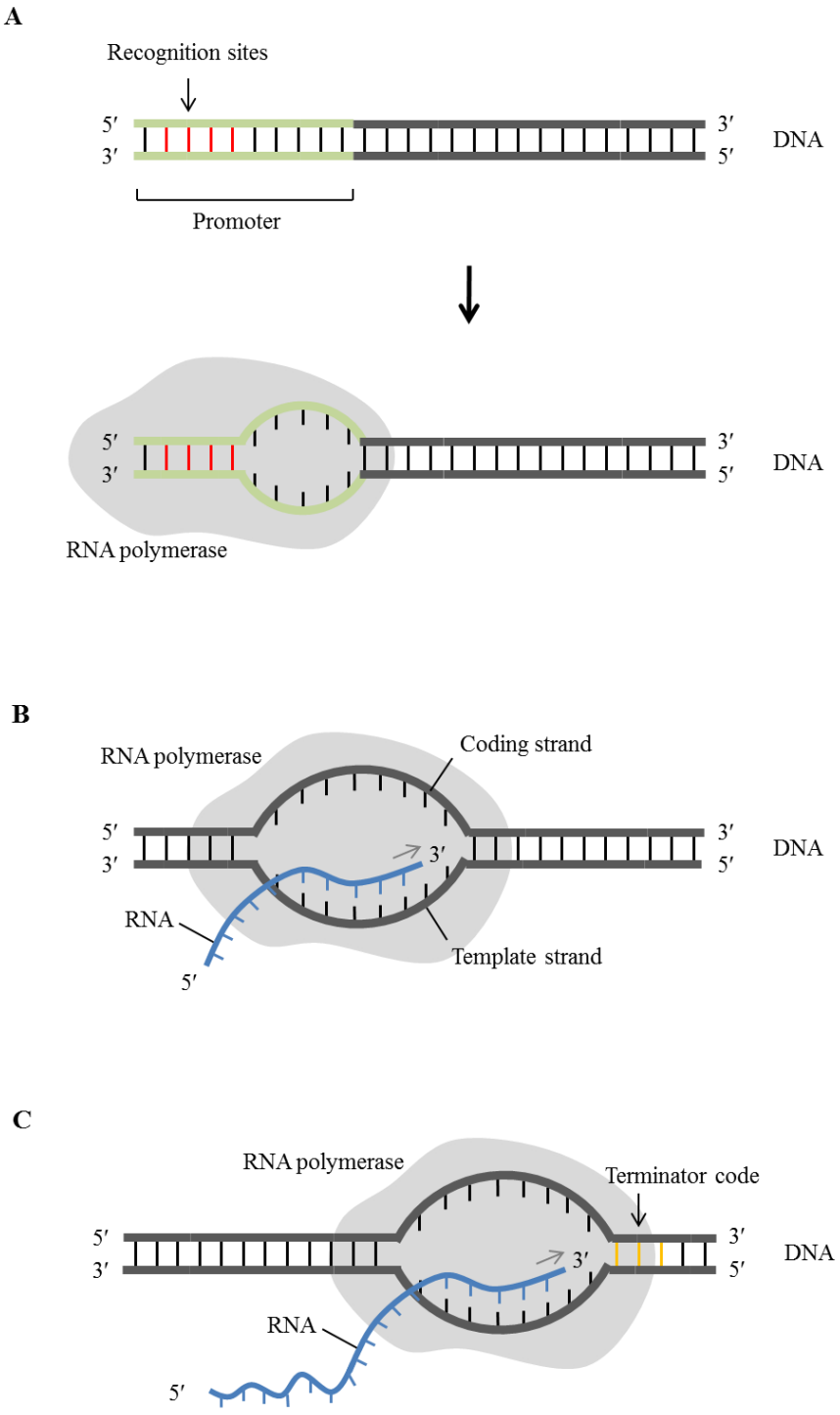


Figure 1.30. Stage of transcription by RNA polymerase. **A**) Initiation. **B**) Elongation. **C**) Termination.

In contrast to most known RNA polymerases, RNA polymerases evolved from bacteriophages T7, T3, SP6 and K11 are relatively simple enzymes, consisting of one subunit. They can carry out transcription in the absence of additional protein factors.[277–279] This family of RNA polymerases is also characterised by higher (compared to bacterial RNA polymerases) rates of RNA synthesis and specificity towards their promoters. Therefore, they are widely used for the synthesis of specific transcripts as well as for studying the transcription mechanism. Among them, bacteriophage T7 RNA polymerase (T7 RNAP) is the most extensively studied. Crystal structure studies of T7 RNAP suggested that there are significant similarities between T7 RNAP and some DNA polymerases. They all resemble a right hand where sub-domains consist of palm, thumb and fingers.[280,281]

Wild-type T7 RNAP was found to accept 2′-modified triphosphates (2′-F-NTP, and 2′-NH_2-UTP) as substrates during transcription.[156,282] T7 RNAP can also accept 2′-OMe-NTPs during transcription in combination with other deoxynucleotides.[283] With these encouraging results, directed evolution of T7 RNAP was studied to improve its tolerance towards the sugar modified substrates. For a T7 RNAP bearing the Y639F mutation the yields of transcripts containing 2′-F-, 2′-NH_2-, 2′-OMe- and 2′-thio-pyrimidines are higher than the wild type enzyme.[162,284] Later studies reported that a double mutant (Y639F:H784A) of T7 RNAP efficiently uses 2′-OMe- and 2′-N_3-pyrimidines as substrates.[163]

1.5.4.3 Primer extension reaction

The primer extension reaction is a technique in which a short oligonucleotide called primer (18–22 bases) is annealed to the 3′-end of a complementary DNA strand (template DNA) forming duplex. A DNA polymerase is then employed to incorporate natural or modified triphosphates to the 3′-end of the primer and extends until it reaches the 5′-end of the template DNA (Figure 1.31). In order to use modified triphosphates as substrates to produce a full-length primer product, conditions such as the concentration of enzyme and divalent ions, incubation time and temperature can be optimised. For divalent ions, Mg^{2+} can be substituted by Mn^{2+} which results in increased tolerance of modified nucleotides by DNA polymerase.[285,286] Depending on the application, the primer can be labelled with a fluorescent tag at the 5′-end for ease of analysis.

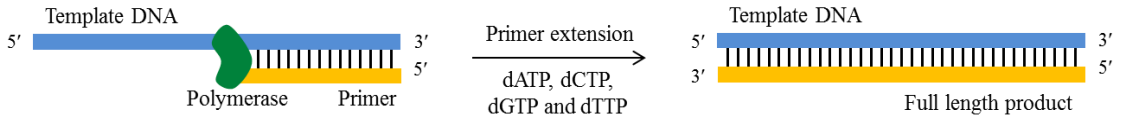


Figure 1.31. Primer extension reaction using a DNA polymerase (green) and four natural nucleotides.

1.5.4.4 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a method used to amplify a single (or multiple copies) of a DNA sample to generate thousands to millions of copies of this sequence. This technique was discovered by Kary Mullis in 1983.[287] Unlike the primer extension reaction in which only a single extended primer is produced by the polymerase, PCR amplifies both DNA strands from a very small amount of template DNA (single- or double-strand). In a PCR reaction, a buffered solution containing the desired DNA sample, two primers (complementary to the 3′ ends of each of the sense and anti-sense strands of the sample) are mixed with a thermostable DNA polymerase and

deoxynucleoside triphosphates (Figure 1.32). Then, through a series of 20–30 thermal cycling steps where the dsDNA is denatured, the primer is annealed and then elongated, the sample is amplified.

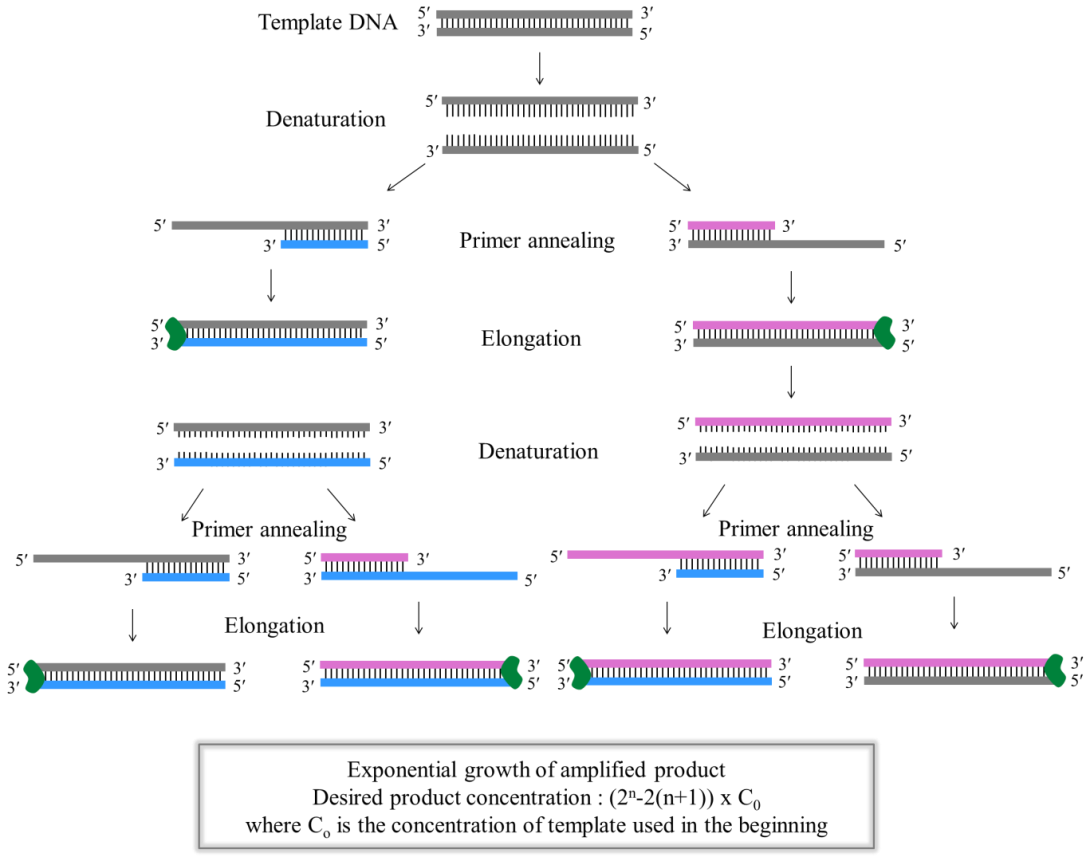


Figure 1.32. PCR using the DNA duplex as a template. Denaturation at 94–98 ℃ for 0.5–1 min; Annealing at 50–60 ℃ (depends on primers T_m) for 20–30 sec; Elongation at 65–72 ℃ for 1–3 min. The pink and blue lines represent the amplified products and DNA polymerase is green.

However, not all DNA polymerases can tolerate modified nucleotides as substrates during PCR. If PCR amplification using modified nucleic acids is unsuccessful, the modified nucleotides cannot be used in SELEX. Here, the primer extension reaction provides an easy system in which to optimise final PCR conditions.[288] Templates contain various modified nucleotide incorporation sites provide a thorough test for the ability of a polymerase to use a modified triphosphate during enzymatic synthesis. This

is important to test using templates containing consecutive modified nucleotide incorporation sites. This is because ssDNA libraries will inevitably contain numerous sequences including consecutive modified sites. Optimised conditions used during the primer extension reaction are adjusted for PCR. Even after linear extension optimisation, it may be necessary to optimise the number of PCR cycles, the concentration of enzymes and concentration of modified nucleotides.

1.6 Aptamer validation

It is important to confirm the target binding properties (e.g. affinity and kinetics) of selected aptamers before their use in further research. In order to characterise the binding properties of an aptamer and its target, a number of biophysical techniques have been developed to measure aptamer-target binding affinity.[29]

- Isothermal Titration Calorimetry (ITC) measures the heat released from the aptamer-target complex on the formation and is one of the most commonly used methods to determine the affinity, stoichiometry and thermodynamic properties of aptamer-target complexes.[289]

- MicroScale Thermophoresis (MST) is also a thermodynamic based technique; the thermophoresis-movement of molecules in a temperature gradient is detected using attached fluorophores to quantify the biomolecular interaction.[290]

- Flow cytometry is a laser-based technique to characterise aptamer-target formation. The fluorescent intensity changes when the fluorescently labelled aptamer binds to its target and therefore, fluorescent intensity indicates the binding affinity.[28,291]

- Surface plasmon resonance (SPR) is also commonly used in the characterisation of aptamer binding kinetics and affinity.[292]

- Electrophoretic mobility shift assay (EMSA) is an affinity gel electrophoresis technique to study aptamer-target complexes. A wide range of qualitative and quantitative analyses for aptamer-target interactions can be applied.[293]

In the next, two biophysical techniques, SPR and EMSA which will be used to characterise the aptamer developed in this study, and its target interactions, will be introduced.

1.6.1 Surface plasmon resonance (SPR)

Surface plasmon resonance (SPR) is an optical biosensing technology to measure a difference in refractive index of the medium at a metal surface. It is used to monitor the binding of analyte molecules to molecules immobilised on the metal surface. The metal film is located at the interface of two dielectric media. When a light beam hits the higher refractive index medium (prism) and travels from the prism to the lower refractive index (air or solution of interest), the light is bent towards the plane of the interface. When the incidence angle is greater than the critical angle, the light is totally reflected at the interface and travels back to the high refractive index medium; this is called total internal reflection (TIF) (Figure 1.33).[294]

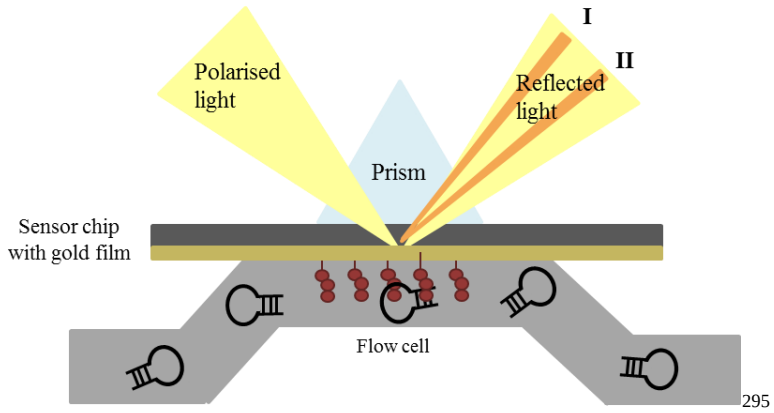


Figure 1.33. Basic SPR instrument configuration for measuring the binding of an analyte molecule to a receptor molecule. A critical angle generated when the polarised light hits the surface of the gold film (**I**). Changes in the critical angle when the analyte bind to the surface (**II**). This figure was constructed based on pictures and information given in reference.[295]

BIAcore™ developed the first commercial instrument in 1991.[296] The surface typically consists of a thin gold film (48–50 nm) on glass support which forms a flow cell where an aqueous solution is passed over continuously. A self-assembled monolayer (SAM) is deposited on the metal and it provides a linker layer between the metal surface and the next layer. A non-cross-linked carboxymethylated dextran is bound on the SAM and provides a hydrophilic environment. A linear chain of 1,6-linked glucose units exhibits a very low non-specific adsorption of biomolecules. As each glucose is modified with a carboxyl group, a ligand can be covalently ligated to the dextran via amide, thiol or aldehyde chemistry. SPR users can immobilise their ligands to the gold surface using a suitable chip from a range of commercially available sensor chips.

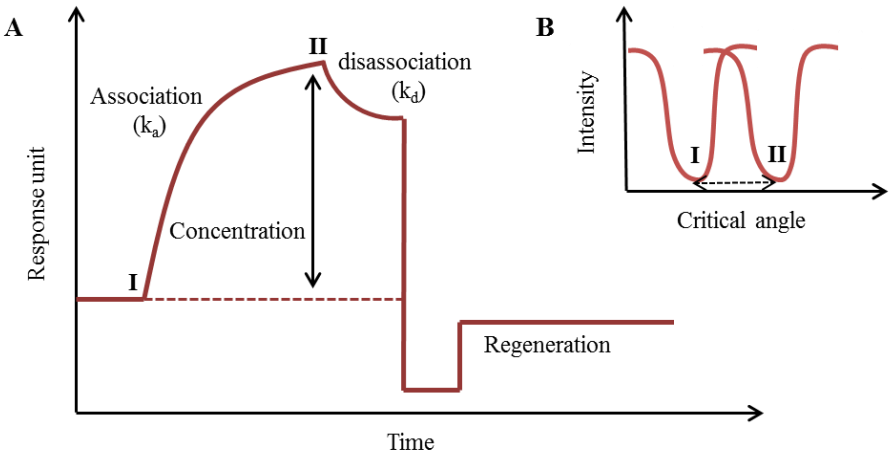


Figure 1.34. Sensorgram data from SPR. **A**) Changes in the response unit create a sensorgram when analyte bind to and dissociate from the surface of the sensor chip. **B**) Change in the critical angle of incident light from the angle I to II on the binding of an analyte molecule to a receptor molecule. This figure was constructed based on pictures and information given in reference.[292]

There are two flow cells on the sensor chip; one is functonalised with a ligand of interest and the other one is left unbound or bound with the negative target as a control. At the beginning of the experiment, only the aqueous solution without the analyte molecule is flowed through and therefore, the response unit (RU) values correspond to the starting critical angle I (Figure 1.34 **B**). Then, the analyte molecule is injected in the aqueous solution through both flow cells. Upon binding of the analyte to the ligand, the refractive index at the gold surface changes and RU increases. This is the association phase during which the binding site of the ligand is occupied by the analyte molecules and thus, the shape of the curve can be used to measure the rate of association (k_a) (Figure 1.34 **A**). When binding reaches equilibrium, called steady-state, the RU value changes to the final critical angle II (Figure 1.34 **B**). The binding affinity (K_D) can be measured when the RU value has reached its maximum (R_{max}). After the analyte injection is finished, the dissociation phase begins during which binding sites become unoccupied and hence, the shape of this curve can be used to measure the rate of

dissociation (k_d) (Figure 1.34 **A**). The surface can then be regenerated and returned to its critical angle I to inject the next analyte.[292] Due to its real-time and label-free features, SPR is used in the areas of biochemistry, biology and medical science applications as an important tool for characterising and quantifying biomolecular interaction.[297,298]

1.6.2 Electrophoretic mobility shift assay (EMSA)

The electrophoretic mobility shift assay (EMSA) is a regularly used method to detect protein-nucleic acid interactions.[299–301] This method resolves protein-nucleic acid complexes on a non-denaturing (native) polyacrylamide gel due to the electrophoretic mobility difference between a protein-nucleic acid complex and free nucleic acid. High detection sensitivity is typically attained by using radioisotopic-labelled nucleic acid[302] at low concentrations (0.1 nM or less) in a small sample volume (20 µL or less). Nucleic acids of interest can also be covalently or non-covalently labelled with fluorophores[303,304] or biotin[305,306] and imaged by fluorescence, chemiluminescence and/or chromophore deposition, respectively. An unlabelled nucleic acid can also be used in the assay and detected by staining with chromophores or fluorophores after gel electrophoresis.

Various sizes of nucleic acids ranging from short to long oligonucleotides,[307,308] and a wide range of structures including single-stranded, duplex, triplex,[309] quadruplex,[310] and even small circular DNAs[311] have been used in the assay. Protein sizes can be varied from small peptide to large transcription factor which has a MW of more than 10,000,000 Da.[293]

The most significant limitation of the mobility shift assay is that nucleic acid-protein complexes are not at chemical equilibrium during the electrophoresis step. Detection of

complexes is hindered due to fast dissociation during electrophoresis. The rapid dissociation of the protein-nucleic acid complexes is happened when significant heat generated during the running of the gel. It is therefore recommended to run the polyacrylamide gel in a cold room (4 ℃). The ionic strength and pH of running buffer play significant roles in the protein-nucleic acid stability. The buffers with high salt concentration (higher than 300 mM) may disrupt the complexes and induce heat during the electrophoresis. In addition, low molecular weight co-factors (e.g. glycerol or ethylene glycol) can be added to the gel and running buffers to enhance the stability of protein-nucleic acid complexes.[293,312,313] Short electrophoresis times, however, give a close approximation to binding patterns especially when complexes are significantly stable in the gel.[314,315] In the literature, both polyacrylamide and agarose gels have been used for EMSA,[316–318] but polyacrylamide gels provide better resolution for protein-nucleic acid complexes of 500,000 MW or less. The concentration of gel used to monitor the migration is most critical in EMSA and therefore, must be carefully optimised. Low concentrations of gel (e.g. 5% w/v acrylamide) can be first used and the concentration is systematically increased to observe the stable complex which may be disrupted by long electrophoresis times. The common electrophoresis buffers such as Tris-borate–EDTA, Tris-acetate–EDTA or Tris-Gly can be used in the assay and the components can be adjusted to optimise the stability of complexes.[293]

1.7 Objectives of the research

The ultimate goal of this study is to develop novel aptamers which bridge the gap between antibodies and conventional aptamers in terms of target affinity and selectivity. The aim is to bring further innovation in nucleic acid chemistry into the aptamer field by introducing chemical diversity to aptamers, placing emphasis on functions that are not currently accessible to antibodies or any other biological and biomimetic systems. Our study is based on the hypothesis that combining nucleic acid and protein-like or halogen bonding interactions into aptamer development will create a new class of ligands which possess superior target binding affinity and selectivity.

The preliminary goal was to synthesise modified C2′-ribothymidine triphosphates and C5-propargylamino-dUTPs functionalised with various hydrophobic modifications and halogen bonding substituents which can be used to expand the chemical diversity of DNA aptamers. As PCR amplification is an essential step during SELEX, it is essential to examine the acceptance of modified triphosphates by DNA polymerases during enzymatic DNA synthesis. With modified triphosphates that are compatible with DNA polymerase, each step of the SELEX process, such as library and primers design, PCR amplification, ssDNA generation, selection methodology and biophysical analysis of aptamer sequences need to be optimised. With our interest in therapeutic aptamer development, HER2 extracellular epitopes are employed as targets during SELEX. Finally, the well-established SELEX method is applied to the selection of ssDNA aptamers containing halogen bearing modifications to study the σ-hole effect which drives halogen bonding forces.

Chapter 2

Synthesis of C2′ sugar modified thymidine triphosphates and their enzymatic reactions

2 Synthesis of C2′ sugar modified thymidine triphosphates and their enzymatic reactions

2.1 Background

Typically, aptamers are developed using naturally occurring nucleotides. In therapeutic applications, however, a major challenge for natural nucleic acid-based aptamers is poor stability in biological media. They are susceptible to degradation by exo- and endonucleases, which leads to inadequate bioavailability. To circumvent these shortcomings, structural modifications are often used to enhance nuclease resistance.[319,320] While unmodified RNA and DNA aptamers have short half-lives of seconds to hours in human serum, their stability can be greatly increased using combinations of modifications. For example, the first FDA approved aptamer, Macugen, –which contains 2'-fluoro-substituted pyrimidines, 2'-O-methyl-substituted purines, 3'-inverted dT cap and a 5'-PEG conjugate has half-life of 131 h in human serum.[153]

For 2'-sugar modified triphosphates to be used as substrates in SELEX, they should be compatible with DNA or RNA polymerases. 2′-NH_2-pyrimidine,[321] 2′-F-pyrimidine,[71] and 2′-OMe-nucleotides[73] have been successfully incorporated during SELEX (as discussed in Chapter 1, section 1.3.1). However, the introduction of bulky substituents at 2′-position is still challenging.[165,166]

As discussed in Chapter 1 (section 1.5.4.1), the divalent cation also plays an essential role in DNA synthesis by DNA polymerases by catalysing nucleotide addition to a growing primer strand and excision of incorrectly incorporated dNTPs via intrinsic 3′–5′ exonuclease activity.[246–249] However, Mn^{2+} can substitute for Mg^{2+},[246,322,323] and following observations have been made:

1) Several groups have reported that substituting Mn^{2+} for Mg^{2+} affects the base selectivity of DNA polymerases by decreasing the fidelity of the polymerase.[286,322–325] Beckman *et al.* proved that the fidelity of a DNA polymerase when a very low concentration of Mn^{2+} is present (<1 µM) is similar to when using Mg^{2+}. With increased concentrations of Mn^{2+}, however, a decrease in fidelity is observed.[286]
2) Mn^{2+} was shown to promote the incorporation of incorrect nucleotides by increasing the rate of incorporation (k_{pol}) and by reducing the k_d (increasing the affinity) of the mismatched deoxynucleotide for the polymerase-template complex.[325,326]
3) Wang *et al*. used *Bacillus stearothermophilus* (BST) polymerase to give structural evidence that the C-A mismatch adopts a tautomeric cognate base pair shape in the presence of Mn^{2+}.[327] This base pair is virtually indistinguishable from the canonical, Watson-Crick base pair in double-stranded DNA at the insertion site. In the presence of Mn^{2+}, the triphosphate tail of the incoming dNTP is properly aligned for catalysis and the polymerase is in a closed conformation helping the incorporation of the incorrect base. However, the polymerase adopts an "ajar" (partially closed) conformation and the C-A mismatch forms a non-cognate wobble base pair when the Mg^{2+} was substituted for Mn^{2+}. In the presence of Mg^{2+}, the triphosphate tail of the incoming base is also not properly aligned for catalysis and therefore, the misincorporation is prevented.[327]
4) Metal ions in the active site of polymerases are also required in the exonuclease site to catalyse the 3′–5′ exonuclease activity as shown in several DNA polymerases.[249,326,328–331] Vashishtha *et al.* showed that Mg^{2+} is more effective in

promoting exonuclease activity of polymerase than Mn^{2+}, but the rate of the base excision varied slightly.[326]

As Mn^{2+} is known to increase the tolerance of polymerases for incorporating modified nucleotides more than Mg^{2+}, it has been added to the reaction mixture during enzymatic synthesis. For examples, Burmeister *et al*. discovered a VEGF aptamer using RNA libraries containing 2′-OMe-NTPs. During their enzymatic synthesis, combinations of $MgCl_2$ and $MnCl_2$ were added to increase the transcription yields using the T7 RNA polymerase mutant Y639F.[73] Eremeeva *et al.* performed enzymatic incorporation of various triphosphate analogues of deoxyguanosine using Taq, Vent exo-, Klenow fragment exo- and polIIIα DNA polymerases in the presence both Mn^{2+}/Mg^{2+} or only Mg^{2+}. The incorporation of the modified dGTPs was dependent on the specific modification. Mn^{2+} was essential for incorporation of some modifications, but not others.[332] Veedu *et al.* performed primer extension reactions using LNA-TTP or LNA-ATP as a substrate. Phusion high-fidelity DNA polymerase with supplemented Mn^{2+} successfully yielded the modified products.[333] Taking the above observations into account, Mn^{2+} was added to the reaction mixture to test whether tolerance of DNA polymerases for modified nucleotides during enzymatic synthesis by DNA polymerase was increased.

In this chapter, we investigate the ability of polymerase enzymes to tolerate modifications at the 2'-ribose position on thymidine. First of all, The 2'-aminoethoxy(AmE)-thymidine modification (Figure 2.1 **A**) was introduced into a DNA oligonucleotide template by solid phase synthesis using phosphoramidite chemistry. The aminoethoxy group was chosen with the aim of using high yielding amide bond formation for post-synthetic functionalisation. Initially, the oligonucleotide containing three 2'-AmE-T sites was used as a template to examine whether DNA polymerases can

read-through this modification and successfully incorporate dATP in the complementary strand (Figure 2.2 **A**).

Secondly, a bulkier modification than aminoethoxy group at 2′-sugar position was prepared to increase the ability of DNA polymerases to tolerate the bulky modification. Here, 2′-azidoethoxy(AzE)-thymidine triphosphate (Figure 2.1 **B**) was synthesised and used as substrate during enzymatic synthesis. The natural DNA template containing three dA sites was used in the reaction and 2′-AzE-TTP was added in the reaction instead of dTTP (Figure 2.2 **B**). Primer extension reactions were performed to examine whether 2′-AzE-TTP can be successfully incorporated into the complementary strand. Therefore, in this chapter the compatibility of DNA polymerase was tested with a range of enzymes and various conditions during enzymatic DNA synthesis. If 2′-AzE-TTP is a good substrate for DNA polymerase, copper catalysed 'click' reactions[334] can be used to post-synthetically attach a wide range of functional groups in a modular, diversity oriented synthetic approach. Post-synthetic attachment of additional functional groups to these building blocks is predicted to not only build up the structural diversity, but also increase nucleic acids half-lives in living organisms. The electroneutral 2′-AzE-T unit may additionally prove to be effective at inhibiting exonuclease digestion for *in vivo* applications without further functionalisation. This work highlights an enduring interest in the enzymatic synthesis of oligonucleotides containing 2′-modified ribothymidine using DNA polymerases.

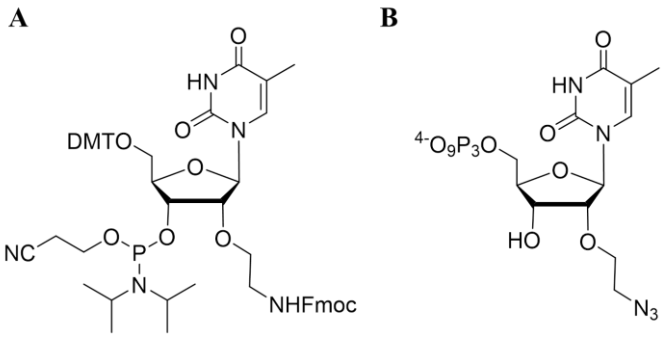


Figure 2.1. **A**) 2′-Aminoethoxy(AmE)-thymidine phosphoramidite. **B**) 2′-Azidoethoxy(AzE)-thymidine-5′-triphosphate.

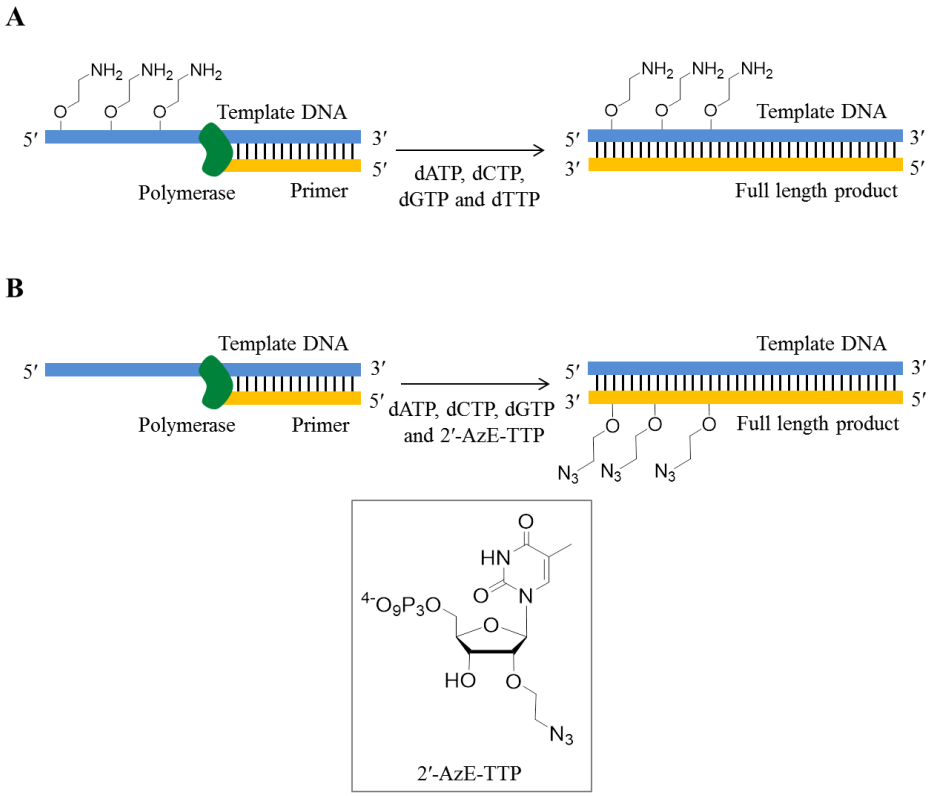


Figure 2.2. **A**) Enzymatic (polymerase-mediated) oligonucleotide synthesis using the 2′-AmE-T modified template. **B**) Enzymatic incorporation of 2′-AzE-TTP using a natural DNA template.

2.2 Primer extension reaction using a 2′-AmE-T modified template

An essential prerequisite for the selection of chemically modified aptamers using SELEX is the ability of DNA polymerases to accurately read-through the modification during the PCR enrichment step which is carried out during each round of the SELEX process.[335,336] In this section, the compatibility of the 2′-AmE-T modification with DNA polymerases was evaluated via a primer extension reaction which uses a short 29 mer 2′-AmE-T modified template, an 18 mer 5′-6-carboxyfluorescein (FAM) labelled primer and natural dNTPs. The 2′-AmE-thymidine modification was incorporated into a DNA oligonucleotide via the modified phosphoramidite monomer (Figure 2.1 **A**) using solid phase synthesis. The 2′-AmE-thymidine phosphoramidite monomer was synthesised by Dr Anna Dysko. This modified 29 mer oligonucleotide (T1, Table 2.1) was used as a template for linear enzymatic synthesis of the complementary unmodified oligonucleotides. FAM was attached at the 5′-terminal of the primer (P1, Table 2.1) to allow easy fluorescent visualisation of the fully extended primer by denaturing polyacrylamide gel electrophoresis (PAGE). The ability of DNA polymerases to read-through the DNA template containing 2′-AmE-T and produce a fully elongated complementary DNA strand was studied by denaturing PAGE analysis. In this technique, the rate of migration of each oligonucleotide at a given voltage and gel composition is primarily determined by its length; hence the appearance of a product band with the same electrophoretic mobility as the complementary template strand indicates successful formation of the fully elongated product.

To begin, three natural DNA polymerases (GoTaq, Klenow fragment and KOD DNA polymerases) were used to produce extended primer products. Gotaq polymerase which is a thermostable family A polymerase, is a native *Thermus aquaticus* (Taq) DNA polymerase with optimised buffer compositions.[337] The enzyme functions efficiently at

a higher temperature (74 ℃) and exhibits 5′–3′ exonuclease activity. KOD DNA polymerase, another high fidelity thermostable DNA polymerase is a recombinant form of *Thermococcus kodakaraensis* KOD DNA polymerase, and is a family B DNA polymerase.[260] This enzyme has strong 3′–5′ exonuclease activity for proofreading that Taq DNA polymerase lacks. Also, the rate of elongation is higher than with Taq DNA polymerase and the PCR accuracy is superior.[260] Another family A polymerase, DNA polymerase I large (Klenow) fragment is a proteolytic product of *Escherichia coli* DNA polymerase.[338] Klenow fragment DNA polymerase has 3′–5′ exonuclease activity, but is deficient in 5′–3′ exonuclease activity, and the enzyme functions at 37 ℃. To ensure the primer P1 anneals with the template T1 and produces the full-length product, the incubation temperature was initially set at 60 ℃ –close to an optimal temperature for the primer (P1) annealing around 52–53 ℃[339]– for all reactions except those involving the Klenow fragment DNA polymerase, which were instead used at 37 ℃. For the positive control reaction, unmodified DNA template (T2, Table 2.1) was used with natural dNTPs (Figure 2.3, lanes **1a**, **2a** and **3a**). However, incubation of the control reactions for 1 h resulted in product digestion when KOD DNA polymerase was used (Figure 2.3, lane **2a**) due to its strong 3′–5′ exonuclease activity. For the negative controls, modified template T1 was used in the absence of dNTPs (Figure 2.3, lanes **1b**, **2b** and **3b**) meaning that no DNA product can be formed.

Of the three DNA polymerases, KOD DNA polymerase produced longer extension products (Figure 2.3, lane **2c**) compared to the GoTaq and Klenow fragment polymerases. In the reaction using the Klenow fragment DNA polymerases, primer elongation stops before or at the point of the first 2′-AmE-T modification on the template (Figure 2.3, lane **1c**). Majority of primer extended products terminates at the point of the first 2′-AmE-T modification when GoTaq polymerase was used in the

reaction (Figure 2.3, lane **3c**). In contrast, denaturing PAGE gel analysis of the product mixture from the reaction using KOD DNA polymerase suggested that this polymerase was able to read-through the first and second modifications on the template to produce some longer elongation products, just before the final modification on the template (Figure 2.3, lane **2c**). Nonetheless, fully extended products were not observed when the modified template T1 and dNTPs were incubated with any of the three DNA polymerases.

Table 2.1. Sequences of the primers (P) and templates (T) used in the primer extension assays and PCR

	Sequence (5′- 3′)
P1	**F**GGTTATGTGTGTCGGCAG
T1	CAG**T**CAC**T**G**T**ACTGCCGACACACATAACC
T2	CAGTCACTGTACTGCCGACACACATAACC
T3	CAGTCACAAAACTGCCGACACACATAACC
P2	GCATTCGAGCAACGTAAG
P3	GGTTATGTGTGTCGGCAG
T4	GGTTATGTGTGTCGGCAGTATTGTCAGTGTGAATTCCAGAG TGTGAGATTGTGTGCTGGCGATCTTACGTTGCTCGAATGC

F: 5'-6-FAM (Fluorescein); **T**: 2′-AmE-T

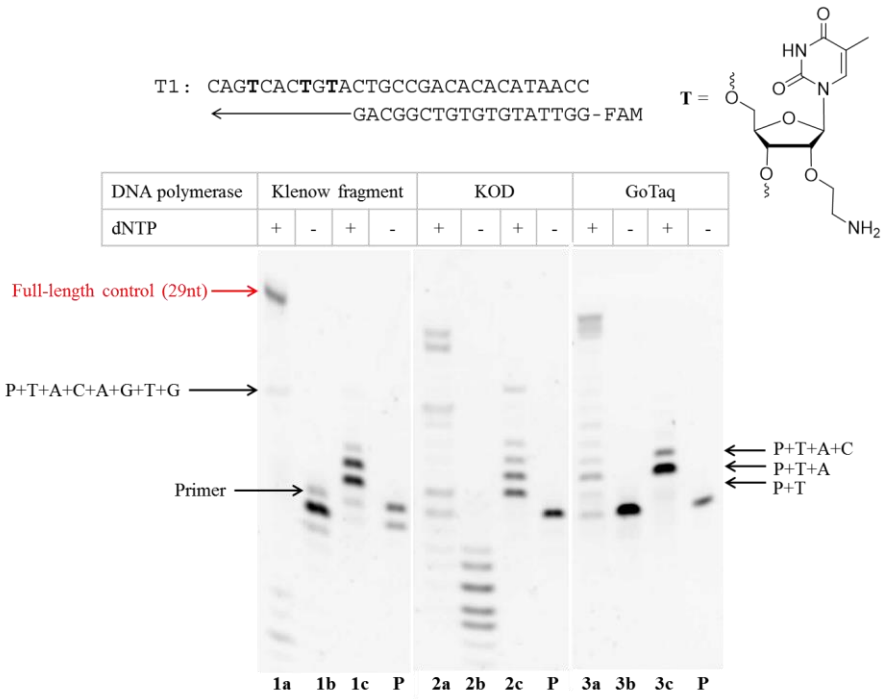


Figure 2.3. Screening of DNA polymerases for their ability to read-through a modified DNA template T1 using natural dNTPs. The positions of the 2′-AmE-T bases are highlighted in bold. **1a-1c**) Klenow fragment DNA polymerase (0.5 U, 37 ℃, 1 h). **2a-2c**) KOD DNA polymerase (0.5 U, 60 ℃, 1 h). **3a-3c**) GoTaq DNA polymerase (0.5 U, 60 ℃, 1 h). 3.2 nmol of each deoxynucleotide triphosphate and 1 mM $MgCl_2$ was used unless otherwise stated. Lanes **1a**/**2a/3a**, positive control, unmodified template T2 with natural dNTPs; lanes **1b**/**2b/3b**, negative control, modified template T1 without dNTPs; lanes **1c**/**2c/3c**, modified template T1 with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

As KOD DNA polymerase showed the most promising result (Figure 2.3, lane **2c**), further attempts to optimise the reaction conditions in order to generate the full-length product were carried out using this enzyme. The incubation temperature was initially set at 60 ℃, which is close to the optimal for primer annealing temperature (52–53 ℃). However, KOD DNA polymerase is a thermostable polymerase which functions efficiently at 72 ℃ (according to the supplier and the literature[172]). Therefore, the primer extension reaction was examined at 60 ℃, 65 ℃, 70 ℃ and 75 ℃ (Figure 2.4, lanes **1b**, **2a**, **2b** and **2c** respectively) in the supplied KOD buffer. In the case of KOD polymerase,

1 mM $MgCl_2$ was added separately to the reaction mixture. Incubation at 65 ℃ seems to slightly improve the yield of longer extension products (Figure 2.4, lane **2a**) compared to the reaction carried out at 60 ℃ (Figure 2.4, lane **1b**) due to reduced polymerase activity at 60 ℃. Also, the reaction at 65 ℃ is more efficient than 70 ℃ (Figure 2.4, lane **2b**), which is more close to the optimal temperature for KOD DNA polymerase. This confirms that forming a stable primer-template duplex is also important for KOD DNA polymerase (and polymerases in general). The reaction at 75 ℃ produced multiple bands which migrate more slowly than the product band (Figure 2.4, lane **2c**). This high MW smear is frequently observed for extension reactions using KOD DNA polymerase, especially when the reactions are incubated for long time periods or at high temperature, suggesting that non-specific primer extension were occurred.

The primer was digested in the absence of dNTPs due to the strong 3′–5′ exonuclease activity of KOD polymerase (Figure 2.4, lane **N**). The enzyme's 3′–5′ exonuclease activity, however, is reported to be dependent on the concentration of Mg^{2+}.[340–342] Therefore, to investigate the effect of Mg^{2+} concentration on the efficiency of the extension reaction, its concentration was varied between 0.5 and 4 mM at 60 ℃. The positive control showed a dim band indicating the full-length product (Figure 2.4, lane **C+**). This is caused by a gel 'smiling' effect which is usually due to uneven heat distribution through the gel during electrophoresis.[343] The experiment suggests that the optimal concentration of Mg^{2+} is 1.5 mM (Figure 2.4, lane **1c**). With increased Mg^{2+} concentration (4 mM), the reaction was inhibited and the majority of the elongated products was terminated after one or two nucleotide(s) incorporation (Figure 2.4, lanes **1c**–**1f**), while at lower concentrations of 0.5–1mM the efficiency also seems to be reduced (Figure 2.4, lanes **1a** and **1b**).

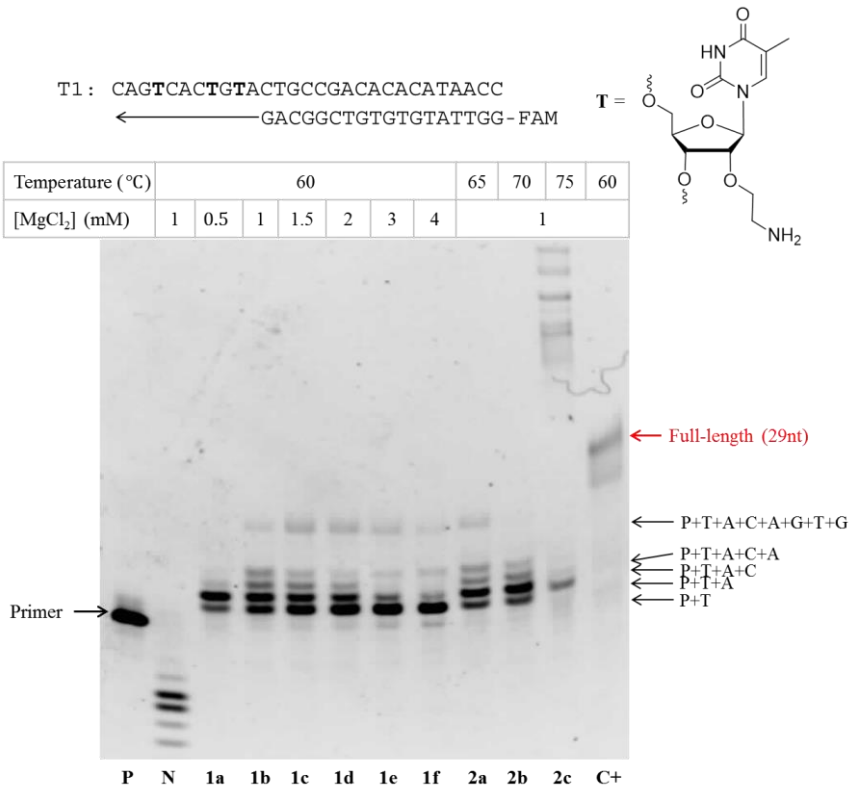


Figure 2.4. Optimisation of incubation temperature (60–75 ℃) and divalent cation concentrations, $MgCl_2$ (1–4 mM) for a read-through of modified DNA template T1 using natural dNTPs. The positions of the 2′-AmE-T bases are highlighted in bold. Reactions were performed using 0.5 U KOD DNA polymerase (60 ℃, 1 h), 3.2 nmol of each deoxynucleotide triphosphate and 1 mM $MgCl_2$ unless otherwise stated. Lanes **1a-1f**, modified template T1 with natural dNTPs and additional $MgCl_2$; lanes **2a-2c**, modified template T1 with natural dNTPs at various incubation temperature; lane **C+**, positive control, unmodified template T2 with natural dNTPs; lane **N**, negative control, modified template T1 without dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

In addition, a time course study of the reaction using the modified template T1 was conducted using dNTPs, 0.5 U KOD DNA polymerase and 1.5 mM $MgCl_2$ at 65 ℃. Increasing the reaction time from 1 h to 4 h gave no discernible improvements in the yields of longer extension products, and evidence of increased primer digestion by the polymerase was observed when the incubation time was extended (Figure 2.5, lanes **1a**–**1c**).

As discussed in section 2.1, addition of manganese ions (Mn^{2+}) is known to increase the tolerance of many different types of polymerase to different modified ribose moieties.[283] The linear extension of the modified template T1 using KOD DNA polymerase at 65 ℃ was therefore carried out in the presence of a varying concentration of Mn^{2+} (Figure 2.5, lanes **3a**–**3c**). Interestingly, the addition of Mn^{2+} cations was found to significantly improve the degree of formation of longer extension products (*i.e.* extended upto seven nucleotides incorporation) (Figure 2.5, lane **3a**). This result indicates that Mn^{2+} cations stimulate the elongation process and overcome premature termination. A possible rationale is that the presence of Mn^{2+} might induce a small change in the steric arrangement of the bound nucleoside triphosphate in the complex with DNA polymerases.[283] The Mn^{2+} concentration of 1 mM (Figure 2.5, lane **3a**) found to be efficient and higher concentrations were strongly inhibitory (Figure 2.5, lanes **3b** and **3c**). However, at 1 mM Mn^{2+}, the reaction terminates just before the last modification and does not go to completion to form the full-length extension products. For the positive control experiments, the unmodified DNA template T2 was used with natural dNTPs and 1.5 mM $MgCl_2$, which produced the 29 mer fully extended product (Figure 2.5, lane **C+**). There is one extra band above the product band, which is assigned to overhanging dA at the 3′-ends (Figure 2.6). It is well known that non-templated addition of nucleotides at the 3′-end is carried out by prokaryotic and eukaryotic DNA polymerases.[344–346]

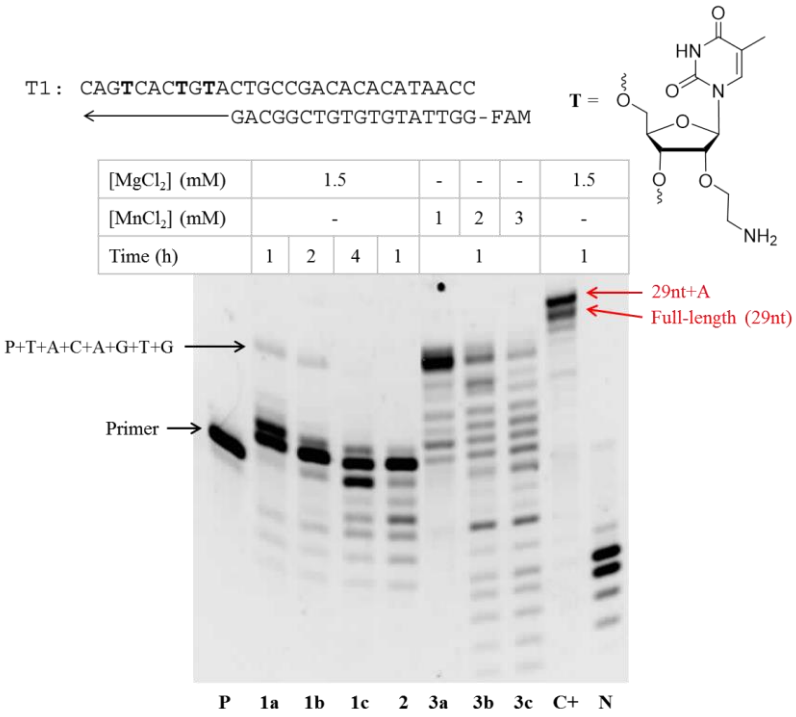


Figure 2.5. Optimisation of time (1–4 h) and divalent cations, $MnCl_2$ (1–3 mM) for a read-through of modified DNA template T1 using natural dNTPs. The positions of the 2′-AmE-T bases are highlighted in bold. Reactions were performed using 0.5 U KOD DNA polymerase (65 ℃, 1 h), 3.2 nmol of each deoxynucleotide triphosphate and 1.5 mM $MgCl_2$ unless otherwise stated. Lanes **1a-1c**, time course experiment using modified template T1 with natural dNTPs; lane **2**, modified template T1 with dTTP only (absence of dA, dG and dC); lanes **3a-3c**, modified template T1 with natural dNTPs and additional $MnCl_2$; lane **C+**: positive control, unmodified template T2 with natural dNTPs; lane **N**, negative control, modified template T1 without dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

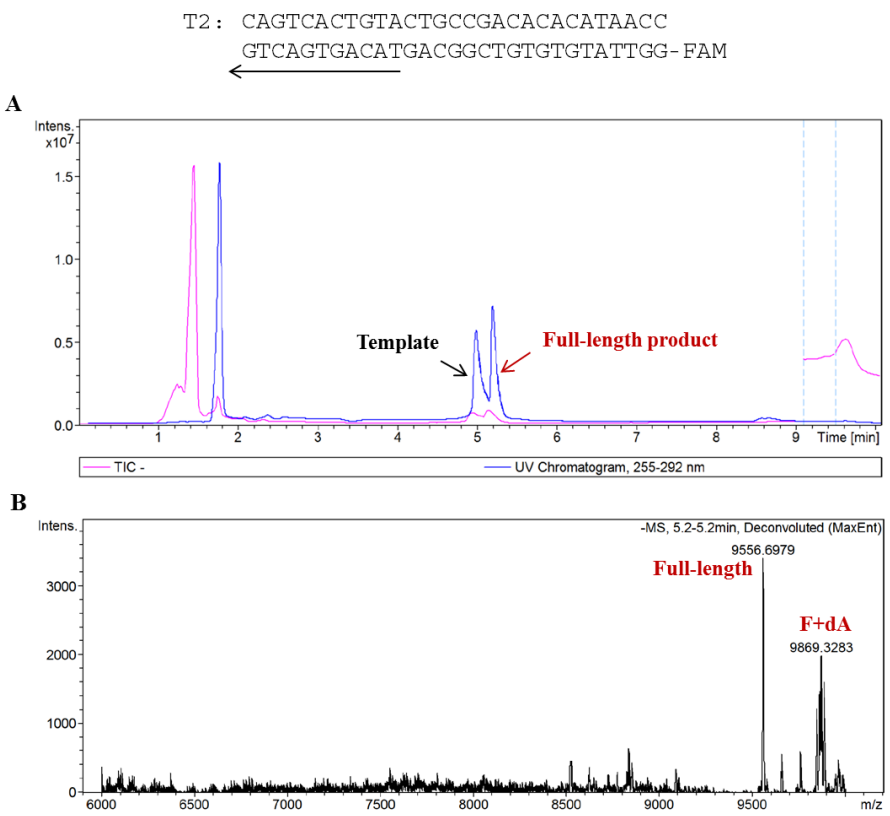


Figure 2.6. MS analysis of primer extension reaction using normal dNTPs and template T2. **A**) HPLC chromatogram and **B**) ES̄mass spectrum. The reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 30 min), 3.2 nmol of each deoxynucleotide triphosphate and 1.5 mM $MgCl_2$. Fully extended primer product: Calc., 9556.0 Da; found, 9556.7 Da (full-length product) and 9869.3 Da (full-length product plus dA, F+dA).

KOD XL DNA polymerase is well-known as an efficient enzyme for accepting base-modified and sugar-modified nucleotides as substrates.[180,182,183,347] KOD XL DNA polymerase is a mixture of wild-type KOD DNA polymerase and a mutant form, KOD exo- DNA polymerase,[262] that is deficient in 3′–5′ exonuclease activity. This feature makes this enzyme more tolerant towards modified substrates. For this purpose, the use of KOD XL DNA polymerase in the primer extension reaction was also investigated.

Based on the results of the prior optimisation studies using KOD DNA polymerase, the primer extension reactions using KOD XL DNA polymerase were carried out at 65 ℃

in the presence of additional 1 mM $MnCl_2$. The reaction time was varied from 30 min to 3 h (Figure 2.7, lanes **1a**–**1e**) and the results were analysed on denaturing PAGE. KOD XL DNA polymerase demonstrated higher read-through of the modified template as incubation time was increased, due to lack of 3′–5′ exonuclease activity of KOD XL polymerase. Conversion of the primer to the desired product was greatest after 3 h (Figure 2.7, lane **1e**). Still, some premature termination was apparent after 3 h and therefore, the amount of KOD XL DNA polymerase in the reaction was increased from 0.5 U to 3 U in an attempt to make reactions proceed further towards completion (Figure 2.7, lanes **2a**–**2c**).

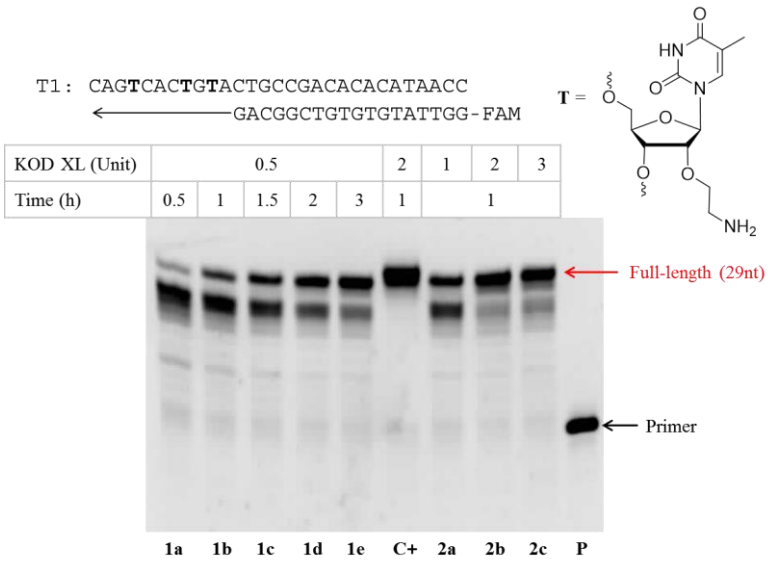


Figure 2.7. Optimisation of time (0.5–3 h) and polymerase concentration (2–3 U) for a read-through of modified DNA template T1 using natural dNTPs. The positions of the 2′-AmE-T bases are highlighted in bold. Reactions were performed using 0.5 U KOD XL DNA polymerase (65 ℃), 3.2 nmol of each deoxynucleotide triphosphate and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using modified template T1 with natural dNTPs; lanes **2a-2c**, modified template T1 with natural dNTPs and increased KOD XL polymerase concentration; lane **C+**, positive control, unmodified template T2 with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

The enzymatic reaction catalysed by the DNA polymerase can be modelled as shown in Equation 2.1. The enzyme (*E*) directly interacts with the substrate (*S*) in a stoichiometric manner to form, an intermediate complex (*ES*), which exists in thermodynamic equilibrium with the free enzyme and substrate. The *ES* complex then converts to product.

A

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} P + E$$

B

$$V_0 = \frac{V_{max}[S]}{K_m + [S]} \qquad K_m = \frac{k_{-1} + k_2}{k_1} \qquad V_{max} = K_{cat}[E]_0$$

Equation 2.1. **A**) Reaction scheme for the single substrate enzymatic reaction. *S* is the substrate; *E* is the polymerase; *ES* is the substrate and polymerase complex intermediate; *P* is the product; k_1, k_{-1}, k_2 are the reaction constants for each reaction. **B**) Michaelis-Menten kinetic equation for a single substrate incorporation reaction. [*S*] is the concentration of substrate; $[E]_0$ is the concentration of polymerase; V_0 and V_{max} are the initial reaction rate and the maximal reaction rate respectively.

K_m is the Michaelis constant and represents the affinity between the enzyme and its substrate and V_{max} is the maximal velocity for the enzymatic reaction. K_{cat} is defined as catalytic constant and is the number of catalytic reactions per unit time. Based on this Michaelis–Menten kinetics equation (Equation 2.1 **B**),[348,349] the maximum rate of reaction, V_{max} is proportional to the concentration of enzyme, $[E]_0$. It is therefore expected that increasing the amount of added enzyme will enhance the yield of the fully extended product. As seen in Figure 2.7, the most intense product band on the gel

(indicated by a red arrow) was achieved after 1 h when 2 U or 3 U KOD DNA polymerase was added (Figure 2.7, lanes **2b** and **2c**).

Using 2 U KOD XL DNA polymerase, time course experiments from 30 min to 7 h were performed to enhance the read-through (Figure 2.8, lanes **1a**–**1e**). After 2–3 h, the full-length product was observed as the predominant species by denaturing PAGE analysis (Figure 2.8, lane **1c**). Further increasing the amount of polymerase to 3 U and 4 U improved the read-through of modified template, producing full-length bands within 1 h (Figure 2.8, lanes **2a** and **2b**). However, some truncated sequences were still observed presumably due to impurities in the DNA template, which was accordingly re-purified by denaturing PAGE.

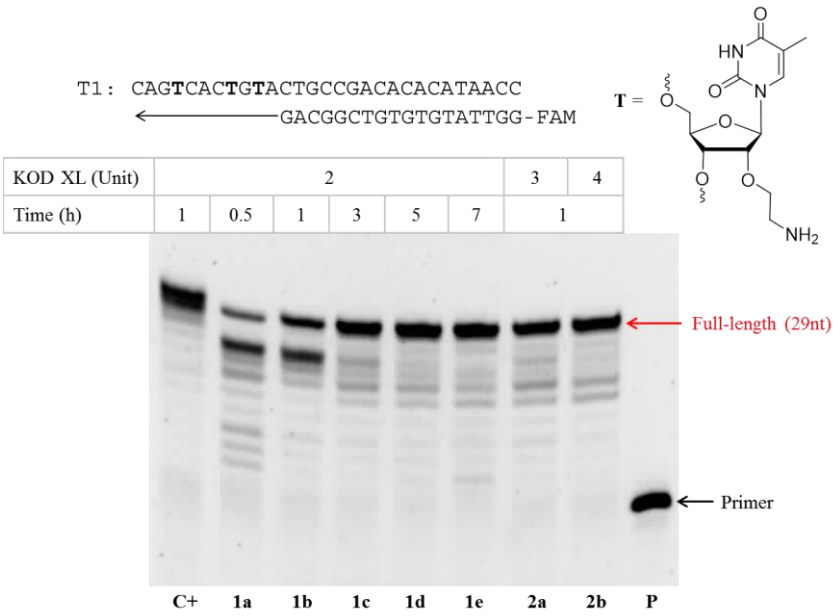


Figure 2.8. Optimisation of time (0.5–7 h) and polymerase concentration (3–4 U) for a read-through of modified DNA template T1 using natural dNTPs. The positions of the 2′-AmE-T bases are highlighted in bold. Reactions were performed using 2 U KOD XL DNA polymerase (65 ℃), 3.2 nmol of each deoxynucleotide triphosphate and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using modified template T1 with natural dNTPs; lanes **2a**/**2b**, modified template T1 with natural dNTPs and increased KOD XL polymerase concentration; lane **C+**, positive control, unmodified template T2 with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

After template re-purification, a final primer extension reaction using 2 U KOD XL DNA polymerase with additional 1 mM $MnCl_2$ was carried out, which produced a single product band (29 mer), when analysed by denaturing PAGE (Figure 2.9, lane **1b**). However, some degradation was observed after an extended 4 h incubation time (Figure 2.9, lane **1c**). Accordingly, an intermediate incubation time of 2 h appears to be optimal.

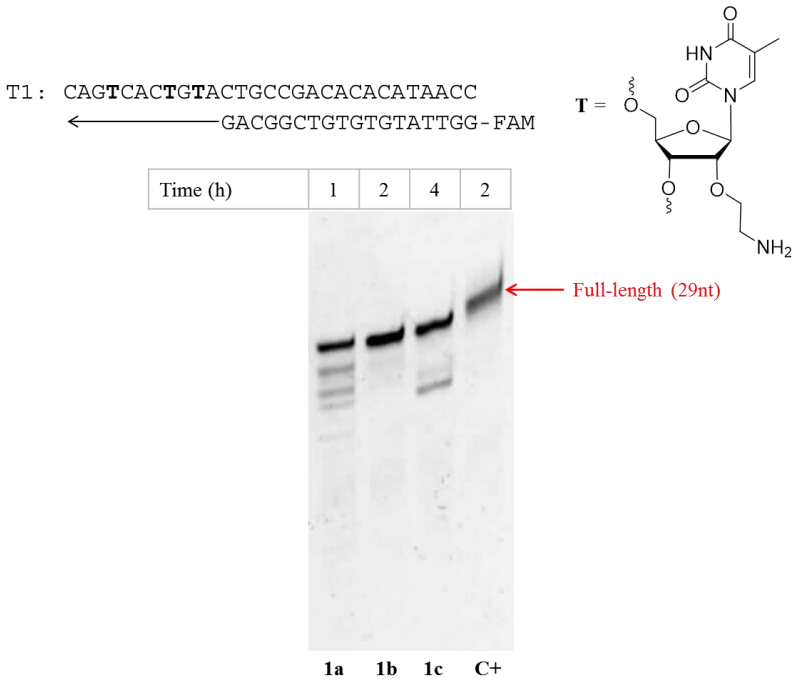
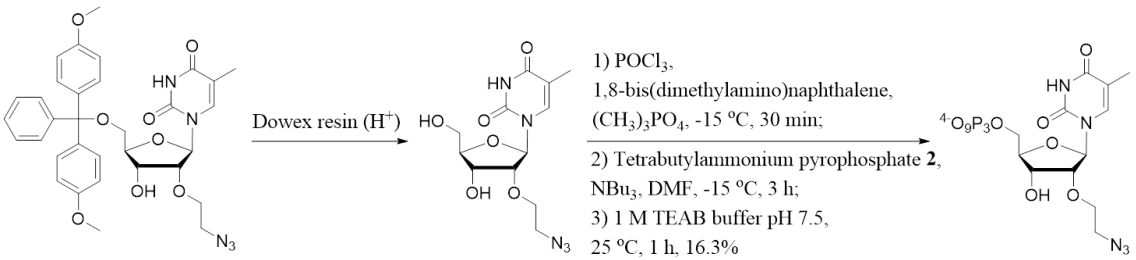


Figure 2.9. Final primer extension of purified template T1 using natural dNTPs over time (1–4 h). The positions of the 2′-AmE-T bases are highlighted in bold. Reactions were performed using 2 U KOD XL DNA polymerase (65 ℃), 3.2 nmol of each deoxynucleotide triphosphate and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1c**, time course experiment using modified template T1 with natural dNTPs; lane **C+**, positive control, unmodified template T2 with natural dNTPs. (20% Denaturing PAGE)

2.3 Preparation of 2′-AzE-TTP

As discussed in Chapter 1, section 1.5.3, Yoshikawa triphosphate synthesis method is one-pot reaction and the optimised reaction conditions produce higher yield of triphosphosphate.[239,240] In this chapter, 2′-Azidoethoxy-thymidine-5′-triphosphate (2′-AzE-TTP) was synthesised from the 4,4′-dimethoxytrityl (DMT) protected nucleoside using Yoshikawa's method as shown in Scheme 2.1.[236] The 5′-DMT protecting group of 2′-AzE-thymidine is initially cleaved to provide the 5′-OH nucleoside **1**, which is then converted to a phosphorodichloridate intermediate by treatment with phosphoryl chloride in trimethyl phosphate. *In situ* reaction of this intermediate with tris(tetrabutylammonium)hydrogen pyrophosphate **2**, which was synthesised and aliquoted inside a glovebox under argon due to its high hygroscopic nature, afforded the modified triphosphate target. The triphosphorylation reaction was performed at -15 ℃ to minimise by-product formation such as 3′-phosphate and 3′, 5′-diphosphate. 2′-AzE-thymidine-5′-triphosphate **3** was obtained in 16.3% yield after purification by RP-HPLC. Its identity and purity were confirmed by UPLC-MS characterisation (see Figure S 8-1 in appendix 8.4).



1 **3**

Scheme 2.1. Synthetic route used to prepare 2′-AzE-thymidine-5′-triphosphate.

2.4 Primer extension reactions using 2′-AzE-TTP

Primer extension reactions (see Chapter 1, section 1.5.4.3) provide an efficient biochemical method to produce chemically modified DNA of defined sequence. This can be achieved by substituting one of the four canonical nucleotides (dATP, dTTP, dCTP and dGTP) with a modified triphosphate. Here, 2′-AzE-TTP was utilised in primer extension assays in place of dTTP to evaluate the ability of this modified triphosphate to act as a substrate for the commercially available DNA polymerases, KOD, GoTaq, DNA polymerase I large (Klenow) fragment and KOD XL (mutant form of KOD) DNA polymerase.

An unmodified DNA template (T2, Table 2.1) has three adenine bases in the extension region which was designed to add three 2′-AzE-TTPs on the complementary strand. Successful linear copying of this template using dATP, dGTP, dCTP and 2′-AzE-TTP will produce a 29 mer full-length product and the canonical dNTPs will be used for a positive control experiment to compare the incorporation efficiency of the modified substrate.

With the encouraging results from the previous optimised primer extension reaction testing read-through ability of a modified template, here four DNA polymerases were screened to examine incorporation efficiency of the modified triphosphate (Figure 2.10). KOD XL DNA polymerase incorporated the modified triphosphate (2′-AzE-TTP) most efficiently to yield a small amount of the full-length products, and a significant quantity of partially extended, intermediate length product (Figure 2.10, lanes **4a**–**4c**). The full-length product containing three 2′-AzE-TTP modifications was expected to migrate more slowly on the gel than the unmodified control (Figure 2.10, lane **C+**) due to its higher MW.

Degradation of the product was observed after 4 h incubation (Figure 2.10, lane **4c**), and a 2 h reaction time gave more efficient incorporation of the modified nucleotide than a shorter 1 h reaction (Figure 2.10, lanes **4a** and **4b**). The lack of 3′–5′ exonuclease activity of KOD XL DNA polymerase makes the enzyme superior for modified nucleotide incorporation. This is supported by published studies on enzymatic synthesis of modified nucleotides using KOD XL DNA polymerase.[182–184,262]

The gel results suggest that the Klenow fragment DNA polymerase incorporated 2′-AzE-T slightly less efficiently than did KOD XL DNA polymerase after 1 h incubation time, producing significant quantities of partially extended, intermediate length products, but no trace of the full-length extended product (Figure 2.10, lanes **2a** and **2b**). Degradation was detected after a 4 h reaction time due to the enzyme's 3′–5′ exonuclease activity (Figure 2.10, lane **2c**). The performances of the GoTaq DNA polymerase (Figure 2.10, lanes **1a**–**1c**) and KOD DNA polymerase were significantly worse (Figure 2.10, lanes **3a**–**3c**). For these polymerases, the extension reactions appeared to stop after the first 2′-AzE-T incorporation.

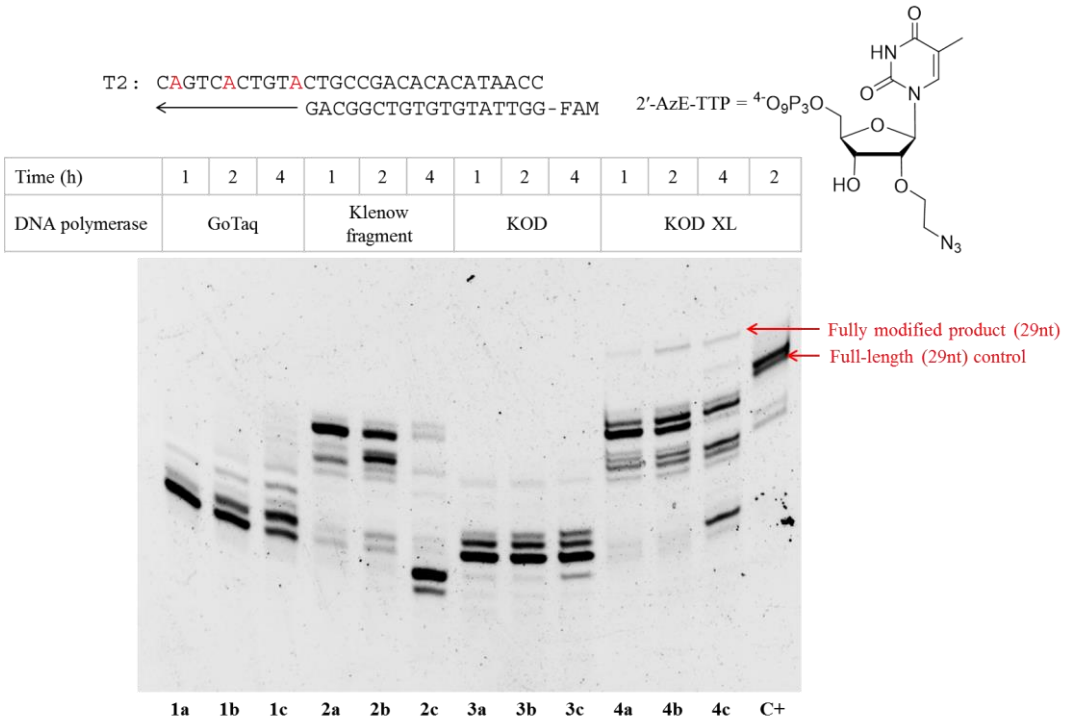


Figure 2.10. Time course (1–4 h) study of 2′-AzE-TTP incorporation using template T2 and primer P1 in the presence of; **1a-1c**) GoTaq (2 U, 60 ℃), **2a-2c**) Klenow fragment (2 U, 37 ℃), **3a-3c**) KOD (2 U, 60 ℃) and **4a-4c**) KOD XL (2 U, 60 ℃) DNA polymerase. 2′-AzE-TTP incorporation sites are highlighted in red. 3.2 nmol of dATP, dCTP, dGTP and 2′-AzE-TTP was added to the reaction. For KOD and KOD XL reactions, 1 mM $MnCl_2$ were supplemented. Lane **C+**, positive control, an unmodified template T2 with natural dNTPs. (20% Denaturing PAGE)

In light of these results, further optimisation of the extension reaction using 2 U of KOD XL DNA polymerase was conducted by varying the temperature (60 ℃, 65 ℃ and 72 ℃) and incubation time (30 min to 3 h). Primer extension reactions were performed at 60 ℃ and 65 ℃, which are lower than polymerase optimum temperature (70–74 ℃) but close to the annealing temperature of primer-template duplex.[339] In addition, the reaction at 75 ℃, which is close to its optimum temperature of the polymerase was attempted. These experiments revealed that the best conditions for 2′-AzE-TTP incorporation by KOD XL DNA polymerase are 65 ℃ for 2–3 h, which produced the most intense bands corresponding to the full-length product on the denaturing PAGE gel (Figure 2.11, lanes

2c and **2d**). When the reaction was performed at 60 ℃, the intensity of premature termination bands was stronger than the reaction at 65 ℃ (Figure 2.11, lanes **1d** and **2d**) due to reduced activity of polymerase. Consequently, the intensities of bands corresponding to the full-length extension product were slightly weaker for reactions performed at 60 ℃. Higher reaction temperature (72 ℃) resulted in the least efficient incorporation of modified nucleotides with KOD XL DNA polymerase (Figure 2.11, lanes **3a**–**3d**). This is probably due to partial denaturation of the duplex products under higher temperature (72 ℃) of the primer extension, which might affect the incorporation efficiency of 2′-AzE-TTP. This can be proved by melting temperature experiments, but T_m measurement was not carried out. Instead, we focused on performing the reaction at 65 ℃ to incorporate 2′-AzE-TTP more efficiently.

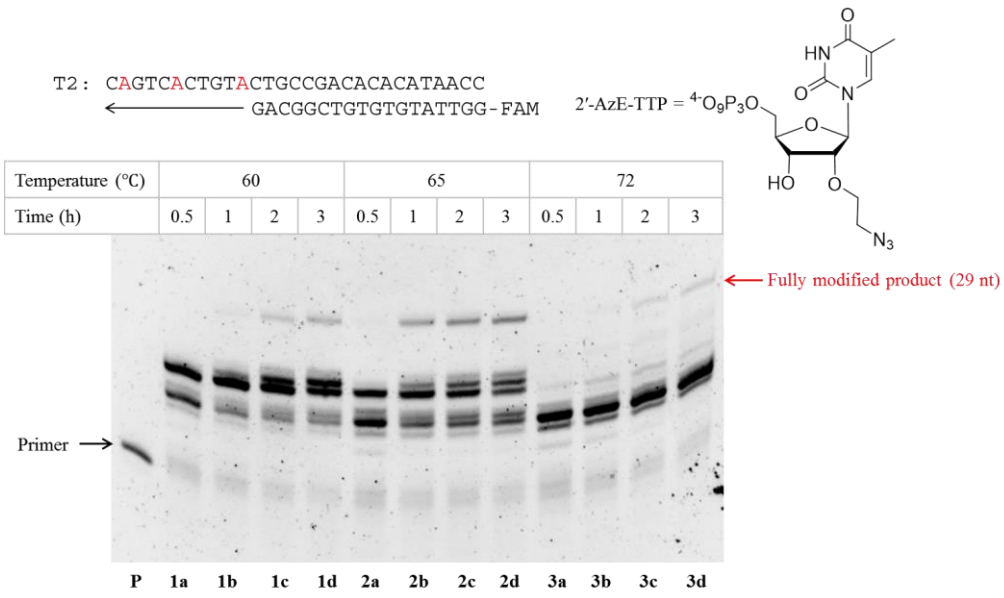


Figure 2.11. Optimisation of 2 U KOD XL DNA polymerase incorporation of 2′-AzE-TTP over time (0.5–3 h) at different temperatures (60 ℃, 65 ℃ and 72 ℃) using template T2. 2′-AzE-TTP incorporation sites are highlighted in red. dATP, dCTP, dGTP, 2′-AzE-TTP (3.2 nmol each) and 1 mM $MnCl_2$ were added unless otherwise stated. Lanes **1a-1d,** time course experiment at 60 ℃; lanes **2a-2d**, time course experiment at 65 ℃; lanes **3a-3d**, time course experiment at 72 ℃; lane **P**, primer P1. (20% Denaturing PAGE)

A higher concentration of KOD XL DNA polymerase was added (3 U) which should increase the rate of formation of the full-length product, as discussed in section 2.2 (Figure 2.12, lanes **1a**–**1d**). The yield was not improved in comparison to the reaction in which 2 U of the enzyme was used (Figure 2.11, lanes **2a**–**2d**). This suggests that the polymerase concentration is not the limiting factor but that fundamentally the polymerase cannot accept the modified substrate as well as was hoped. Prolonged incubation (3 h) ultimately led to the degradation of the primer (Figure 2.12, lane **1d**). A time course study with the unmodified template T2 and canonical dNTPs produced full-length product after 1 h and no degradation was observed after 3 h (Figure 2.12, lanes **C1**–**C3**).

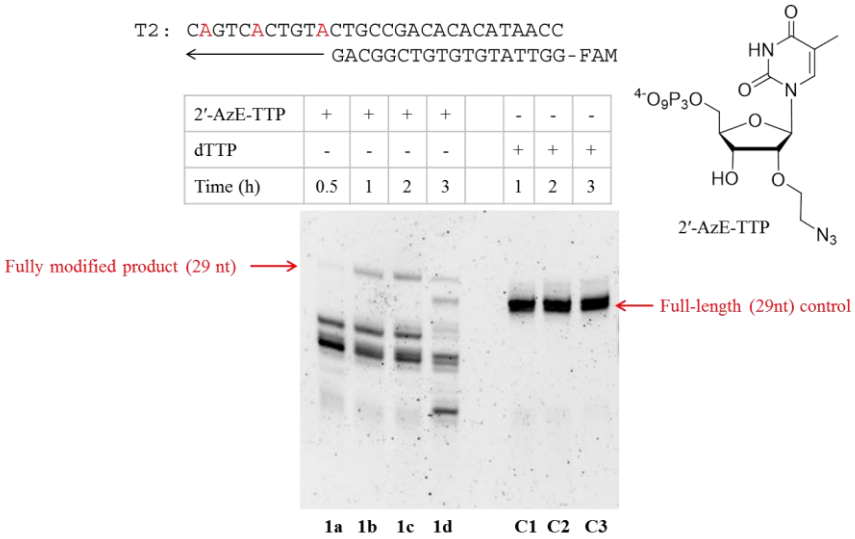


Figure 2.12. Optimisation of 3 U KOD XL DNA polymerase incorporation of 2′-AzE-TTP over time (0.5–3 h, 65 °C, lanes **1a-1d**) using template T2 and primer P1. 2′-AzE-TTP incorporation sites are highlighted in red. dATP, dCTP, dGTP, 2′-AzE-TTP (3.2 nmol each) and 1 mM $MnCl_2$ were used unless otherwise stated. Lanes **C1-C3**, positive control, time course experiment using unmodified template T2 with natural dNTPs. (20% Denaturing PAGE)

To enhance the polymerase binding of 2′-AzE-TTP, the concentration of the modified substrate, 2′-AzE-TTP, in the reaction mixture was increased from 3.2 nmol to 12.8

nmol (Figure 2.13, lanes **1**–**4**). The results of this experiment indicate that the primer extension efficiency does not significantly improve with increasing concentration of the modified substrate.

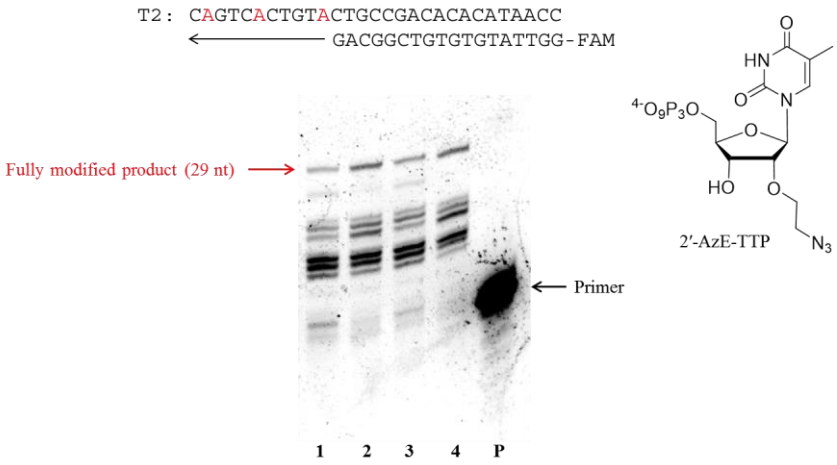


Figure 2.13. Optimisation of 2 U KOD XL DNA polymerase (65 ℃, 2 h) incorporation of 2′-AzE-TTP by varying substrate concentration (3.2–12.8 nmol). Template T2 was used in the reaction. 2′-AzE-TTP incorporation sites are highlighted in red. dATP, dCTP, dGTP (3.2 nmol each) and 1 mM $MnCl_2$ were used unless otherwise stated. Lane **1**, reaction with 3.2 nmol of 2′-AzE-TTP; lane **2**, reaction with 5.4 nmol of 2′-AzE-TTP; lane **3**, reaction with 6.4 nmol of 2′-AzE-TTP; lane **4**, reaction with 12.8 nmol of 2′-AzE-TTP; lane **P**, primer P1. (20% Denaturing PAGE)

When the primer extension reaction was carried out in the presence of only an unmodified dTTP (absence dA, dG and dC), at 65 ℃ with addition of 1 mM Mn^{2+}, one or two additional nucleotides were efficiently incorporated to the 3′-end of partially extended primer (Figure 2.14, lanes **1a** and **1c**). This is presumably caused by the non-templated addition of the second nucleotide.[344–346] In contrast, for the analogous reaction using the modified with 2′-AzE-TTP only, the major product band corresponding to the unextended primer was observed (Figure 2.14, lanes **1b** and **1d**). A very small amount of extended primer to which one nucleotide has been added was also detected, along with some degradation products arising from the 3′–5′ exonuclease

activity of the enzyme. This confirms that the enzyme's tolerance for the modified 2′-AzE-TTP is significantly lower than for canonical dTTP.

The reaction with various concentrations of $MnCl_2$ was tested to see if the tolerance of the polymerase for 2′-AzE-TTP increased. However, increased concentrations of Mn^{2+} inhibited the reaction (Figure 2.14, lanes **2b**–**2e**) and only the reaction with 1 mM $MnCl_2$ showed the desired full-length band, which again formed in low yield (Figure 2.14, lane **2a**). The results are consistent with literature suggesting higher concentrations of Mn^{2+} are strongly inhibitory in enzymatic reactions.[350]

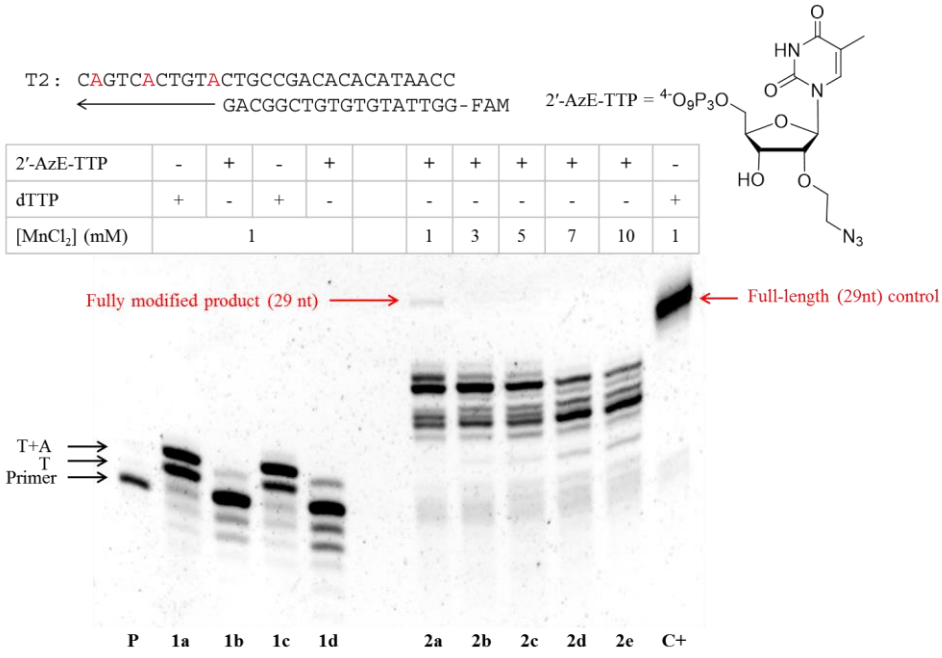


Figure 2.14. Optimisation of 2 U KOD XL DNA polymerase (65 ℃, 2 h) incorporation of 2′-AzE-TTP by varying $MnCl_2$ concentration (1–10 mM). Template T2 was used and 2′AzE-TTP incorporation sites are highlighted in red. dATP, dCTP, dGTP (3.2 nmol each), 2′-AzE-TTP (5.4 nmol) and 1 mM $MnCl_2$ were used unless otherwise stated. Lanes **1a**/**1c**: natural dTTP only (absence of dA, dG and dC) for 15 and 30 min respectively; lanes **1b**/**1d**, 2′-AzE-TTP only (absence of dA, dG and dC) for 15 and 30 min respectively; lanes **2a**-**2e**, primer extension reaction with varied $MnCl_2$ concentration; lane **C+**, positive control, unmodified template T2 with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

Inefficient incorporation of 2′-AzE-TTP is probably due to the steric effects at the polymerase binding site which probably preclude incorporation of 2′-AzE-TTP. DNA polymerases showed a stringent selectivity for dNTPs over NTPs due to steric interference between a bulky side chain of an active site residue (steric gate) and the 2′-OH group on the ribose ring of an incoming NTP. The amino acid residues involved in sugar discrimination is usually glutamate for family A polymerases and tyrosine or phenylalanine for family B polymerases. Reducing the size of this side chain at the 'steric gate' position can reduce discrimination against NTP incorporation by several orders of magnitude.[270,351]

2.5 PCR using 2′-AzE-TTP

With optimal conditions acquired from the primer extension reactions, a PCR experiment was attempted using an 81 mer unmodified template (T4, Table 2.1) and KOD XL DNA polymerase. During PCR, 2′-AzE-TTP is incorporated in place of canonical dTTP, and the modified products act as templates for the subsequent PCR cycle (Figure 2.15). Thus, 2′-AzE-TTP in the template must be tolerated by DNA polymerase, functioning as a thymidine analogue in terms of base pairing/stacking properties, and must also be accepted in the enzyme's active site to produce the modified 81 base pair (bp) amplicon. These features make PCR with modified nucleotides more challenging than primer extension reactions. Consistent with the primer extension results, incorporation of modified substrates is more demanding than read-through of the modified template.

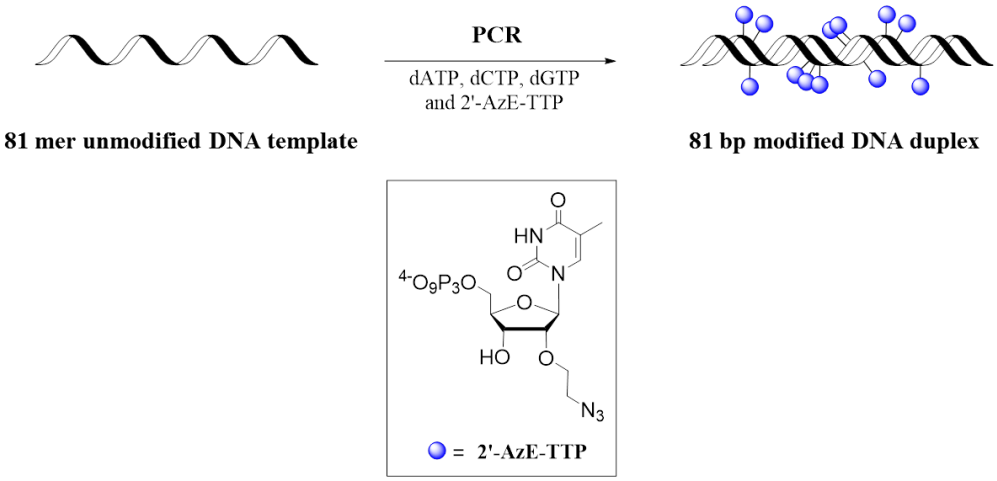


Figure 2.15. PCR using dATP, dCTP, dGTP and 2′-AzE-TTP.

An unmodified 81 mer single-stranded DNA, T4 was used as a PCR template with forward and reverse primers (P2 and P3, Table 2.1) to produce 81 bp modified dsDNA. Enzymatic synthesis of the modified nucleic acids using DNA polymerase in the absence of the natural substrates generally incorporates modified nucleotides. However, primer extension using 2′-AzE-TTP was not successful. Therefore, the PCR reaction was carried out using mixtures of 2′-AzE-TTP and canonical dTTP in different ratios to examine whether 2′-AzE-TTP competes with dTTP during the thermal cycling. The incorporation of modified nucleotides in the presence of natural counterparts is required for some application to control the reactive groups present in DNA. For example, Cox *et al*. used a mixture of amine-modified dUTP and dTTP in the same reaction to synthesise DNA with reduced number of the reactive groups. Consequently, the density of fluorophores introduced to the modified nucleotides was controlled and therefore, fluorescence quenching in the labelled DNA was avoided.[352]

As controls, PCR reactions with 100% of dTTP and 100% of the 2′-AzE-TTP were conducted (Figure 2.16, lanes **C+** and **4** respectively). Amplicons containing the

modification will only be obtained when the 2′-AzE-TTP competes with dTTP for incorporation by the polymerase. After PCR, the samples were run on a 2% agarose gel with a 25 bp reference ladder as a marker (Figure 2.16, lane **L**). When 100% of 2′-AzE-TTP was used, no PCR amplicon was detected (Figure 2.16, lane **4**). Conversely, when different ratios of dTTP and 2′-AzE-TTP were used, amplicons were detected (Figure 2.16, lanes **1**–**3**). Incorporation of 2′-AzE-TTP would be expected to produce bands which migrate slightly more slowly than the unmodified sequence in lane **C+** due to the higher MW of the modification. However, lanes **1**–**3** all migrated at the same speed, and this suggests that none of the modified TTP was incorporated during PCR (Figure 2.16).

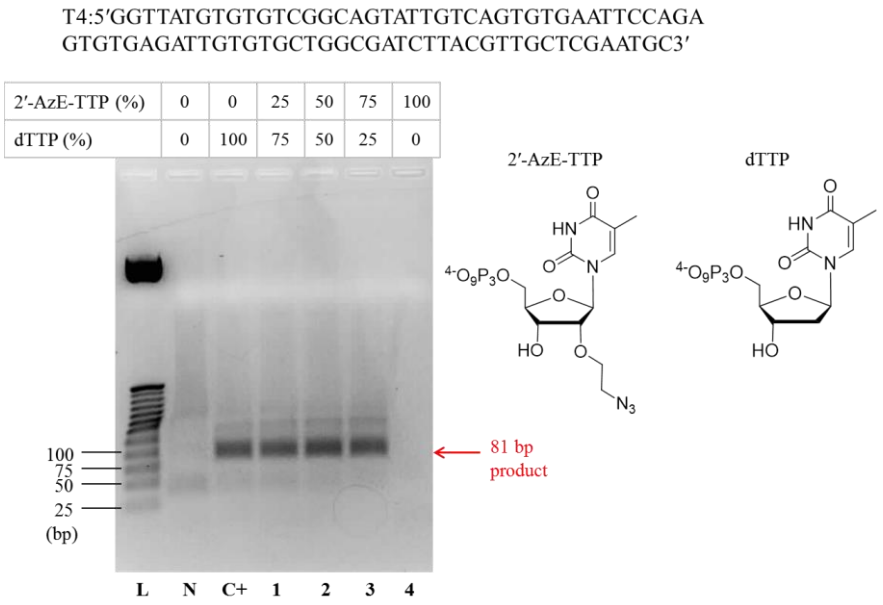
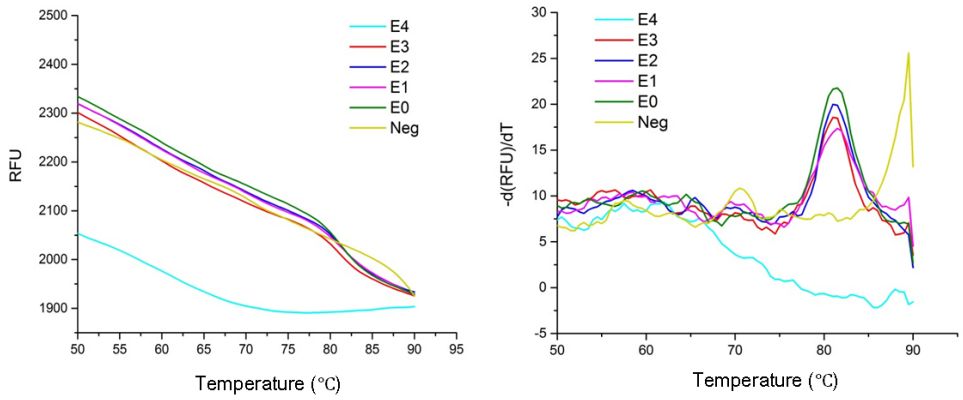


Figure 2.16. Analysis of PCR products from reactions using various ratios of 2′-AzE-TTP to dTTP (total = 5 nmol) and 1 U KOD XL DNA polymerase (25 cycles). Template T4 and two primers P2 and P3 were used in the reaction. Lane **L**, DNA ladder (25 base pair); lanes **1**-**4**, PCR amplicons when the different ratio of 2′-AzE-TTP and dTTP added; lane **N**, negative control reaction without a template; lane **C+**, positive control reaction with natural dNTPs. (2% Agarose gel)

Melting temperature experiments performed directly after the completion of the PCR reaction using different ratios of dTTP and 2′-AzE-TTP confirmed that, as indicated on an agarose gel (Figure 2.16), when 100% 2′-AzE-TTP was added to the reaction, an amplicon was not produced. Hence, a T_m could not be determined (Figure 2.17, **E4**). In contrast, for reactions using different ratios of dTTP to 2′-AzE-TTP, amplicons (**E1**, **E2** and **E3**) all displayed the same T_m as the unmodified control (**E0**) at 81.5 ℃.



Samples	Neg	E0	E1	E2	E3	E4
2′-AzE-TTP (%)	0	0	25	50	75	100
TTP (%)	100	100	75	50	25	0

Figure 2.17. Fluorescence melting curves (left) and first derivatives (right) of PCR amplicons from reactions using 2′-AzE-TTP and 1 U KOD XL DNA polymerase. **Neg**, negative control reaction without template; **E0**, positive control reaction with natural dNTPs; **E1**-**E4**, different ratios of 2′-AzE-TTP to natural dTTP (total = 5 nmol).

2.6 Primer extension reactions using 2′-ethynyl-UTP

Due to a failure to incorporate 2′-AzE-TTP into DNA, the use of another 2′-sugar modified nucleotide, 2′-ethynyl-UTP, in the primer extension reaction was attempted. 2′-ethynyl-UTP was initially synthesised by Dr Marius Haugland. Compared to an azidoethoxy moiety at C2′ position, the ethynyl group is less bulky and still capable of

undergoing click chemistry reactions for further functionalisation.[334] An initial time course experiment carried out with 2 U KOD XL DNA polymerase and supplemented 1 mM $MnCl_2$ using T2 and T3 templates (Table 2.1) gave poor incorporation (Figure 2.18 **A**), especially for the newly synthesised template T3. This contains six modified UTP incorporation sites, four of which are conseceutive, making the enzymatic reaction very challenging.

When the reaction using T2 was continued for 2 h, the product started to degrade (Figure 2.18 **A**, lane **1e**) and the reaction did not go to completion after shorter periods of up to 1 h (Figure 2.18 **A**, lane **1d**). The reaction using the more challenging template T3, which requires successive 2′-ethynyl-UTP incorporation at the start of the elongation site, did not incorporate a single nucleotide (Figure 2.18 **A**, lanes **2a**–**2e**).

Using the less challenging template T2, the amount of Mn^{2+} (0–10 mM) was systematically varied in an attempt to increase the tolerance of the DNA polymerase for the modified substrate. Supplementation of 0.5–1 mM $MnCl_2$ in the reaction mixture afforded the greatest yield of longer extension products, but the full-length product was still not detected (Figure 2.18 **B**, lanes **3b** and **3c**). Higher concentrations of Mn^{2+} in the reaction appeared to suppress the primer extension reaction (Figure 2.18 **B**, lanes **3d**–**3f**).

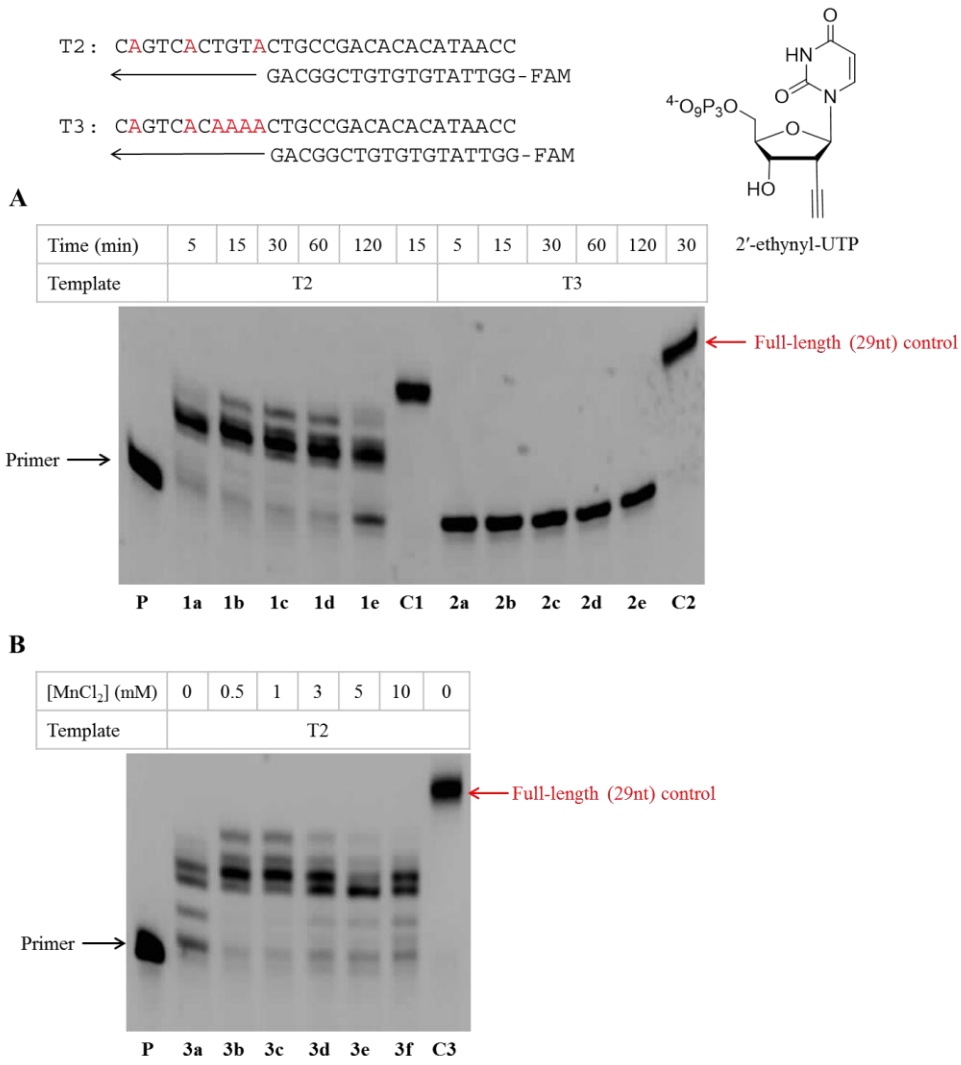


Figure 2.18. Optimisation of 2 U KOD XL DNA polymerase incorporation of 2′-ethynyl-UTP over time (5–120 min) for two different DNA templates (T2 and T3). 2′-ethynyl-UTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ for 30 min with 3.2 nmol of dATP, dCTP, dGTP and 2′-ethynyl-UTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using T2 template; lanes **2a-2e**, time course experiment using T3 template; lanes **3a-3f**, different amount of $MnCl_2$ added in the reaction using T2 template; lanes **C1**/**C3**, positive control, T2 template with natural dNTPs; lane **C2**, positive control, T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

2.7 Conclusions

Initially, a modified template containing 2′-AmE-T was used in the primer extension reaction to test the ability of DNA polymerases to read the 2′-AmE-T modification and incorporate dATP in the antisense strand. Among four different polymerases used in the reaction, KOD XL DNA polymerase was the most compatible with this modification probably due to the lack of 3′–5′ exonuclease activity of this enzyme. Optimal results were obtained with addition of 1 mM $MnCl_2$ to the reaction mixture.

Then, the ability of DNA polymerases to incorporate 2′-modified triphosphates was investigated. A modified triphosphate, 2′-AzE-TTP, was also synthesised by adaptation of the Yoshikawa method, and obtained in 16% yield. Primer extension reactions using this 2′-AzE-TTP in place of canonical dTTP were attempted using a canonical DNA template that contains dispersed dA sites. Even after several carerful attempts to optimise the reaction conditions, 2′-AzE-TTP could not be successfully incorporated into DNA by primer extension. Attempts to incorporate this triphosphate into an amplification product during PCR were also unsuccessful. It is known that bulky modifications at the 2′-sugar position is limited by the steric gate of DNA polymerase,[166,270,351] which may explain the low tolerance of polymerase enzyme for this modified triphosphate.

Efforts to compare the incorporation efficiency of a less bulky modification, 2′-ethynyl-UTP, suggested that this modification is also poorly tolerated by DNA polymerases under the optimised reaction conditions. In particular when a challenging unmodified template with clustered dA sites was used, the polymerase did not incorporate a single modified UTP into the complementary strand. DNA polymerases therefore appear to have a very low tolerance for modified triphosphates containing 2′-sugar modifications.

Although some 2′-sugar modified nucleotides such as 2′-F-, 2′-NH_2-dNTPs are accepted by T7 RNA polymerase, the bulkier modifications such as 2′-OMe- and 2′-N_3-dNTPs are not easily accepted by wild-type RNA polymerase and mutant T7 RNA polymerase is required to accept these bulky modifications.[353] With regards to the previous studies using 2′-sugar modified building blocks, 2′-AzE-TTP appears not to be a good substrate for DNA polymerases. Further studies to synthesis DNA or RNA containing 2′-AzE-TTP can be tried using wild type or mutant T7 RNA polymerase. Yet, it is outside of the scope of this study.

Chapter 3

Synthesis of C5 base modified dUTP and their enzymatic reactions

3 Synthesis of C5 base modified dUTP and their enzymatic reactions

3.1 Background

Another important aspect in nucleic acid aptamer generation is the expansion of chemical and structural diversity by incorporation of modified nucleobases. The objectives of introducing modifications are to enhance the strength and variety of intermolecular interactions in which the aptamers can partake beyond those which are available in the canonical nucleotide structures. Over decades, a large number of base modifications have been used to increase the binding affinity and specificity of aptamers towards their targets.[187,288,354]

When incorporating modified triphosphates, high proofreading capability of DNA polymerases is required to prevent the incorporation of incorrect dNTPs. Surprisingly, family B DNA polymerases tend to tolerate non-canonical modified nucleotides as substrates, and a large number of modified nucleotides have been enzymatically introduced on DNA strands.[176,257]

The C5 position of deoxyuridine is the most frequently modified site owing to the relative ease of introducing chemical modifications at this position, as well as the minimal disruptive effects of modifications at this position on overall duplex stability. The enzymatic incorporation efficiency of unnatural nucleotides depends on several factors. The most important factors affecting the substrate properties of the nucleotide are thought to be the length and rigidity of the linker and the nature of the chemical group.[258,288,355] Based on the literature, deoxyuridine triphosphates bearing an alkynyl linkage at the C5 position are well tolerated by DNA polymerases.[257,356–358] A propargylamino linkage at C5 position of deoxyuridine (Figure 3.1) has been used in

our group to add functional groups via amide bond formation.[359] Modified triphosphates with a propargylamino linker were successfully synthesised by primer extension and PCR.[359] The success of this approach is probably because the rigid alkynyl linkage protrudes from the major groove of DNA in its double-stranded form. Therefore, the modification is firmly oriented in a specific direction and base-pairing is not perturbed by the modification.[360] The efficiency of enzymatic incorporation of such amide derivatives by DNA polymerases will be discussed in this chapter.

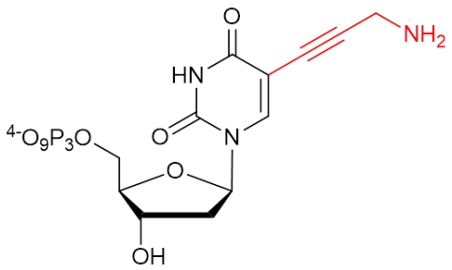


Figure 3.1. 2′-deoxyuridine-5′-triphosphate with propargylamine as a linker at C5 position.

Two distinct types of modifications were incorporated via this labeling method. The first approach focuses on incorporation of hydrophobic amino acid side chains in order to increase the hydrophobic characteristics of DNA to more closely mimic antibody-target binding interactions. The major difference between antibody-antigen interactions compared to naturally occurring protein-protein binding forces is that antibodies predominantly have aromatic amino acids in their binding region. The hydrophobic, aromatic and van der Waals forces that these aromatic groups participate in are generally very short range interactions and therefore, effectively pull together two surfaces which are complementary in shape to generate strong binding.[335,336] Taking this into account, aptamers with hydrophobic modifications may exhibit increased target binding affinities with additional antibody-like character. This hypothesis is exemplified by the recent work by Janjic and co-workers at Somalogic who used hydrophobic

modifications to develop aptamers with very high affinity and selectivity for different epitopes of proteins.[184]

In addition, this chapter will investigate the development of aptamers which exploit a class of non-covalent sigma (σ) hole interactions, called halogen bonding (XB). Halogen bonding project was done in collaboration with Prof. Paul Beer. I would like to express my appreciation to Prof. Paul Beer for his expertise and support given through the collaborative work.

The XB is formally defined as an interaction which "occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity".[363] In the commonly invoked σ-hole model of XB, the distribution of electron density surrounding a covalently bonded halogen atom is predicted to be anisotropic, resulting in the presence of a localised eletropositive crown –the σ-hole–, which imparts directional Lewis acid (X) character to the halogen atom, enabling it to attract a variety of electron rich Lewis bases (Y) (Figure 3.2 **A**). XB interactions are highly directional, with R–X···Y bond angles close to 180°, and are comparable in strength to hydrogen bonding.[364] The strength of XB is also tunable as it depends on several intrinsic properties of the XB donor moiety. Firstly, the halogen atom itself can be selected from Cl, Br and I, where iodine has the highest polarisability, and therefore the largest σ-hole, and chlorine has the lowest polarisability (the least polarisable halogen, fluorine does not participate in XB interactions, except in a small number of unusual cases).[365] Secondly, the greater electron-withdrawing ability of the scaffold that the halogen atom is covalently bonded to increases the size of the σ-hole, leading to greater interaction energy.[366] While the purely electrostatic σ-hole model provides a conceptually simple

rationalisation for most properties of XB interactions, it is still controversial. Alternative models which take into account additional contributions from charge transfer, dispersion and induction interactions have also been proposed, and there is ongoing debate about the relative importance of each contribution.

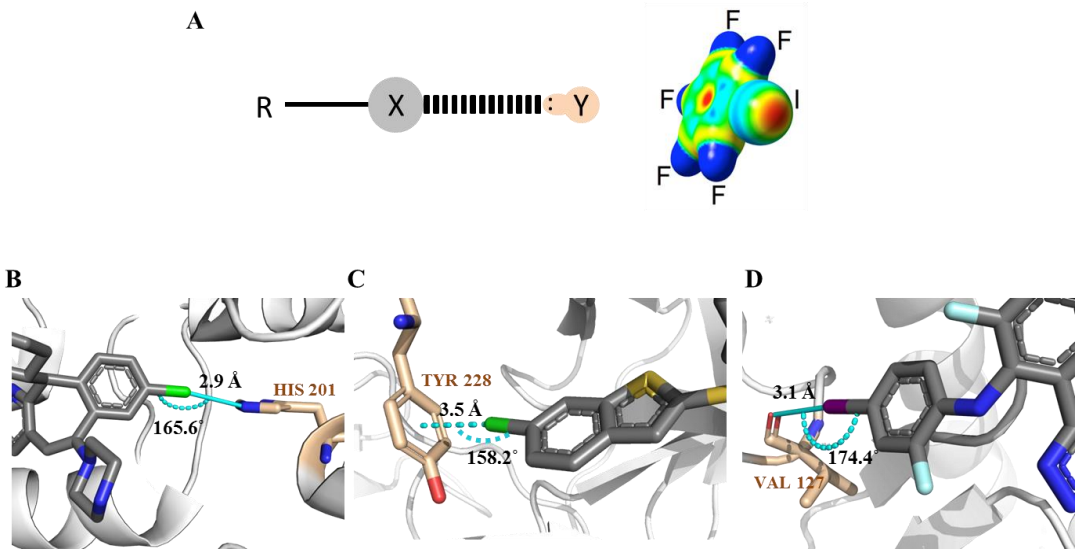


Figure 3.2. Examples of halogen bonds in protein-ligand complexes. **A)** Schematic representation of a halogen-bonding interaction (X = Cl/Br/I; Y = Lewis base) and regions of concentrated negative electrostatic potential (blue) and depleted potential (red) on pentafluoroiodobenzene.[367] Picture was reused from originally published paper[367] and permitted by IntechOpen (see appendix 8.3). **B)** Compound forming Cl···N contact with the nitrogen of HIS 258 in FPT (PDB code 2BED).[368] **C)** Compound forming Cl···π contact with the phenyl ring of TYR 228 in coagulation factor Xa (PDB code 1NFU).[369] **D)** Compound forming I···O contact with the carbonyl oxygen of Val 127 in MEK1 (PDB cod 3EQB).[370] The structural presentations were edited by me using PyMOL program.[371]

XB interactions have been investigated theoretically and in the solid state since the 1950s, and have been used as a tool in solid state crystal engineering applications for around two decades. More recently, investigations into their capacity to drive solution-phase assembly processes have received intensive attention within the supramolecular chemistry field,[372] but the potential of XB interactions is less widely appreciated among the biological research community.[373] Nonetheless, retrospective analyses of

crystallographic data files from the Protein Data Bank have revealed the existence of XB interactions in a significant percentage of solid-state complexes of proteins and halogenated ligands (Figure 3.2 **B**, **C** and **D**).[366,374,375] A smaller number of XB interactions between proteins and halogenated ligands have also been classified in solution, and there is now significant interest in the rational design of drug molecules which interact with their targets via halogen bonding interactions. In the nucleic acids field, XB interactions have been mostly overlooked,[376] with one important exception, Ho and co-workers observed that a bromouracil modification, which was introduced into a decamer oligonucleotide for crystallographic phasing purposes, stabilises the formation of a Holliday junction in the solid state via the formation of an XB interaction between the bromine atom and a phosphate oxygen acceptor.[377] They went on to investigate the effect of different halouracil XB donor groups on the stability of the Holliday junction in solution and in the solid state.[378–381] Although XB interactions are currently unknown in the aptamer field, hundreds of examples in which the amino acid chains of a protein accept XBs from halogenated ligands are now known.[366,374,375] Therefore, for an aptamer-target application it is foreseeable that amino acid residues in proteins could potentially act as Lewis bases to form strong halogen bonding interactions with halogenated nucleic acids. Importantly, while alternative intermolecular interactions such as hydrogen bonding are often unable to compete effectively when exposed to the solvent in biologically relevant, aqueous environments, there is growing evidence that XB interactions are comparatively resistant to competing polar, protic solvents[382–384] and this property makes XB interactions very attractive in the context of aptamer-target interactions. In this chapter, we investigate the ability of polymerase enzymes to tolerate modified nucleoside triphosphates bearing hydrophobic

characteristics or halogen bonding substituents for uses in SELEX to expand the interactions between an aptamer and its target.

3.2 Preparation of modified triphosphates derived from 5-PA-dUTP

As introduced in section 3.1, 5-propargylamino-dUTP (5-PA-dUTP, Figure 3.1) is a good substrate for DNA polymerases and acts as a linker to introduce functional groups on the primary amine.[356,357] 5-PA-dUTP is produced from 5-iodo-dUTP (5-I-dUTP) by a Sonogashira cross-coupling reaction.[385] 5-I-dUTP was first synthesised from its commercially available nucleoside (5′-OH) using optimised Yoshikawa (Scheme 3.1)[236] and Ludwig-Eckstein (Scheme 3.2)[241] methods to increase the yield of the triphosphorylation reaction. The low yield of 2′-AzE-dUTP in chapter 2 was attributed to the very hygroscopic properties of tris(tetrabutylammonium)hydrogen pyrophosphate **2** (Figure 3.3 **A**), combined with the water sensitive nature of the Yoshikawa reaction. In order to avoid this problem, an alternative pyrophosphate salt was tried. In 2015, Korhonen *et al*. reported a synthetic route for making the less hygroscopic salt tris[bis(triphenylphosphoranylidene)ammonium] (PPN) pyrophosphate **4** (Figure 3.3 **B**), which can be used in place of highly hygroscopic tris(tetrabutylammonium) hydrogen pyrophosphate (Scheme 3.1) and the purified triphosphates were obtained in 21–48% yield.[386] However, when PPN pyrophosphate was used in the triphosphorylation reaction 5-I-dUTP **5.1** was obtained in a disappointing yield of 10% mainly due to the formation of by-products from side-reactions at 3′- and, 5′-OH sites.

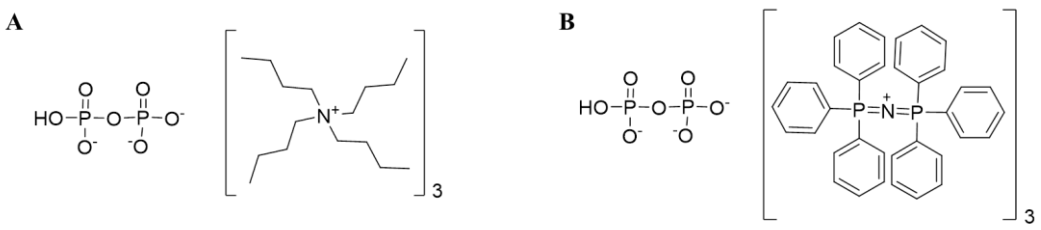
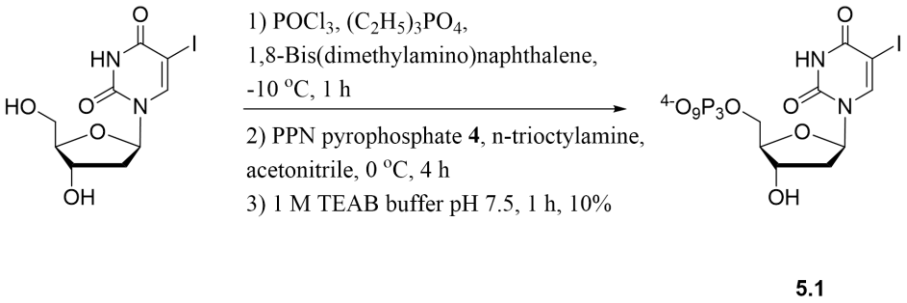
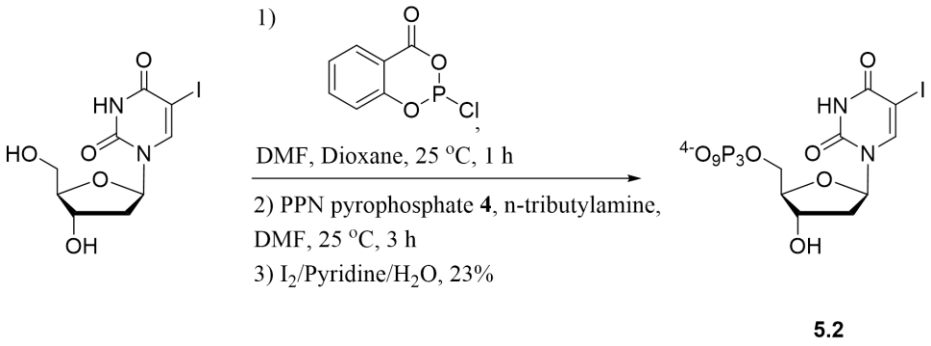


Figure 3.3. **A**) Tris(tetrabutylammonium) hydrogen pyrophosphate **2**. **B**) tris[bis-(triphenylphosphoranylidene)ammonium] (PPN) pyrophosphate **4**.



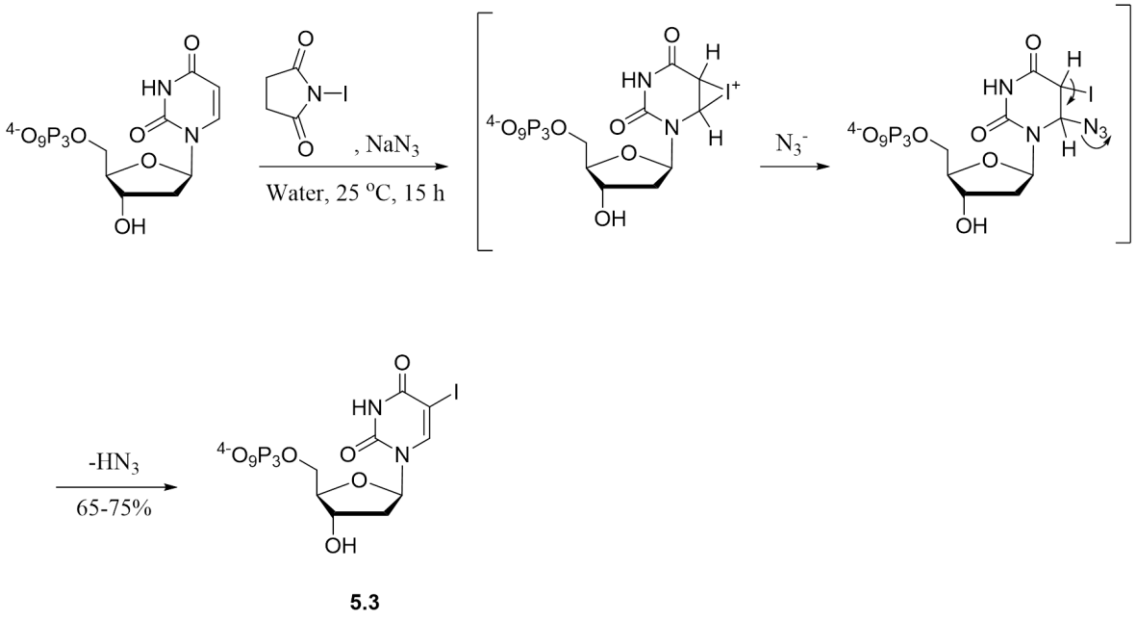
Scheme 3.1. An improved Yoshikawa synthetic route to prepare 5-iodo-2′-deoxyuridine-5′-triphosphate.

For comparison, an alternative triphosphorylation of 5-I-dU **5.2** from nucleoside was carried out using the salicyl phosphochlorite method known as Ludwig-Eckstein reaction (Scheme 3.2).[241] However, the dissolution of the nucleoside in pyridine:dioxane (1:3) solvent mixture was problematic. Analysis of the crude product mixture revealed the presence of both unreacted starting material and by-products, while the desired product formed with a low yield of 13%. The solubility issue was mitigated by dissolving the nucleoside in DMF solution, and this method with PPN pyrophosphate yielded a slightly larger amount of the desired product (23%) after a laborious purification process which involved anion exchange chromatography and RP-HPLC to remove the many sider-products formed during the reaction.



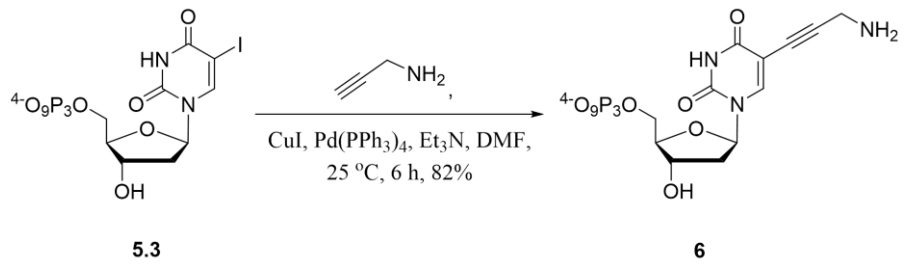
Scheme 3.2. Ludwig-Eckstein synthesis method to prepare 5-iodo-2′-deoxyuridine-5′-triphosphate.

In light of the low yields and poor recovery of 5-I-dUTP, a recently published new method of synthesising this 5-I-dUTP was tested, which avoids the triphosphorylation reaction.[387] Commercially available 2′-deoxyuridine-5′-triphosphate (dUTP) was used as a precursor, and this was treated with *N*-iodosuccinimide and sodium azide at 25 ℃ overnight to afford 5-I-dUTP **5.3** in a high yield of 65–75% after ion exchange purification. An azido group at the C6 position of dUTP plays an important role in these C5 halogenation reactions. The azido group at the C6 position of the intermediates is an excellent leaving group for regeneration of the 5,6-olefinic bond (Scheme 3.3). Kumar *et al*. proved that low yield of iodination of 2′-deoxyuridine was given in the absence of sodium azide.[388] Therefore, the halogenation reaction at the C5 position of dUTP using an electrophilic *N*-iodosuccinimide and sodium azide was shown to be effective synthesis route. More importantly, the extent of by-product formation (3′-monophosphate, 5′-monophosphate and 3′- and 5′-diphosphates), which is the major impediment for Yoshikawa and Ludwig-Eckstein methods was substantially decreased.



Scheme 3.3. A new synthesis method to prepare 5-iodo-2′-deoxyuridine-5′-triphosphate (5-I-dUTP) from 2′-deoxyuridine-5′-triphosphate (dUTP) via making 5-halo-6-azido-5,6-dihydro intermediates.

The propargylamino group was then introduced to 5-I-dUTP **5.3** by a Sonogashira cross-coupling reaction to provide 5-PA-dUTP **6** (Scheme 3.4).[389,390] Thorough degassing of the reagents was found to induce a substantial increase in yield, from 21% to 82%.



Scheme 3.4. Sonogashira coupling reaction for propargylamino group introduction.

3.2.1 Synthesis of modified triphosphates bearing hydrophobic substituents

The amino acid side chains labels phenylalanine **7**, alanine **8**, isoleucine **9** and tryptophan **10** were appended on to the 5-PA-dUTP **6** via amide bond formation (Figure 3.4). The primary amine is a suitable attachment point for active ester derivatives of these amino acids. Commercially available Fmoc-protected amino acid pentafluorophenyl active esters were incubated with 5-PA-dUTP at 45 ℃ for 3 h in 1:1 ratio of 1 M TEAB buffer (pH 8.5) to DMF.[391] The Fmoc group was cleaved by heating at 55 ℃ in aqueous ammonia for 2 h, and the product was purified by RP-HPLC. However, the yields for the amide coupling step were low (13–32%) with a considerable amount of unreacted 5-PA-dUTP starting material being recovered after purification.

Yields were particularly low for phenylalanine **7** and isoleucine **9** which are quite hydrophobic. Therefore, the ratio of 1 M TEAB buffer (pH 8.5) to DMF was changed from 1:1 to 1:2 to increase their solubility in the reaction mixture (Scheme 3.5). 1-Hydroxy-1*H*-benzotriazole (HOBt) was also added to reduce epimerisation in the product.[233,392] Ultimately, the optimised conditions increased the yield to around 50%.

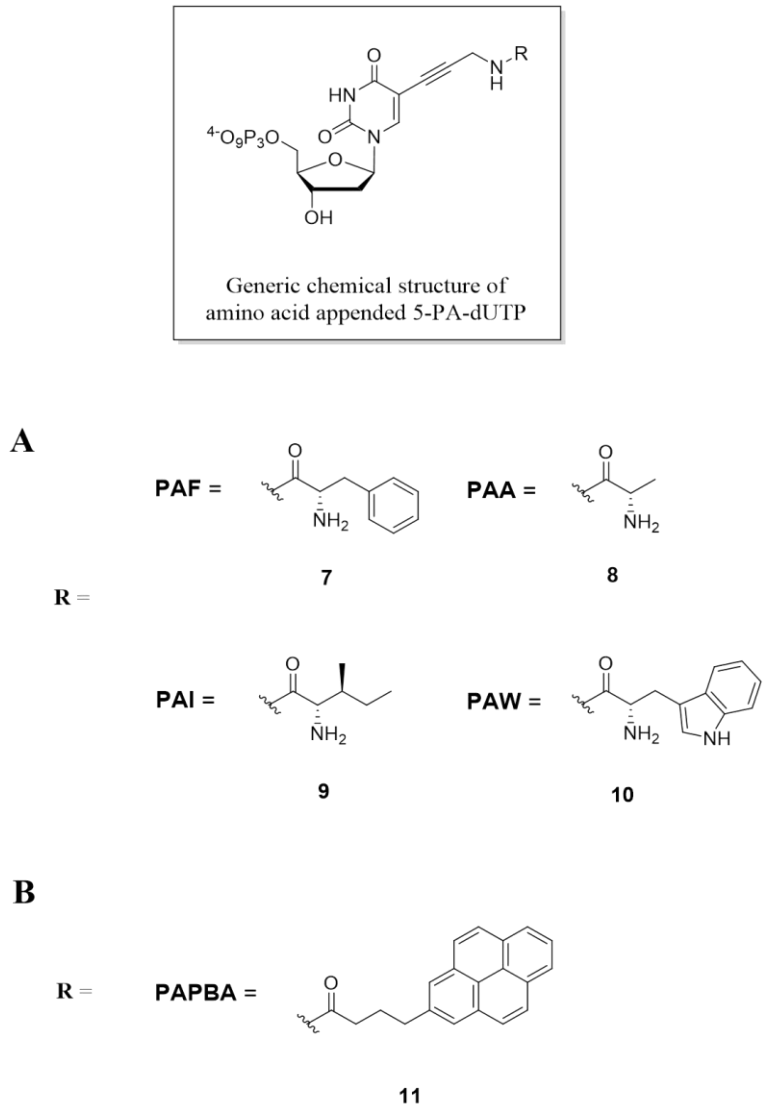
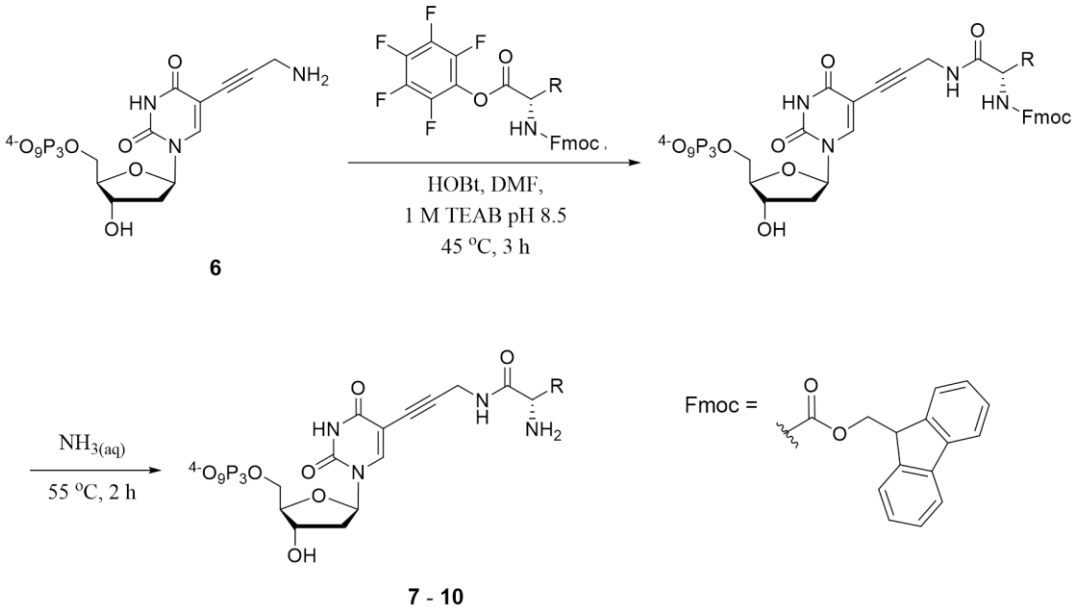
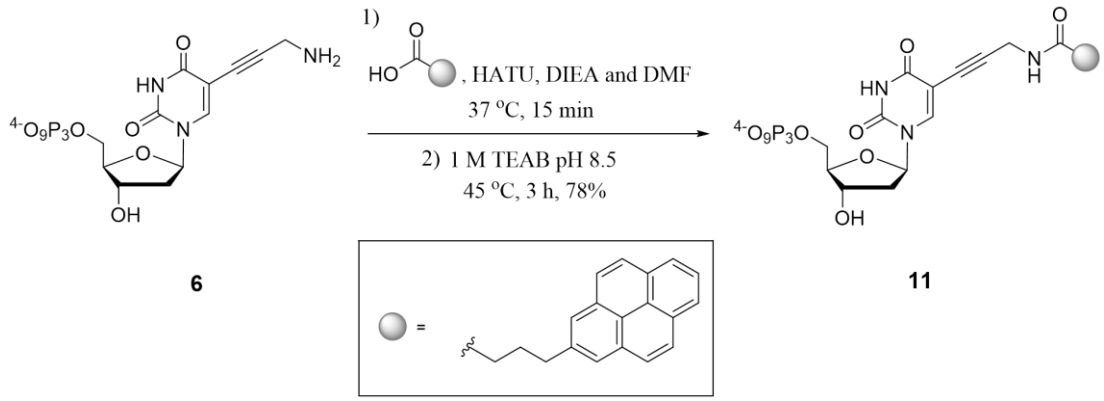


Figure 3.4. Labelling of 5-PA-dUTP with hydrophobic moieties. Labelling condition: **A**) Active ester, HOBt, DMF and 1M TEAB (pH 8.5), 3 h, 45 ℃ and **B**) carboxylic acid, HATU, DIEA, DMF and 1 M TEAB (pH 8.5), 3 h, 45 ℃.



Scheme 3.5. Reaction scheme for Fmoc protected amino acid (R) coupling with 5-PA-dUTP **6**. For structures of R substituents, see Figure 3.4 **A**.

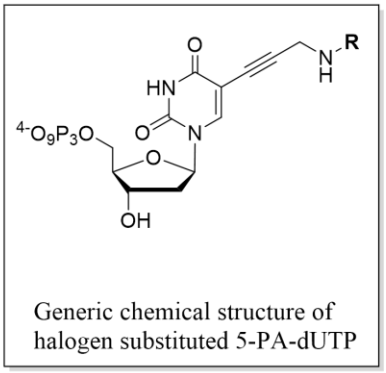
Another modification, pyrene butyric acid **11** which exhibits significant hydrophobicity was coupled to 5-PA-dUTP **6** via a HATU mediated amide coupling reaction (Scheme 3.6).[393] The carboxylic acid form of the compound (synthesised by Dr Anna Dysko) was activated at 37 ℃ for 15 min in the presence of the HATU coupling agent and DIEA in anhydrous DMF before introduction of an aqueous solution of the amine.



Scheme 3.6. Reaction scheme for pyrenebutyric acid labelling with 5-PA-dUTP using HATU reagent.

3.2.2 Synthesis of modified triphosphates bearing halogen bonding substituents

Two different iodine-containing functional groups were appended onto 5-PA-dUTP **6** via amide bond formation as potential XB donor groups. In the first example, **12**, a phenyl ring was used as a scaffold to hold an iodine halogen bond donor group in addition to two inductively electron withdrawing fluorine atoms which serve to increase the polarisation of the electron density surrounding the iodine atom, and hence, the σ-hole size. As a control, a difluoro-substituted phenyl ring which lacks an iodine XB donor substituent was also synthesised (**13**). The second halogen bonding example, **14**, incorporates an iodotriazole XB donor motif. The three electronegative nitrogen atoms in the heterocyclic triazole scaffold withdraw electrons to strongly polarise the C–I bond and thereby increase the σ-hole size (Figure 3.5).



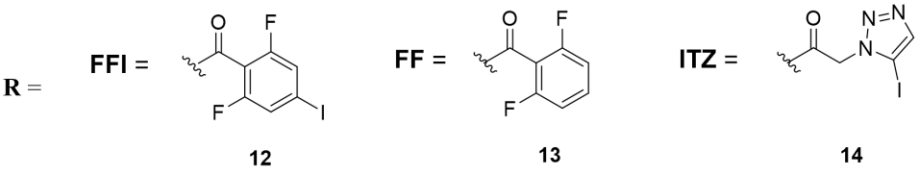
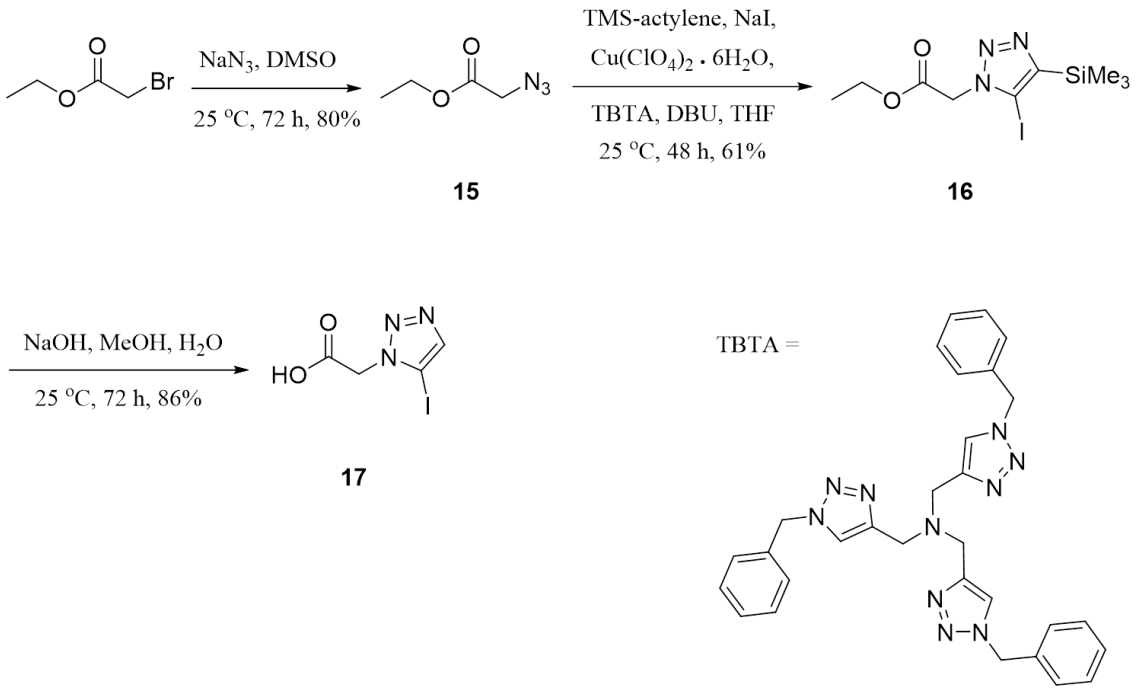


Figure 3.5. Labelling of 5-PA-dUTP with halogen containing moieties. Labelling condition: Carboxylic acid, HATU, DIEA, DMF and 1 M TEAB pH 8.5, 15 h, 25 ℃.

5-PA-dUTP **6** was labelled with 2,6-difluoro-4-iodobenzoic acid, 2,6-difluorobenzoic acid and 5-iodo-1,2,3-triazole-1-acetic acid via HATU coupling reactions. The labelling

reactions were carried out at 25 ℃ for 15 h. The 2,6-difluorobenzoic acid and 2,6-difluoro-4-iodobenzoic acid derivatives were commercially sourced and 5-iodo-1,2,3-triazole-1-acetic acid was chemically synthesised starting from ethyl bromoacetate, which was first converted to ethyl azidoacetate **15** by treatment with sodium azide. Ethyl azidoacetate was then converted to 5-iodo-4-TMS-1,2,3-triazole-1-ethylacetate **16** using an adapted version of the conventional CuAAC 'Click' reaction.[394] Finally, the trimethylsilyl and ethyl ester protecting groups were cleaved using a base-catalysed double hydrolysis reaction, to afford 5-iodo-1,2,3-triazole-1-acetic acid **17** (Scheme 3.7).



Scheme 3.7. Synthesis route for 5-Iodo-1,2,3-triazole-1-acetic acid **17**.

3.3 Primer extension reactions using modified dUTP bearing hydrophobic modifications

In section 3.2.1, hydrophobic amino acids side chains phenylalanine, tryptophan and isoleucine, a less hydrophobic amino acid (alanine), and a strongly hydrophobic moiety with a large aromatic surface, pyrene, were coupled to 5-PA-dUTP to produce a range of modified dNTPs possessing hydrophobic characteristics. As described in section 1.5.4.3, primer extension is an efficient and simple method to introduce modified triphosphates into DNA. The DNA templates, T2 and T3, which were used in chapter 2, were used to investigate incorporation of the new modified triphosphates possessing hydrophobic side chains (Table 2.1). Template T2 contains three separated adenine bases in the extension region and T3 has six adenine bases, four of which occur consecutively at the start of the extension region. In order to accurately copy these template sequences, the DNA polymerase needs to incorporate four modified dUTPs in succession, making this a very challenging sequence. Both templates (T2 and T3) have the same primer binding region where the 6-carboxyfluorescein (FAM) labelled primer binds (P1, Table 2.1). Fully elongated or partially extended products were analysed by 20% denaturing PAGE gel. Natural dNTPs (dATP, dCTP, dGTP and dTTP) were used in positive control experiments.

In Chapter 2, KOD XL DNA polymerase was proved to be the most suitable for the enzymatic production of modified DNA containing modified nucleotides when the buffer was supplemented with 1 mM $MnCl_2$. The effects of Mn^{2+} were also discussed in Chapter 2 (section 2.1). Initially, 5-propargylamino-phenylalanine-2′-deoxyuridine-5′-triphosphate (5-PAF-dUTP) was used in the reaction using 2 U of each KOD and KOD XL DNA polymerases at 65 ℃ for 1 h (Figure 3.6). KOD XL DNA polymerase

produced full-length products within 1 h (Figure 3.6, lanes **1c** and **2c**), and the results were fully consistent with base-modified nucleotide incorporation. The bands corresponding to the full-length modified products have slightly lower electrophoretic mobility compared to the unmodified counterparts (Figure 3.6, lane **C+**) as a result of the higher MW and structure of the modified products. UPLC-MS analysis of the product mixtures confirmed that in each case the masses of the products correlate with the theoretical mass of the full length modified or unmodified extension product. KOD DNA polymerase gave less promising results and was found to degrade this DNA, degradation extending even into the primer (Figure 3.6, lanes **1a**, **1b**, **2a** and **2b**). This is most likely due to enhanced 3′–5′ exonuclease activity of KOD DNA polymerase over the extended incubation time. Reactions with KOD DNA polymerase were attempted in both KOD XL buffer (Figure 3.6, lanes **1b** and **2b**) and KOD DNA polymerase with KOD buffer (Figure 3.6, lanes **1a** and **2a**), but the results were similar in both case. These results indicate that the incorporation of modified triphosphates is not affected by the buffer composition but is more dependent on the 3′–5′ exonuclease activity of DNA polymerase. Hence, further enzymatic reactions were performed with KOD XL polymerase.

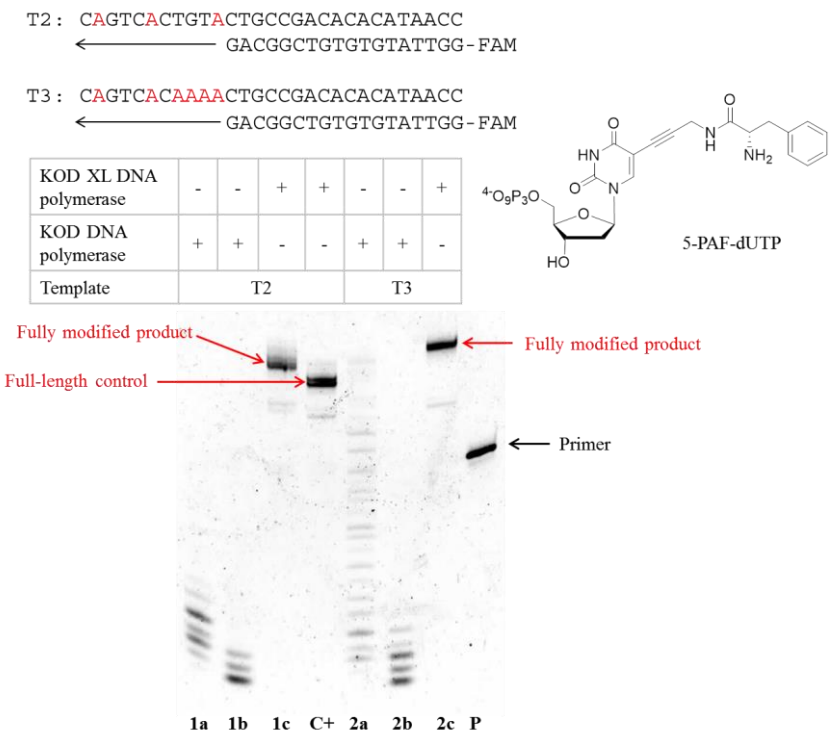


Figure 3.6. Enzymatic incorporation of 5-PAF-dUTP using two different DNA templates (T2 and T3) as a function of polymerases (2 U) and buffers. 5-PAF-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ for 1 h using 3.2 nmol of dATP, dCTP, dGTP and 5-PAF-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a**/**2a**, reactions with KOD DNA polymerase and KOD buffer; lanes **1b**/**2b**, reactions with KOD DNA polymerase and KOD XL buffer; lanes **1c**/**2c**, reactions with KOD XL DNA polymerase and KOD XL buffer; lane **C+**, positive control, T2 template with natural dNTPs using KOD XL DNA polymerase and KOD XL buffer; lane **P**, primer P1. (20% Denaturing PAGE)

A time course experiment revealed that primers were fully extended after 15 min using KOD XL DNA polymerase for both templates T2 and T3; the shortest time frame tested (Figure 3.7, lanes **1a** and **2a**). The extension of modified strands was compared to the positive control strands (all unmodified dNTPs). In order to confirm the full extension of the positive control sample, two reactions containing the unmodified dNTPs were incubated for 60 min and 120 min (Figure 3.7, lanes **C1** and **C2**, respectively). The fully extended primer in lane **2a** migrated more slowly on the denaturing PAGE than that in

lane **1a** due to its higher MW and its structure: the product from extension of template T3 shown in lane **2a** contains six 5-PAF-dUs whereas that from extension of template T2 shown in lane **1a** contains only three 5-PAF-dUs. As mentioned in Chapter 2, an extra band is visible above the product bands due to overhanging dA at the 3′-ends. Non-templated addition of nucleotides at the 3′-ends is carried out by prokaryotic and eukaryotic DNA polymerases.[344–346] The successful formation of the fully extended products containing three and six modifications was confirmed by mass spectrometry (Table 3.1).

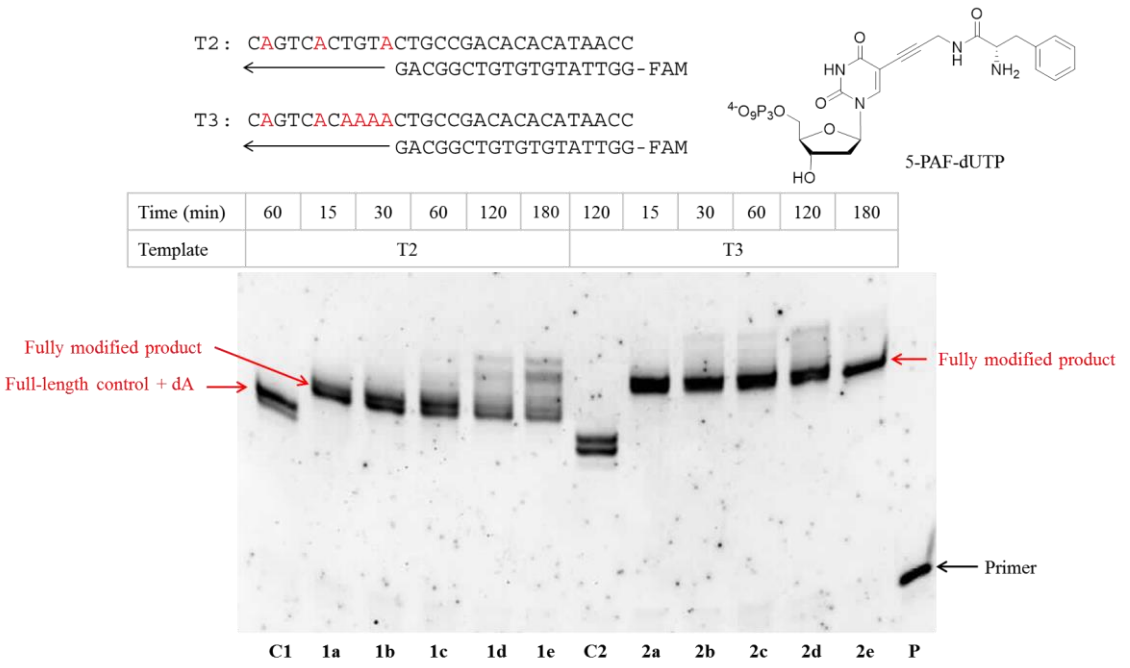


Figure 3.7. Enzymatic incorporation of 5-PAF-dUTP using two different DNA templates (T2 and T3) using 2 U KOD XL DNA polymerase over time (15 min to 3 h). 5-PAF-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-PAF-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e**, time course experiment using a T3 template; lane **C1**, positive control, a T2 template with natural dNTPs; lane **C2**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

Unlike the 2′-sugar modification, C5-base modified nucleotides are well-tolerated by KOD XL DNA polymerase. With this promising result, another modification, 5-propargylamino-tryptophan-2′-deoxyuridine-5′-triphosphate (5-PAW-dUTP) was tested for its incorporation under the same reaction condition as above. A reaction time of 15 min appears adequate to incorporate 5-PAF-dUTP (Figure 3.7, lane **1a**) even with consecutive incorporation (Figure 3.7, lane **2a**). Hence, a shorter incubation time course was tested with 5-PAW-dUTP. Over the times tested (1 min to 30 min), the reaction using template T2 produced a fully elongated product band within 1 min (Figure 3.8, lane **1a**). For the more challenging template T3 a longer incubation of 5 min was required for formation of the full-length product, as the DNA polymerase requires more time to incorporate four contiguous modified bases at the beginning of the extension site (Figure 3.8, lane **2c**). When the reaction using T3 was incubated for only 1 or 2 min, many truncated sequences appeared on the gel (Figure 3.8, lanes **2a** and **2b**). Positive controls were incubated for 15 min and 30 min (Figure 3.8, lanes **C1** and **C2**) to confirm the full extension of the primer so that they could be compared with modified strands band more clearly. Fully elongated product bands on the gel are indicated with red arrows. The primer extension reaction using 5-PAW-dUTP and templates T2 and T3 were also analysed by mass spectrometry (Table 3.1), which confirmed these assignments.

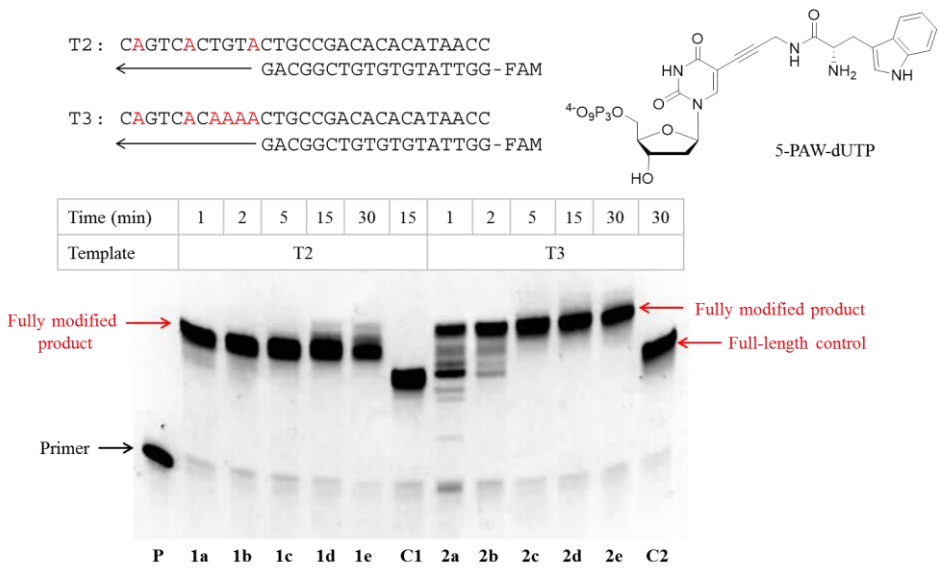


Figure 3.8. Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-PAW-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-PAW-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-PAW-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e**, time course experiment using a T3 template; lane **C1**, positive control, a T2 template with natural dNTPs; lane **C2**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

In addition, 5-propargylamino-isoleucine-2′-deoxyuridine-5′-triphosphate (5-PAI-dUTP) was used in the primer extension reaction as an alternative building block in place of dTTP. The reaction conditions remained the same as above. As isoleucine is less bulky than phenylalanine and tryptophan, 5-PAI-dUTP was incorporated into DNA more rapidly by KOD XL DNA polymerase. The reaction using T2 was accomplished within 1 min (Figure 3.9, lane **1a**) whereas a 2 min incubation time was required for T3 (Figure 3.9, lane **2b**). The reaction was close to completion after 1 min, but a faint trace of lower MW intermediates can be observed, indicating that 2 min is preferable (Figure 3.9, lane **2a**). Mass spectrometry analysis of the products, confirming the formation of fully elongated primers containing three and six 5-PAI-dUTP, is presented in Table 3.1.

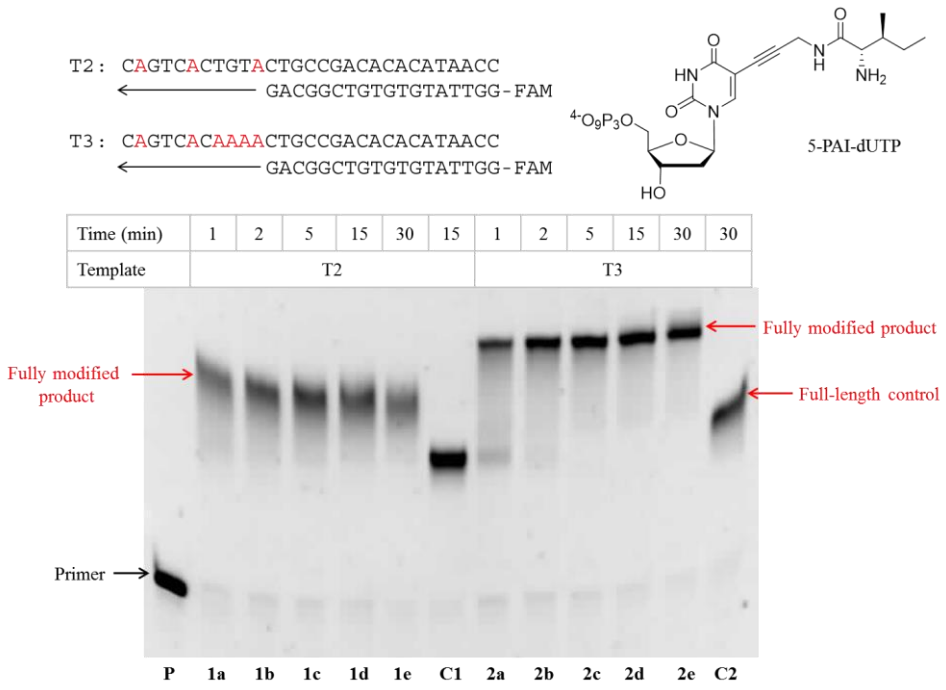


Figure 3.9. Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-PAI-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-PAI-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-PAI-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e**, time course experiment using a T3 template; lane **C1**, positive control, a T2 template with natural dNTPs; lane **C2**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

5-propargylamino-alanine-2′-deoxyuridine-5′-triphosphate (5-PAA-dUTP) was examined in the primer extension reaction under the same reaction conditions. The reaction using T2 reached completion after 1–2 min (Figure 3.10, lanes **1a** and **1b**). The reaction using T3, however, required a longer elongation time of >5 min to remove all traces of incompletely extended intermediate sequences (Figure 3.10, lanes **2a**–**2c**). Although alanine is the smallest of the side-chain modifications tested, it requires the longest reaction time when there are four successive adenines in the template. It is possible that other family B DNA polymerases such as Vent exo- may incorporate 5-PAA-dUTP faster than KOD XL DNA polymerase.[179] Mass spectrometry analysis of

the product mixtures after a 15 min incubation confirmed the formation of fully elongated primers containing three and six 5-PAA-dUTP (Table 3.1).

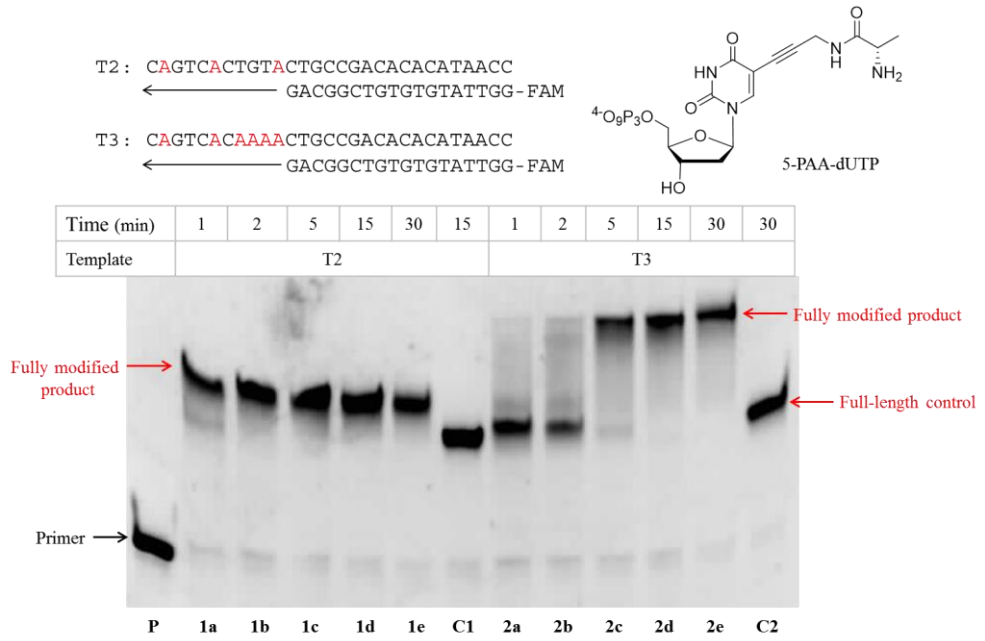


Figure 3.10. Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-PAA-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-PAA-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-PAA-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e**, time course experiment using a T3 template; lane **C1**, positive control, a T2 template with natural dNTPs; lane **C2**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

Finally, the more extreme hydrophobic modification, 5-propargylamino-1-pyrenebutyl-2′-deoxyuridine-5′-triphosphate (5-PAPBA-dUTP) was tested in the primer extension reaction to observe whether this bulky modification can be efficiently incorporated into DNA. Longer incubation times were assessed in this reaction (1 min to 1 h) in anticipation of the more sterically bulky modification taking longer to be incorporated. Over the range of times tested, a fully extended complementary copy of the template T2 was obtained after 15 min (Figure 3.11, lane **1d**). In contrast, the full-length products

were not observed for the more challenging template T3 despite the longer reaction times (Figure 3.11, lanes **2a-2f**). A structural analysis of a Taq DNA polymerase variant (KlenTaq) done by Hottin *et al*. provided a picture of the ternary complex with primer-template DNA.[176] A numbers of base modified nucleotides were bound to the complex and showed how the modified triphosphates are tolerated by DNA polymerase. In order to generate an active enzyme-substrate complex, the finger domain of KlenTaq DNA polymerase changes its conformation from an open state to a closed state and this is achieved by reorientation of O-helix. In contrast to canonical nucleotide binding, one of the amino acids on O-helix, Arg_{660} is reoriented when the modified nucleotide is supplied due to steric hindrance. The residue, Arg_{587} from the thumb domain of the enzyme is also displaced to enable the modification to locate outside of the enzyme through the cavities. The displacement of these two residues highlights the tolerance of modified substrates by DNA polymerase; even bulky groups can be incorporated into DNA if a long linker is used.[176] However, the relatively short linker combined with the large size of the modification of 5-PAPBA-dU may cause steric hindrance in the complex and therefore, may not be incorporated by DNA polymerase. More efficient enzymatic incorporation may be achieved if a longer linker is used.[258,395] Sufficient quantities of products containing 5-PAPBA-dU were generated to confirm the identity of the fully extended products using the T2 template which was also observed by mass spectrometry analysis (Table 3.1). On the other hand, no extension was observed by both gel electrophoresis and mass spectrometry analysis when the T3 template was used in the reaction (Table 3.1).

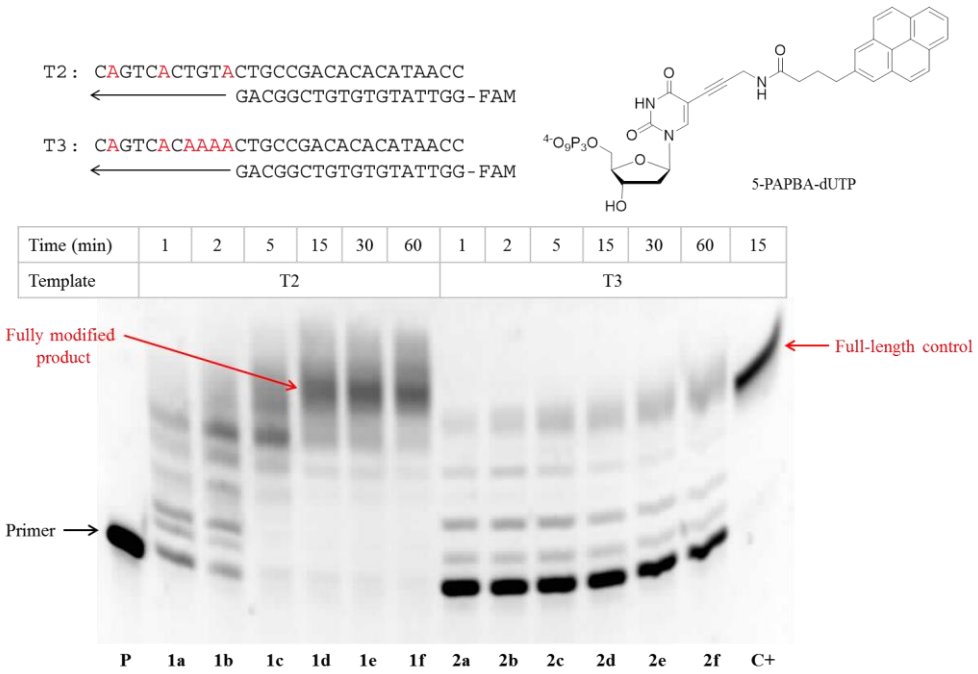


Figure 3.11. Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-PAPBA-dUTP over time (1–60 min) for two different DNA templates (T2 and T3). 5-PAPBA-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-PAPBA-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1f**, time course experiment using a T2 template; lanes **2a-2f**, time course experiment using a T3 template; lane **C+**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% denaturing PAGE)

Table 3.1. UPLC-MS analysis of final primer extension reaction using 5-PAF-, 5-PAW-, 5-PAI-, 5-PAA- and 5-PAPBA-dUTPs. Primer extended products were analysed for both template T2 and T3. For representative mass spectra see appendix 8.5.

Modification	**Template (T2)**		**Template (T3)**	
	Calc. (Da)	**Found (Da)**	**Calc.(Da)**	**Found (Da)**
5-PAF-dUTP	10115.5	10114.7	10671.1	10670.5
5-PAW-dUTP	10232.8	10232.1	10905.7	10905.8
5-PAI-dUTP	10013.5	10012.6 10325.6 (+dA)	10467.1	10780.3 (+dA)
5-PAA-dUTP	9887.3	9886.7	10214.6	10214.3
5-PAPBA-dUTP	10485.0	10484.7	11410.1	Not detectable

3.4 PCR reactions using modified dUTP bearing hydrophobic modifications

The successful enzymatic incorporation of the modified triphosphates 5-PAF-dUTP, 5-PAW-dUTP, 5-PAI-dUTP and 5-PAA-dUTP in primer extension reactions using KOD XL DNA polymerase prompted the study of these modified triphosphates in PCR. It is critically important to test whether the modified nucleotides can be repeatedly incorporated into longer DNA (in this case 81 base pairs), as well as efficient read-through of the modified templates (Figure 3.12). Template T4 and primers P2 and P3 directed incorporation of eleven modified nucleotides in the top strand, and twenty-two modified nucleotides in the bottom strand.

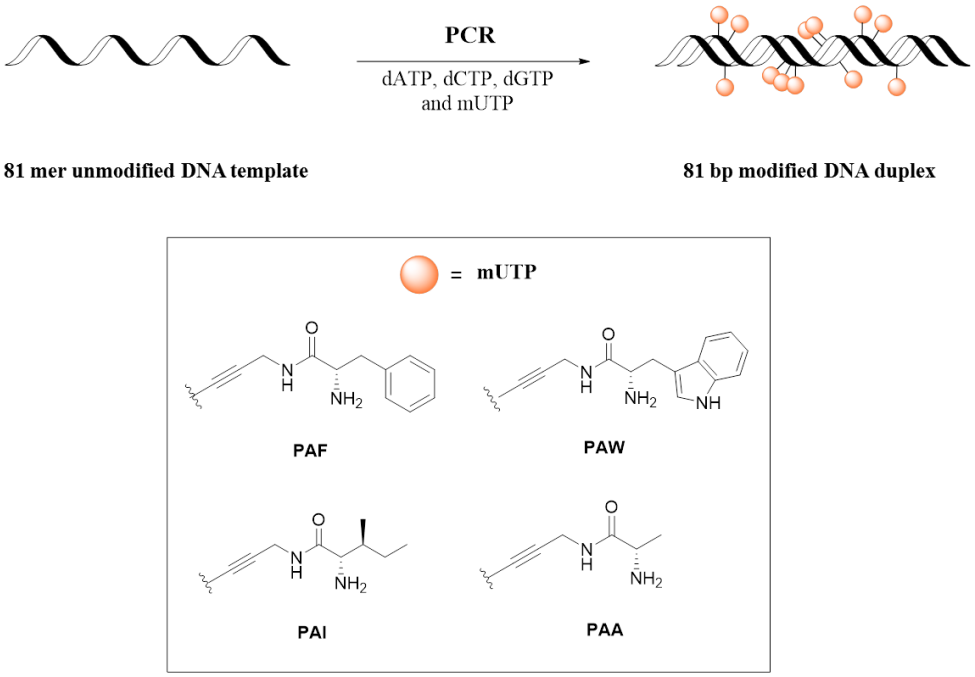


Figure 3.12. PCR using nucleotides containing hydrophobic modifications.

Initially, PCR was carried out using an 81 mer DNA template (T4, Table 2.1) and KOD XL DNA polymerase. The modified triphosphate incorporation sites are well separated

on the PCR template with just two consecutive incorporation sites of modified triphosphate along the 81 mer template. The precursor nucleotide, 5-PA-dUTP was used in PCR, before trying other modifications (Figure 3.13). The use of 5-PA-dUTP in PCR amplification was successful when 100% of 5-PA-dUTP was added. After 20 PCR cycles, the fully modified 81 base pair product was observed on an agarose gel (Figure 3.13, lane **PA**). However, the band intensity of the positive control was very weak due to insufficient number of cycles (Figure 3.13, lane **C+**). Some primer dimerisation occurred during PCR, the nature of which was established by comparison with a negative control reaction performed in the absence of the template (Figure 3.13, lane **N**).

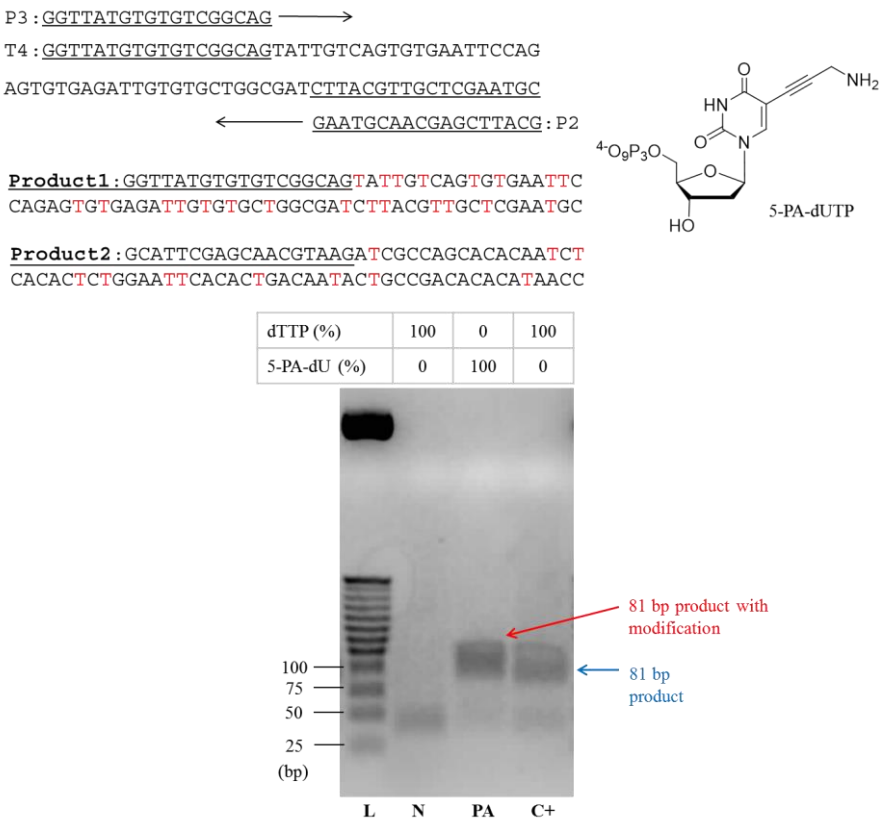


Figure 3.13. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-PA-dUTP, 1 mM $MnCl_2$ and 0.5 U KOD XL DNA polymerase (20 cycles). The positions where 5-PA-dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (25 base pairs); lane **PA**, PCR product containing 5-PA-dUTP; lane **N**, negative control reaction without template T4; lane **C+**, positive control reaction with natural dNTPs. (2% Agarose gel)

The encouraging results of the PCR reaction using 5-PA-dUTP prompted further PCR amplification reactions to be attempted using the modified triphosphates functionalised with hydrophobic side chains. The PCR cycle consisted of a DNA duplex denaturation step (95 ℃ for 30 sec), followed by a primer annealing step at 51 ℃ for 30 sec and an extension step at 72 ℃ for 30 sec. A different number of PCR cycles (20, 25 and 30 cycles) were tried using 5-PAF-dUTP under the same reaction conditions (Figure 3.14). This experiment illustrates that 25 cycles produced a larger quantity of amplicons compared to 20 cycles (Figure 3.14, lanes **F1** and **F2**), as evidenced by the stronger intensity of the distinctive product band on agarose gel. However, KOD XL DNA polymerase usually works less efficiently with a higher number (e.g. 30) PCR cycles, producing higher MW smear bands for both modified products and positive control experiments (Figure 3.14, lanes **C3** and **F3**). Due to the higher MW of the modification compared to unmodified dTTP, all modified 81 base pair PCR amplification products migrated slower than the analogous unmodified positive controls on the gel (Figure 3.14, lanes **F1** and **C1**)

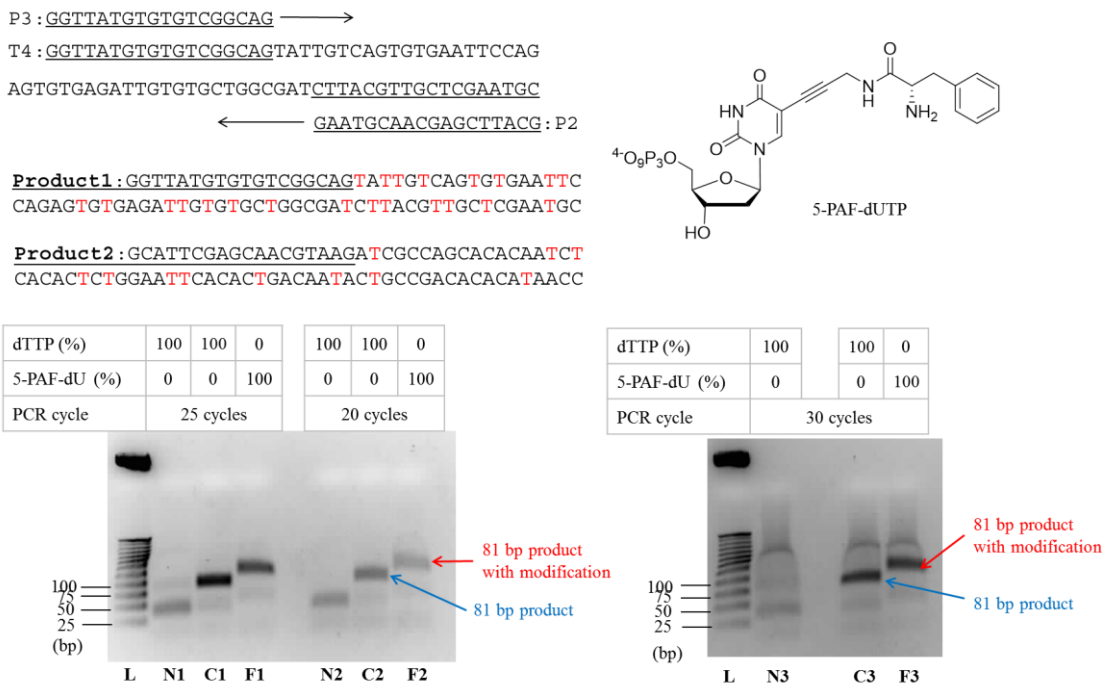


Figure 3.14. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-PAF-dUTP, 1 mM $MnCl_2$ and 0.5 U KOD XL DNA polymerase (20–30 cycles). The positions where 5-PAF-dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (25 base pairs); lanes **F1**/**F2**/**F3**, PCR products containing 5-PAF-dU after 25, 20 or 30 cycles; lanes **N1**/**N2**/**N3**, negative control reaction without template after 25, 20 or 30 cycles; lanes **C1**/**C2**/**C3**, positive control reaction with natural dNTPs after 25, 20 or 30 cycles. (2% Agarose gel)

Consequently, 25 cycles of PCR under the same conditions were conducted with 5-PAW-dUTP, 5-PAI-dUTP (Figure 3.15) and 5-PAA-dUTP (Figure 3.16). The modified products are indicated with a red arrow and positive controls containing only natural dNTPs are indicated with a blue arrow.

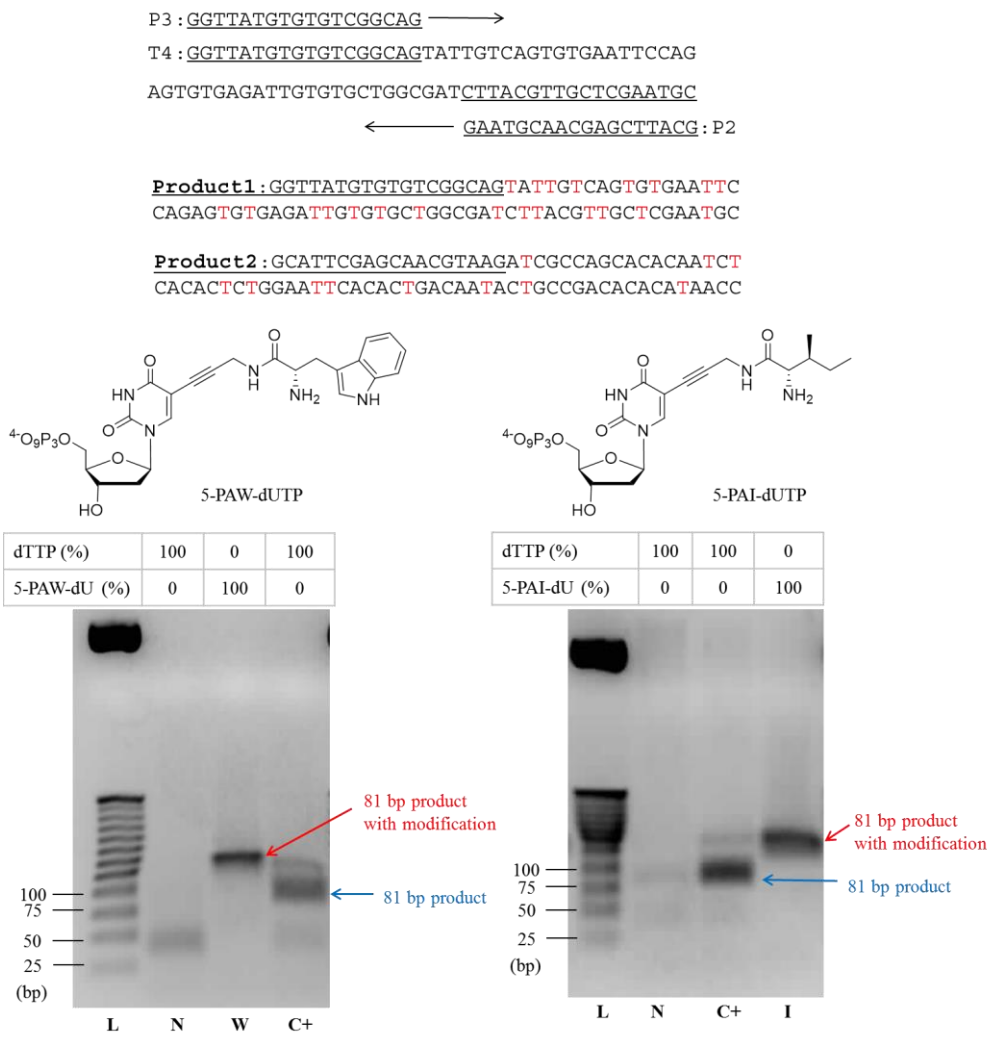


Figure 3.15. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-PAW-dUTP or 5-PAI-dUTP, 1 mM $MnCl_2$ and 0.5 U KOD XL DNA polymerase (25 cycles). The positions where modified dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (25 base pairs); lane **W**, PCR product containing 5-PAW-dU; lane **I**, PCR product containing 5-PAI-dU; lane **N**, negative control reaction without template T4; lane **C+**, positive control reaction with natural dNTPs. (2% Agarose gel)

The PCR reactions were observed to proceed efficiently in the presence of the 5-PAW-dUTP and 5-PAI-dUTP modified triphosphates, but the reaction using 5-PAA-dUTP produced the modified amplification products in a comparatively low yield (Figure 3.16, lane **A**). This result is consistent with the primer extension reaction results, where 5-

PAA-dUTP was shown to require longer reaction time for efficient incorporation into DNA by DNA polymerase (Figure 3.10).

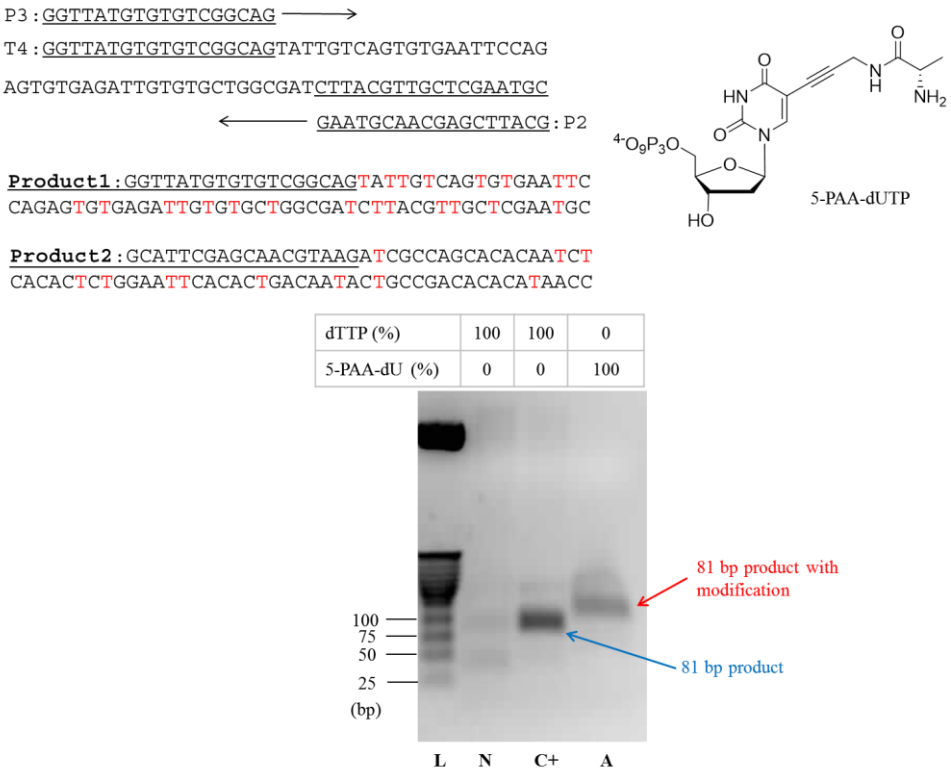


Figure 3.16. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-PAA-dUTP, 1 mM $MnCl_2$ and 0.5 U KOD XL DNA polymerase (25 cycles). The positions where 5-PAA-dUTPs were incorporated are highlighted in red.Lane **L**, DNA ladder (25 base pairs); lane **A**, PCR product containing 5-PAA-dU; lane **N**, negative control reaction without template T4; lane **C+**, positive control reaction with natural dNTPs. (2% Agarose gel)

Figure 3.17 shows the T_m curves (left) and the first derivatives (right) of PCR amplicons containing 5-PAF-dUTP (Figure 3.17 **A**) or 5-PAW-dUTP (Figure 3.17 **B**). These melting temperature experiments were performed on a Bio-Rad CFX96. PCR amplification was carried out using the intercalating dye, SYBR green to provide a fluorescent signal and followed by the melting temperature experiments. The DNA duplexes containing 5-PAF-dU or 5-PAW-dU dissociated at 84 ℃ or 84.5 ℃,

respectively, whereas the corresponding melting temperature for a DNA duplex containing natural dTTP was 80 ℃. The results demonstrate that the 5-PAF-dU and 5-PAW-dU modifications stabilise DNA duplexes. Also, melting curves for the control reactions performed in the absence of the template indicate primer dimerisation (Figure 3.17 **A**, **NTC**).

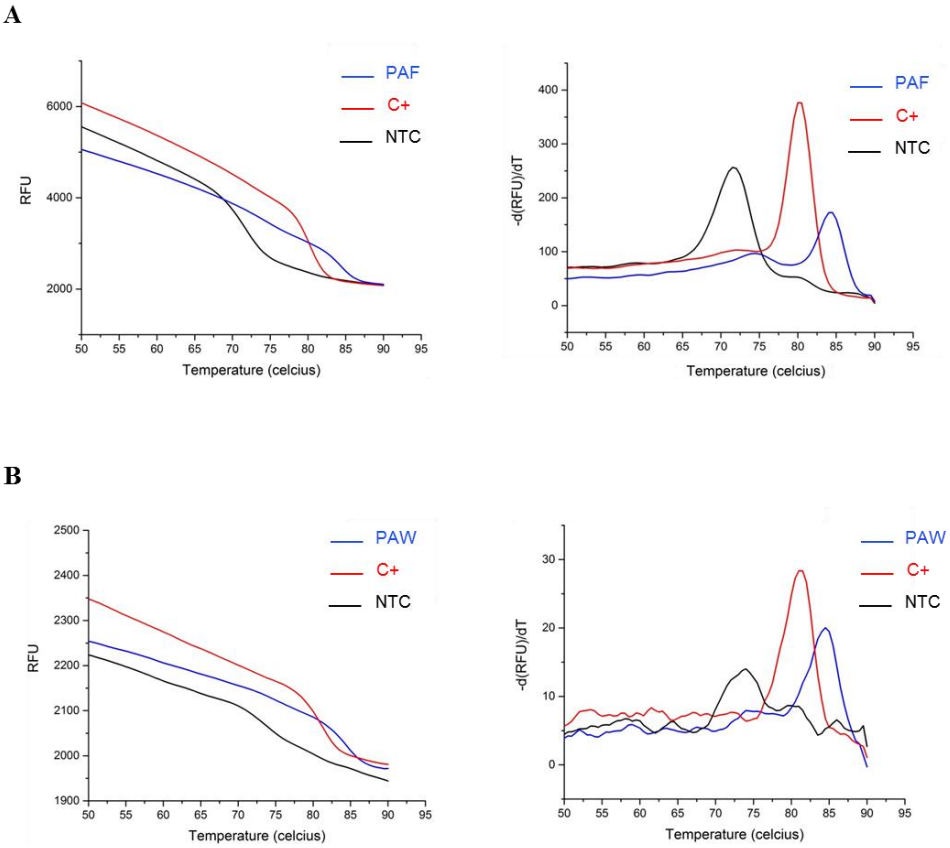


Figure 3.17. Fluorescence melting curves (left) and the first derivatives (right) of PCR products containing **A**) 5-PAF-dUTP (PAF) and **B**) 5-PAW-dUTP (PAW). **C+**, positive controls with natural dNTPs; **NTC**, non-template control. PCR amplification (25 cycles) was performed using T4 and two primers, P2 and P3. 5 nmol of each dATP, dGTP, dCTP and dTTP or modified dUTP, 0.5 U KOD XL DNA polymerase and 1 mM $MnCl_2$ were added to the reaction mixture. A melting temperature experiment was subsequently measured using SYBR green dye in the reaction.

The T_m curve (left) and the first derivatives (right) of PCR amplicons containing 5-PAI-dUTP (Figure 3.18 **A**) or 5-PAA-dUTP (Figure 3.18 **B**) are shown. This melting

temperature experiment was performed using a Lightcycler® 1.5. The PCR amplification was carried out using SYBR green and subsequently, melting temperature experiments were performed. For the derivatives curves, a Gaussian function was used to generate best-fit lines. When 5-PAI-dUTP or 5-PAA-dUTP was used in place of dTTP, the DNA duplex dissociated at 83.5 ℃ and 87.5 ℃, respectively while the analogous canonical DNA duplex dissociates at a lower temperature of 79 ℃, proving that the 5-PAI-dUTP and 5-PAA-dUTP modifications stabilise the duplex.

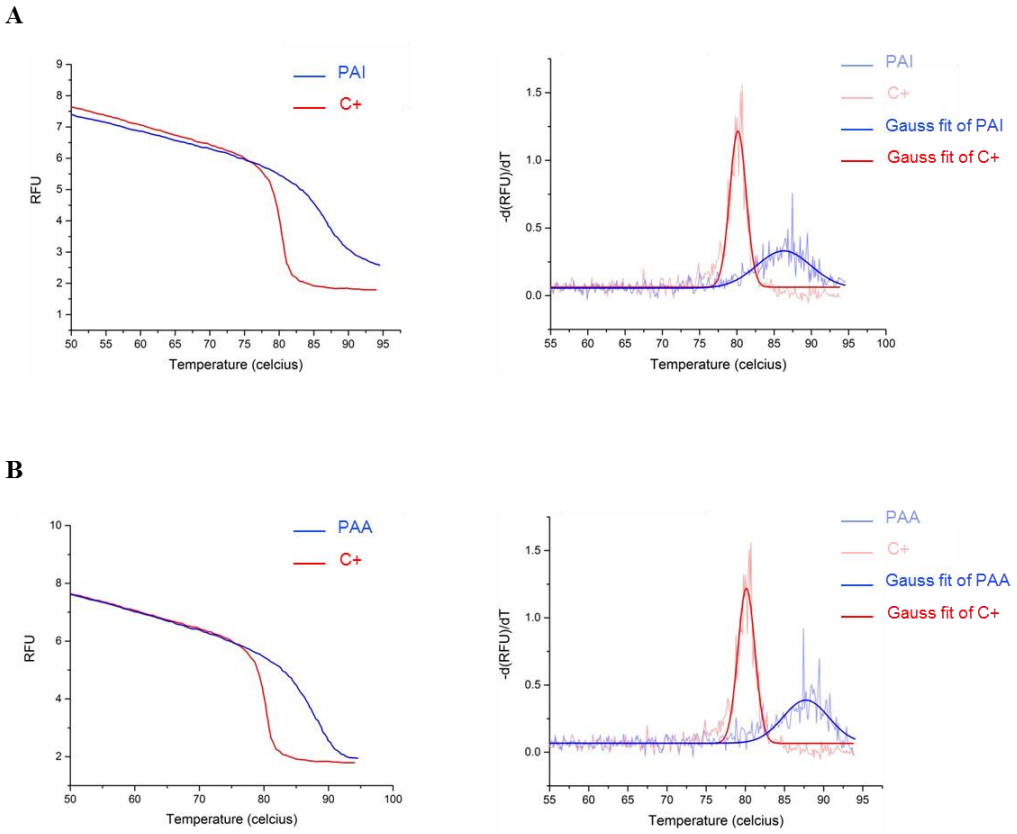


Figure 3.18. Fluorescence melting curves (left) and the first derivatives (right) of PCR products containing **A**) 5-PAI-dUTP (PAI) and **B**) 5-PAA-dUTP (PAA). **C+**, positive controls with natural dNTPs. PCR amplification (25 cycles) was performed using T4 and two primers, P2 and P3. 5 nmol of each dATP, dGTP, dCTP and dTTP or modified dUTP, 0.5 U KOD XL DNA polymerase and 1 mM $MnCl_2$ were added to the reaction mixture. A melting temperature experiment was subsequently measured using SYBR green dye in the reaction.

3.5 Primer extension reactions using modified dUTP bearing halogen bonding substituents

As the primary goal of the enzymatic synthesis of modified DNA is to generate DNA libraries for SELEX, it was necessary to find conditions under which a template-dependent polymerase would accept modified nucleotides as substrates. Various DNA polymerases (GoTaq, Klenow fragment, KOD and KOD XL) were screened primarily in primer extension reaction. KOD XL DNA polymerase with supplementary manganese was previously found to be superior for incorporation of C5 base modified nucleotides into DNA (see section 3.3). As discussed in Chapter 2 (section 2.1), the tolerance of polymerases to modified nucleotides increases when Mn^{2+} is added to the reaction. In this section, halogen bearing moieties, 5-FF-dUTP, 5-FFI-dUTP and 5-ITZ-dUTP were introduced as substrates in the primer extension reaction. The reaction conditions were kept the same as those described in section 3.3.

Before these modifications were added to the reaction, 5-I-dUTP was tested in the enzymatic reaction out of curiosity, to establish whether a bulky iodine atom attached at the C5 position without any intervening linker can be tolerated by DNA polymerases. The first trial was with KOD XL DNA polymerase using the template T2, which produced a fully extended modified product within 5 min of reaction time (Figure 3.19, lane **1c**) whereas a reaction using T3 template did not produce a full length product even after 4 h (Figure 3.19, lane **3e**), indicating that capability of KOD XL DNA polymerase to incorporate this bulky iodine-containing modification attached directly at the C5 position in a serial manner is limited.

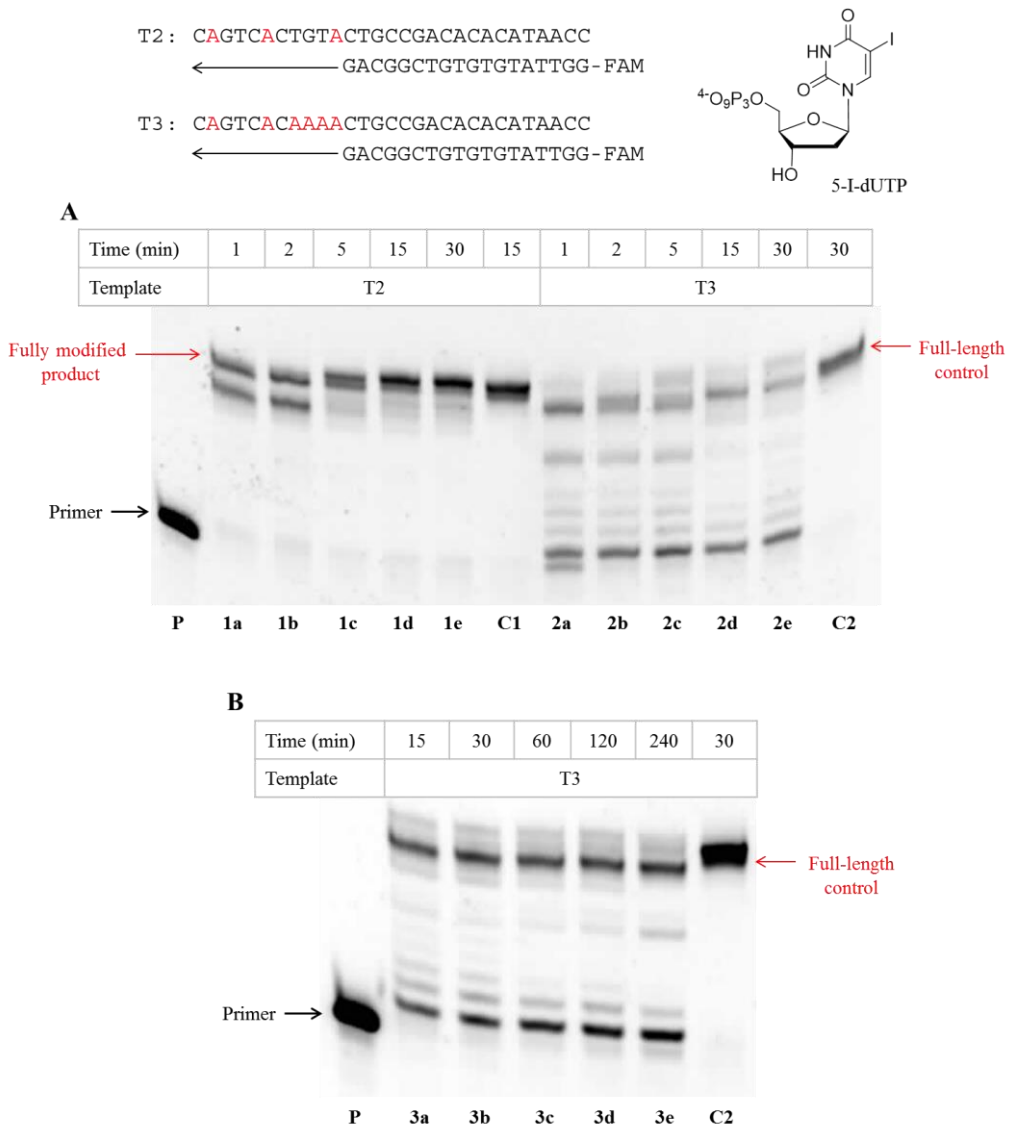


Figure 3.19. **A**) Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-I-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). **B**) Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-I-dUTP over time (15 min to 4 h) for a T3 template. 5-I-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-I-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e** and **3a-3e**, time course experiment using a T3 template; lane **C1**, positive control, a T2 template with natural dNTPs; lane **C2**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

Another attempt to incorporate 5-I-dUTP was conducted with Vent exo- DNA polymerase[265] in presence of both Mg^{2+}/Mn^{2+} or only Mg^{2+} (without Mn^{2+}) (Figure 3.20). 5-I-dUTP was not accepted as a substrate for Vent exo- DNA polymerase, with or

without the additional $MnCl_2$. The reaction appeared to be more strongly inhibited when 1 mM $MnCl_2$ was added, and as a result the major product was the unextended primer (Figure 3.20 **A**). More elongated DNA containing 5-I-dUTPs was detected in the presence of only Mg^{2+} than without Mn^{2+} (Figure 3.20 **B**, lanes **3a**–**3c**).

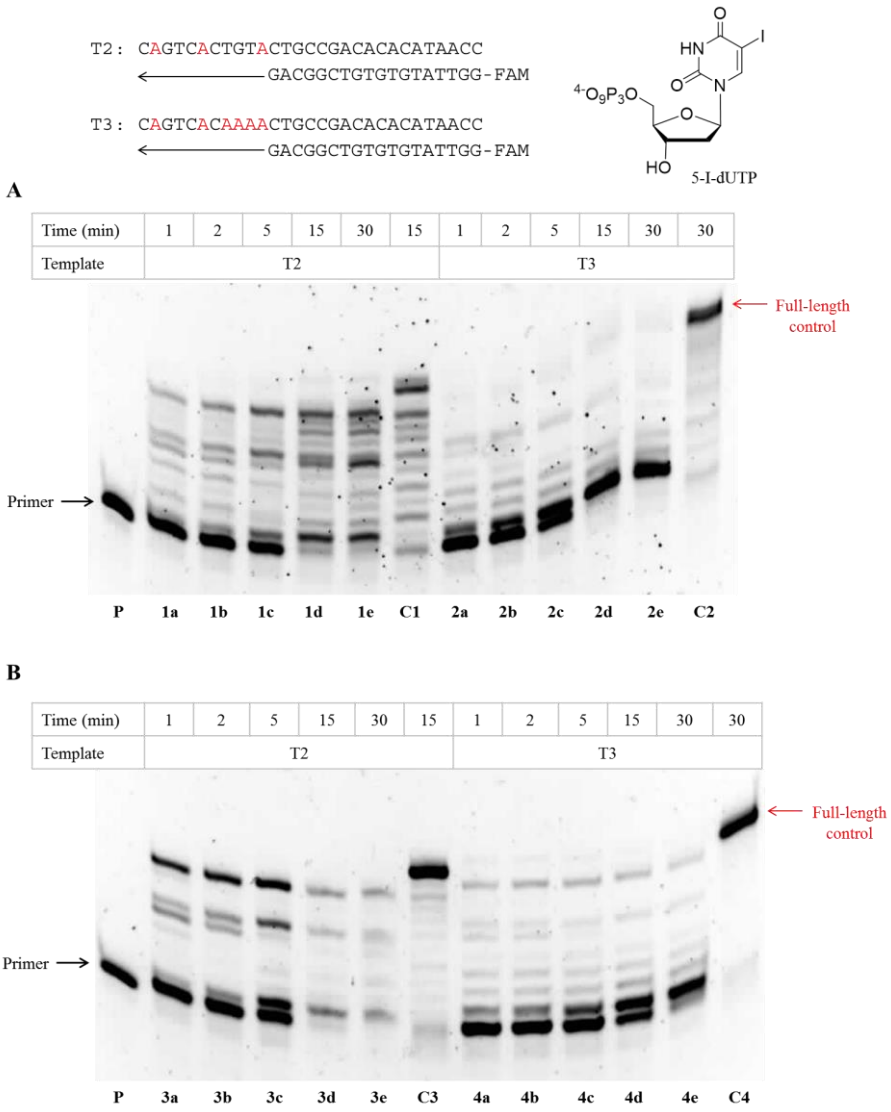


Figure 3.20. Optimisation of 2 U Vent exo- DNA polymerase incorporation of 5-I-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-I-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-I-dUTP unless otherwise stated. **A** is the reaction with additional 1 mM $MnCl_2$. **B** is the reaction without additional $MnCl_2$. Lanes **1a-1e** and **3a-3e**, time course experiment using a T2 template; lanes **2a-2e** and **4a-4e**, time course experiment using a T3 template; lanes **C1**/**C3**, positive control, a T2 template with natural dNTPs; lanes **C2**/**C4**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE).

The other halogen bearing modified triphosphates were added into primer extension reactions using KOD XL DNA polymerase and Vent exo- DNA polymerase. KOD XL DNA polymerase successfully incorporates 5-propargylamino-2,6-difluorobenzyl-dUTP (5-FF-dUTP), 5-propargylamino-2,6-difluoro-4-iodobenzyl-dUTP (5-FFI-dUTP) and 5-propargylamino-iodotriazole-dUTP (5-ITZ-dUTP), using templates T2 and T3. Due to the high MW of the modifications, products containing six modifications ran slower on denaturing PAGE than those containing three modifications, and both migrated slower than positive controls containing natural dNTPs. Figure 3.21 compares the reaction products containing 5-FF-dUTPs when KOD XL and Vent exo- DNA polymerases were used. Both polymerases gave full-length primer extended products using the template T2 after a 15 min incubation period (Figure 3.21 **A**, lane **1d** and **B**, lane **3d**). For template T3, KOD XL DNA polymerase produced the full-length extension product efficiently within 1 min (Figure 3.21 **A**, lane **2a**), whereas Vent exo- DNA polymerase incorporated the modified triphosphate more slowly, requiring 15 mins to produce the full-length product (Figure 3.21 **B**, lane **4d**).

In light of this, only KOD XL DNA polymerase was tested in primer extension reactions using the iodine bearing triphosphate, 5-FFI-dUTP and 5-ITZ-dUTP. When 5-FFI-dUTP was used as a substrate in the reaction, fully elongated primers were observed after 15 min for both the T2 and T3 templates (Figure 3.22, lanes **1d** and **2d**). Compared to 5-FF-dUTP incorporation (Figure 3.21 **A**), 5-FFI-dUTP incorporation using T2 looks similar on the gel (Figure 3.22, lane **1d**). However, consecutive incorporation of 5-FFI-dUTP using the more challenging template T3 required more than 15 min incorporation time (Figure 3.22, lane **2d**). This could be because the iodo-substituent on the phenyl ring increases the steric bulk of the modification compared to

5-FF-dUTP, meaning that KOD XL DNA polymerase requires more time to incorporate modifications when they occur consecutively.

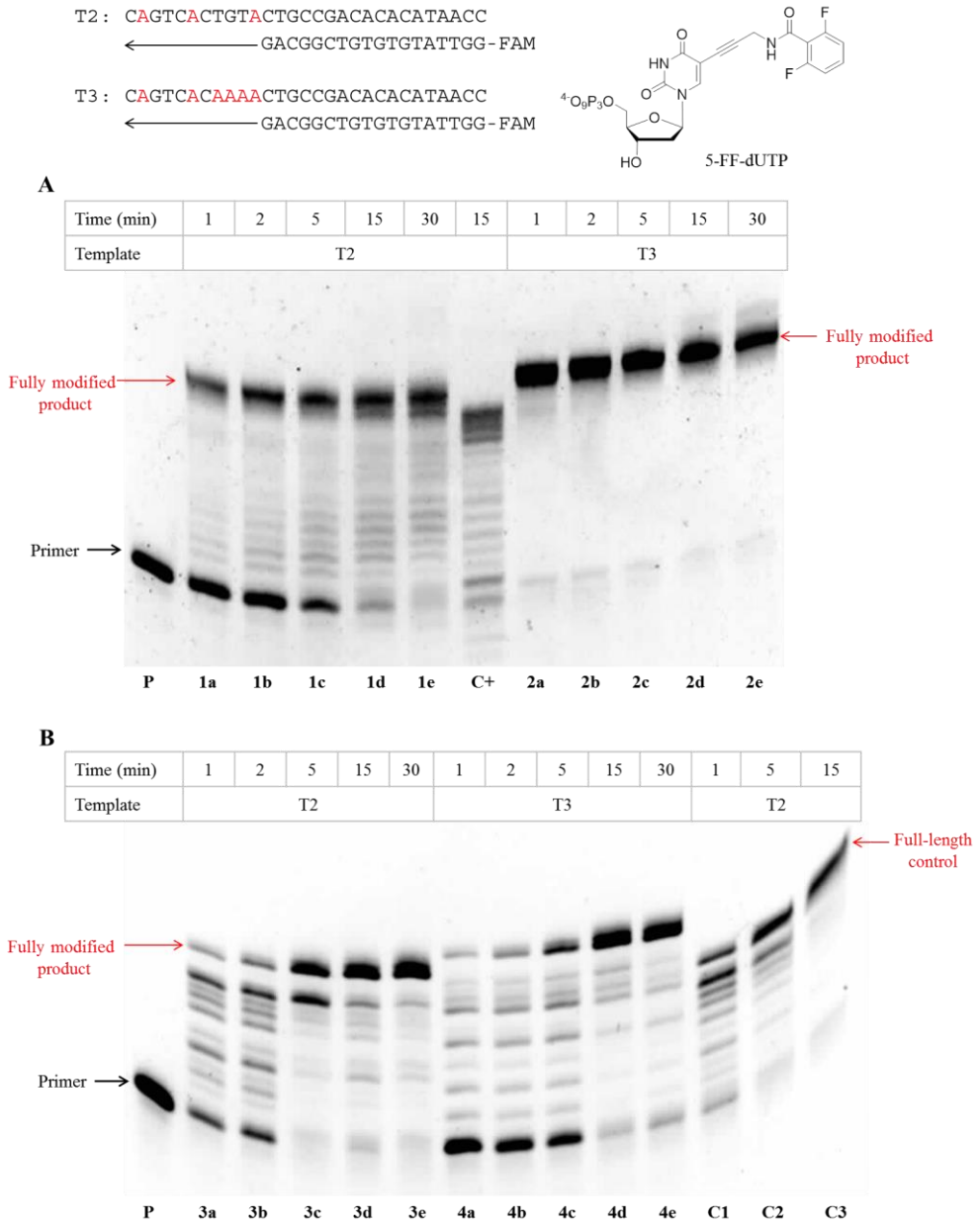


Figure 3.21. Optimisation of 2 U DNA polymerase incorporation of 5-FF-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-FF-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dGTP, dCTP and 5-FF-dUTP unless otherwise stated. **A** is the reaction using KOD XL DNA polymerase with 1 mM $MnCl_2$. **B** is the reaction using Vent exo- DNA polymerase without additional $MnCl_2$. Lanes **1a-1e** and **3a-3e**, time course experiment using a T2 template; lanes **2a-2e** and **4a-4e**, time course experiment using a T3 template; lanes **C+** and **C1-C3**, positive control, a T2 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

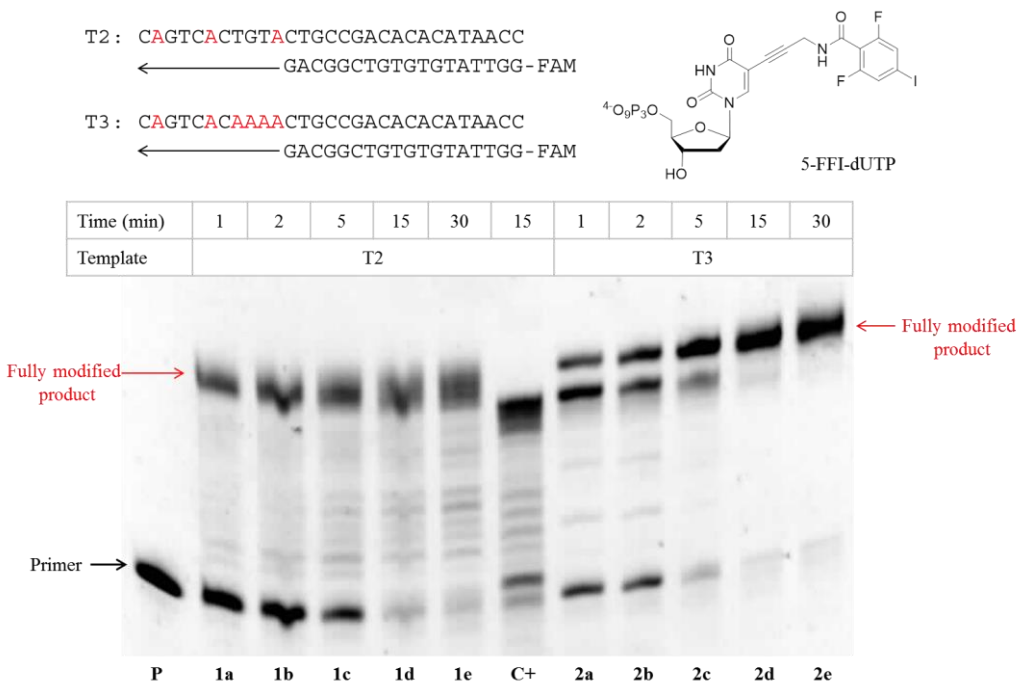


Figure 3.22. Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-FFI-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-FFI-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-FFI-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e**, time course experiment using a T3 template; lane **C+**, positive control, a T2 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

The halogen modified nucleotide 5-ITZ-dUTP was incorporated into DNA within a reaction time of 1–2 min by KOD XL DNA polymerase (Figure 3.23, lanes **1a** and **2b**). Compared to the reactions using 5-FF-dUTP (Figure 3.21) and 5-FFI-dUTP (Figure 3.22), 5-ITZ-dUTP was efficiently tolerated by KOD XL DNA polymerase. Prolonged incubation times of more than 5 min generate a smear of high MW bands on the denaturing PAGE (Figure 3.23, lanes **1c**–**1e** and **2c**–**2e**) a common phenomenon with KOD XL DNA polymerase which is caused by the random addition of extra nucleotides.

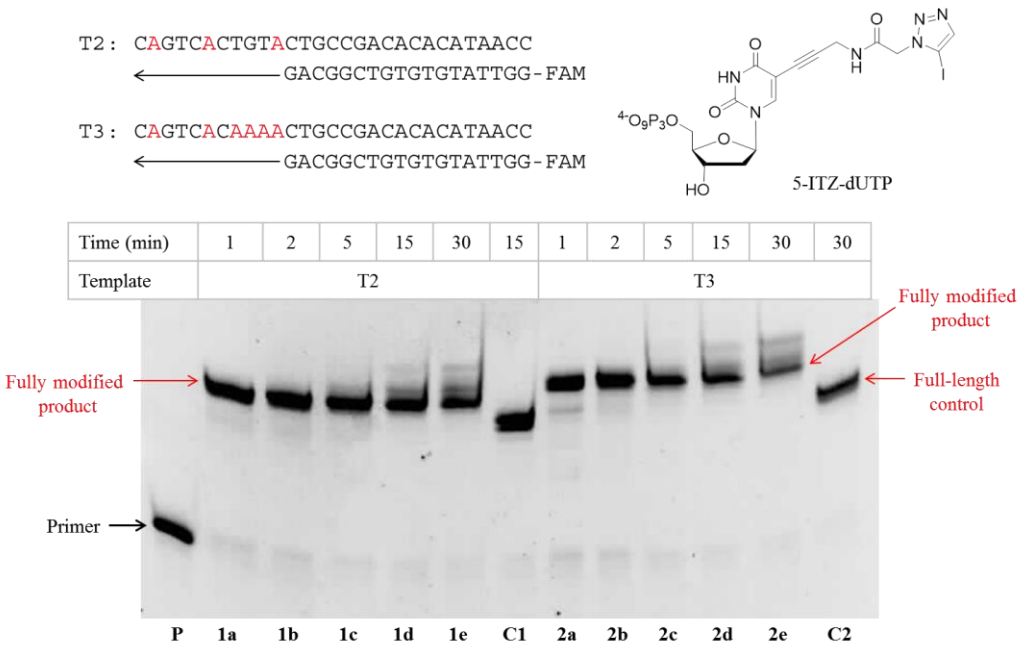


Figure 3.23. Optimisation of 2 U KOD XL DNA polymerase incorporation of 5-ITZ-dUTP over time (1–30 min) for two different DNA templates (T2 and T3). 5-ITZ-dUTP incorporation sites are highlighted in red. Reactions were performed at 65 ℃ with 3.2 nmol of dATP, dCTP, dGTP and 5-ITZ-dUTP and 1 mM $MnCl_2$ unless otherwise stated. Lanes **1a-1e**, time course experiment using a T2 template; lanes **2a-2e**, time course experiment using a T3 template; lane **C1**, positive control, a T2 template with natural dNTPs; lane **C2**, positive control, a T3 template with natural dNTPs; lane **P**, primer P1. (20% Denaturing PAGE)

All of the primer extension reaction products containing halogen modified DNA were analysed by mass spectrometry, which confirmed formation of the expected products (Table 3.2).

Table 3.2. UPLC-MS analysis of final primer extension reaction using 5-I-dUTP, 5-FF-dUTP, 5-FFI-dUTP and 5-ITZ-dUTP. Primer extended products were analysed for both templates T2 and T3. For representative mass spectra see appendix 8.6.

Modification	Template (T2)		Template (T3)	
	Calc. (Da)	Found (Da)	Calc. (Da)	Found (Da)
5-I-dUTP	9892.9	9893.1 10205.6 (+dA)	-	-
5-FF-dUTP	10094.6	10092.5 10405.0 (+dA)	10628.9	10627.5 10941.0 (+dA)
5-FFI-dUTP	10472.3	10471.0 10784.5 (+dA)	11384.3	11383.0 11697.0 (+dA)
5-ITZ-dUTP	10379.3	10693.0 (+dA)	11198.3	11510.5 (+dA)

3.6 PCR reactions using modified dUTP bearing halogen bonding substituents

The positive results of the primer extension reactions led us to investigate PCR amplification using halogenated nucleotides with KOD XL DNA polymerase. 5-FF-dUTP, 5-FFI-dUTP and 5-ITZ-dUTP were used as building blocks in these experiments (Figure 3.24). Template T4 and primers P2 and P3 directed incorporation of eleven modified nucleotides in the top strand, and twenty-two modified nucleotides in the bottom strand (Figure 3.25). Positive control (using all four natural dNTPs) and negative control reactions (similar to positive control but without the addition of template) were conducted in parallel to the modified nucleotide containing PCR reactions. To verify that the modified dUTPs are correctly incorporated in keeping with Watson-Crick base-pairing rules over multiple cycles of PCR, the amplicons were analysed on a 2% agarose gel.

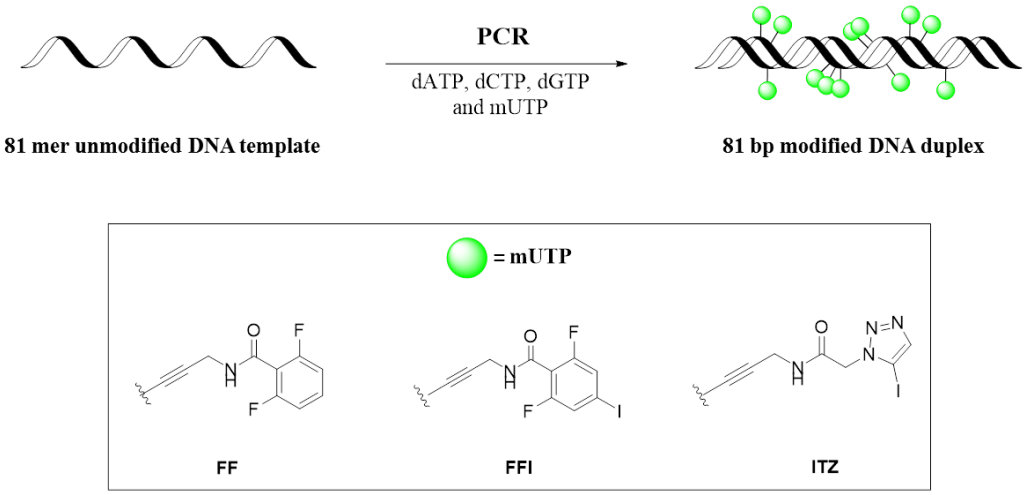


Figure 3.24. PCR using dATP, dCTP, dGTP and halogen modified dUTP (5-FF-dUTP or 5-FFI-dUTP or 5-ITZ-dUTP).

Various time periods for the extension step of the PCR cycle were evaluated using 5-FF-dUTP. Extension at 65 ℃ for 2 min produced non-specific amplifications (Figure 3.25 **A**, lanes **1** and **2**). This phenomenon also occurred for positive controls (Figure 3.25 **A**, lane **C1**). When the extension time was decreased to 30 sec, the yield of the product containing modification was much higher (Figure 3.25 **C**, lane **5**) than for an extension time of 1 min (Figure 3.25 **B**, lanes **3** and **4**), with a lower proportion of by-products. During this experiment, the concentration of 5-FF-dUTPs was varied (2.5 nmol, 5 nmol and 10 nmol) with 5 nmol of dATP, dCTP and dGTP in the reaction mixture. Irrespective of the amount of modified nucleotide added, the amount of amplicons produced is similar. The intensity of products observed by agarose gel analysis appears comparable when 5 or 10 nmol of 5-FF-dUTPs were added (Figure 3.25 **A**, lanes **1** and **2**). Similar to positive controls (Figure 3.25 **C**, lane **C3**), a sufficient amount of amplification products were amplified when 5 nmol of modified nucleotide was added (Figure 3.25 **C**, lane **5**).

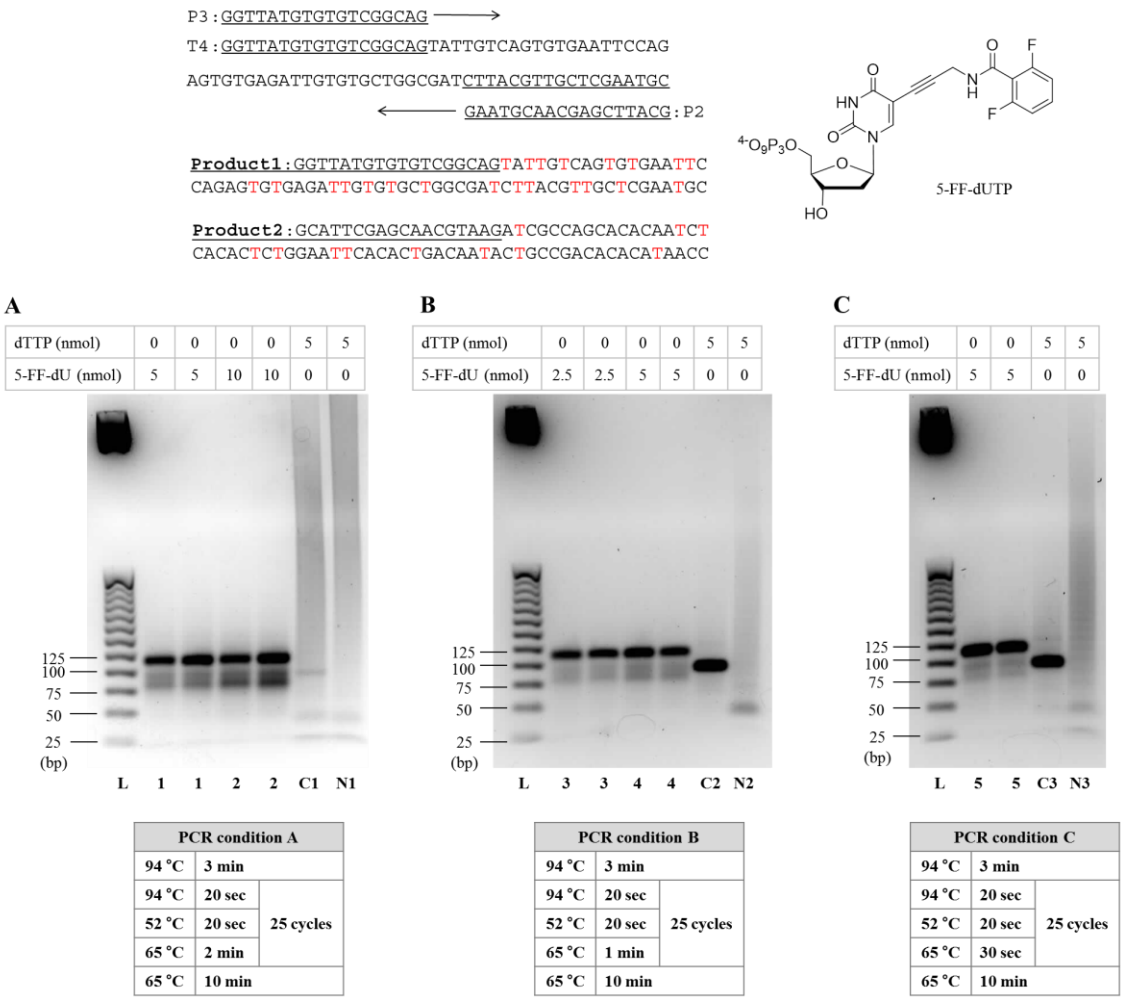


Figure 3.25. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and different amount (2.5 nmol, 5 nmol and 10 nmol) of 5-FF-dUTP, 1 mM $MnCl_2$ and 0.5 U KOD XL DNA polymerase (25 cycles). The positions where 5-FF-dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (25 base pairs). Lanes **1**/**4**/**5**, PCR amplicons using 5 nmol 5-FF-dU; lane **2**, PCR amplicons using 10 nmol 5-FF-dU; lane **3**, PCR amplicons using 2.5 nmol 5-FF-dU; lanes **N1**/**N2**/**N3**, negative control reaction without template T4; lanes **C1**/**C2**/**C3**, positive control reaction with natural dNTPs. (2% Agarose gel)

KOD XL DNA polymerase also efficiently incorporates 5-ITZ-dUTP over 25 PCR cycles (Figure 3.26). Consistent with the results shown in Figure 3.25, a shorter extension time produced fewer non-specific products and therefore a higher yield of the desired modified products (Figure 3.26 **B**, lanes **3** and **4**). The modified DNA

amplification products showed a mobility shift on agarose gel due to the mass increase associated with the modification. Products containing modifications are indicated with a red arrow. In constrast, when Vent exo- DNA poylymerase was used instead of KOD XL DNA polymerase, 5-ITZ-dUTP was incorporated less efficiently and the reaction yielded only a small amount of the desired amplification product (Figure 3.26 **C**, lane **5**). In this reaction (Figure 3.26 **C**), the annealing temperature increased to 60 ℃ and the reaction mixture incubated at 70 ℃ for 2 min instead, in order to reduce the primer dimerisation. However, this was not efficient and KOD XL DNA polymerase appears superior for PCR amplification using 5-ITZ-dUTP.

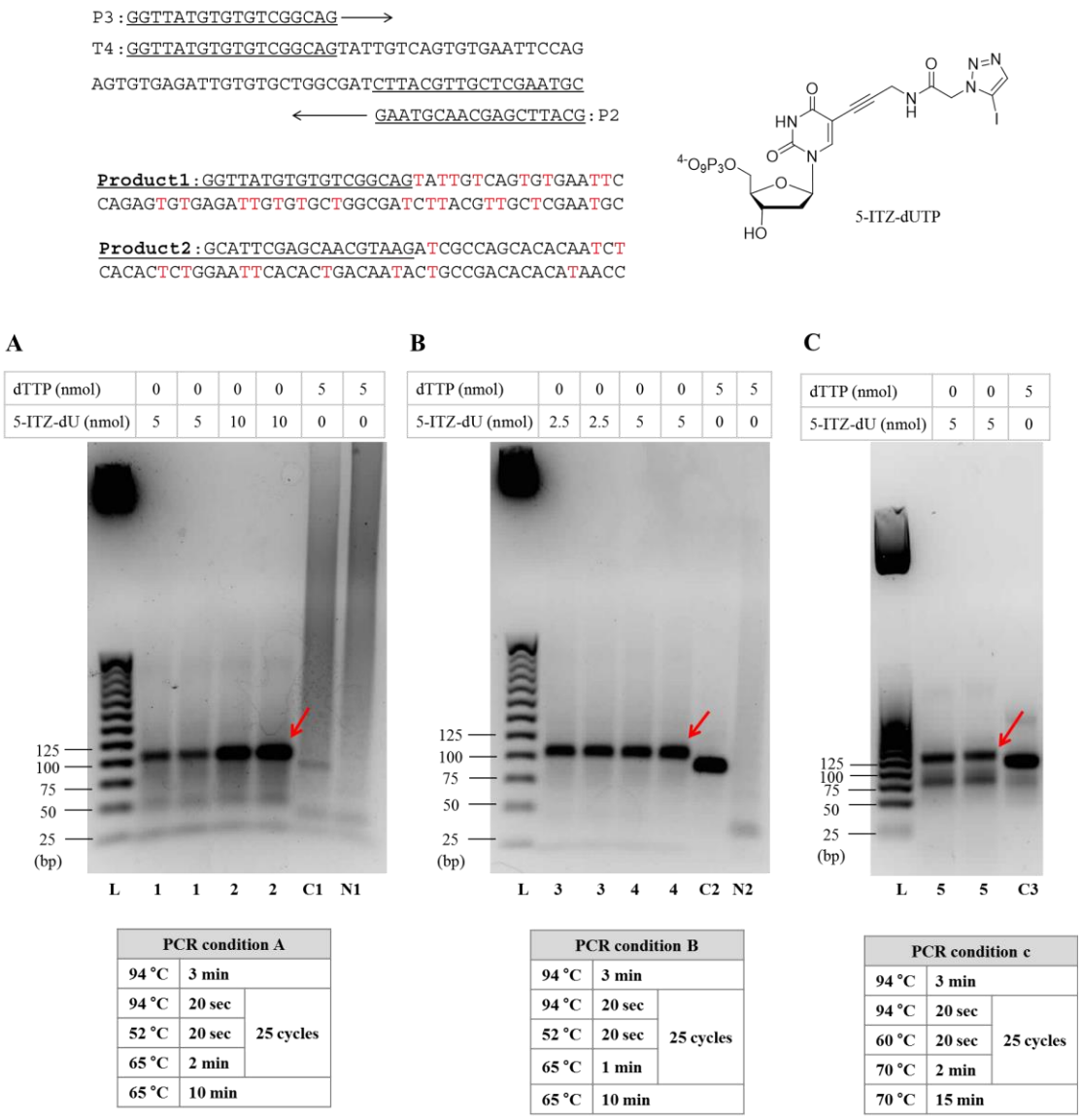


Figure 3.26. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and different amount (2.5 nmol, 5 nmol and 10 nmol) of 5-ITZ-dUTP and 0.5 U DNA polymerase (25 cycles) unless otherwise stated. The positions where 5-ITZ-dUTPs were incorporated are highlighted in red. **Condition A/B**) KOD XL DNA polymerase was used with 1 mM $MnCl_2$. **Condition C**) Vent exo- DNA polymerase was used without additional $MnCl_2$. Lane **L**, DNA ladder (25 base pairs); lanes **1**/**4**/**5**, PCR amplicons using 5 nmol 5-ITZ-dU; lane **2**, PCR amplicons using 10 nmol 5-ITZ-dU; lane **3**, PCR amplicons using 2.5 nmol 5-ITZ-dU; lanes **N1**/**N2**, negative control reaction without template T4; lanes **C1**/**C2**/**C3**, positive control reaction with natural dNTPs. (2% Agarose gel)

PCR experiments using the modified triphosphate 5-FFI-dUTP proved to be more challenging than the primer extensions with the same substrates, and no amplicons were

detected by 2% agarose gel analysis of the product mixture when KOD XL DNA polymerase was used.

Melting experiments on the amplicons showed that the 5-FF-dUTP modification preserved the stability of the DNA duplex (Figure 3.27 **B**), whereas the 5-ITZ-dUTP increased the stability of the duplex by 5 ℃ relative to unmodified DNA (Figure 3.27 **A**). Nonetheless, DNA duplex containing 5-FFI-dUTP was not formed (Figure 3.27 **C**).

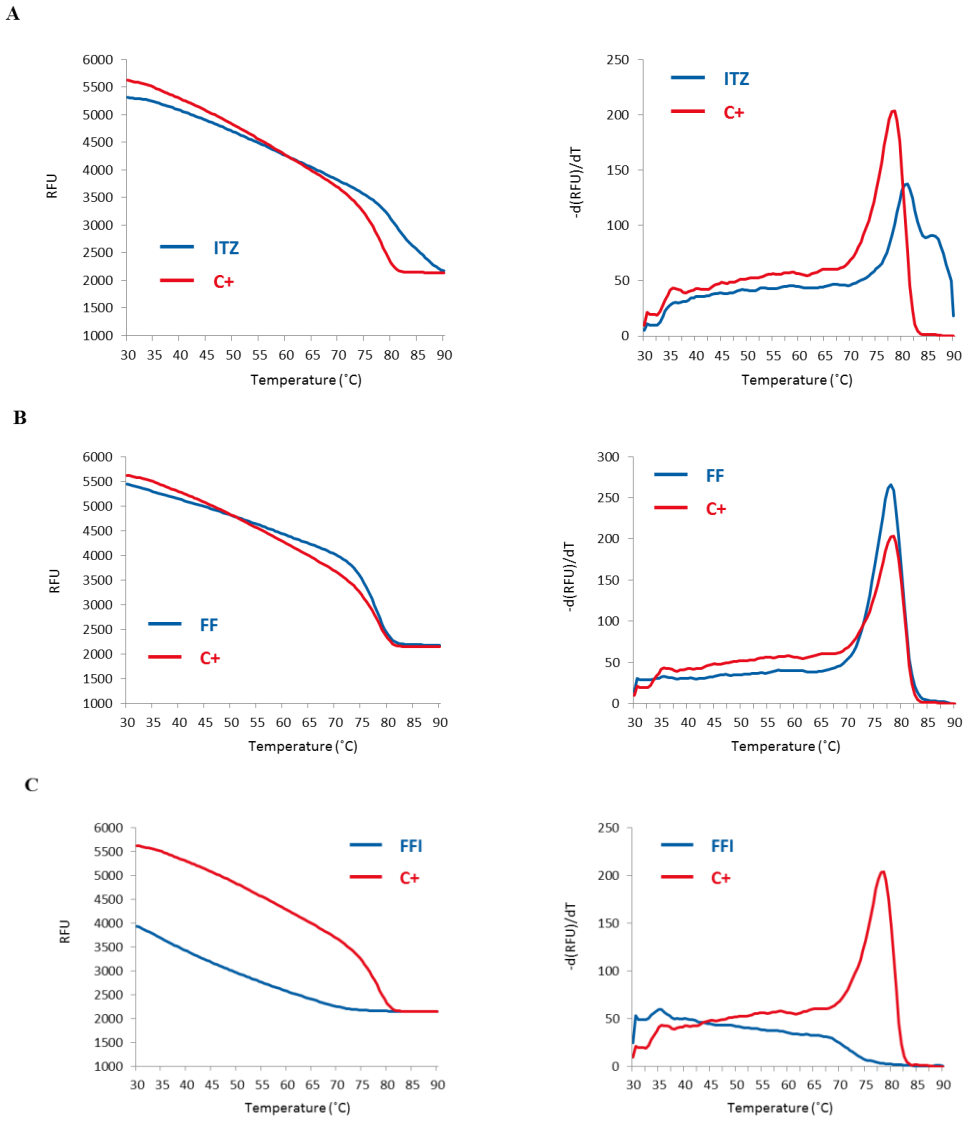


Figure 3.27. Fluorescence melting curves (left) and the first derivatives (right) of PCR products containing **A**) 5-ITZ-dUTP (ITZ), **B**) 5-FF-dUTP (FF) and **C**) 5-FFI-dUTP (FFI). **C+**, positive controls with natural dNTPs.

In order to incorporate 5-FFI-dUTP into DNA, several attempts to vary the PCR conditions were tried. An increased amount (10 nmol) of 5-FFI-dUTP was added to the reaction mixture, which was incubated at 65 ℃ for 2 min and 3.5 min using KOD XL DNA polymerase. However, these experiments again failed to produce amplicon bands (Figure 3.28 **A**, lane **1** and **B**, lanes **2** and **3**) on the gel, whereas positive control experiments showed strong bands in the expected region (Figure 3.28 **A**, lane **C1** and **B**, lane **C2**).

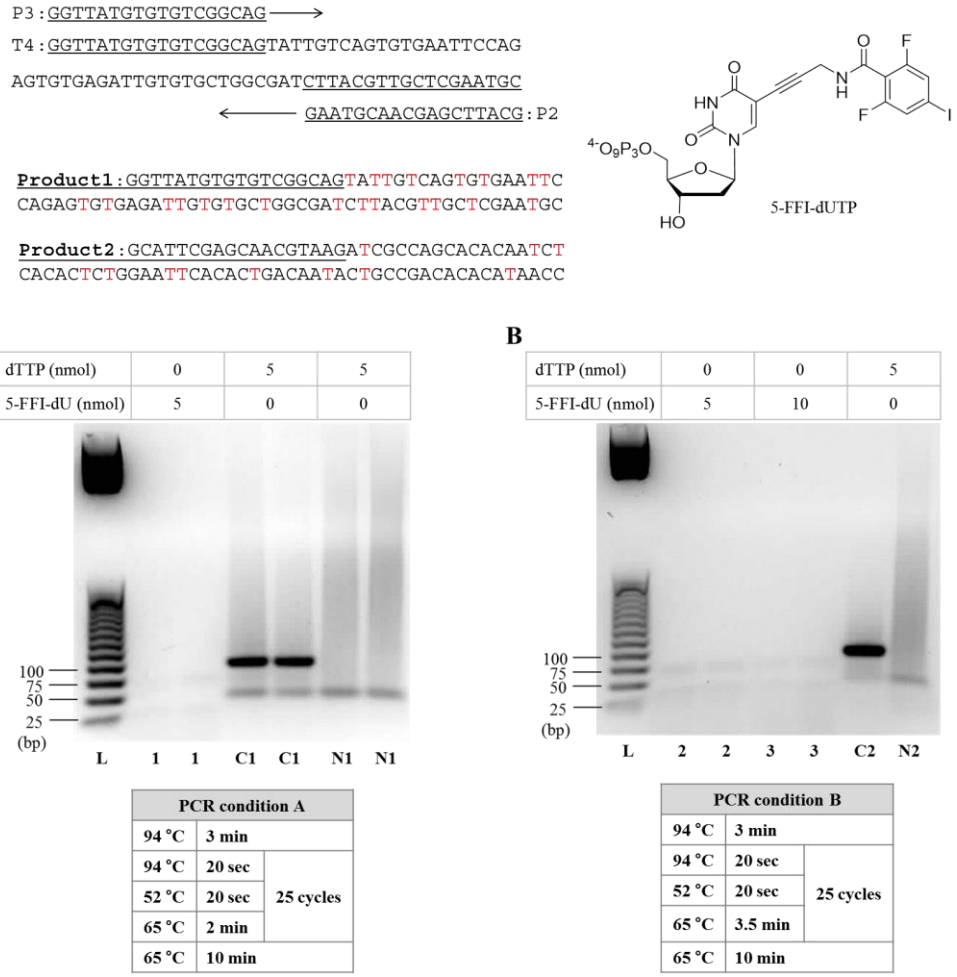
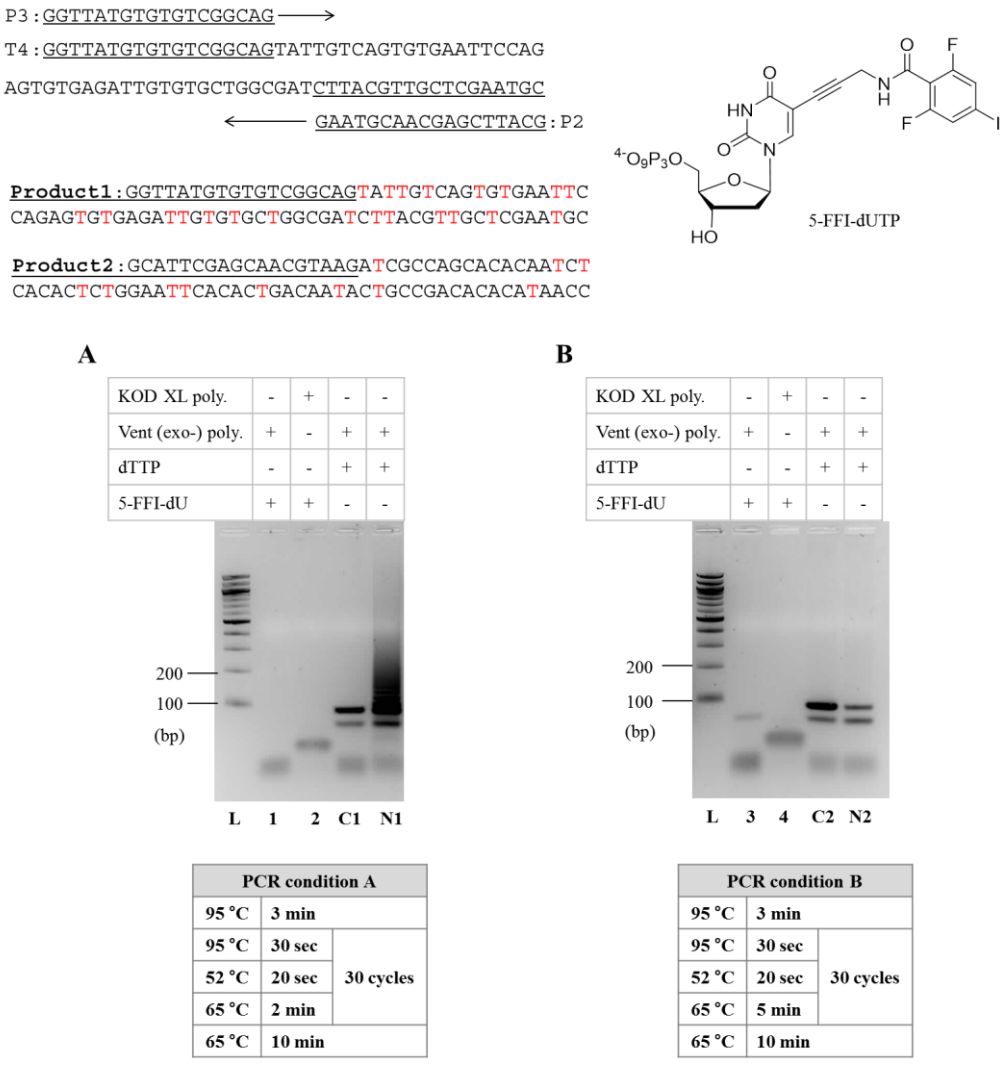


Figure 3.28. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and different amount (5 nmol and 10 nmol) of 5-FFI-dUTP, 1 mM $MnCl_2$ and 0.5 U KOD XL DNA polymerase (25 cycles). The positions where 5-FFI-dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (25 base pairs). Lanes **1**/**2**, PCR amplicons using 5 nmol 5-FFI-dU; lane **3**, PCR amplicons using 10 nmol 5-FFI-dU; lanes **C1**/**C2**, positive control reaction with natural dNTPs; lanes **N1**/**N2**, negative control reaction without template T4. (2% Agarose gel)

The number of PCR cycles was increased from 25 to 30 using KOD XL DNA polymerase and the extension time was also increased to 5 min at 65 ℃ (Figure 3.29 **A**, lane **2** and **B**, lane **4**). The same conditions were also applied using Vent exo- DNA polymerase in place of KOD XL DNA polymerase (Figure 3.29 **A**, lane **1** and **B**, lane **3**). However, both KOD XL and Vent exo- DNA polymerases did not efficiently

incorporate the modified substrate with 5-FFI-dUTP probably because of bulky iodine moiety. Primer dimerisation, as well as possible contamination of the PCR reagents, led to non-specific amplification (Figure 3.29, lanes **N1** and **N2**).



PCR condition A		
95 °C	3 min	
95 °C	30 sec	30 cycles
52 °C	20 sec	
65 °C	2 min	
65 °C	10 min	

PCR condition B		
95 °C	3 min	
95 °C	30 sec	30 cycles
52 °C	20 sec	
65 °C	5 min	
65 °C	10 min	

Figure 3.29. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-FFI-dUTP, 0.5 U of DNA polymerase (30 cycles). The positions where 5-FFI-dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (100 base pairs). Lanes **1**/**3**, PCR amplicons using Vent exo- DNA polymerase without additional $MnCl_2$; lanes **2**/**4**, PCR amplicons using KOD XL DNA polymerase with 1 mM $MnCl_2$; lanes **C1**/**C2**, positive control reaction with natural dNTPs; lanes **N1**/**N2**, negative control reaction without template T4. (2% Agarose gel)

The extension temperature was raised to 72 ℃ and the reaction mixture was incubated for a short period of 1 min to avoid the formation of non-specific products, as well as a longer period of 5 min to accommodate the slower reaction kinetics, but PCR amplification was again not observed (Figure 3.30, lanes **2** and **4**). Vent exo- DNA polymerase was found to perform slightly better, producing small amounts of amplification product incorporating the modified 5-FFI-dUTP, but still in low yield (Figure 3.30, lanes **1** and **3**). Positive and negative controls were non-specifically amplified with a 5 min PCR extension time (Figure 3.30, lanes **C2** and **N2**).

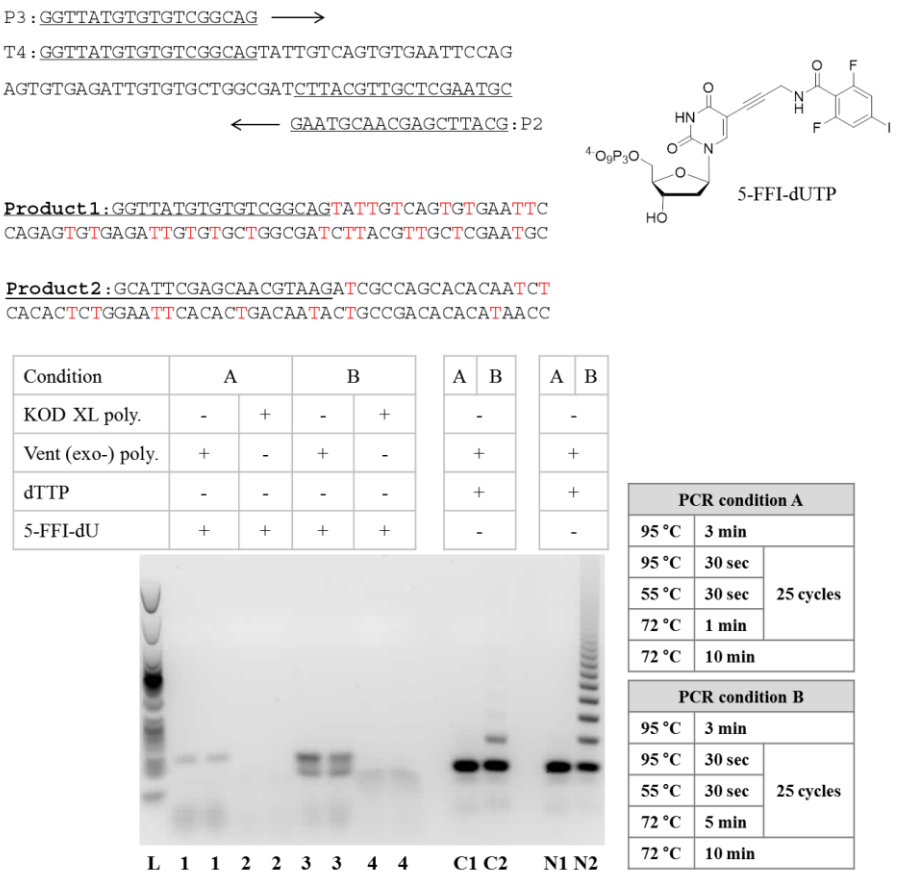


Figure 3.30. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-FFI-dUTP, 0.5 U of DNA polymerase (25 cycles). The positions where 5-FFI-dUTPs were incorporated are highlighted in red. Lane **L**, DNA ladder (Low MW). Lanes **1**/**3**, PCR amplicons using Vent exo- DNA polymerase without additional $MnCl_2$; lanes **2**/**4**, PCR amplicons using KOD XL DNA polymerase with 1 mM $MnCl_2$; lanes **C1**/**C2**, positive control reaction with natural dNTPs; lanes **N1**/**N2**, negative control reaction without template T4. (2% Agarose gel)

Phusion high-fidelity (HF) DNA polymerase is also known to efficiently incorporate modified triphosphates,[333] and Q5 Hot Start enzyme is known to reduce primer-dimer formation. These two DNA polymerases were tested in parallel with Vent exo- and KOD XL DNA polymerases. Disappointingly, after 30 cycles, none of the enzymes tested were able to generate the full-length PCR amplification product incorporating 5-FFI-dUTP, as confirmed by analysis using agarose gel (Figure 3.31, lanes **1a**, **1b, 2a**, **2b**,

3, **4a** and **4b**). When PCR products were analysed on denaturing PAGE gel, only short fragments of DNA were observed especially when 5-FFI-dUTP was added as substrate (Figure S 8-22 in appendix 8.7).

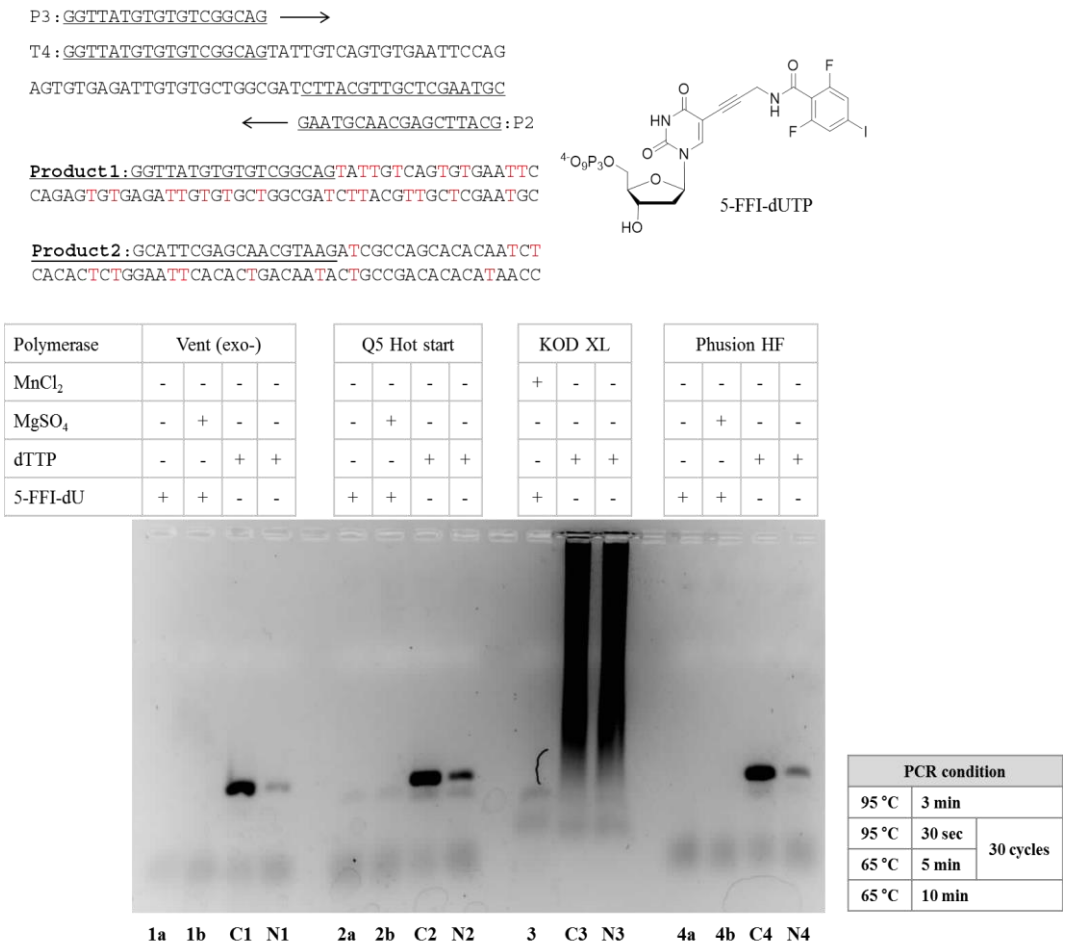
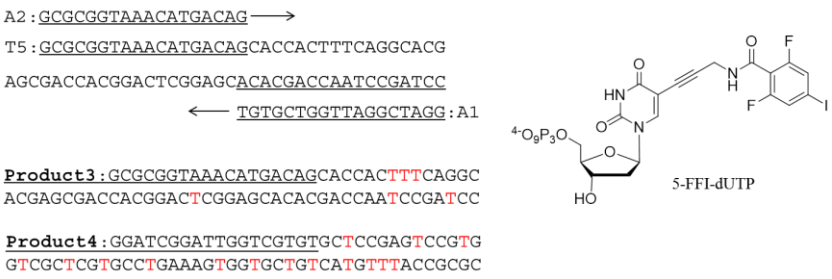


Figure 3.31. Analysis of PCR products using 81 mer PCR template T4, two primers P2 and P3, 5 nmol of dATP, dCTP, dGTP and 5-FFI-dUTP, 0.5 U of DNA polymerase (30 cycles). The positions where 5-FFI-dUTPs were incorporated are highlighted in red. Lane **1a**, PCR amplicons using Vent exo- DNA polymerase without additional Mn^{2+}; lane **1b**, PCR amplicons using Vent exo- DNA polymerase with 5 mM $MgSO_4$; lane **2a**, PCR amplicons using Q5 Hot Start DNA polymerase without additional Mn^{2+}; lane **2b**, PCR amplicons using Q5 Hot Start DNA polymerase with additional 5 mM $MgSO_4$; lane **3**, PCR amplicons using KOD XL DNA polymerase with additional 1 mM $MnCl_2$; lane **4a**, PCR amplicons using Phusion HF DNA polymerase without additional Mn^{2+}; lane **4b**, PCR amplicons using Phusion HF DNA polymerase with additional 5 mM $MgSO_4$; lanes **C1-C4**, positive control reaction with natural dNTPs; lanes **N1-N4**, negative control reaction without template T4. (2% Agarose gel)

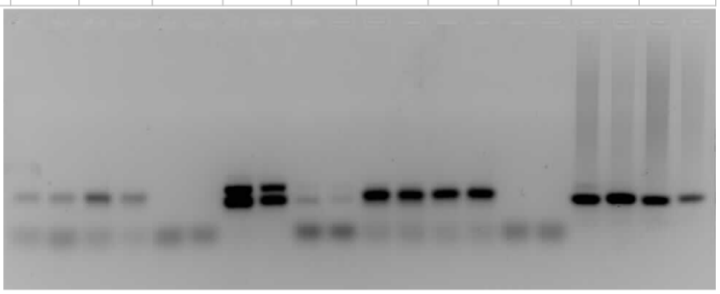
To further investigate PCR using 5-FFI-dUTP and various DNA polymerases, we carried out another PCR experiment using a shorter template (Table 3.3) which includes fewer incorporation sites for the modified triphosphate but still retains three consecutive incorporation sites. The new 74 mer DNA template T5 and primers A1 and A2 direct incorporation of sixteen modified nucleotides in the top strand and six modified nucleotides in the bottom strand, with a single stretch of three successive modified residues (Figure 3.32). Four DNA polymerases, Vent exo-, Q5 Hot Start, KOD XL and Phusion HF were screened with different combination of divalent cations (Mn^{2+} or Mg^{2+}). For the reactions carried out with additional 5 mM $MgSO_4$ or 1 mM $MnCl_2$, distinctive bands are observed on the agarose gel which are consistent with the formation of the target PCR amplicons (Figure 3.32, lanes **1b**, **2b**, **3** and **4b**), but when the samples were re-anaylsed by denaturing PAGE, none of the modified products were found to be full-length (Figure S 8-23 in appendix 8.7). The excess Mg^{2+} in the reaction may enhance the enzyme activity, thereby resulting in non-specific amplification.

Table 3.3. 74 mer PCR template (T) and forward and reverse PCR primers (A)

	Sequence (5′- 3′)
A1	GGATCGGATTTGGTCGTG
A2	GCGCGGTAAACATGACAG
T5	GCGCGGTAAACATGACAGCACCACTTTCAGGCACGAG CGACCACGGACTCGGAGCACACGACCAAATCCGATCC



Condition	A					B				
$MnCl_2$	-	-	+	-	-	-	-	+	-	-
$MgSO_4$	-	+	-	-	-	-	+	-	-	-
dTTP	-	-	-	+	+	-	-	-	+	+
5-FFI-dU	+	+	+	-	-	+	+	+	-	-



1 1 2 2 3 3 C1 C1 N1 N1 4 4 5 5 6 6 C2 C2 N2 N2

PCR condition A		
95 °C	3 min	
95 °C	30 sec	30 cycles
55 °C	30 sec	
72 °C	1 min	
72 °C	10 min	

PCR condition B		
95 °C	3 min	
95 °C	30 sec	30 cycles
55 °C	30 sec	
72 °C	5 min	
72 °C	10 min	

Figure 3.32. Analysis of PCR products using 74 mer PCR template T5, two primers A1 and A2, 5 nmol of dATP, dCTP, dGTP and 5-FFI-dUTP, 0.5 U of DNA polymerase (30 cycles). The positions where 5-FFI-dUTPs were incorporated are highlighted in red. Lane **1a**, PCR amplicons using Vent exo- DNA polymerase without additional Mn^{2+}; lane **1b**, PCR amplicons using Vent exo- DNA polymerase with additional 5 mM $MgSO_4$; lane **2a**, PCR amplicons using Q5 Hot Start DNA polymerase without additional Mn^{2+}; lane **2b**, PCR amplicons using Q5 Hot Start DNA polymerase with additional 5 mM $MgSO_4$; lane **3**, PCR amplicons using KOD XL DNA polymerase with additional 1 mM $MnCl_2$; lane **4a**, PCR amplicons using Phusion HF DNA polymerase without additional Mn^{2+}; lane **4b**, PCR amplicons using Phusion HF DNA polymerase with additional 5 mM $MgSO_4$; lanes **C1-C4**, positive control reaction with natural dNTPs; lanes **N1-N4**, negative control reaction without template T5. (2% Agarose gel)

One possible reason for the difficulty of PCR amplification using 5-FFI-dUTP could be that the modified template was not read-through by the DNA polymerase. If the

modified DNA strands are poor templates for polymerisation, exponential amplification of the product cannot occur. In order to prove this, the oligonucleotide containing 5-FFI-dUTP at various positions can be synthesised by solid phase method and used as a template for a linear copying reaction. An alternative explanation could be that DNA duplexes containing 5-FFI-dUTP are very stable so that they are not denatured during PCR. In this case, PCR will not produce any amplicon. Further work on this modified triphosphate was not carried out because enzymatic DNA synthesis using 5-ITZ-dUTP was successful and this modified nucleotide can therefore be used in SELEX to discover aptamer containing halogen bonding interactions.

3.7 Conclusions

Initially, attempts were made to synthesise a triphosphate derived from 5-iodo-deoxyuridine nucleoside using the optimised Yoshikawa and Ludwig-Eckstein methods. Due to the indiscriminate reaction at both 5′- and 3′-OH positions of the nucleoside, these synthetic routes produced extensive amounts of by-products and therefore, the final yields after the triphosphorylation were impractically low. A new synthetic protocol to produce 5-I-dUTP was evaluated using natural dUTP as a precursor, and the yield of the 5-I-dUTP was increased to 82%. A Sonogashira coupling reaction using 5-I-dUTP successfully produced 5-PA-dUTP and this compound was labelled with various hydrophobic and halogen-containing modifications via amide coupling reactions. The modified triphosphates were purified and used as building blocks in enzymatic primer extension and PCR reactions.

The modified triphosphates 5-PAF-dUTP, 5-PAW-dUTP, 5-PAI-dUTP and 5-PAA-dUTP, which are functionalised with hydrophobic substituents, were tolerated during

primer extension reactions and during PCR using KOD XL DNA polymerase with 1 mM $MnCl_2$. In contrast to the 2′ sugar modified triphosphates, discussed in the previous chapter, these base modified triphosphates were shown to be compatible with engineered DNA polymerases unless extremely bulky modifications were present. For examples, the results showed that the pyrene containing 5-PAPBA-dUTP was not incorporated into DNA by KOD XL DNA polymerase; this can be rationalised on steric grounds.

For halogen bearing modifications, 5-FF-dUTP and 5-ITZ-dUTP were efficiently incorporated into DNA during primer extension reactions. Compared to Vent exo- DNA polymerase, KOD XL DNA polymerase with 1 mM $MnCl_2$ was found to be superior for incorporation of these modified triphosphates. 5-FF-dUTP and 5-ITZ-dUTP also produced PCR amplicons with high yields even when the extension step of the PCR cycle was reduced to less than 1 min. Although primer extension reactions using 5-FFI-dUTP could produce full-length products especially when the incubation time was extended, PCR amplicons incorporating 5-FFI-dUTP were not attained using short (74 mer) and long (81 mer) template sequences. Four different DNA polymerases (Vent exo-, KOD XL, Q5 Hot start and Phusion HF) were tested in an attempt to improve the efficiency of PCR using 5-FFI-dUTP, but in all cases the correct amplification products were not observed by denaturing PAGE analysis. The poor performance of 5-FFI-dUTP compared to 5-FF-dUTP in PCR reactions suggests that the presence of the additional, bulky iodine atom in 5-FFI-dUTP may sterically inhibit the efficient processing of this substrate by polymerase enzymes. Unlike 5-FFI-dUTP, 5-ITZ-dUTP showed high efficiency of incorporation during enzymatic synthesis and can therefore be taken forward for the discovery of modified DNA aptamers via SELEX, in order to study the effects of halogen bonding on aptamer-target interactions.

Chapter 4

DNA aptamer for HER2 extracellular epitopes

4 DNA aptamer for HER2 extracellular epitopes

4.1 Background

Prior to performing SELEX using unnatural substrates, optimisation of selection techniques is necessary. This includes the amplification of the DNA library by PCR, single-stranded (ss) DNA generation and the selection methodology. Therefore, in this chapter, the goal is to optimise these steps while discovering DNA aptamers for a human epidermal growth factor receptor 2 (HER2).

HER2 is a member of the epidermal growth factor receptor family which also contains receptors 1, 3, and 4 (EGFR, HER3 and HER4). Mutation and dysregulation of these EGFR family members have been correlated with cancer development and progression. HER2 is a 184 kDa transmembrane glycoprotein with tyrosine kinase activity,[396] and consists of an extracellular binding domain, a single transmembrane-spanning domain, and a long cytoplasmic tyrosine kinase domain.[397] HER2 receptor is activated by homo- or heterodimerisation with another family member. These dimers generate intracellular signals that are significantly stronger than signals originating from other HER combinations (non-HER2 containing). When HER2-EGFR heterodimers are formed, signaling proceeds through phosphorylation events initiated from either of the activated receptors. On the other hand, HER2-HER3 heterodimer signaling proceeds only through the activated HER2 receptor, as HER3 does not exhibit kinase activity itself, making HER2 as an important factor.[398] In fact, few HER2 molecules exist at the cell surface in normal cells forming few heterodimers and therefore, growth signals are relatively weak and controllable. However, the overexpression of HER2 triggers the formation of multiple HER2 heterodimers causing the strong cell signaling[399,400] which resulted in the deregulation of cell proliferation (mitogen-activated protein kinase, MAPK pathway)

and cell survival (phosphoinositide 3-kinase).[401,402] Hence, targeted therapies for HER2 receptors aim to control the signaling pathways which play a significant role in cancer development and progression.

Monoclonal antibodies targeting human HER2 result in cell growth inhibition both *in vitro* and in patients with breast tumours.[403] The most well-known HER2 monoclonal antibodies are trastuzumab (Herceptin) and pertuzumab (Perjeta), and an affibody (small engineered mimic of monoclonal antibody[404]) composed of 58 amino acid residue (ZHER2) is also known (Figure 4.1 **A**). Trastuzumab has an anti-proliferative effect on cells transformed by overexpression of HER2 by binding to the C-terminal of domain IV (Figure 4.1). Binding at this site prevents the extended domain II loop of the inactive states of the HER1 and HER3 receptors interacting with HER2.[397,405,406] In contrast, pertuzumab inhibits the growth of tumours displaying both low and high HER2 receptor levels by binding near the centre of domain II of HER2 (Figure 4.1). Upon the binding to this domain, dimerisation with another HER-receptor is blocked.[407] Finally, the affibody (ZHER2) does not exhibit any biological effects.[408–410] ZHER2 instead works as a tracer or a carrier without affecting HER2 function, even in the presence of therapeutic monoclonal antibodies. Hence, ZHER2 can be used for imaging of HER2 or as a targeting system for delivery of therapeutic agents in parallel to ongoing therapy using either of the monoclonal antibodies (Figure 4.1 **B**).[409,411]

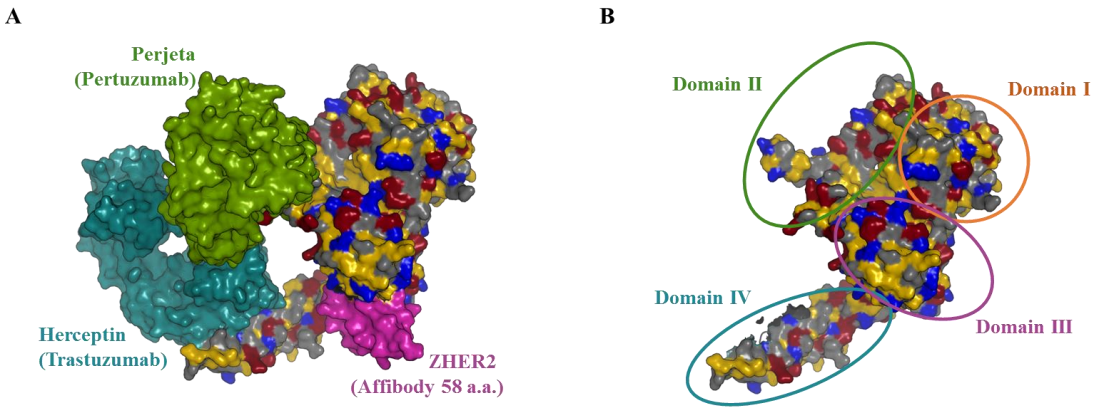


Figure 4.1. **A**) HER2 extracellular domain complexes with monoclonal antibodies (Perjeta (green), PDB code 1S78[407] and Herceptin (blue), PDB code 1N8Z[397]) and affibody (ZHER2 (magenta), PDB code 3MZW[412]). **B**) Structure of the HER2 extracellular domain and its domains. Domains I (orange), II (green), III (magenta) and IV (blue) are coloured accordingly.[397] The figure and the structural presentations were generated using PyMOL program.[371]

As an alternative therapeutic modality, several aptamers targeting HER2 have been reported, mostly RNA-based.[413–417] Purified HER2 protein[414,416–418] or a mouse cell line[415] were used as targets during SELEX. In 2011, Kim and Jeong discovered a 34 mer RNA aptamer which was specific to the extracellular domain of ErbB2 protein with a binding affinity of 3.5 nM.[414] Thiel *et al*. found RNA aptamers using novel HER2 positive cell internalised SELEX and the K_D values were 45–85 nM.[415] In 2012, Liu *et al.* found an 86 mer DNA aptamer which bound to an epitope peptide of HER2 with a K_D of 18.9 nM and this aptamer also bound to the extracellular domain of HER2 protein with a K_D of 316 nM.[417,419] Since HER2 is a proven target with important biological effects it would be good to target it for our SELEX optimisation.

4.2 Target selection

One of the epitopes selected (peptide **1**, Table 4.1) is part of domain II, the binding site of pertuzumab (Figure 4.2) and also the site of HER-receptor dimerisation. Short peptides were designed as targets to help direct aptamers to the desired site on the protein. Peptide **1** was chosen due to its important biological role and its location away from the negatively charged cell membrane. Peptide **1** is also well-structured, potentially allowing shape complementary binding with aptamers. The amino acid sequence of peptide **1** has a good mixture of polar and non-polar amino acids maximising the potential interactions with DNA bases, and only a few negatively charged amino acid chains, which should reduce repulsive interactions between the aptamers and target (Figure 4.2 **A**).

Table 4.1. Sequences of peptide **1** and **2** were purchased from Peptide Protein Reseach ltd and their masses are analysed by MS (Figure 4.15).

	Sequences (N terminus – C terminus)	Calc. (Da)	Found (Da)
Peptide **1**	Ac-KPALVTYNTDTFESMPNPEGR-NH_2	2408.65	2408.63
Peptide **2**	Ac–KGDPLNNTTPVTGASPG–NH_2	1666.79	1666.88

Ac: Acetyl; NH_2: amide

The second epitope (peptide **2**, Table 4.1) was selected due to its similar amino acid sequence to peptide **1**, but also due to its being structurally disordered and located on the opposite side of the HER2 extracellular domain (Figure 4.2 **B**). The two peptides can, therefore, be used as negative targets for each other in SELEX and the binding characteristics of enriched sequences from two individual SELEX can be compared and analysed.

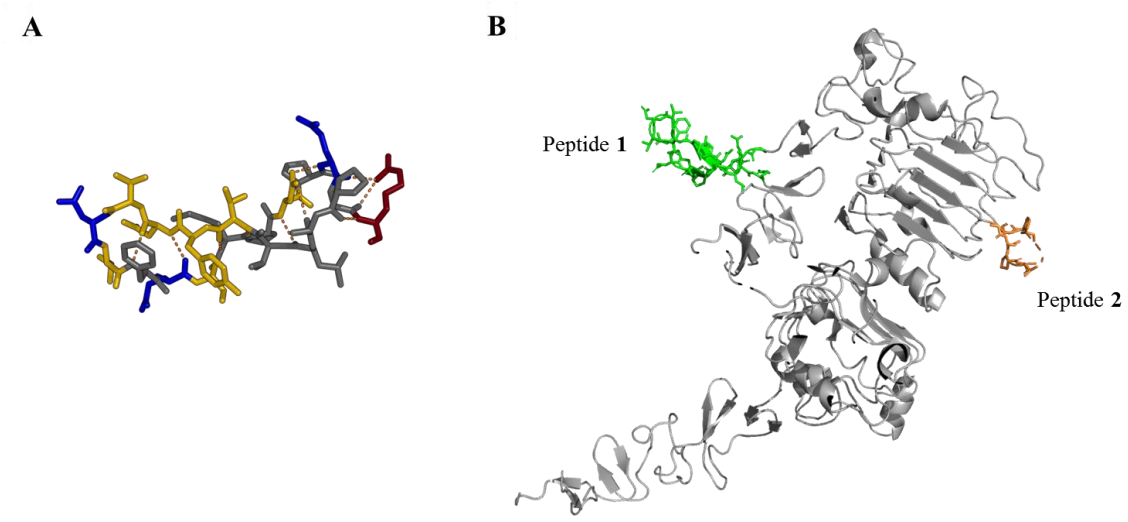


Figure 4.2. **A**) Structure of peptide **1** shown in the sticks (PDB code 1S78).[407] Polar amino acids (yellow), non-polar (dark grey), negatively charged amino acid (blue) and positively charged amino acid (red). The figure was generated using PyMOL program.[371] **B**) Ribbon diagram of HER2 (grey) with peptide **1** in domain II (green) and peptide **2** in domain I (orange) (PDB code 1N8Z).[397] Peptide **2** is structurally disordered and therefore it is not fully visible in crystal structure. The structural representation was made using PyMOL program.[371]

4.3 ssDNA library design

The design of the input library is an important starting point for the SELEX experiment. The sequence space and structural diversity are related to the length of the random region (N bases) in the DNA library. As discussed in section 1.2.2, the random region in the DNA library can consist of 4^N possible sequences. There is no fixed number of nucleotides in the random region, but the diversity of the sequences and the desired complexity of the tertiary structures formed have to be balanced to make SELEX successful. Very long libraries can raise the possibility of PCR artifact formation and not all possible sequences will exist in the library. Commonly, 30–50 nucleotides are chosen for the randomised regions. In this chapter, N was selected to be 40. In addition to this, two fixed primer sequences also need to be added to each end of the random region for amplification during SELEX. As a rule, 20 mer primers were generally

synthesised for uses in SELEX.[420] When designing primers, primer dimer formation and self-complementarity should be avoided, and the melting temperatures of the two primers should be similar.[421–423]

4.3.1 ssDNA generation method

The purity of ssDNA and the recovery yield from dsDNA is very important in SELEX. As described in section 1.2.3.2, there are several methods available for ssDNA generation. Here, lambda exonuclease digestion and streptavidin-magnetic beads separations methods were explored. These two methods are commonly used in SELEX and are reportedly fast and high yielding.[105,110] ssDNA containing non-canonical nucleotides should give similar yields and purities to canonical DNA during ssDNA generation. Therefore, the same ssDNA generation protocol should be applicable to both canonical and non-canonical ssDNAs. Table 4.2 shows primers and templates used for ssDNA generation experiments.

Table 4.2. Primers (P) and template (T) used for ssDNA generation experiments

	Sequence (5′- 3′)
P2	GCATTCGAGCAACGTAAG
P3	GGTTATGTGTGTCGGCAG
P4	**P**GCATTCGAGCAACGTAAG
P5	**B**GGTTATGTGTGTCGGCAG
T6	GGTTATGTGTGTCGGCAGTATTGTCAGTGTGAATTCCAGAG TGTGAGATTGTGTGCTGGCGATCTTACGTTGCTCGAATGC

P: Phosphate; **B**: Hexyl-biotin

4.3.1.1 Lambda (λ) exonuclease digestion

5′-phosphate reverse primer P4, 5′-hydroxyl forward primer P3 and 81 mer DNA template T6 were used in the PCR (Table 4.2). 5-PAF-dUTP was used instead of natural dTTP to form a modified DNA duplex and as a positive control, PCR with natural dNTP to form canonical DNA was also performed (Figure 4.3 **A**). PCR products were extracted using phenol/chloroform and isolated by ethanol precipitation. dsDNA was then incubated with λ-exonuclease before measuring the thermal stability of any residual duplex DNA. The melting temperature profiles show the canonical dsDNA melting transition but it is of far lower intensity than the untreated sample (Figure 4.3 **B1**) suggesting ssDNA generation using natural dNTP is efficient. However, DNA containing 5-PAF-dUTP still showed a strong duplex transition indicating that the full digestion of the modified ssDNA was incomplete (Figure 4.3 **B2**). This could be because the DNA containing the modified triphosphate is not a good substrate for λ-exonuclease. Hence, a streptavidin-magnetic beads separation method was explored as an alternative method of ssDNA generation containing modified and unmodified nucleotides.

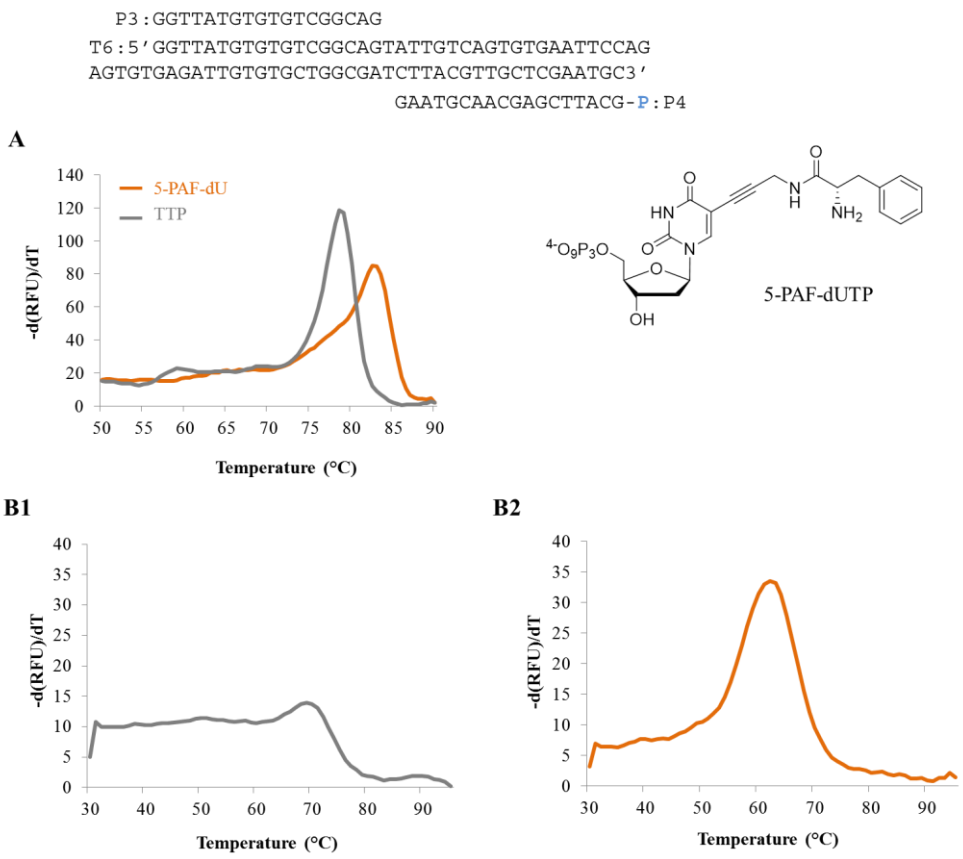


Figure 4.3. **A**) Melting temperature experiment with PCR amplicons containing natural dTTP (grey line) and 5-PAF-dUTP (orange line) in place of dTTP. 5′-phosphate reverse primer P4, 5′-hydroxyl forward primer P3 and 81 mer DNA template T6 were introduced in the reaction mixture. **B1**) Melting temperature experiments with canonical PCR products after lambda exonuclease digestion. **B2**) Melting temperature experiments with modified PCR products after lambda exonuclease digestion.

4.3.1.2 Streptavidin-magnetic beads separation

Prior to applying this method to modified DNA, biotinylated unmodified PCR products were produced using 5′-hydroxyl reverse primer P2, 5′-biotinylated forward primer P5 and 81 mer template T6 (Table 4.2). PCR products were then incubated with streptavidin magnetic beads (Dynabeads M-280 streptavidin, Thermo Fisher Scientific) to capture the biotinylated strand in the presence of low salt or high salt buffer. The desired complementary stand was eluted with NaOH. The buffer containing the

complementary strand was neutralised using HCL before loading on a 12% native polyacrylamide gel (Figure 4.4). When high salt buffer was used, biotin labelled DNA was quantitatively captured (Figure 4.4, Lane **1a**) with no high MW band (81 bp dsDNA) observed on the gel.

In contrast, a large amount of dsDNA was still unbound in low salt buffer (Figure 4.4, lanes **2a** and **3a**). After capture, the desired strands were displaced by alkaline treatment (20 mM NaOH) as recommended in the literature,[102,424] however, both dsDNA and ssDNA were eluted (Figure 4.4, lanes **1c** and **2c**). Some studies have suggested that pre-conditioning the beads in NaOH can prevent contamination of the eluate with streptavidin-biotinylated strand.[102,424] This is because NaOH can release the more labile streptavidin monomers before the beads are used to capture the biotinylated dsDNA. However, pre-conditioning the beads (20 mM NaOH, 1 h) produced ssDNA as well as streptavidin-biotinylated dsDNA (Figure 4.4, lane **3c**). Finally, the beads were treated with formamide to remove all oligonucleotide from the beads and confirm that dsDNA on the beads was successfully converted into ssDNA (Figure 4.4, lanes **1d**, **2d** and **3d**). ssDNA was observed, probably the biotinylated strand which ran slower than the complementary strand with little contamination with dsDNA (Figure 4.4, lanes **2d** and **3d**). However, the biotinylated strands are not ideal for use in SELEX as the biotin can be involved in target binding.[103]

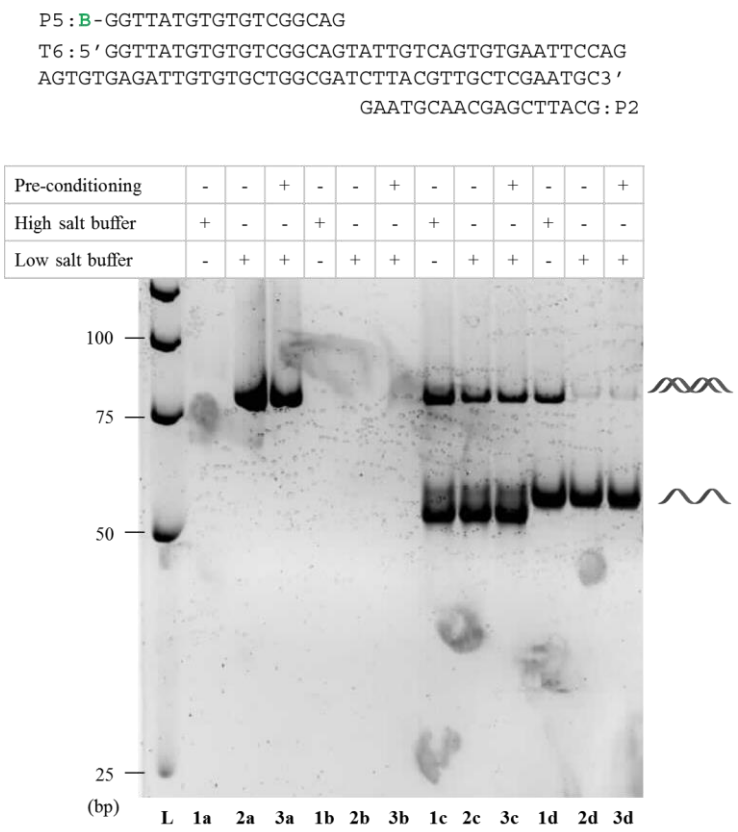


Figure 4.4. Optimisation of streptavidin magnetic bead separation method after PCR amplification. 5′-hydroxyl reverse primer P2, 5′-biotin forward primer P5 and 81 mer DNA template T6 were introduced in the reaction mixture. Lane **L**, 25 base-pair DNA ladder; lanes **1a**/**2a**/**3a**, supernatant after binding to streptavidin-magnetic beads; lanes **1b**/**2b**/**3b**, 3rd wash after binding; lanes **1c**/**2c**/**3c**, complementary strands after alkaline treatment (20 mM NaOH); lanes **1d**/**2d**/**3d**, biotinylated strand after formamide treatment. High salts buffer contains 10 mM HEPES pH 7.5, 2 M NaCl and 0.01% TWEEN-20. Low salts buffer contains 10 mM HEPES pH 7.5, 150 mM NaCl and 0.01% TWEEN-20. Pre-conditioning was done with 20 mM NaOH at 25 ℃ for 1 h. All other conditions are the same unless otherwise stated. (12% native polyacrylamide gel electrophoresis stained with 0.5× SYBR Gold).

Pre-conditioning beads for longer hours (24 h) also did not give a big improvement in the purity of ssDNA (Figure 4.5, lane **1c**). The problems of this ssDNA generation method are not only due to the release of streptavidin monomer as reported in the literature,[103] but also incomplete dsDNA dissociation.

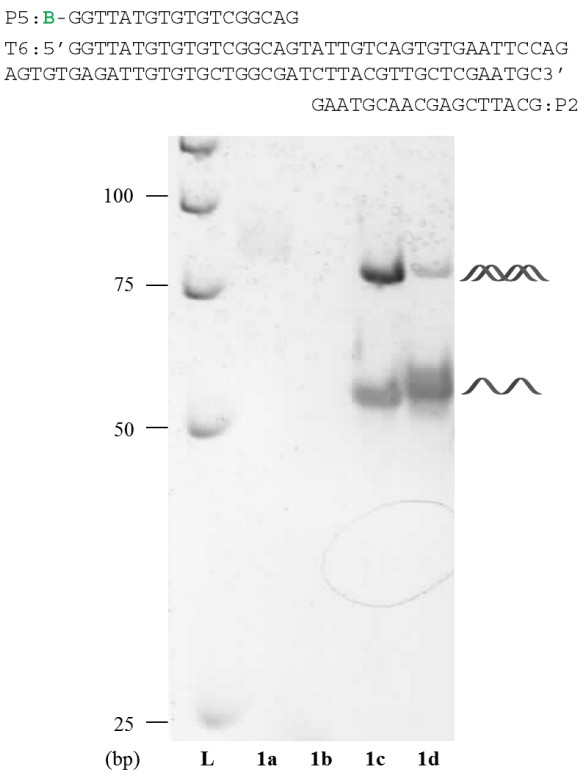


Figure 4.5. Optimisation of streptavidin magnetic beads separation method after PCR amplification. 5′-hydroxyl reverse primer P2, 5′-biotin forward primer P5 and 81 mer DNA template T6 were introduced in the reaction mixture. Lane **L**, 25 base-pair DNA ladder; lane **1a**, supernatant after binding to streptavidin-magnetic beads; lane **1b**, 3rd wash after binding; lane **1c**, complementary strands after alkaline treatment (20 mM NaOH); lane **1d**, biotinylated strand after formamide treatment. A high salt buffer which contains 10 mM HEPES pH 7.5, 2 M NaCl and 0.01% TWEEN-20 was used for binding. Pre-conditioning was done with 20 mM NaOH at 25 ℃ for 24 h. (12% native polyacrylamide gel electrophoresis stained with 0.5× SYBR Gold).

In order to fully dissociate the dsDNA into ssDNA, the concentration of NaOH was increased to 200 mM. Pre-conditioning beads were also carried out with 200 mM NaOH, instead of 20 mM used previously to wash away weakly bound streptavidin monomers.[101] Based on previous results, the high salt buffer was used to capture dsDNA. Figure 4.6 compares the yield of ssDNA with and without pre-conditioning steps. Dramatically decreased amount of dsDNA was observed after alkaline treatment and without pre-conditioning (Figure 4.6, lane **1c**). In summary, preparation of ssDNA

based on streptavidin magnetic beads in the high salt buffer and detachment with 200 mM NaOH without pre-conditioning the beads produced a high yield of ssDNA.

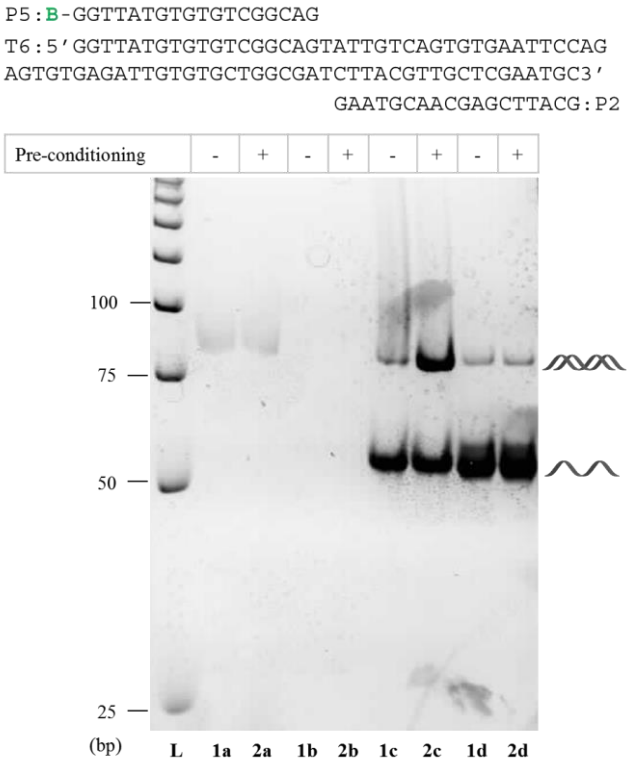


Figure 4.6. Optimisation of streptavidin magnetic beads separation method after PCR amplification. 5′-hydroxyl reverse primer P2, 5′-biotin forward primer P5 and 81 mer DNA template T6 were introduced in the reaction mixture. Lane **L**, 25 base-pair DNA ladder; lanes **1a**/**2a**, supernatant after binding to streptavidin-magnetic beads; lanes **1b**/**2b**, 3rd wash after binding; lanes **1c**/**2c**, complementary strands after alkaline treatment (200 mM NaOH); lanes **1d**/**2d**, biotinylated strand after formamide treatment. High salts buffer (10 mM HEPES pH 7.5, 2 M NaCl and 0.01% TWEEN-20) was used in capturing step. Pre-conditioning of beads with 200 mM NaOH at 25 ℃ for 1 h was done. (12% native polyacrylamide gel electrophoresis stained with 0.5× SYBR Gold)

Small amounts of biotinylated strands released from the beads can regenerate dsDNA after neutralisation and thereby may limit the diversity of ssDNA pools. In addition, detached streptavidin possibly acts as a target during SELEX and therefore, this method was not used further.

In summary, both lambda exonuclease and streptavidin-biotin ssDNA preparation methods do not seem appropriate. Therefore, size-separation by gel electrophoresis was ultimately used (see section 4.4.2). This method is more time-consuming and expensive than the other two methods, but the purity of ssDNA after extraction is known to be high.[108,110]

4.3.2 ssDNA library and primers

As discussed in section 4.3, the ssDNA library T7 was designed for SELEX using canonical nucleotides (Table 4.3). It consisted of a 40 nucleotide (nt) randomised region, flanked by two primer binding regions (Figure 4.7). The random region consists of an equal amount of all four nucleotides (N). Forward primer P6 is comprised of 20 mer with 5′-FAM and reverse primer P7 has 20 mer with a hexaethylene glycol (HEG) spacer and an extension with polyA tails at the 5′-end (Table 4.3) for ease of size separation of the desired strand by a gel electrophoresis. The desired strand used in selection is FAM-labelled DNA which can be easily visualised and quantified by fluorescence intensity, even with low amounts of DNA (see section 4.4.3). The melting temperatures of the primers were designed to be similar (64 ℃ for P6 and 64.8 ℃ for P7) in the presence of 50 mM NaCl and 2 mM $MgCl_2$ (Q5 high fidelity DNA polymerase buffer, calculated by a New England Biolabs *T*m calculator). Homo- and hetero-dimerisation tests on the primers using the IDT oligo analyser confirmed that the maximum base pairings are two which should eliminate primer dimer formation during PCR. A starting ssDNA library containing a randomised region and two primers were prepared by solid-phase synthesis (see section 1.5.1).

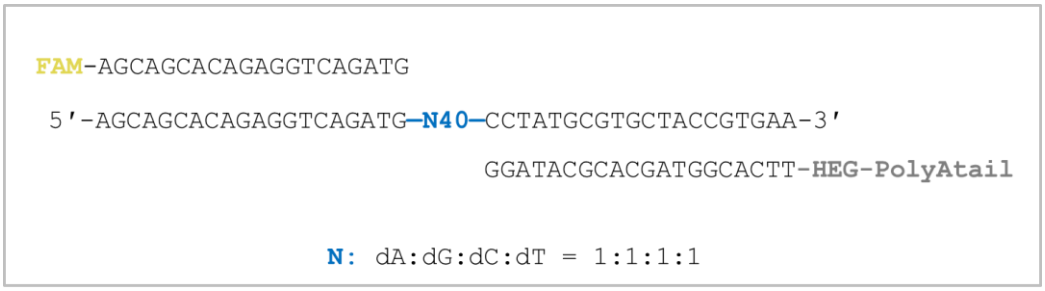


Figure 4.7. Schematic representation of ssDNA library using non-modified substrates. The library is composed of a 40 mer randomised region, flanked by FAM labelled 20 mer primer and PolyA tailed 40 mer primer binding sites. The random region in ssDNA library was synthesised ideally with an equal distribution of dA, dG, dC and dT.

Table 4.3. SELEX primers (P) and initial ssDNA library (T)

	Sequence (5´-3´)
P6	**F**AGCAGCACAGAGGTCAGATG
P7	AAAAAAAAAAAAAAAAAAAA**H**TTCACGGTAGCACGCATAGG
T7	AGCAGCACAGAGGTCAGATGNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNCCTATGCGTGCTACCGTGAA

F: 5′-6-FAM (Fluorescein); **H**: Hexaethylene Glycol (HEG);
N: Random sequences of A, G, C or T (25% chance of each base)

4.4 Amplification of DNA library and optimisation

The most crucial steps during SELEX are to amplify the enriched library for the next round of selection. This consists of PCR amplification to generate multiple copies of the selected sequences and ssDNA generation for use in the next round of selection.

4.4.1 PCR amplification with natural dNTPs

For PCR, Vent exo- DNA polymerase, commonly used to amplify DNA that contains modified triphosphates,[257] and Q5 Hot Start DNA polymerase, which works efficiently for natural DNA with high fidelity, were tested. Initially, the two DNA polymerases were tested in a non-templated control (NTC) reaction to identify any contamination of PCR reagents. In SELEX, contamination with sequences that are not from the desired DNA pools can lead to selection failure. Non-specific amplification appeared only after 30 cycles for both DNA polymerases (Figure 4.8 **A** and **B**). This is typical for PCR and therefore it was assumed there is no contamination in the PCR reagents.[96,425–427] Q5 Hot Start DNA polymerase was selected for SELEX using dNTPs as this enzyme is cost-effective and, can be handled at room temperature minimising non-specific amplification. The PCR amplification using ssDNA library T7 and two primers P6 and P7 was performed using Q5 DNA polymerase and the successful amplification was achieved after 15–20 cycles (Figure 4.8 **C**). Therefore, 25 cycles in PCR were used during SELEX to avoid non-specific amplification and achieve efficient library amplification (Figure 4.9).

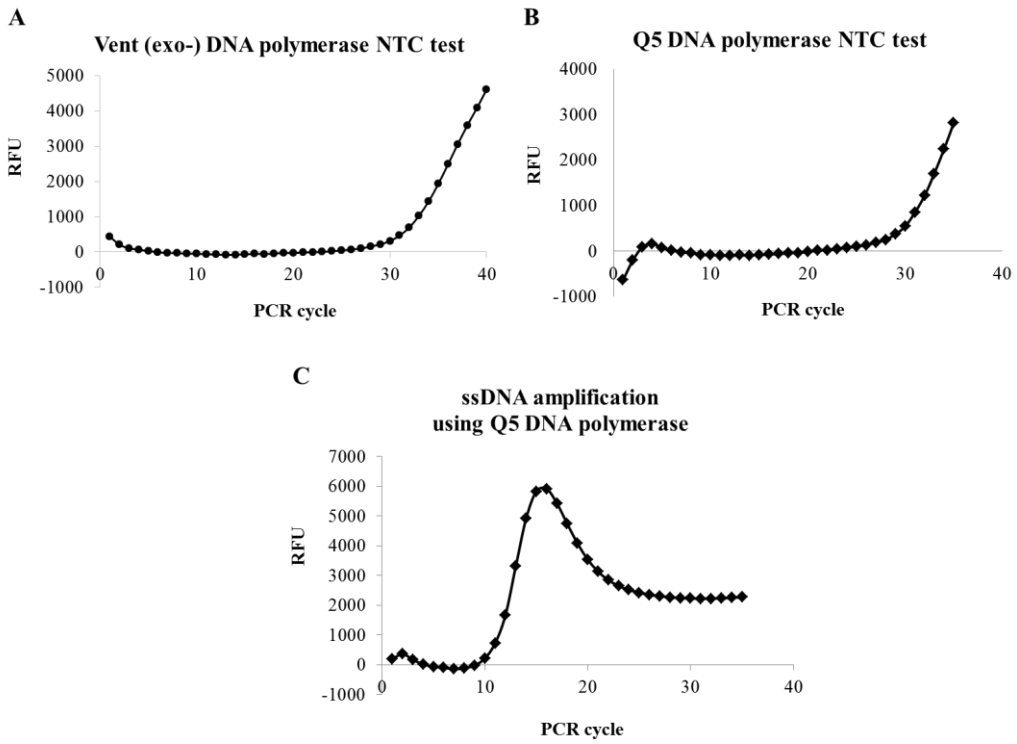


Figure 4.8. Q-PCR non-templated control (NTC) test by using Vent exo- DNA polymerase (**A**) and Q5 DNA polymerase (**B**). **C**) ssDNA Q-PCR amplification of ssDNA library T7 using Q5 DNA polymerase.

Initial denaturation	Denaturation	Annealing	Extension	Final extension	Cooling
98 ℃	98 ℃	65 ℃	72 ℃	72 ℃	4 ℃
30 sec	10 sec	20 sec	30 sec	2 min	10 sec
			x24		

Figure 4.9. PCR cycling program for SELEX. Initial denaturation at 98 ℃ for 30 sec; 25 cycles of denaturation at 98 ℃ for 10 sec, annealing at 65 ℃ for 20 sec and extension at 72 ℃ for 30 sec; final extension at 72 ℃ for 2 min and hold at 4 ℃ for 10 sec. Note that for Q-PCR quantification, fluorescence was measured at the end of the extension (camera).

4.4.2 Size separation for ssDNA

The subsequent step is to generate ssDNA which is required for the next round of SELEX. As discussed in section 4.3.1, lambda exonuclease digestion and streptavidin magnetic beads separation methods proved problematic. Hence, size separation based methods were used through the design of primers with different lengths. This method, although lower-yielding, gives better separation of ssDNA from dsDNA and avoids contamination with streptavidin. Reverse primer P7 (Table 4.3) contains a hexylethylene glycol spacer and a 5′ extension of 20 dA nucleotides which makes this antisense strand migrates slower on the denaturing PAGE gel than the sense strand. Moreover, the forward primer P6 (Table 4.3) is labelled at the 5′ position with fluorescein (FAM) which enables selective visualisation and excision of the strand of interest. Denaturing PAGE gel in Figure 4.10 shows good separation of the 80 mer FAM labelled desired strands and 100 mer antisense strands after PCR amplification.

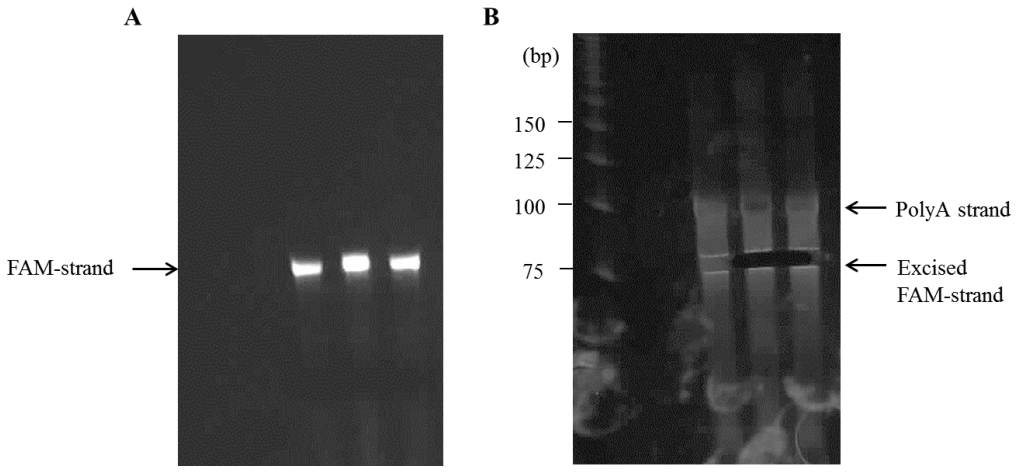


Figure 4.10. An example of FAM labelled product extraction from 10% denaturing PAGE after PCR amplification. 5′-polyA tail reverse primer P7, 5′-FAM forward primer P6 and ssDNA libraries were used in PCR. The streaky bands represent non-specific amplification and truncated (not fully elongated or excess primer) sequences during PCR. **A**) Fluorescence detection of FAM-strands. **B**) UV detection of dsDNA at 260 nm with both FAM-strand (lower bands around 80 mer strand) PolyA-strand (upper bands indicated by an arrow) after SYBR Gold staining (0.5×).

4.4.3 Quantification of ssDNA by fluorescence measurement

The inaccurate molar extinction coefficient of a random library makes ssDNA concentration hard to determine. This is important as the concentration of target vs ligand affects the recovered aptamer affinity constant (K_D value).[428] However, through fluorescently modified primers, ssDNA can be easily quantified by fluorescence intensity. Very low amounts of ssDNA (50 fmol) can also be measured, far lower than the detectable limits of UV absorbance.[114] The absolute amount of the desired sequences was calculated by comparing the fluorescence to a calibration curve generated using a known concentration of the FAM-labelled primer (Figure 4.11).

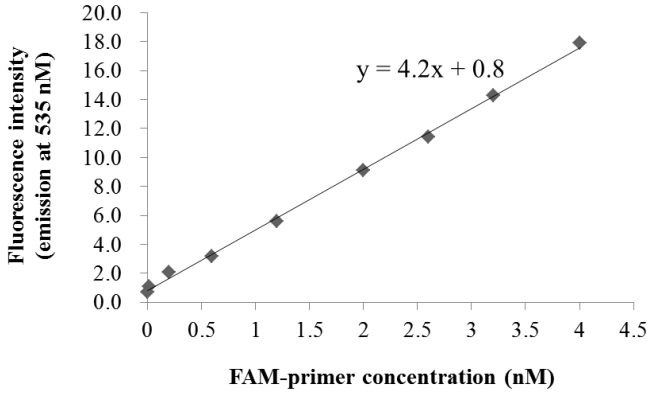


Figure 4.11. A standard fluorescence intensity curve for the FAM-labelled primer concentration from 20 pM to 4 nM.

Fluorescence intensity of ssDNA pool for each round of SELEX was measured using a fluorimeter and ssDNA concentration kept at 160 nM for each round or adjusted suitably.

4.5 Adaptation of the selection method

For selection, the target peptides were immobilised on beads, a method that has yielded high-affinity aptamers in literature.[429] The major challenge of performing SELEX with peptides is partitioning the bound and unbound sequences due to the relatively small size of the peptides. The mass difference between the unbound sequences and the peptide–ssDNA complexes is insufficient to separate by size. Small molecule SELEX typically involve immobilisation of one of the biomolecules.[118,430] Therefore, immobilisation of the peptides on magnetic beads was selected as it should allow rapid selections and only a small amount of target needs to be used during selection. This makes the selection competitive and therefore, high affinity aptamers should be discovered using this method.[114,430–432]

Published aptamers are commonly selected using Tris-HCl buffer,[114,433–436] HEPES buffer,[156,184] or phosphate buffered saline (PBS) pH 7.5 buffer.[115,437,438] Here, PBS was selected as it eliminates free amines from the buffer, which can complicate immobilisation and subsequent affinity testing.

4.5.1 Peptide immobilisation and quantification

Quantification of peptide concentration is required to design a SELEX procedure. The amount of targets incubated with ssDNA is an important parameter. Concentration affects detectable affinity constants of aptamers. If lower concentrations of targets are used, higher affinity aptamers will be selected, but if too low, nothing will be retained for amplification. Hence, the stringency of the selection can be increased by adjusting the amount of targets to discover aptamers with high binding affinity in the later rounds of SELEX.

There are several assays to quantify the protein/peptide concentration.[439–441] The intrinsic UV-absorbing properties of aromatic tyrosine (Y) or tryptophan (W) around 280 nm can be exploited when sufficient protein/peptide is available (20–3,000 µg).[442,443] However, during SELEX, peptides are immobilised on a solid support in very low quantities and peptide **2** does not contain any tyrosine or tryptophan in its sequence (Table 4.1). Alternatively, amino groups can be labelled using various fluorescent probes, a common technique to quantitate a small amount of protein/peptide (0.05–25 µg). Lysine residues (K) or a free N-terminus on protein/peptides can react with dyes to generate fluorescence. This should give a linear response with increasing protein concentration. Three dyes are frequently used – o-phthalaldehyde (OPA),[444] Fluorescamine,[445,446] and 3-(4-carboxybenzoyl)quinoline-2-carboxyaldehyde (CBQCA).[447–449] Among them, fluorescamine reacts directly with the amine functional group, whereas OPA and CBQCATM require additional reagents. Therefore, for simplicity, target peptide concentration was assayed using fluorescamine.

Both peptide **1** and **2** have N-terminal lysine (K) and C-terminal amide. Due to the resonace effect of C-terminal amide, fluorescamine preferentially reacts faster with the primary amino groups of the lysine side chains (Table 4.1). Therefore, a serial dilution of peptide **1** was prepared starting from 0.05 µM to a maximum concentration of 20 µM to generate a standard curve against which samples are compared (note this concentration was based on the solid mass of peptide provided, and the MW).

Once the standard curve for quantification was set, 4.15 nmol of peptide **1** was then attached to the 5.9 mg carboxylic acid beads (Dynabeads M-270 carboxylic acid, Thermo Fisher Scientific) by amide bond formation via EDC/NHS coupling, according to the manufacturer's recommendation. After immobilisation of peptide **1**, the

supernatant was collected and the fluorescence intensity was measured to check the concentration of free peptides. As a control, 4.15 nmol free peptide **1**, before immobilisation, was also measured (Figure 4.12). Based on the standard curve, the initial concentration of peptide **1** in buffer was approximately, 20 µM and the supernatant collected after immobilisation was estimated to be 3 µM indicating the majority of peptide **1** was immobilised on the carboxylic acid magnetic beads.

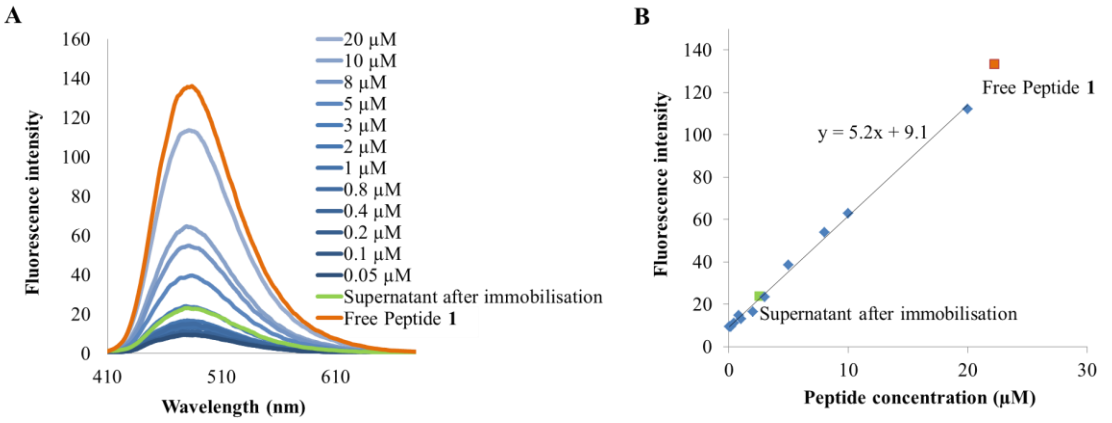


Figure 4.12. **A**) Immobilised peptide **1** on carboxylic acid magnetic beads and quantification by fluorescein assay. **B**) A calibration curve of peptides at 475 nM and two values fitted with an equation, y=5.2x+9.1 where y values are known. The fluorescence intensity of free peptide **1** before immobilisation is coloured in orange. The fluorescence intensity of the supernatant after peptide immobilisation is coloured in light green.

After immobilisation, beads were washed 4 times with buffer and the fluorescence intensity of the 4th wash was analysed to confirm unbound peptide **1** is completely washed away (Figure 4.13). Deactivation of the unreacted carboxylic acids was done by adding 1M ethanolamine-HCl (pH 8.5) to the beads before washing 5 times to remove excess ethanolamine. The beads containing immobilised peptide were kept at 4 ℃ in buffer. To test stability (SELEX typically takes 3–5 weeks), the peptide in the buffer

solution after 3 weeks of storage was determined and this indicated that the hydrolysis of peptides (loss of peptide from beads) occurs over time (red line, Figure 4.13).

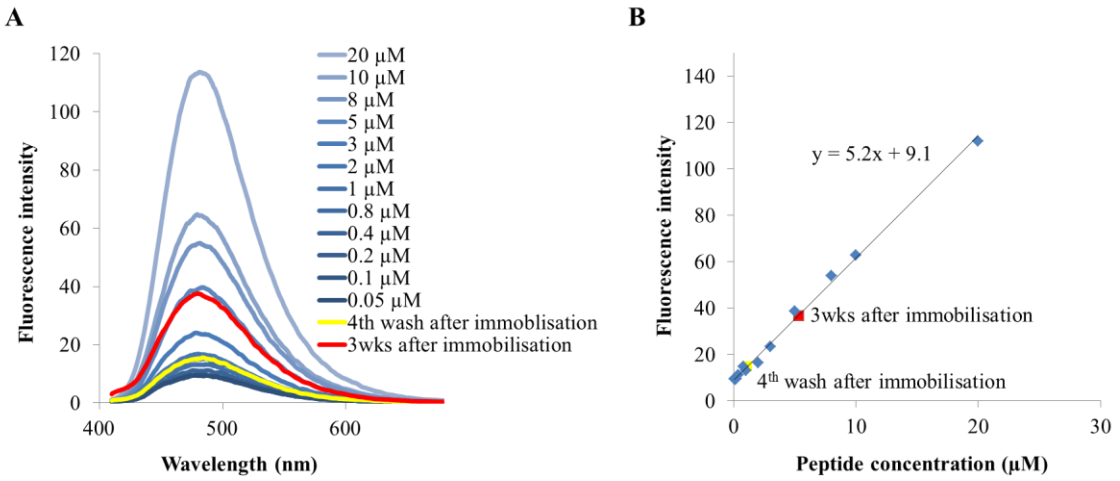


Figure 4.13. **A**) Immobilised peptide **1** on carboxylic acid magnetic beads and quantification by fluorescein assay. **B**) A calibration curve of peptides at 475 nM and two values fitted with the equation, y=5.2x+9.1 where y values are known. The fluorescence intensity of 4th wash after the peptide **1** immobilisation is coloured in yellow. The fluorescence intensity of the immobilised peptides kept in a storage buffer for 3 weeks is coloured in red.

Attempts to find the best storage conditions for the immobilised peptides were made and the results were compared with free peptide **1** (Figure 4.14). Four different storage conditions were tested: 1) condition A – B/W buffer (40 mM HEPES pH 7.5, 125 mM NaCl, 5 mM KCL, 1mM $MgCl_2$ and 1 mM $CaCl_2$) at -20 ℃; 2) condition B – is the same B/W buffer but kept at 4 ℃; 3) condition C – 70% ethanol at -20 ℃; 4) condition D – 40 mM HEPES pH 7.5 with 1 mM EDTA at 4 ℃. Results suggested that immobilised peptide on beads kept at -20 ℃ in B/W buffer or at 4 ℃ in HEPES buffer containing EDTA were the most stable (Figure 4.14, **A1** and **D1**). However, the manufacturer's recommendation is to keep the magnetic beads at 4 ℃ (*i.e.* unfrozen). Therefore, in order to keep the immobilised peptide in solution (non-frozen state) at -20

℃, 70% ethanol was added (condition C) and the sample kept at -20 ℃. However, condition C showed the highest rate of hydrolysis of peptides for three days (Figure 4.14, **C1**). Surprisingly, free peptide **1** in all conditions did not show any significant degradation (Figure 4.14, **A2**, **B2**, **C2** and **D2**).

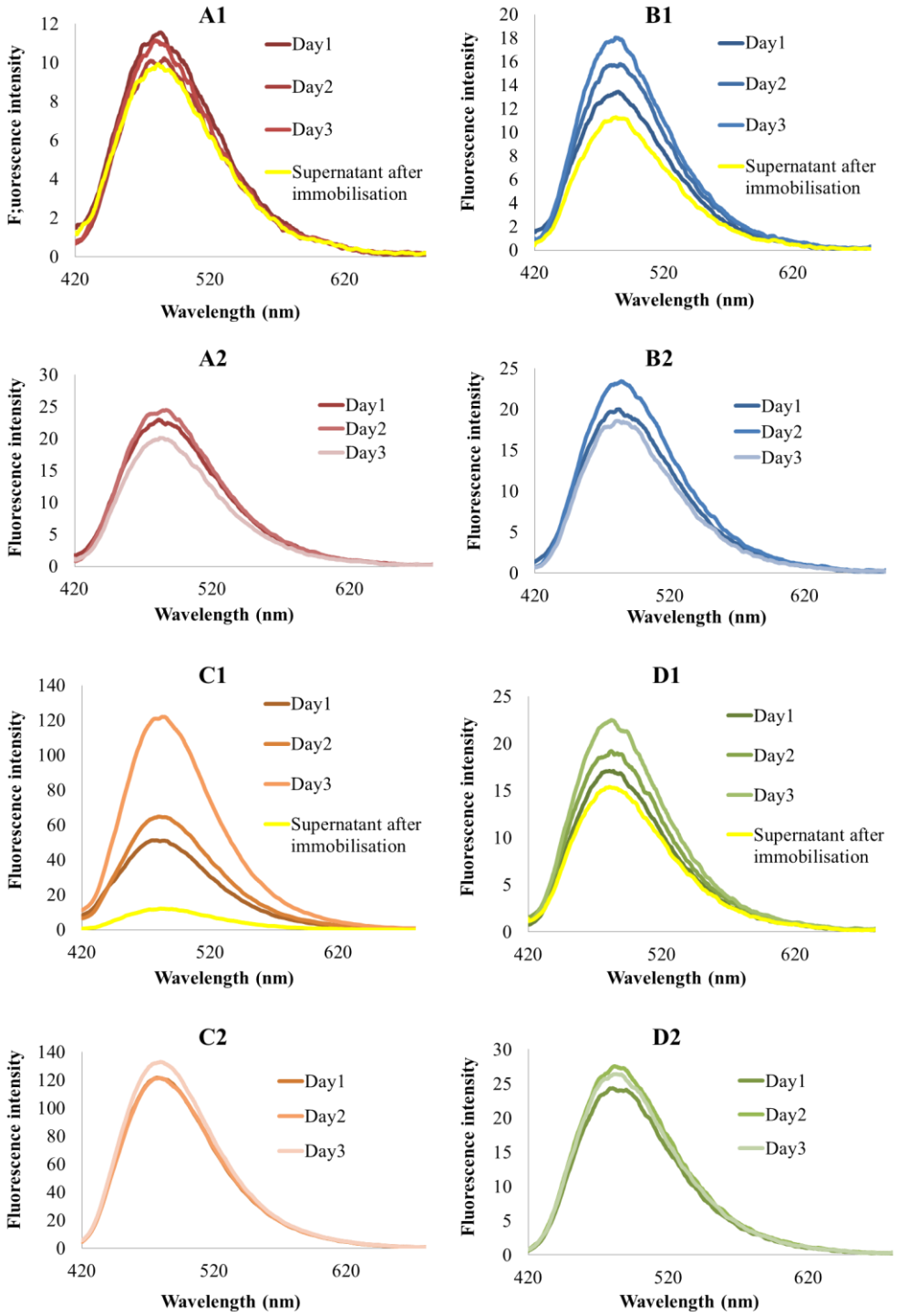
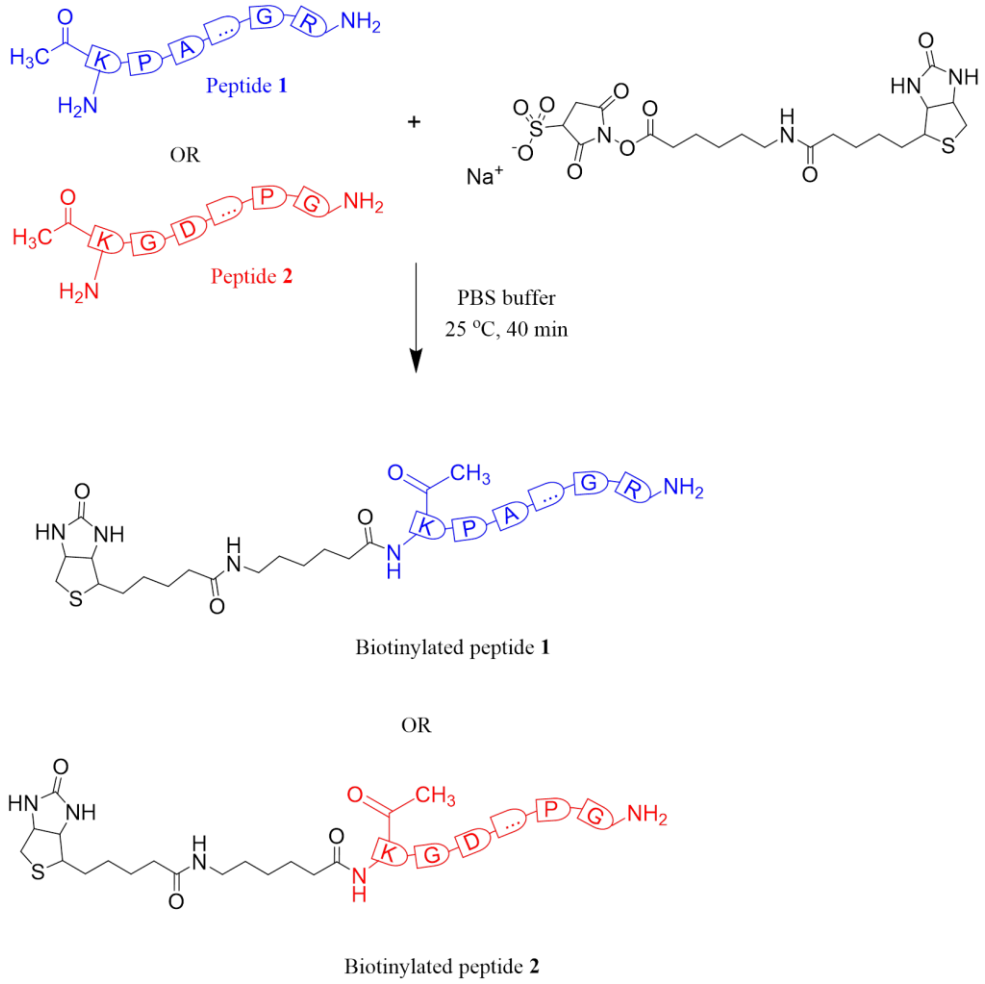


Figure 4.14. Stability of immobilised peptides and free peptides in different storage conditions. Condition A: immobilised peptide **1** (**A1**) and free peptide **1** (**A2**) in B/W buffer at -20 ℃. Condition B: immobilised peptide **1** (**B1**) and free peptide **1** (**B2**) in B/W buffer at 4 ℃. Condition C: immobilised peptide **1** (**C1**) and free peptide **1** (**C2**) in 70% ethanol at -20 ℃. Condition D: immobilised peptide **1** (**D1**) and free peptide **1** (**D2**) in 40 mM HEPES with 1 mM EDTA buffer at 4 ℃. When the peptide is degraded into amino acids fragments, free amines at N-terminus are reacted with the fluorescamine and therefore, the fluorescence intensities become higher.

Therefore, we decided to use peptides in solution during selection and use streptavidin-coated magnetic beads to capture the ssDNA-peptide complexes after selection. In order to do this, the peptides were biotinylated at the primary amine of the lysine residue with an amine-reactive biotin *N*-hydroxysuccinimide ester. Here, we used Sulfo-NHS-LC-Biotin (Thermo Fisher Scientific) to add biotin to target peptides (Scheme 4.1).



Scheme 4.1. Labelling of peptides with sulfo-NHS-LC-Biotin reagent.

After purification by a Pierce™ C18 spin column, the biotinylated peptide was characterised by mass spectrometry (Figure 4.15).

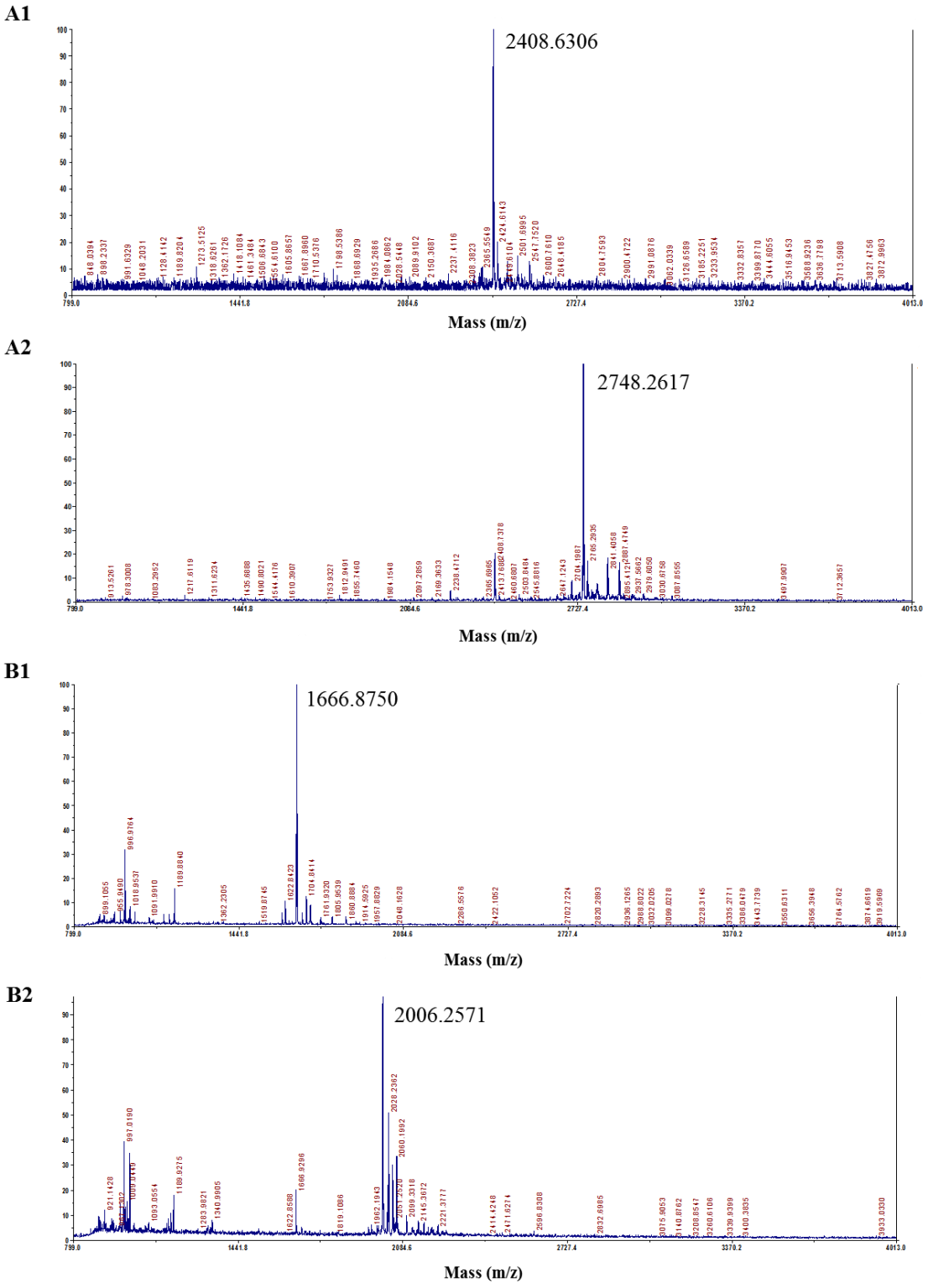


Figure 4.15. **A1**) Crude mass spectrum of peptide **1**. Calc. 2408.6; found. 2408.6 (g/mol). **A2**) Biotinylated peptide **1**. Calc. 2749.3; found. 2748.3 (g/mol); **B1**) Crude mass spectrum of peptide **2**. Calc. 1666.8; found. 1666.9 (g/mol). **B2**) Biotinylated peptide **2**. Calc. 2006.2; found. 2006.3 (g/mol).

The exact concentration after purification is unknown but the recovery rate is assumed to be 100%. The concentration of biotinylated peptide **1** (BP1) was assumed to be 500 µM and the concentration of biotinylated peptide **2** (BP2) was assumed to be 620 µM. A final concentration of 10 µM of the peptide in 100 µL HBS-EP buffer (supplied by GE healthcare) was prepared for both BP1 and BP2. In order to cross-validate the concentration, the biotinylated peptide flowed over a SPR chip coated with streptavidin. This allows calculation of analyte binding capacity (R_{max} value for maximum response), since this will determine the expected response range (Equation 4.1). The following equation assumes a 1:1 binding stoichiometry:

$$R_{max}(\mathrm{RU}) = \frac{\text{Analyte MW}}{\text{Ligand MW}} \times \text{immobilised ligand level (RU)}$$

Equation 4.1. R_{max} equation for the maximum response of analyte on SPR.

According to the equation above, the theoretical R_{max} is 106.8 RU for FC1 and 74.2 RU for FC2 where streptavidin (ligand) MW is assumed to be 53,000 Da. Around 2,000RU streptavidin was coated onto flow cell 1 and 2 of a CM5 chip. After immobilisation of peptides on a SPR chip, the binding observed for FC1 was 104 RU whereas 72 RU was observed for FC2 (Figure 4.16). The theoretical R_{max} values matched the experimental values for both BP1 and BP2, suggesting the bulk of the sample is biotinylated.

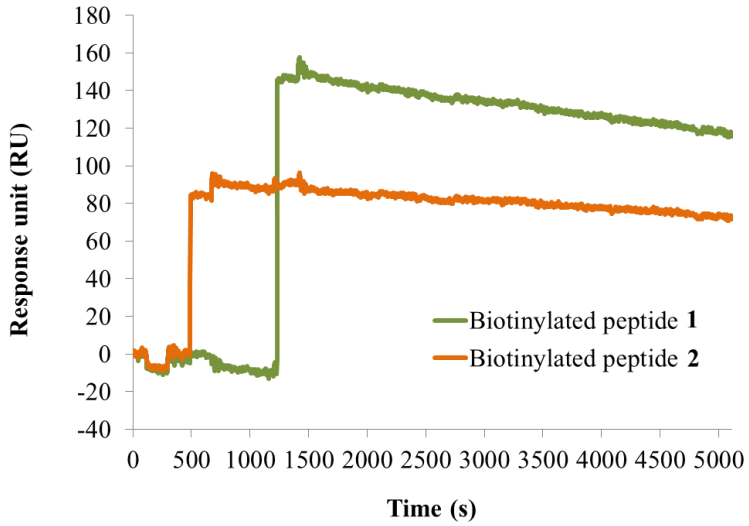


Figure 4.16. Immobilisation of biotinylated peptide **1** (BP1) and **2** (BP2) on FC1 and FC2, respectively. 104 RU for FC1 and 72 RU for FC2 were observed. The theoretical R_{max} is 106.8 RU for FC1 and 74.2 RU for FC2.

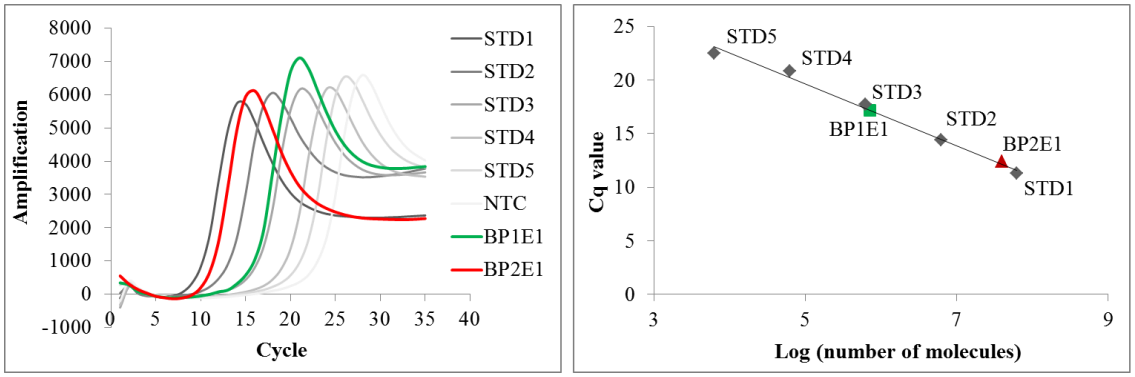
4.5.2 Selection

After establishing the required techniques and materials for SELEX, two SELEX protocols were performed in parallel for peptide target BP1 or BP2 using an unmodified DNA library. PBS buffer and the same starting ssDNA library T7 were used for both targets. ssDNA and targets incubated in free solution and streptavidin magnetic beads (Dynabeads Myone Streptavidin T1, Thermo Fisher Scientific) were used to capture the ssDNA-BP complexes.

In order to induce folding of the ssDNA for shape complementary binding, libraries were denatured at 95 ℃ for 15 min and shock frozen in liquid nitrogen. This aids unimolecular oligonucleotide folding as the samples slowly thaw to RT. 2 µM of the ssDNA library T7 was incubated with 1 µM of BP1 and 1.2 µM of BP2, respectively in 1 mL PBS buffer at 25 ℃ for 2 h. 250 µg of non-conjugated streptavidin beads was added to the mixture and incubated for another 20 min at 800 rpm to capture the ssDNA-BP complexes and free BP. The beads were thoroughly washed three times (shaken for 5 min) with 200 µL PBS buffer to remove the non-specific and weakly

bound DNA. The enriched ssDNA was then eluted with 100 μL PBS buffer by heat denaturation (95 ℃, 15 min). In order to eliminate the non-specific binding to the immobilised support, the eluted ssDNA (150 μL PBS buffer) was then incubated with 250 μg non-conjugated streptavidin beads (BPC1 or BPC2, negative selection). The supernatant was collected after 20 min incubation. The eluted ssDNA after the first round of SELEX was then determined by fluorimetry.

Quantitative real-time polymerase chain reaction (Q-PCR) was used for the absolute quantification of target-bound oligonucleotides during SELEX.[450] The amplification curve of eluted ssDNA from BP1 from round 1 (abbreviation: BP1E1) reached the maximum amplification around 20–22 cycles of PCR and 15–17 cycles for BP2E1 (eluted ssDNA from BP2 from round 1) (Figure 4.17). Therefore, the optimum would be around 25 cycles to avoid over-amplification which can lead to non-template specific amplification and subsequently to a complete loss of the desired aptamer. Each round of SELEX was monitored by Q-PCR. The concentrations of the eluted ssDNA are 10–20 pM in the early stage of selection and the eluted ssDNA was quantified by using Q-PCR standard curves, although it is challenging to quantify by fluorescence intensity due to the low amount of the ssDNA. The concentration of the oligonucleotides to be used in the next round of selection was adjusted according to the previous results. The concentrations of oligonucleotide used for the selection are 0.1–1 μM and therefore, these were also quantified by fluorescence intensity. SELEX rounds were then repeated until the binding recovery rate plateaus and did not increase further.



Sample	Concentration (per 1µL)	Number of molecules
STD1	1 nM	6.23E8
STD2	100 pM	6.23E7
STD3	10 pM	6.23E6
STD4	1 pM	6.23E5
STD5	0.1 pM	6.23E4

Figure 4.17. Q-PCR amplification results for the first round of SELEX (left) and a standard curve of cq values vs log (number of molecules) (right). STD1–STD5 represents standard samples of a known quantity. NTC is non-templated control. BP1E1 is a green curve indicating a number of ssDNA molecule eluted from round 1 selection using BP1. BP2E1 is a red curve indicating a number of ssDNA molecule eluted from round 1 selection using BP2.

Recovery rates are calculated as follows; the number of molecules eluted from the positive (or negative) target divide by the number of molecules in the incubated ssDNA pool (see appendix 8.8 for data). Trends in binding rate over the rounds were observed and also the binding between the positive and negative targets was compared (Figure 4.18).

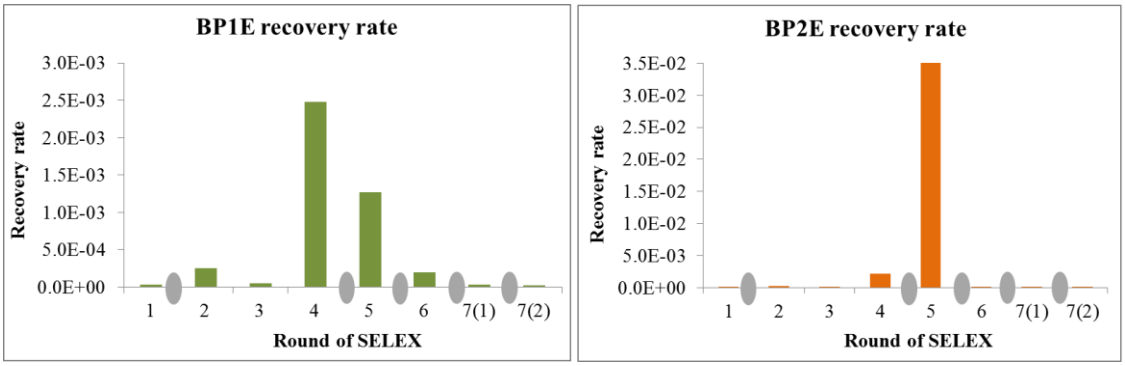


Figure 4.18. The recovery rate of ssDNA bound to BP1 (left) and BP2 (right) determined in each selection round by absolute quantification using Q-PCR. The bar graph shows the progress in the selection of BP1 and BP2 specific aptamers. Negative selection (streptavidin beads/protein) was introduced after round 1 and re-introduced from rounds 5, 6 and 7 as indicated with grey circles.

Note that the eluted DNA was PCR amplified (see section 4.4.1) and ssDNA generation was performed as described in section 4.4.2. The desired gel extracted ssDNA was exchanged into PBS buffer by using a 10 kDa SnakeSkin dialysis tubing (Thermo Fisher Scientific). This dialysis method was used for ssDNA pool 1–3. However, dialysis caused the loss of ssDNA and therefore, low quantities of initial ssDNA pools were used from round 2–4 of SELEX (Table 4.4). For example, the concentration of BP1 pool 1, before the dialysis, was 1.9 µM, but it decreased to 0.6 µM after dialysis overnight. The fluorescence intensity of PBS outside the dialysis tube was checked and this confirmed the loss of ssDNA. Therefore, the dialysis method was changed to a combination of Amicon ultracentrifugal filtration (10 kDa MWCO, MerckMillipore) and Sephadex illustra NAP columns (GE Healthcare). The new buffer exchange method maintained a high yield of ssDNA from round 5.

In order to enhance the number of sequences that bind specifically to the desired sample, streptavidin beads were used for the negative selection in the first round. The subsequent rounds of selection were carried out with BP (positive targets) only to enrich

the population before the negative selection was started again from round 5 onwards. Before performing the positive selection in round 5, the initial ssDNA was incubated with 400 µg streptavidin magnetic beads and the unbound sequences were then incubated with positive targets. During round 6 and 7, 10 µM of streptavidin (New England Biolabs) was used as the negative target. The unbound sequences were recovered using 30 kDa MWCO Amicon ultracentrifugal filters. The streptavidin protein and the complexes were captured and the unbound sequences were filtered through the membrane. This is almost 100–200 equivalents more streptavidin relative to peptide used in the positive selection (Table 4.4). This increases the stringency of selection and should ensure that sequences that bind to streptavidin were eliminated prior to binding with BP targets.

As shown in Figure 4.18, the sequences were enriched at rounds 4 and 5 for BP1 and round 5 for BP2. However, the negative selection rounds appear to remove a large proportion of the sequences enriched during rounds 5 and 6, suggesting the enriched sequences bound to the streptavidin, instead of BP targets. The MW of BP1 and BP2 are about 2.4 kDa and 1.6 kDa, respectively whereas the MW of streptavidin is about 53 kDa, so this is not surprising. The much larger surface area of streptavidin favours the selection of aptamers for it instead of for the BPs. In 2002, Tahiri-Alaoui *et al*. discovered a 2′-F modified RNA aptamer binding to streptavidin that was pre-saturated with biotinylated peptide. Their finding elucidated that aptamers can also bind to the non-biotin-binding region of streptavidin with a high binding affinity.[156] In addition, a lack of the negative selection from rounds 2–4 may reduce the number of sequences that bind to peptides. Instead, those sequences bound to streptavidin beads would be enriched during this period.

Table 4.4. SELEX conditions from rounds (R) 1 to 7 for **A**) BP1 and **B**) BP2. BP1 represents a biotinylated peptide **1** and BP2 represents a biotinylated peptide **2** (positive target = pos). SA beads are streptavidin magnetic beads and SA proteins are streptavidin proteins (negative target = neg).

A) BP1 SELEX								
	R1	R2	R3	R4	R5	R6	R7	
ssDNA pool	2 µM	160 nM	0.5 µM	0.5 µM	1 µM	1 µM	200 nM	100 nM
Volume (selection buffer)	1 mL	50 µL	50 µL	50 µL	600 µL	150 µL	100 µL	100 µL
BP1	1 µM 1 nmol	200 nM 10 pmol	200 nM 10 pmol	200 nM 10 pmol	100 nM 10 pmol	50 nM 7.5 pmol	100 nM 10 pmol	50 nM 5 pmol
SA beads	250 µg	-	-	-	400 µg	-	-	-
SA protein	-	-	-	-	-	1.5 nmol	1 nmol	1 nmol
Incubation time	Pos 2 h ↓ Neg 20 min	Pos 1h	Pos 1h	Pos 1h	Neg 1 h ↓ Pos 1h	Neg 1 h ↓ Pos 1h	Neg 1 h ↓ Pos 1h	Neg 1 h ↓ Pos 1h

B) BP2 SELEX								
	R1	R2	R3	R4	R5	R6	R7	
ssDNA pool	2 µM	160 nM	0.5 µM	0.5 µM	1 µM	1 µM	200 nM	100 nM
Volume (selection buffer)	1 mL	50 µL	50 µL	50 µL	200 µL	150 µL	100 µL	100 µL
BP2	1.2 µM 1.2 nmol	250 nM 12.5 pmol	200 nM 10 pmol	200 nM 10 pmol	100 nM 10 pmol	50 nM 7.5 pmol	100 nM 10 pmol	50 nM 5 pmol
SA beads	250 µg	-	-	-	400 µg	-	-	-
SA protein	-	-	-	-	-	1.5 nmol	1 nmol	1 nmol
Incubation time	Pos 2 h ↓ Neg 20 min	Pos 1h	Pos 1h	Pos 1h	Neg 1 h ↓ Pos 1h	Neg 1 h ↓ Pos 1h	Neg 1 h ↓ Pos 1h	Neg 1 h ↓ Pos 1h

4.5.3 Melting temperatures from rounds 1–7 of SELEX

The melting temperature of dsDNA depends on the sequences composition and length. Hence, qualitative estimation of the diversity of each SELEX round can be made by determining the melting temperature of the dsDNA product after PCR in the presence of the intercalating dye, EvaGreen. Once, the temperature approaches the melting

temperature of the dsDNA the intercalating dyes are released resulting in a decrease of fluorescence intensity. As expected, in the first few rounds of SELEX lower melting temperatures were observed due to the hybridisation of the primer sites only rather than the entire sequence (Figure 4.19). In round 4, melting profiles showed that the diversity of the pool started to drop (the dsDNA melting transition at 60 ℃ decreased). During SELEX cycle 5–7, distinctive peaks at higher temperatures were observed and only board, shallow peaks were present at low melting temperatures. This suggests the majority of sequences were fully matched in terms of base pairing, indicating sequence enrichment.

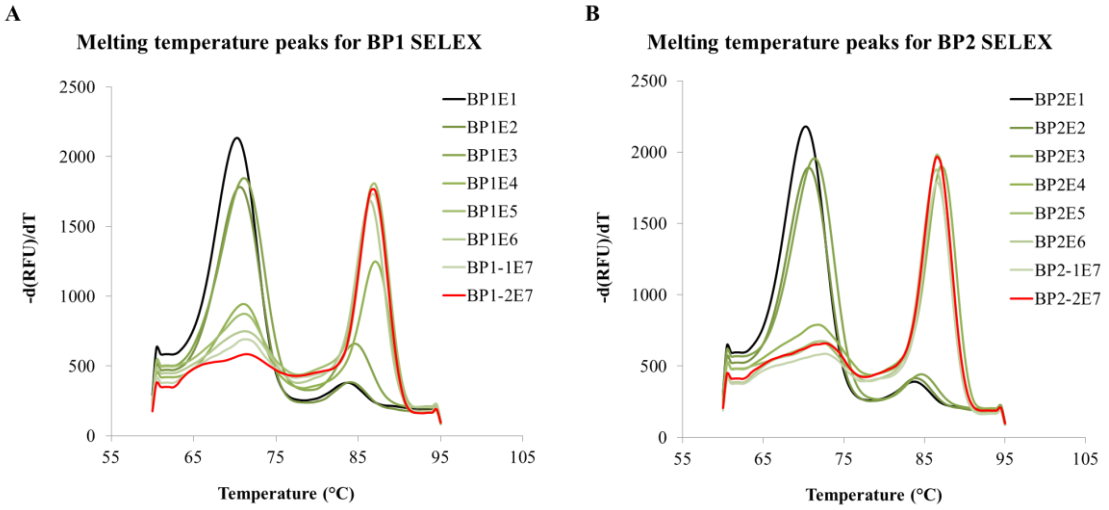


Figure 4.19. Fluorescent melting temperatures of Q-PCR products for **A**) BP1 and **B**) BP2 SELEX. E1, ssDNA eluted from round 1; E2, ssDNA eluted from round 2. E3, ssDNA eluted from round 3; E4, ssDNA eluted from round 4, E5 ssDNA eluted from round 5, E6, ssDNA eluted from round 6; E7, ssDNA eluted from round 7.

4.6 Binding of aptamer candidates

Next, each ssDNA pool from different rounds of SELEX for BP1 and BP2 were analysed by surface plasmon resonance (see section 1.6.1). If a clear binding response is seen, this is a good indication SELEX steps have worked. Peptide **1** and **2** were manually immobilised onto two individual CM5 chips via amide coupling.

For BP1 SELEX, eluted ssDNA from round 5 showed the second highest recovery rate and eluted ssDNA from round 6 showed a dramatic decrease in recovery rate (Figure 4.18 **A**). In order to validate these sequences pools on SPR, ssDNA eluted from BP1 (positive target) during round 5 and 6 were prepared by PCR and ssDNA was isolated by PAGE. These were denoted as pool5P and pool6P, respectively. In contrast, ssDNA eluted from streptavidin magnetic beads or streptavidin protein (negative target) during round 5 and 6 were assigned as pool5N and pool6N, respectively.

Non-specific binding of sequences was observed for BP1 and BP2 (see appendix 8.9 and 8.10 for SPR data) due to the high salt concentration in the sample after the gel extractions step. Additional desalting steps were therefore performed using 10 kDa molecular weight cut-off (MWCO) Amicon ultracentrifugal filter (10 times wash with deionised water), followed by a Sephadex illustra NAP-10 column to minimise non-specific binding. This time, pool5 and pool6 from BP1 SELEX (pool5P and pool6P, Figure 4.20) were used to analyse binding to peptide **1** on SPR. ssDNA pools that bound to the negative selection target were also prepared (pool5N and pool6N, Figure 4.20) for comparison to sequences that were recovered from positive peptide **1** selection. After additional desalting, the non-specific binding was dramatically decreased (Figure 4.20 **A** and **B**) and there was a binding response from pool5P (Figure 4.20 **C**). According to Equation 4.1, the expected R_{max} for peptide **1** is approximately 4,200 RU where the MW

of the 80 mer ssDNA is 24,707 Da, peptide **1** is 2,408 Da and the immobilised RU value was 410. The observed maximum binding level of pool5P to peptide **1**, however, was 30 RU which is very low.

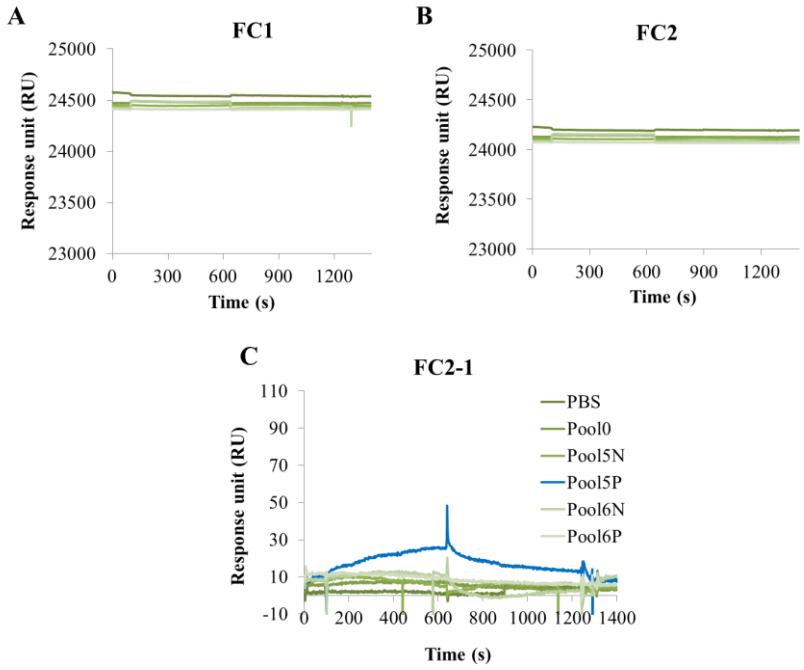


Figure 4.20. BP1 SELEX starting library, pool5N, pool5P, pool6N and pool6P binding to peptide **1** after desalting. **A**) ssDNA pools binding to reference (FC1). **B**) ssDNA pools binding to peptide **1** (FC2). Note 104 RU was immobilised on the FC1 **C**) Reference subtracted binding response of the ssDNA pool (FC2-1). Note that 410 RU was immobilised on FC1.

Our hypothesis is that during SELEX, sequences were enriched for streptavidin instead of the peptides due to the small size of peptides compared to streptavidin. To test this, streptavidin was immobilised on FC2 of a CM5 chip with FC1 left as a reference (Figure 4.21 **A**). 0.5 µM of ssDNA pools (pool5 and pool6) from BP1 SELEX were prepared and tested. Pool5N and pool6N showed binding to streptavidin because they are sequences recovered from negative streptavidin only selection. However, pool6P also showed the binding to streptavidin (Figure 4.21 **B**) which means the enriched sequences were selected for streptavidin binding. The maximum binding of pool6P was

~100 RU (Figure 4.21 **C**) with the estimated theoretical R_{max} value being approximately 1,100RU. This suggests the selected aptamers have a stronger affinity for streptavidin than to the target peptides.

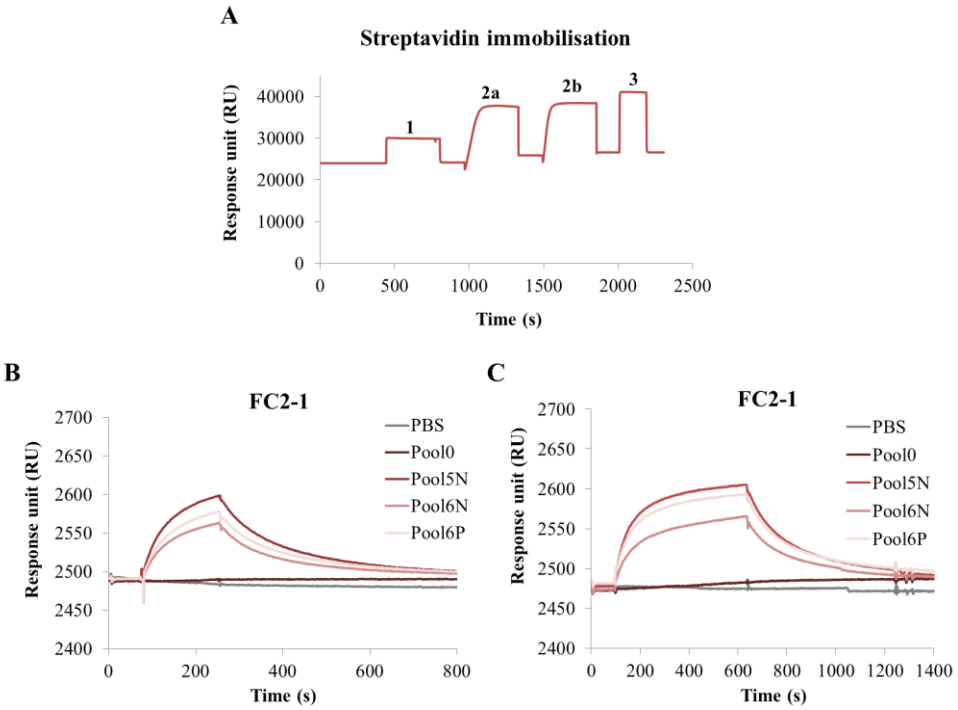


Figure 4.21. **A**) Streptavidin immobilisation to FC2 on CM5. **1**, EDC/NHS activation; **2a**, the first injection of streptavidin; **2b**, the second injection of streptavidin; **3**, 1 M ethanolamine-HCl (pH 8.5) deactivation. **B**) Binding analysis of BP1 SELEX ssDNA pools with short contact time (180 s). **C**) Binding analysis of BP1 SELEX ssDNA pools with long contact time (560 s). Note that 2412 RU was immobilised on the active flow cell (FC1) and that reference subtracted binding responses are shown.

4.7 Conclusions

In this chapter, a canonical DNA library was used to establish SELEX methods for future use with a non-canonical DNA library. The targets of interest were two different epitopes found in HER2 extracellular domain, with peptide **1** being biologically important and the site of binding for the drug pertuzumab, while peptide **2** is an unstructured control target of similar composition. For SELEX, the best ssDNA generation method uses a FAM-labelled primer, which enables easier quantification of DNA using fluorescence as well as aiding gel visualisation, and another primer with HEG and poly-A tails for size separation by denaturing PAGE. Selection against immobilised peptide targets was attempted, but the immobilised peptides were not stable in buffers. Therefore, biotinylated peptides were incubated with ssDNA and the complexes were captured by streptavidin magnetic beads to partition between unbound ssDNA and bound ssDNA. It was found that desalting gel extracted ssDNA by dialysis tubing was tedious and led to a significant loss of material. A faster and high yield desalting step was achieved using a ultracentrifugal filter and NAP columns which helps to shorten the selection time.

In total, 7 rounds of selections were performed for both BP1 and BP2 targets. Aptamer enrichment was observed by Q-PCR in round 4 and 5 for both selections, but the majority of the aptamers were lost after the negative selection using streptavidin. Melting temperature experiments supported sequence enrichment after round 4 for BP1 SELEX. However, SPR binding analysis showed that the enriched aptamers from BP1 SELEX had a higher affinity towards streptavidin and no (or very low) binding to their target peptides. The main reasons for the inability to obtain aptamers for the peptides are probably the relatively small size of the peptides compared to streptavidin and also aptamer bound to non-biotin-binding region of streptavidin.[156] A lack of the negative

selection at the early stage of selection may also reduce the number of sequences binding to peptides. However, the SELEX protocol was established for use with modified DNA libraries. This was employed in Chapter 5 with streptavidin as a model target.

Chapter 5

Selection of halogen bearing DNA aptamer against streptavidin

5 Selection of halogen bearing DNA aptamer against streptavidin

5.1 Background

In chapter 4, selection techniques were established using canonical DNA libraries. Aptamer selection was successful, but to streptavidin immobilised on the solid-support rather than the expected HER2 peptide target. This enabled proof of concept studies to evaluate the power of halogen bonding in a biomolecule-biomolecule context that are described in this chapter. Streptavidin is a well characterised, thermally stable, easy-to-handle protein and unmodified aptamers that bind to it are known, providing ideal controls.[114,115,434,451–454] Therefore, streptavidin immobilised on magnetic beads was used as the target during the SELEX work in this chapter.

Streptavidin, a 52.8 kDa homo-tetrameric protein (Figure 5.1 **A**) originally isolated from *Streptomyces avidinii*, is widely used in bioanalysis and purification, because of its extremely strong non-covalent interactions with the natural ligand biotin ($K_D = 10^{-14}$–10^{-16} M). In the absence of biotin, the binding pocket is open and filled with solvent molecules. Biotin binding displaces the solvent and allows biotin to form up to eight hydrogen bonds as well as van der Waals interaction with the protein (Figure 5.1 **B**).[453,455] Subsequent conformational changes result in loops 3 and 4 from the β strand (Ser45 to Ser52) covering the binding pocket, strongly stabilising the streptavidin-biotin interaction (Figure 5.1 **B**).[456]

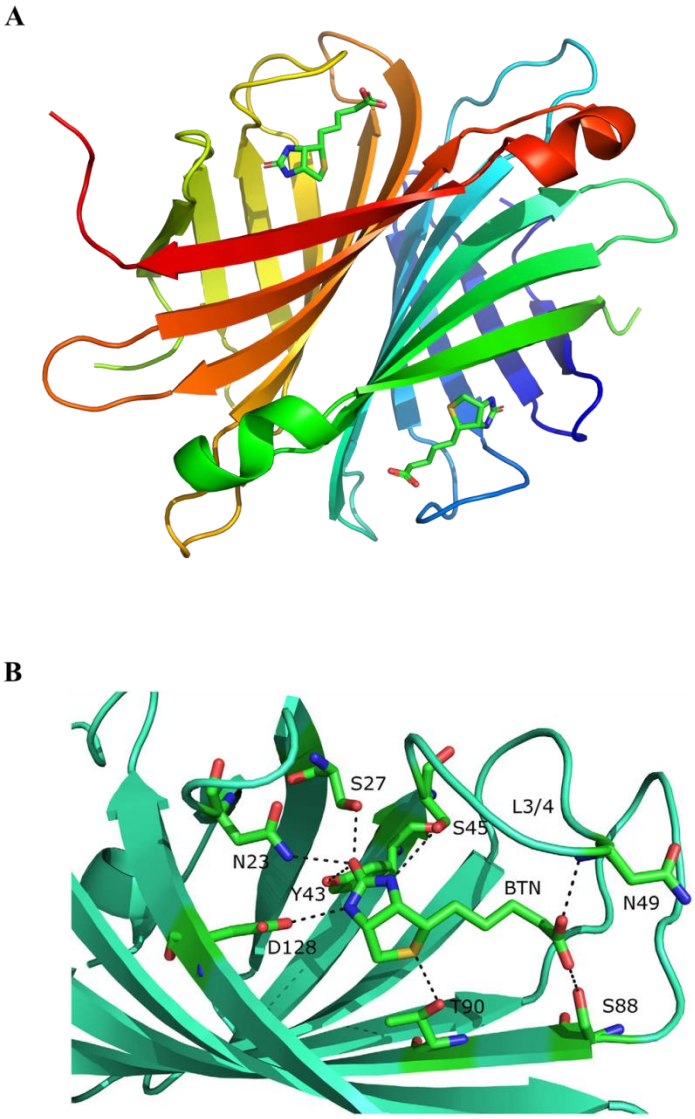


Figure 5.1. **A**) The secondary structure elements of streptavidin (PDB code 3RY2)[454] are indicated in cartoon form and bound biotin molecules are shown as sticks. This figure was generated using PyMOL.[371] **B**) Zoom-in of biotin binding site of wild-type streptavidin. Eight hydrogen bonds to biotin are marked as dashed lines. Picture reuse was permitted by the Springer Nature (see appendix 8.11).[455]

However, the strong multi-valent interactions between streptavidin and biotin sometimes limit their downstream applications. As a result, modification of streptavidin-biotin interactions has been explored.[457] Streptavidin has four binding sites that multiple biotinylated ligands can bind to with high affinity. However, this can cause

target clustering *in vivo* which may change the ultimate target function. For example, streptavidin was used to label a biotinylated cell surface receptor and as a result, cellular signalling was perturbed and receptor internalisation was induced due to receptor crosslinking. The use of streptavidin in cell biology is therefore restricted to fixed cells (not live cells).[457] To solve this problem, Howarth *et al.* re-engineered streptavidin to inactivate three of the binding sites and leave only one active site.[458] In addition, streptavidin has been re-engineered to lower the binding affinity to biotin, which is useful for applications such as affinity purification. The biotinylated ligand can then be recovered under mild elution conditions and used for further steps. To do this, O'Sullivan *et al.* introduced mutations at loop 7 and 8 between β7 and β8 and as a result, the binding affinity is reduced (K_D = 20 nM) by increasing the biotin dissociation rate. The binding is lower than for wild-type streptavidin, but is still high enough to be useful in bioseparation.[459]

Alternatively, RNA or DNA aptamers that bind to streptavidin have been developed by several research groups.[114,115,156,433,434,452,460,461] Especially, DNA aptamers share a common secondary structure motif (bulge-hairpin) with several key conserved nucleotides and have dissociation constants of 40–85 nM (Figure 5.2). It was shown that these aptamers compete with biotin for streptavidin binding, indicating the aptamers bind in or near the biotin-binding pocket.[156,434] Ruigrok *et al*. performed native mass spectrometry to study the stoichiometry of the streptavidin–aptamer complex and discovered that two aptamer moieties simultaneously bind to streptavidin. They also used SAXS analysis to show that one aptamer occupies two biotin binding pockets. The two conserved guanines (G) in the loop and bulge (Figure 5.2) interact with the loops in streptavidin that cover the two biotin pockets.[452]

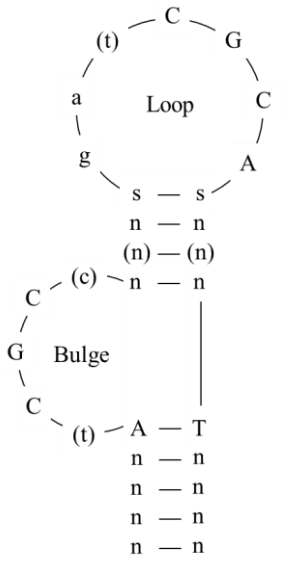


Figure 5.2. Aptamers to streptavidin. The secondary structure of the predicted stem-bulge-stem-loop sequences that were found by three different research groups.[114,115,434,452] Conserved nucleotides in the majority of sequences are written in capital and conserved nucleotides within family/families are written in lower cases. Nucleotides that are absent in some families were written in lower case with brackets, n represents one of four nucleotides and s is for C or G.

In order to increase target binding affinities of aptamers, the non-essential nucleotides which are irrelevant in aptamer binding can be removed from the aptamer. This can be done based on secondary structure predictions and sequence truncation by using a trial-and-error process. Not every nucleotide in aptamer sequences chosen by SELEX plays a critical role in target binding. Nucleotides in the primer sites have been shown to occasionally be involved in binding, but more often they have little or no effect. Therefore, sequences with truncations at the 5′- or 3′-end are synthesised to identify the minimal binding sequence.[462] As an example, the most studied thrombin binding aptamer was initially chosen from a library containing 60 mer random region with 50 mer primer sites. However, only 15 nucleotides are required for essential binding.[463] Another example is the first approved therapeutic RNA aptamer against VEGF. This

aptamer was selected from 61 and 71 nucleotides libraries, but only 27 nucleotides were chosen for its use as a drug.[61,71]

In Chapter 3, various modified dTTP analogues were synthesised and validated as substrates for DNA polymerases. In this chapter, the aim is to expand the interactions that aptamers can make with their targets. Therefore, 5-ITZ-dUTP is used in place of dTTP (Figure 5.3) during SELEX. 5-ITZ-dUTP was shown to be a good substrate for KOD XL DNA polymerase (see section 3.5) and the iodo-substituted triazole is anticipated to make halogen bonding interactions with the target which cannot occur in natural DNA (see section 3.1). Aptamers identified in SELEX will then be compared to control aptamers and the effect of the halogen modification will be studied by SPR experiments. This work is novel as halogen bonding in a biomolecule-biomolecule context has not been demonstrated previously.

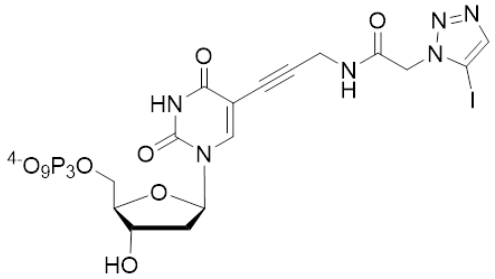


Figure 5.3. The chemical structure of 5-propargylamino-iodotriazole-dUTP (5-ITZ-dUTP) which is used in streptavidin aptamer SELEX.

5.2 Design of the 5-ITZ-dU library

Based on the SELEX procedure established in Chapter 4 (section 4.3.2), the same forward and reverse primers were used and the length of ssDNA random region was kept to a 40 mer. The key difference is that all thymidines of the random region were substituted with 5-ITZ-dU nucleotides (Figure 5.4). Therefore, the starting unmodified ssDNA library (T8, Table 5.1) was enzymatically replicated using 5-ITZ-dUTP and a primer with a poly-T tail (P8, Table 5.1). After size separation by denaturing PAGE, the desired modified ssDNA library was isolated (Figure 5.5). The enzymatically synthesised ssDNA libraries should contain an equal distribution of dA, dG, dC and 5-ITZ-dU and it was used as a starting library for SELEX.

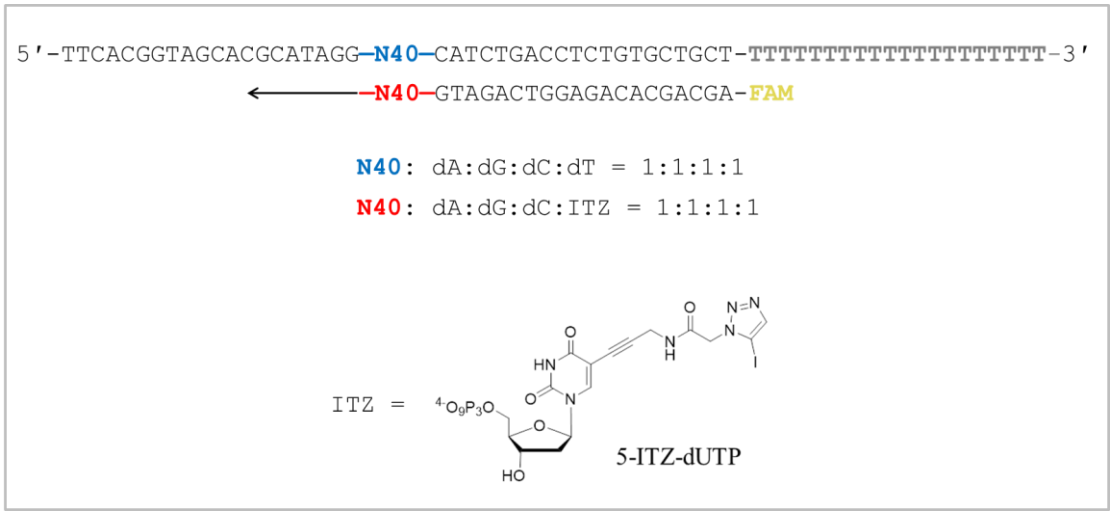


Figure 5.4. Schematic representation of ssDNA library generation using modified 5-ITZ-dUTP. The library is composed of a 40 mer randomised region with a poly-T tail at the 3′-end. A FAM labelled 20 mer primer was bound to the 3′-end primer site before linear copying along the direction of the arrow. The random region in the copied ssDNA library contained an equal distribution of dA, dG, dC and ITZ (5-ITZ-dUTP) and was used for SELEX.

Table 5.1. Primers (P) and DNA library template (T) used for streptavidin aptamer SELEX.

	Sequence (5´- 3´)
P6	**F**AGCAGCACAGAGGTCAGATG
P8	TTTTTTTTTTTTTTTTTTTT**H**TCACGGTAGCACGCATAGG
T8	TTCACGGTAGCACGCATAGGNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNCATCTGACCTCTGTGCTGCTTTTTTTTTTTTTTTTTTT

F: 5'-6-FAM (Fluorescein); **H**: Hexaethylene Glycol (HEG);
N: Random sequences of A, G, C or T (25% chance of each base)

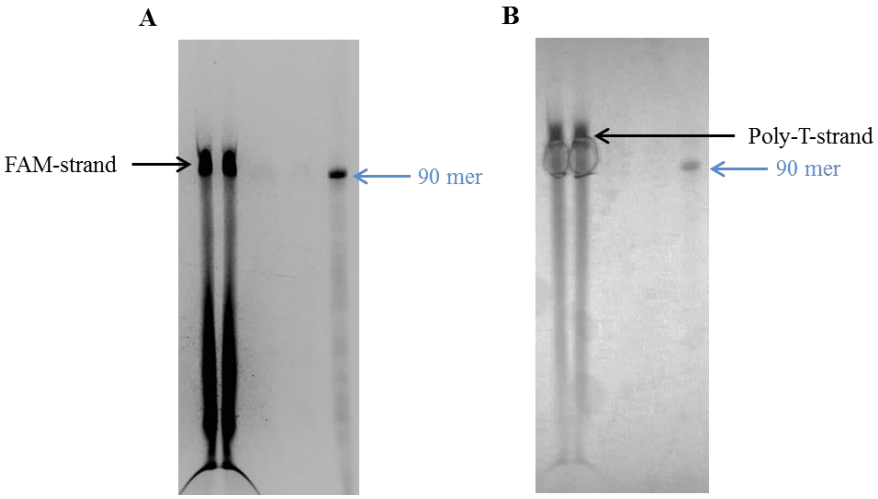


Figure 5.5. An example of FAM labelled 80 mer containing 5-ITZ-dU, which was extracted from the 10% denaturing PAGE gel after the primer extension reaction. ssDNA template T8 and FAM primer P6 were used in the reaction. The streaky bands represented non-specific amplification and truncated (not fully elongated or excess primer) sequences during enzymatic synthesis. **A**) Fluorescence detection of FAM- labelled strands with excess primers as the lowest MW band. **B**) UV detection of dsDNA at 260 nm with both FAM-strands (lower bands around 80 mer strand) and poly-T-strands (upper bands around 100 mer indicated by an arrow).

5.3 PCR amplification of 5-ITZ-dU library and ssDNA generation

In Chapter 3 (section 3.6), 5-ITZ-dU modified PCR amplicons were successfully generated using KOD XL DNA polymerase supplemented with 1 mM $MnCl_2$. Therefore, for SELEX, these conditions were used.

Melting temperature for this specific buffer was determined by mixing SELEX library T8 and FAM primer P6 (Table 5.1) and monitoring fluorescence thermal denaturation. As shown in Figure 5.6, the melting temperature is 71 ℃. Therefore, an annealing temperature of 65 ℃ was chosen and extension performed at 70 ℃ which is similar to the manufacturer's recommendation. The extension time was 1 min, which is slightly longer than normal, in order to ensure efficient incorporation of modified bases (Figure 5.7); various 5-ITZ-dUTP time course experiments in Chapter 3 (section 3.6) demonstrated that an extension time of 1 min was optimal and that prolonged extension times produced non-specific amplification products.

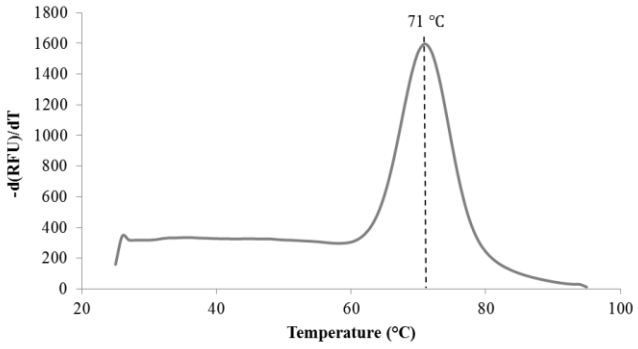


Figure 5.6. Melting temperature of SELEX template T8 and FAM primer P6 (Table 5.1). Melting temperature experiment was performed in KOD XL buffer supplemented with 1 mM $MnCl_2$. 2.5 µM of T8 and 0.5 µM of P6 were mixed with 1× EvaGreen (fluorescent dye).

Initial denaturation	Denaturation	Annealing	Extension	Final extension	Cooling
98 ℃	98 ℃	65 ℃	70 ℃	70 ℃	4 ℃
30 sec	30 sec	20 sec	1 min	5 min	30 sec
	←				
			x29		

Figure 5.7. PCR cycling programmed for streptavidin SELEX containing modified nucleotides. Initial denaturation at 98 ℃ for 30 sec; 30 cycles of denaturation of 98 ℃ for 30 sec, annealing at 65 ℃ for 20 sec and extension at 70 ℃ for 1 min; final extension at 70 ℃ for 5 min and hold at 4 ℃ for 30 sec. Note that for Q-PCR quantification, fluorescence was measured at the end of the extension (camera). Fluorescence was not monitored in normal SELEX.

A KOD XL DNA polymerase PCR master mix was made for consistency in amplification during each round of selection and kept at -20 ℃. Unlike Q5 DNA polymerase, the KOD XL DNA polymerase mixture was supplemented with 1 mM $MnCl_2$ for efficient incorporation of the modified triphosphate during enzymatic reactions. However, manganese in the PCR mixture inactivated the enzyme after storage for two days at -20 ℃. The colour of the mixture containing 1 mM $MnCl_2$ changed from a clear solution to a light brown solution suggesting that the oxidation state of manganese has changed. This was confirmed by comparing the PCR amplicons of the first round of SELEX (Figure 5.8 **A**–denaturing PAGE and **B**–agarose). The product bands were not observed when the PCR master mix contained $MnCl_2$ and had been stored for two days (Figure 5.8 **A**, lane 2 and **B**, lane 4). However, when $MnCl_2$ was added just before the reaction to the PCR master mix stored at -20 ℃, PCR products were observed (Figure 5.8 **A**, lane 1 and Figure 5.8 **B**, lane 5). Therefore, freshly prepared 1 mM $MnCl_2$ was added to the PCR master mix before PCR amplification during SELEX.

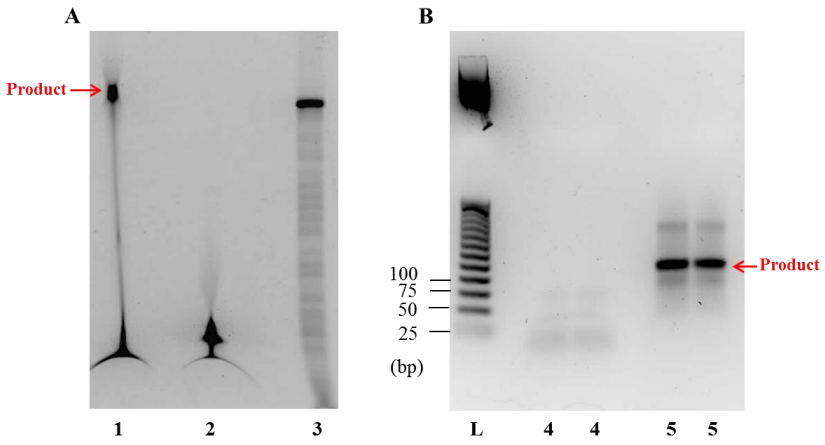


Figure 5.8. PCR products of ssDNA eluted from round 1 selection analysed on **A**) 10% denaturing PAGE and **B**) 2% agarose gel using a PCR mixture that was stored at -20 ℃ for two days with $MnCl_2$ (lanes **2** and **4**) and when $MnCl_2$ was separately added just before the reaction (lanes **1** and **5**). Lane **3** is a reference 80 mer oligonucleotide and lane **L** is the 25 bp DNA ladder. The successfully amplified product is indicated with red arrows.

5.4 Selection of a modified DNA aptamer for streptavidin

Streptavidin-coated magnetic beads were used as positive targets (Dynabeads M-280 streptavidin, Thermo Fisher Scientific). In order to prevent, non-specific sequence enrichment caused by binding to the magnetic beads, carboxylic acid magnetic beads (Dynabeads M-270 carboxylic acid, Thermo Fisher Scientific) were used as negative targets. In Chapter 4, the importance of performing the negative selection in the early stage of selection was highlighted. Hence, in this chapter, one round of selection involves negative selection first to eliminate sequences that non-specifically bind to the magnetic beads before the unbound sequences were then incubated with the desired streptavidin target beads.

The initial SELEX library was prepared by a linear copying of the template using modified nucleotides and followed by size separation ssDNA displacement by denaturing PAGE (Figure 5.9 **A**). ssDNA libraries containing 10^{13}–10^{14} molecules were

then incubated with magnetic beads for 1 h in 1× Phosphate-buffered saline (PBS) buffer (Figure 5.9 **B**) and negative target bound sequences were removed using a magnet (Figure 5.9 **C**). Unbound sequences were then collected and incubated with streptavidin beads for another 1 h (Figure 5.9 **D**). ssDNA-streptavidin complexes were captured using a magnet (Figure 5.9 **F**) and unbound sequences for positive targets were removed (Figure 5.9 **E**). ssDNA was eluted from the target by heat denaturation at 95 ℃ for 15 min and then PCR amplification of eluted DNA was performed. The desired strands were displaced by gel electrophoresis for uses in the next round of selection (Figure 5.9 **G**).

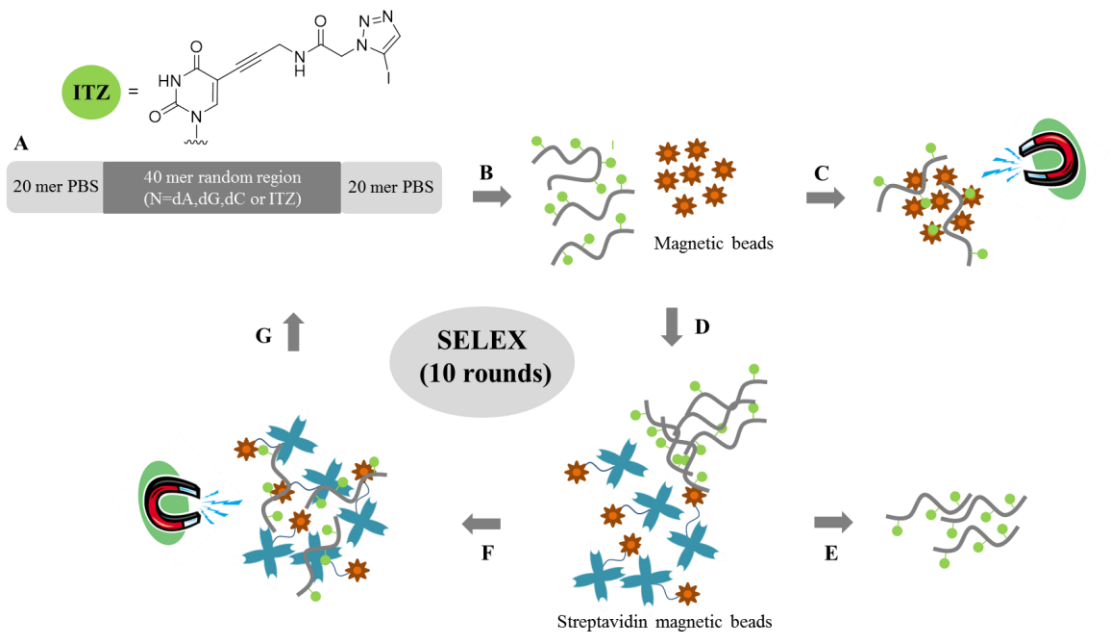


Figure 5.9. A schematic diagram of streptavidin SELEX. **A**) ssDNA library preparation containing 5-ITZ-dU. **B**) Negative selection with carboxylic acid magnetic beads. **C**) Removal of negative target bound sequences. **D**) Streptavidin magnetic beads incubation with unbound sequences (positive selection). **E**) Removal of unbound sequences. **F**) Capture of streptavidin-ssDNA complexes. **G**) Elution of ssDNA, PCR amplification and single strand displacement for the next round of selection.

In total, 10 rounds of selection were performed to discover the enriched sequences for streptavidin. To increase the selection pressure, the quantity of magnetic beads and the concentration of ssDNA were reduced during the SELEX process. The selection conditions from rounds 1 to 9 are shown in Table 5.2.

Table 5.2. Streptavidin SELEX conditions from round 1 (R1) to round 9 (R9)

		R1	R2	R3	R4	R5	R6	R7	R8	R9
ssDNA	Concentration	1 μM	0.5 μM	0.6 μM	0.5 μM	0.5 μM	0.5 μM	0.3 μM	0.3 μM	0.3 μM
	Volume	500 μL	500 μL	500 μL	300 μL	200 μL	300 μL	500 μL	500 μL	500 μL
	Number of molecules	3×10^{14}	1.5×10^{14}	1.8×10^{14}	9×10^{13}	6×10^{13}	9×10^{13}	9×10^{13}	9×10^{13}	9×10^{13}
Targets	Positive (Streptavidin)	100 μL (1 mg)	50 μL (0.5 mg)	100 μL (1 mg)	50 μL (0.5 mg)	100 μL (1 mg)	100 μL (1 mg)	100 μL (1 mg)	100 μL (1 mg)	100 μL (1 mg)
	Negative (Dynabeads)	100 μL (1 mg)	50 μL (0.5 mg)	100 μL (1 mg)	50 μL (0.5 mg)	50 μL (0.5 mg)	50 μL (0.5 mg)	100 μL (1 mg)	100 μL (1 mg)	100 μL (1 mg)

After 9 rounds of selection when enrichment of aptamers appeared strong (see section 5.4.1), the amplified ssDNA was divided into three groups and selected under different conditions (Table 5.3).

In round 10A, DNA libraries were incubated with streptavidin and biotin for 1 h in order to select for ssDNA that binds to non-biotin binding sites or that binds to streptavidin more specifically than biotin. In round 10B, ssDNA-streptavidin complexes were washed with 1 × PBS containing 20% acetonitrile to reduce the number of sequences that bind via hydrophobic interactions and to select ssDNA that binds to its target via halogen bonding. In round 10C, the concentration of ssDNA and the quantity of target were greatly decreased (3/10 dilution for ssDNA and 1/5 dilution for the target) to discover ssDNA that binds to its target with higher binding affinity and specificity.

Table 5.3. Streptavidin SELEX conditions for round 10 (R10)

		R10A	R10B	R10C
ssDNA	Concentration	0.2 µM	0.2 µM	60 nM
	Volume	500 µL	500 µL	500 µL
	Number of molecules	6×10^{13}	6×10^{13}	1.8×10^{13}
Targets	Negative (Dynabeads)	100 µL (1 mg) + 0.5 µL Biotin	100 µL (1 mg)	20 µL (200 µg)
	Positive (Streptavidin)	100 µL (1 mg) + 0.5 µL Biotin	100 µL (1 mg)	20 µL (200 µg)

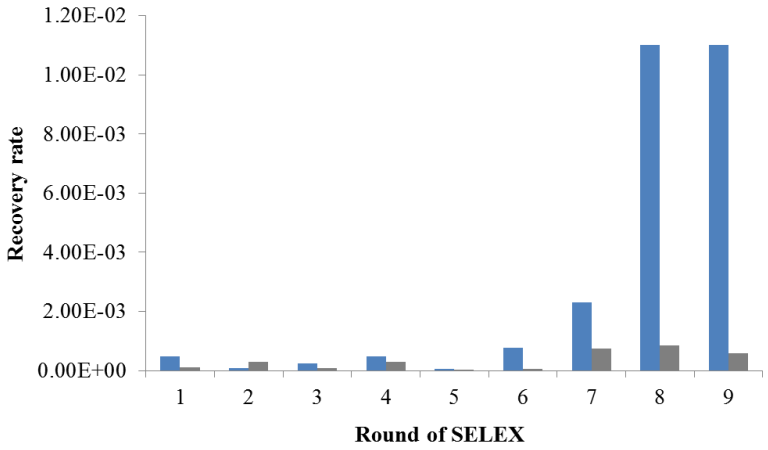
5.4.1 Q-PCR quantification

During SELEX, enrichment of target-specific aptamers was monitored by Q-PCR as described in Chapter 4 (section 4.5.2). The recovery rates from rounds 1 to 9 were measured and the recovery rates between the positive (streptavidin beads) and negative (magnetic beads) targets were compared. A table showing quantification of the data for each round can be found in the appendix (see appendix 8.12) with Figure 5.10 **A** showing a visual representation of the data for rounds 1 to 9. Sequences specific to streptavidin were enriched significantly from round 6 and reached a maximum at rounds 8 and 9 (blue bars). There was always some non-specific binding towards the negative targets (grey bars), but the recovery rate of these sequences relative to the positive target (blue bars) is negligible at later rounds suggesting SELEX was successful (Figure 5.10 **A**).

However, the Q-PCR experiments to monitor the selection progress were performed using Q5 DNA polymerase. This enzyme has not been yet validated for reading DNA containing base modifications. Q-PCR experiments were therefore repeated at the end of SELEX using both KOD XL DNA polymerase, a suitable enzyme for modified DNA amplification, and as control Q5 DNA polymerase. The PCR cycle in which

fluorescence can be detected is termed the quantitation cycle (Cq). This was used to compare the two polymerases. The Cq value for each round was then compared relative to the value for round 6 for each polymerase (Figure 5.10 **B**). No big differences were observed from rounds 1 to 5 when comparing the two DNA polymerases. However, for rounds 8 and 9 bigger differences in the Cq values were observed. Lower Cq values mean higher initial copy numbers of the target. For Q5 DNA polymerase, lower Cq was observed compared to KOD XL polymerase suggesting there were more sequences containing the modification in rounds 8 and 9. Once more this suggests SELEX was successful.

A



B

Round	KOD XL Cq value	ΔCq (Round 6 – Round x)	ΔCq (Round 6 – Round x)	Q5 Cq value
1	9.365	3.69	3.09	10.61
2	13.5	-0.445	-0.585	14.285
3	11.93	1.125	1.52	12.18
4				
5	15.515	-2.46	-2.355	16.055
6	13.055	0	0	13.7
7	11.825	1.23	1.52	12.18
8	9.565	3.49	2.68	11.02
9	9.35	3.705	2.715	10.985

Figure 5.10. **A**) A bar graph of the recovery rate of ssDNA bound streptavidin beads (positive target, blue) and magnetic beads (negative target, grey) from rounds 1 to 9 as determined by Q-PCR using Q5 DNA polymerase. **B**) A table of Cq value comparison when KOD XL or Q5 DNA polymerase were used during Q-PCR analysis. ΔCq is Cq value of round 6 (yellow box) minus Cq value of round x.

5.5 Identification of aptamer candidates

Sequences recovered from SELEX rounds 0 (*i.e*. initial library), 5, 6, 7, 8, 9 and 10 were analysed by NGS in order to gain a better understanding of the enrichment process. Sequencing experiments and data analysis were performed on an Illumina MiSeq instrument by Dr Arun Shivalingam.

5.5.1 Base composition

Firstly, the base composition of the random region was analysed in order to detect any bias in the original library and changes in composition as SELEX proceeds. For each round, the total number of dA, dC, dG or 5-ITZ-dU bases in the random region (40 mer) for all sequences was counted. This was then divided by the number of sequences observed and then divided by the mean length of the random region (40 mer) (Equation 5.1). This gives the average base frequency at any position of the random region for each round.

$$\text{Base frequency} = \frac{\left(\dfrac{\text{Total number of each base in the random region}}{\text{Number of sequences observed}}\right)}{\text{Mean length of the random region (40 mer)}}$$

Equation 5.1. An equation to calculate the average base frequency at any position of the random region (40 mer) for rounds 0 and 5 to 9.

For the initial library (*i.e*. round 0), a roughly equal base composition was observed with more enrichment of C and a slight reduction in G. This is possibly due to errors of monomer mixing during solid phase synthesis, slightly better coupling of some phosphoramidites than others, and/or enzymatic bias when generating the initial modified library by linear copying (see section 5.2). From rounds 5 to 10, the amount of 5-ITZ-dU did not show a big change (Figure 5.11). This confirmed that there was no

significant selection pressure against 5-ITZ-dU during SELEX.

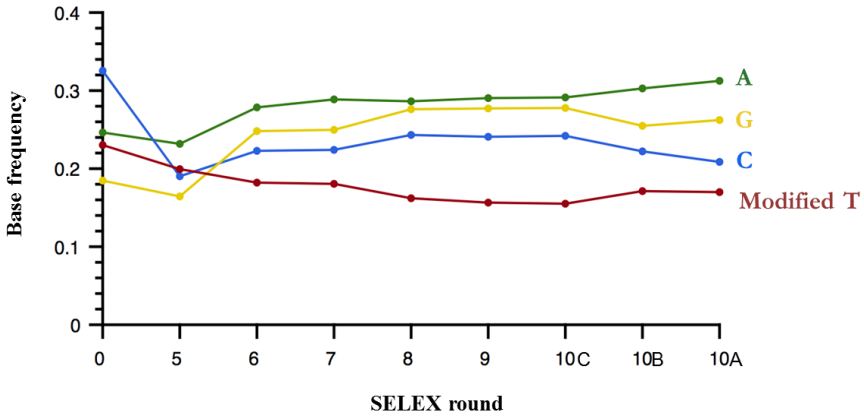


Figure 5.11. Base composition of the random region of the ssDNA library per SELEX rounds. dA (green), dG (yellow), dC (blue) and 5-ITZ-dU (red).

5.5.2 Unique sequence enrichment

Unique sequence enrichment is the most commonly used method for analysing aptamer sequencing data. This involves identifying the unique sequences for a specific SELEX round and counting how many times it is observed. This frequency can then be compared to other rounds of SELEX that have the same sequences. To visualise these results, the 12 most enriched sequences (y-axis, sequences in Table 5.4) were plotted against SELEX round (x-axis) with the change in frequency of the sequence represented as a heat-map (Figure 5.12 **A**). The first thing to note is that enriched sequences were mainly observed in rounds 8 and 9, which is consistent with the Q-PCR results (Figure 5.10 **A**). Next, only sequence U9 and U12 are conserved over multiple continuous rounds (9 and 10C for U9; 8 and 9 for U12), potentially suggesting SELEX was not successful. However, if the sequences in Table 5.4 were considered, it is clear that all enriched sequences are minor variations of each other. The base consensus sequences in Figure 5.12 **B** clearly shows that aptamer sequences are G-rich and potential G-quadruplexes. Importantly, conserved critical T positions at 23 and 35 were observed (Figure 5.12 **B**). This is interesting because modified aptamer selection for streptavidin has not identified G-quadruplex motifs before suggesting the 5-ITZ-dU modification may affect binding.[434]

Table 5.4. 12 most enriched sequences by unique sequence enrichment analysis.

Number	Sequences (5′-3′)
U1	AAACGTAAGGGGCTTGACGCGTCCGGCGCAGGACTCAAAA
U2	GCGGAAAGGCAGGGAAAGGGGATGGGGCGGGACATAGG
U3	GCGGAAAGGCAGGGAAGGGGGATGGGGCGGGACATTAGGG
U4	GCGGGAAGGCAGGGAAGGGAAATGGGGCGGGACATTTGGG
U5	GCGGGAAGGCAGGGAAGGGGAATGGGGCGGGACATTGGG
U6	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATATGAA
U7	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATATGGG
U8	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATCTAAA
U9	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATCTGGG
U10	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTAAA
U11	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTGGG
U12	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTTGGG

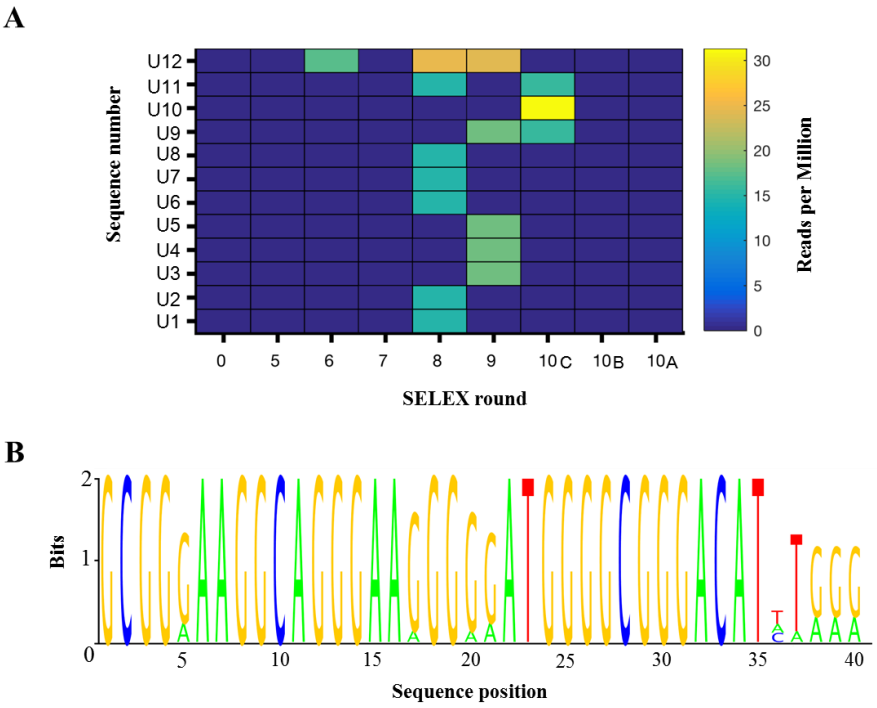


Figure 5.12. NGS analysis using the unique sequence enrichment method. **A**) A colour-coded heat-map of frequency each sequence (y-axis, sequences in Table 5.4) against SELEX round (x-axis). **B**) Base consensus sequences in the 40 mer random regions.

5.5.3 G4 enrichment

Based on the unique sequence enrichment method, sequences that fit the canonical G-quadruplex motif (Figure 5.13 **A**) were counted per round and plotted (Figure 5.13 **B**). There was a clear enrichment of this motif during SELEX. The initial library (*i.e*. round 0) contained only five G4 forming sequences per 1 million sequenced ssDNA oligonucleotides, but this increased to approximately 29,000 per 1 million sequenced ssDNAs by round 9 (Figure 5.13 **B**). This is noteworthy as G-quadruplex sequences are usually poorly amplified by PCR.[464,465]

The selection stringency was increased in 10A–C and the effects on G-quadruplex frequency are also shown in Figure 5.13 **B**. The amount of G-quadruplexes was maintained after dilution of target and library in round 10C, suggesting the enriched G-quadruplexes bound to the target with high affinity. When the aptamer-streptavidin complexes were washed with 20% acetonitrile in PBS buffer (round 10B), a significant decrease in G4 sequences reads was observed suggesting they bind through hydrophobic interactions. Finally, incubating the sequences with the streptavidin and biotin (round 10A) caused a decrease in reads suggesting some of these G-quadruplexes compete with biotin or that biotin inhibits the sequences from binding to the streptavidin. Finally, the frequency of i-motifs (Figure 5.13 **B**), C-rich sequences that might be enriched if G-rich duplex DNA is carried over into other rounds, was significantly diminished by rounds 5 to 9.

In this analysis, the consensus G-quadruplex sequence (Figure 5.13 **C**) was G1 (Table 5.5). This sequence is almost identical to U12 (Table 5.4) except the terminal dG at the 3′-end is not highly conserved and was therefore omitted.

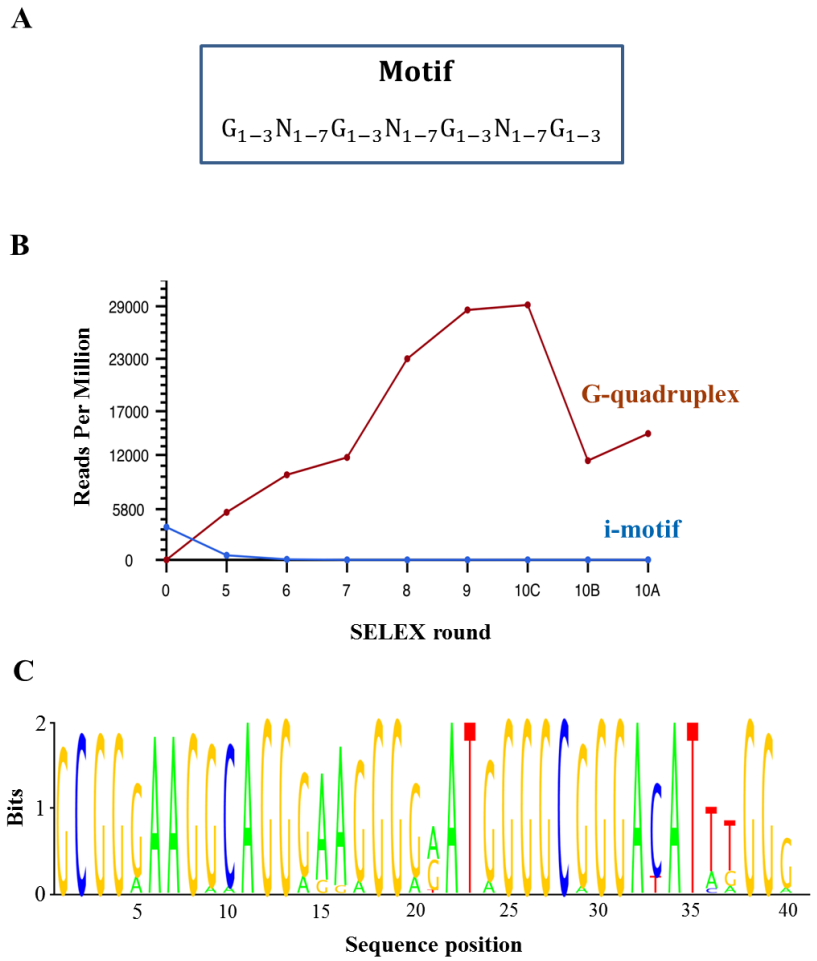


Figure 5.13. NGS analysis by a G4 enrichment method. **A**) The generic G-quadruplex folding motif. **B**) Frequencies of G-quadruplex (red line) and i-motif (blue line) observed in rounds 0 and 5 to 10. **C**) Base consensus sequences in the 40 mer random regions.

Table 5.5. The consensus sequence from G4 enrichment analysis.

Number	Sequences (5′-3′)
G1	GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTTGG

5.5.4 NMer cluster

In order to discover non-G-rich sequences, the random regions were split into smaller sequences (termed 'NMers') and clustered to find novel motifs. 8 of the most abundant sequences are shown in Table 5.6 and their change in enrichment over rounds is shown in Figure 5.14. All 'NMer' contained 5-ITZ-dU (Table 5.6) suggesting the modification plays a role in binding. Among these sequences, N1 has the highest enrichment in round 9 and 10C. N3 is also interesting as it has three 5-ITZ-dU (Table 5.6), and the spacing between the modification is similar to G-quadruplex sequence G1 (Table 5.5) suggesting it is a possible binding motif.

Table 5.6. 8 most abundant sequences selected by NMer cluster analysis.

Number	Sequences (5′-3′)
N1	CCACGCGGCCGGCGCAGGCCTGGG
N2	GACGCGTCCGGCGCAGGACTC
N3	ACACGCACCGGTCGCAGGTTGGC
N4	CCGAAGCGCACCGGTCGCAGGTCA
N5	CACCGGTCGCAGGT
N6	GAAGGCAGGGAAGGGGGATGGGGC
N7	CACGCACCGATCGCAGGTTGGCC
N8	CGCGTCCGGCGC

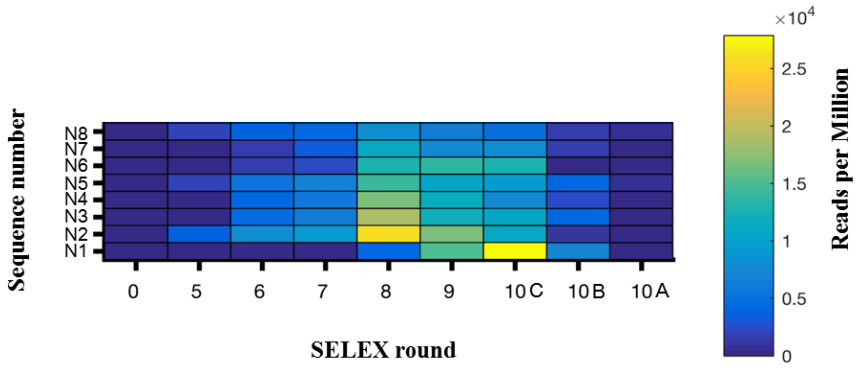


Figure 5.14. NGS analysis NMer cluster methods. A colour-coded heat-map of frequency each sequence (y-axis, sequences in Table 5.6) against SELEX round (x-axis).

5.6 Validation of aptamer candidates

The potential aptamers were selected from 40 mer random libraries flanked by two 20 mer primer sites. However, not every nucleotide is involved in target binding. The advantages of removing unnecessary nucleotides are that it makes aptamer size smaller, potentially making synthesis and structural studies easier. It can also improve binding affinities, but this varies depending on the sequences.[466–468] Hence, the potential G-quadruplex sequences were prepared with and without primer sites.

AG1-ITZ (Table 5.7) and its primer-flanked variant (AG1L-ITZ, Table 5.7) were chosen as this sequence was the consensus G-quadruplex motif. Therefore, it was anticipated to have a high binding affinity towards streptavidin. AU4-ITZ (Table 5.7) was also chosen as the guanine at position 21 (and to a lesser extent position 20) does not appear to be strongly conserved (Figure 5.13 **C**).

In addition, two sequences from NMer cluster analysis (AN1-ITZ and AN3-ITZ, Table 5.7) were also prepared for binding characterisation. AN1-ITZ was the most enriched sequence by NMer cluster analysis and consists of 24 mer with only one 5-ITZ-dU. AN3-ITZ is a less enriched sequence of similar length (23 mer), but it contains three 5-ITZ-dUs and therefore potentially has a higher chance of displaying halogen bonding.

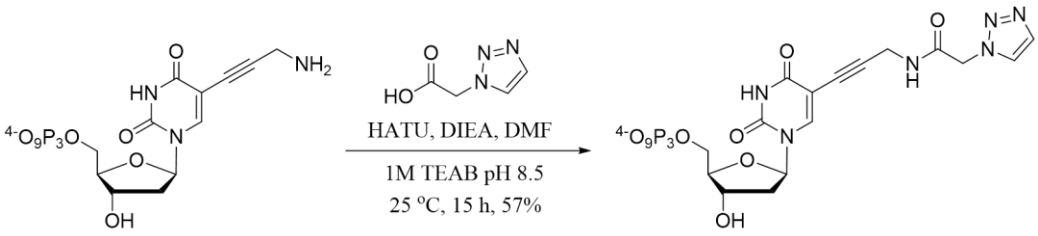
Table 5.7. Potential aptamer sequences selected after NGS analysis.

	Sequence (5′- 3′)
AG1-ITZ	**F**GCGGGAAGGCAGGGAAGGGGGA**T**GGGGCGGGACA**TTT**GG
AU4-ITZ	**F**GCGGGAAGGCAGGGAAGGGAAA**T**GGGGCGGGACA**TTT**GGG
AG1L-ITZ	**F**AGCAGCACAGAGGTCAGATGGCGGGAAGGCAGGGAAGGGGG A**T**GGGGCGGGACA**TTT**GGCC**T**A**T**GCG**T**GC**T**ACCG**T**GAA
AN1-ITZ	CCACGCGGCCGGCGCAGGCC**T**GGG
AN3-ITZ	ACACGCACCGG**T**CGCAGG**TT**GGC

F: 5'-6-FAM (Fluorescein); **T**:5-ITZ-dUTP;

* Primer binding sites are underlined

As controls, aptamer sequences were also produced with dTTP and 5-propargylamino-prototriazole-dUTP **18** (5-PTZ-dU) in place of 5-ITZ-dU. This compound was synthesised using 5-PA-dUTP **6** as a precursor. Commercially available 1*H*-1,2,3-triazole-1-acetic acid was labelled to the primary amine of 5-PA-dUTP via HATU coupling reaction (Scheme 5.1). It was expected that comparison of 5-PTZ-dU and 5-ITZ-dU or dTTP and 5-ITZ-dU modified aptamer would clear differences affinity of the aptamer, providing evidence for halogen bonding with the target.



6 **18**

Scheme 5.1. Synthesis of 5-propargylamino-prototriazole-dUTP (5-PTZ-dUTP).

Finally, further positive controls – streptavidin DNA aptamers found in the literature that share a conserved bulge-hairpin secondary structure[434,452] – were prepared (Table 5.8).

Table 5.8. Published streptavidin DNA aptamer sequences.[434,452]

	Sequence (5′- 3′)
AC1	ATTGACCGCTGTGTGACGCAACACTCAAT
AC2	GGGAACGCACCGATCGCAGGTTTCCC

The aptamer sequences were enzymatically and chemically synthesised and their binding affinities and kinetics were determined by surface plasmon resonance (SPR) and electrophoretic mobility shift assay (EMSA) as described in the following sections.

5.6.1 Synthesis of aptamer candidates

The selected aptamers were prepared using various chemical and enzymatic methods to discover the most efficient synthetic route. The first method relies on enzymatic synthesis (linear copying using the modified triphosphate and denaturing PAGE purification), the second uses solid phase synthesis using the modifications as phosphoramidite monomers, and the third involves post-synthetic labelling of oligonucleotides containing propargylamino sites with a carboxylic acid derivative of iodotriazole.

5.6.1.1 Enzymatic synthesis

During SELEX, modified ssDNA was prepared using KOD XL DNA polymerase and the modified 5-ITZ-dUTP. However, a large amount of enzymes is required to produce larger quantities of the aptamers and KOD XL is more expensive than most polymerases. Therefore, the use of Vent exo- DNA polymerase was re-evaluated. In Chapter 3, Vent exo- DNA polymerases failed to incorporate the halogen substituted triphosphate efficiently, when the reaction was supplemented with 1 mM $MnCl_2$. In the absence of $MnCl_2$, Vent exo- DNA polymerase works in a similar manner to KOD XL DNA polymerase although longer incubation times were required (see Chapter 3, Figure 3.20 and Figure 3.21). Therefore, aptamer candidates in this chapter were synthesised using Vent exo- DNA polymerase without additional $MnCl_2$. The desired products were successfully synthesised as discussed below.

Table 5.9. Primers (P) and templates (T) used in enzymatic synthesis of G-quadruplex aptamer sequences. T9 is for AG1-ITZ and T10 is for AU4-ITZ.

	Sequence (5′- 3′)
P9	**F**GCGGGAAGGCAGGGAAGG
T9	CCAAATGTCCCGCCCCATCCCCCTTCCCTGCCTTCCCGC
T10	CCCAAATGTCCCGCCCCATTTCCCTTCCCTGCCTTCCCGC
F: 5'-6-FAM (Fluorescein)	

AG1-ITZ was synthesised by primer extension reaction using primer P9 and template T9 (Table 5.9). In the reaction mixture, 5-ITZ-dUTP was added instead of dTTP. The primer P9 was labelled with 5′-FAM to visualise the fully extended complementary sequence, which migrated slower on the gel than the template due to its high MW of modifications (Figure 5.15 **A**). However, the separation between the template and the

product was not good as seen by the band streaking (Figure 5.15 **A**). When the excised product bands are analysed by UPLC-MS, both template and full-length product peaks were detected indicating that the two strands were not fully dissociated during denaturing PAGE. This was probably due to the increased melting temperature of the GC-rich dsDNA product, with the T_m increased further by the 5-ITZ-dU modification as discussed in Chapter 3 (section 3.6, Figure 3.27). It is also possible that the G-quadruplex is not fully denatured by denaturing PAGE. In order to separate non-denatured strands, further purification was performed by RP-HPLC. Various running buffers such as HAA (75–90% buffer B, 20 min), ammonium acetate (20–35% buffer B, 20 min) or TEAB buffer (20–35% buffer B, 20 min) were used and the RP-HPLC column heated to 55 ℃ to improve dsDNA denaturation. However, it was insufficient. The HPLC trace and deconvoluted mass spectrum still showed a smaller peak corresponding to the template and a larger peak for the product (Figure 5.15 **B** and **C**).

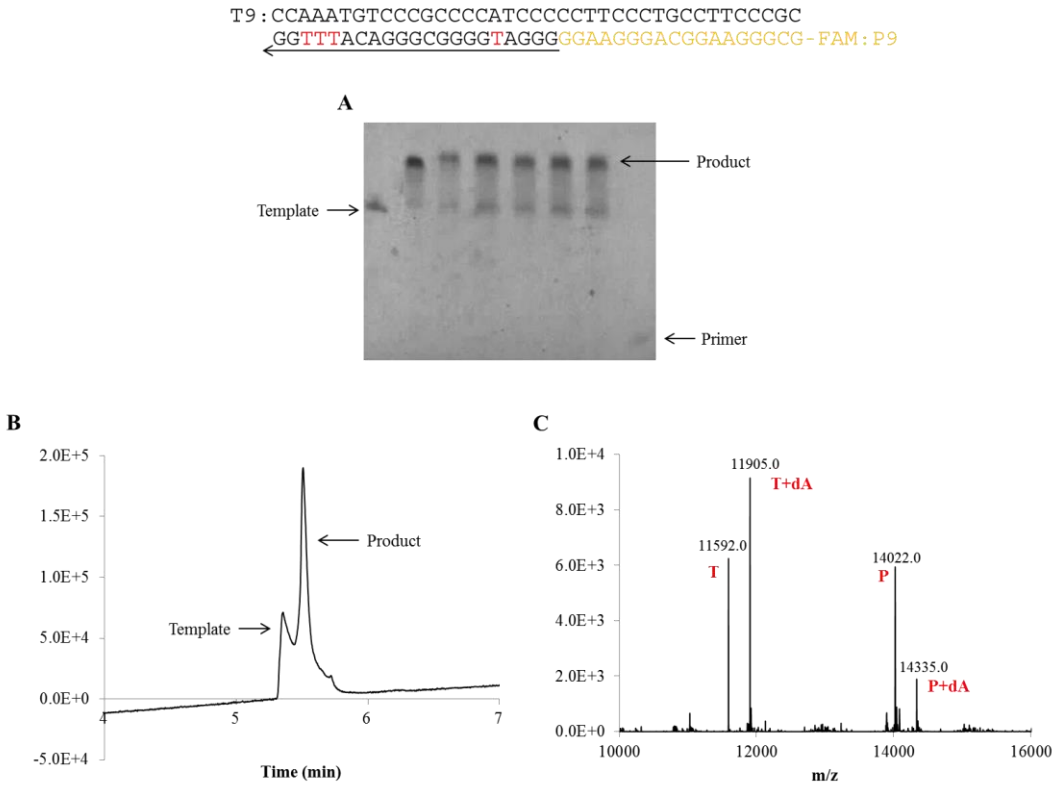


Figure 5.15. Primer extension reaction was performed using 2 U Vent exo- DNA polymerase (65 ℃, 2 h), 5 nmol of primer P9 and template T9 and 250 nmol of dATP, dGTP, dCTP and 5-ITZ-dUTP. The FAM-labelled primer is highlighted in yellow and the 5-ITZ-dUTP incorporation site are coloured in red. **A**) Size separation of two strands on 12% denaturing PAGE gel. **B**) HPLC trace of the product after gel extraction purification and further purification by RP-HPLC. **C**) The final mass result after gel extraction and further purification by RP-HPLC. Template (**T**) strand calc. 11592.0 and found. 11592.0 and 11905.0 (**+dA**). Product (**P**) strand calc. 14022.0 and found.14022.0 and 14335.0 (**+dA**).

On the other hand, AU4-ITZ was successfully synthesised by the primer extension, gel extraction (Figure 5.16 **A**) and RP-HPLC was used to remove the residual template (Figure 5.16 **B** and **C**) but the recovery was very low (Theoretical yield: 5.0 nmol and the actual yield after purification: 0.9 nmol). In general, the enzymatic synthesis of the aptamer sequences involved multiple purification steps and therefore the overall process was tedious and low yielding.

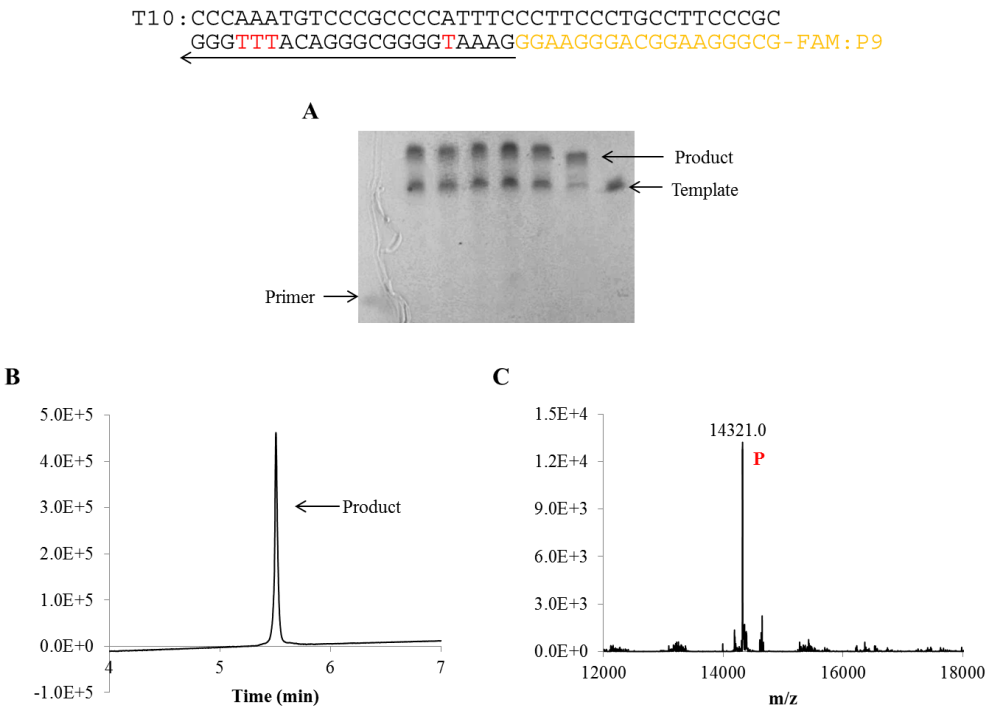


Figure 5.16. Primer extension reaction was performed using 2 U Vent exo- DNA polymerase (65 ℃, 2 h), 5 nmol of primer P9 and template T10 and 250 nmol of dATP, dGTP, dCTP and 5-ITZ-dUTP. The FAM-labelled primer is highlighted in yellow and the 5-ITZ-dUTP incorporation site are coloured in red. **A**) Size separation of two strands on 12 % denaturing PAGE gel. **B**) HPLC trace of the product after gel extraction purification and further purification by RP-HPLC. **C**) The final mass result after gel extraction and further purification by RP-HPLC. Product (**P**) strand calc. 14318.7 and found.14321.0.

5.6.1.2 Denaturing PAGE using a basic buffer system

In order to solve the dsDNA product denaturation problem, more basic conditions were tested in denaturing PAGE. AG1-ITZ sequences were synthesised with forward and reverse primers to produce AG1L-ITZ (Table 5.7). AG1L-ITZ is G-quadruplex sequences consists of 79 nucleotides with 9 positions of 5-ITZ-dU sites. It was enzymatically synthesised using Vent exo- DNA polymerase and dTTP or 5-PTZ-dU in place of 5-ITZ-dU (AG1L-T or AG1L-PTZ, respectively), followed by denaturing PAGE. Template T11 and primer P6 (Table 5.10) were used.

Table 5.10. AG1 sequences with full-length primers regions and the primer (P) and template (T) used to synthesise AG1L containing modifications (ITZ for 5-ITZ-dU, PTZ for 5-PTZ-dU and T for dTTP).

	Sequence (5′- 3′)
P6	**F**AGCAGCACAGAGGTCAGATG
T11	TTCACGGTAGCACGCATAGGCCAAATGTCCCGCCCCATC CCCCTTCCCTGCCTTCCCGCCATCTGACCTCTGTGCTGCT

F: 5'-6-FAM (Fluorescein)

Next, 10 mM NaOH was added to the linear copying mixture and heated at 95 ℃ for 15 min before purification on a denaturing gel running with 1× Tris-Borate-EDTA (TBE, pH 8.3) buffer. It was hoped this would provide more effective denaturation of the template and the product strands. As shown in Figure 5.17 **A**, however, there was no separation between the template and product due to the high melting temperature of long GC-rich sequences bearing 9 of 5-ITZ-dUs. Therefore, a more basic buffer system – Tris-CAPS (TC, pH9.6) buffer was used to help denature the dsDNA.[469] The primer extension reaction products were pre-denatured using alkaline, followed by heat treatment as described above and the gel was prepared using 1× TC buffer. All other conditions (except the gel casting buffer and running buffer) were kept the same and the separation between the template and products was far better as shown in Figure 5.17 **B** (Theoretical yield: 30.0 nmol and actual yield after TC-PAGE: 15.7 nmol). Further purification by RP-HPLC was not required and the desired products were verified by UPLC-MS (Table 5.11).

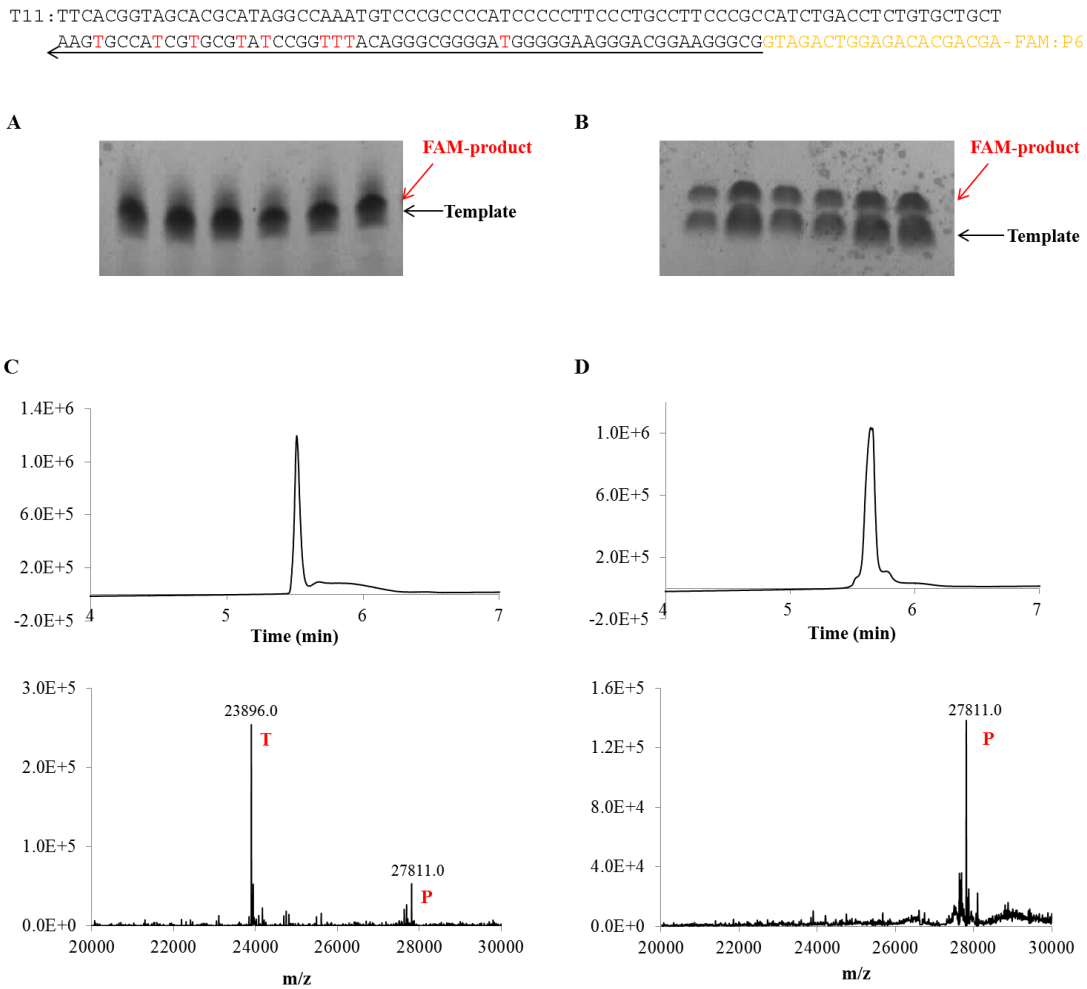


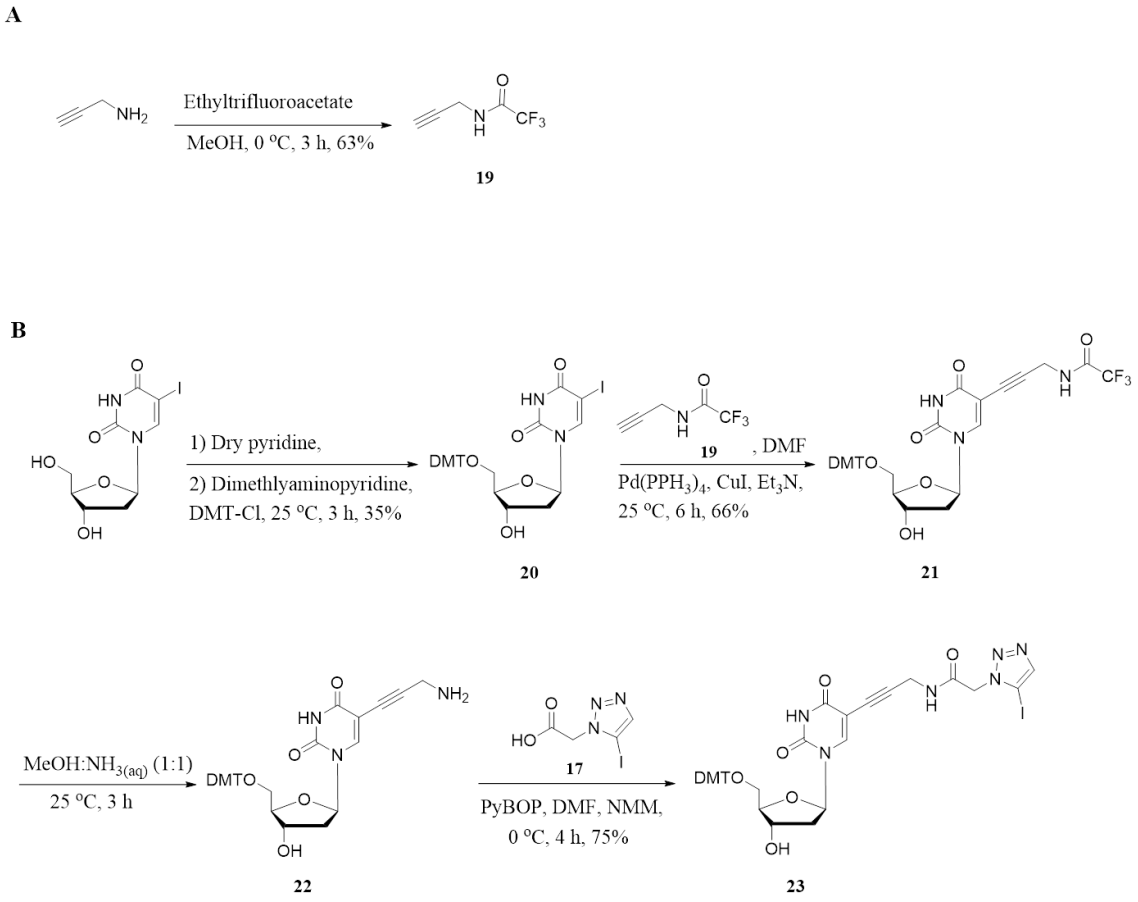
Figure 5.17. 8% denaturing PAGE gel after primer extension using template T11 and primer P6. The reaction performed at 65 ℃ for 2 h using Vent exo- DNA polymerase with dATP, dGTP, dCTP and 5-ITZ-dUTP. **A**) Gel electrophoresis using 1× Tris-Borate-EDTA buffer. **B**) Gel electrophoresis using 1× Tris-CAPS buffer. **C**) HPLC trace and mass of the gel extracted products from A. Template (**T**) strand calc. 23893.4 and found. 23896.0. Product (**P**) strand calc. 27809.5 and found. 27811.0 **D**) HPLC trace and mass of the gel extracted product from B. Product (**P**) strand calc. 27809.5 and found. 27811.0.

Table 5.11. Mass spec results of AG1L containing modifications. Representative mass spectra in appendix 8.13.

Modification	AG1L	
	Calc. (Da)	**Found (Da)**
dTTP (T)	25343.5	25347.0
5-PTZ-dUTP (PTZ)	26676.3	26673.0
5-ITZ-dUTP (ITZ)	27809.5	27811.0

5.6.1.3 Solid phase synthesis

Due to low yields and laborious purification required for enzymatically synthesised aptamers, direct incorporation of the modification during solid phase synthesis was considered. In order to do this, the 5-ITZ-dU phosphoramidite was required. The starting material was 5-iodo-2′-deoxyuridine (Scheme 5.2). The 5′-OH of the nucleoside was protected with a dimethoxytrityl group **20** (35% yield). Next, propargylamine was protected using ethyl trifluoroacetate to give **19** in 63% yield. TFA-aminopropyne **19** was then coupled to the C5 of the 5′-DMT-5-iodo-2′-deoxyuridine **20** via the Sonogashira reaction (66% yield) to produce 5′-DMT-5-propargyl-TFA-2′-deoxyuridine **21**. The TFA protecting group of compound **21** was then deprotected in methylamine/ethanol (33% v/v). However, by-products were produced when monitoring the reaction by TLC (by-product was not the starting material). Full deprotection of compound **22** was achieved using concentrated aqueous ammonia in methanol (50% v/v). Finally, the 5-iodo-1*H*-1,2,3-triazole-1-acetic acid **17** (ITZ-acetic acid, section 3.2.2) was coupled to the propargylamine via an amide coupling reaction using PyBOP to produce DMT-5-ITZ-dU **23** (75% yield).



Scheme 5.2. Synthesis of TFA-aminopropyne **19** (**A**) and DMT-5-ITZ-dU **23** (**B**).

This monomer **23** was then converted into the phosphoramidite (Figure 5.18) by Dr Asha Brown and Dr Daniel Singleton, before solid-phase synthesis to generate AN1-ITZ and AN3-ITZ (Table 5.7) using an extended coupling time for the 5-ITZ-dU phosphoramidite. In addition, acetyl dC and DMF dG phosphoramidites were used for fast deprotection of the oligonucleotides. This is to allow for the poor stability of iodotriazole in ammonia solution. Solid-phase synthesis was an effective method for the introduction of three iodotriazole modifications. Other aptamer sequences (AG1-ITZ, AU4-ITZ and AG1L-ITZ) have four or nine modification sites (Table 5.7) and controls containing 5-PTZ-dU (instead of 5-ITZ-dU) were also required. The direct synthesis of both monomers in sufficient quantities for all these sequences was impractical.

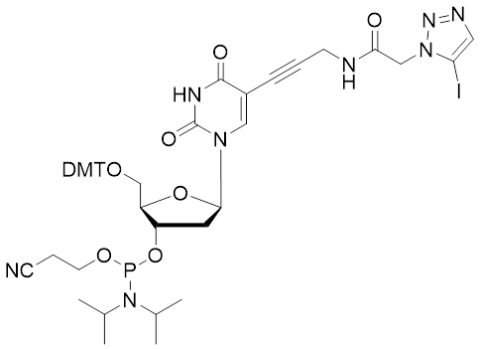


Figure 5.18. DMT-5-ITZ-dU phosphoramidite synthesised by Dr Asha Brown and Dr Daniel Singleton.

5.6.1.4 Post-synthesis conjugation

The final approach considered was post-synthetic labelling of the oligonucleotide to introduce the iodotriazole modification. There are a number of methods commonly used for post-synthetic oligonucleotide functionalisation including amine coupling or copper-catalysed azide-alkyne cycloaddition.[470,471] Since our modified triphosphate has propargylamine as a linker and the reaction worked well during phosphoramidite synthesis, the amide coupling reaction was used directly on the oligonucleotide. The 5-PA-dU phosphoramidite was already available in the group in large quantities and kindly provided by Sarah Walsh and ATD bio. The oligonucleotide containing the 5-PA-dU at the sites of interest was synthesised by solid-phase chemistry (Table 5.12). Oligonucleotide S-G1 (Table 5.12) is the same as AG1-ITZ (Table 5.7) but includes few additional nucleotides from its primer sites at the 3′- and 5′-ends. Importantly, S-G1 contains only one 5-PA-dU site, which allows the efficiency of labelling using 5-iodo-1*H*-1,2,3-triazole-1-acetic acid (ITZ-acetic acid) to be easily evaluated.

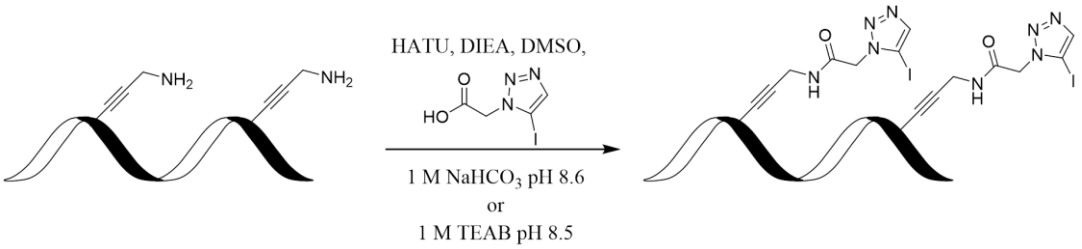
Table 5.12. Templates (S or T) used to synthesis aptamer candidates via post-synthesis conjugation method.

	Sequence (5′- 3′)
S-G1	**F**AGATGGCGGGAAGGCAGGGAAGGGGGA**X**GGGGCGGGACATTTGGGCCTA
T-G1	**F**GCGGGAAGGCAGGGAAGGGGGA**X**GGGGCGGGACA**XXX**GG
T-U4	**F**GCGGGAAGGCAGGGAAGGGAAA**X**GGGGCGGGACA**XXX**GGG
T-N1	CCACGCGGCCGGCGCAGGCC**X**GGG
T-N3	ACACGCACCGG**X**CGCAGG**XX**GGC
T-G1b	**F**GCGGGAAGGCAGGGAAGGGGGA**X**GGGGCGGGACA**XXX**GG**Y**
T-U4b	**F**GCGGGAAGGCAGGGAAGGGAAA**X**GGGGCGGGACA**XXX**GGG**Y**

F: 5'-6-FAM (Fluorescein); **X**: 5-PA-dUTP; **Y**: Propanol

* Primer binding sites are underlined

ITZ-acetic acid was synthesised as described in Chapter 3 (section 3.2.2) and coupled to the 5-PA-dU oligonucleotide amine via amide coupling using HATU (Scheme 5.3). HATU was successfully used in Chapter 3 to label different modification onto 5-PA-dUTP. Hence, the reaction was performed using 1:1 ratio of DMF:Buffer(1M $NaHCO_3$ or TEAB, pH 8.6) at 25 ℃ and 50 ℃ for 4 h. The reaction was also tried at 50 ℃ to accelerate the reaction rate. However, the reaction did not work as the oligonucleotide precipitated from the reaction. The ratio of DMF:Buffer was changed to 2:3 to increase the solubility of the oligonucleotide, but still the reaction did not go to the completion (Figure 5.19).



Scheme 5.3. Amide coupling reactions using HATU reagents. The oligonucleotide contains 5-PA-dUTP in place of dTTP and ITZ-acetic acid was labelled on the oligonucleotides.

S-G1:FAM-AGATGGCGGGAAGGCAGGGAAGGGGGA**X**GGGGCGGGACATTTGGGCCTA

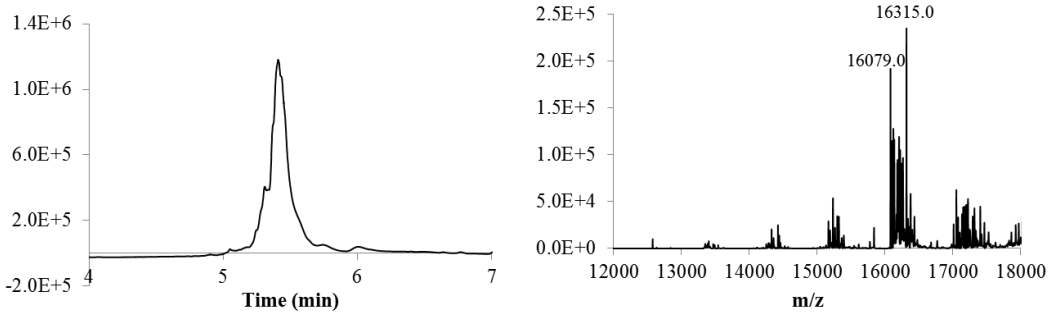
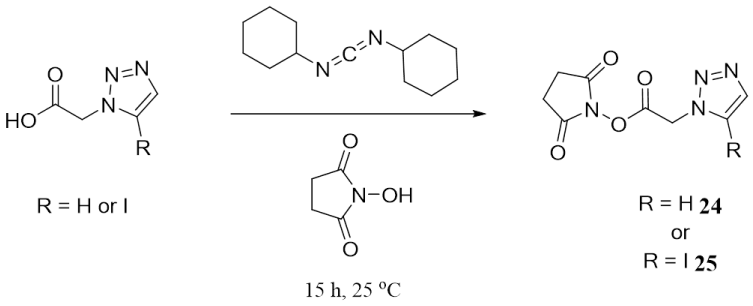


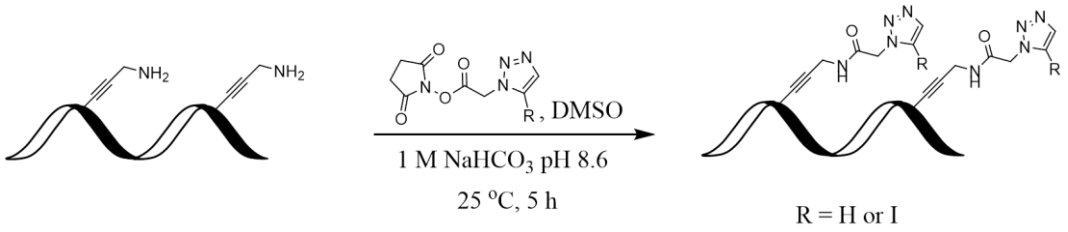
Figure 5.19. HPLC trace and mass spectrum after HATU coupling using template S-G1 and ITZ-acetic acid. The sequence written as **X** represents 5-PA-dU. Template strand calc. 16079.0 and found. 16079.0. Product strand calc. 16314.0 and found. 16315.0.

In order to push the reaction to completion, the ITZ-acetic acid and proto-triazole derivate were converted into the pre-activated NHS esters **24**/**25** via the addition of *N*-hydroxysuccinimide via *N*,*N*′-dicycloheylcarbodiimide (DCC) coupling, followed by the filtration of the excess reagents and dicyclohexylurea by-product (Scheme 5.4). Some small peaks corresponding to minor impurities were observed in the aliphatic region of the NMR spectrum. However, owing to its sensitivity towards hydrolysis, the NHS ester was used directly in the next step without attempting further purification. NHS esters are extremely useful for bioconjugate applications due to their good balance between reactivity and relative stability towards hydrolysis.[472]



Scheme 5.4. Synthesis of NHS ester of 5-proto- or 5-iodo-1*H*-1,2,3-triazole-1-acetic acid (PTZ- or ITZ-acetic acid). R = H for PTZ-acetic acid or I for ITZ-acetic acid 25.

Before labelling at multiple sites in the oligonucleotide, T-N1 (Table 5.12) which only contains one 5-PA-dU site was reacted with PTZ-NHS ester **24** (Scheme 5.5). The reaction was performed under basic conditions (1M $NaHCO_3$ pH 8.6) to keep the primary amine deprotonated and reactive. A large excess of the NHS ester was added to drive the reaction to completion because the active ester hydrolyses in aqueous solutions.



Scheme 5.5. Labelling of 5-iodo- or 5-proto-1*H*-1,2,3-triazoles (ITZ and PTZ, respectively) on the oligonucleotides containing propargylamine on dUTP via NHS ester coupling reaction.

After labelling, the products were purified using Sephadex illustra NAP columns (GE Healthcare) and analysed by UPLC-MS and the exact masses were determined (Figure 5.20 **A**). In addition, T-N3 (Table 5.12) which contains three 5-PA-dU units was reacted with PTZ-NHS **24** ester to produce AN3-PTZ via the procedure discussed above. After labelling, the products were purified by Sephadex illustra NAP columns and

characterised by UPLC-MS (Figure 5.20 **B**). The products were used as controls to validate the halogen bonding effects in the aptamer. The template T-U4b (Table 5.12) which contain four modifications were then reacted with ITZ-NHS ester **25** and the product was successfully produced (Figure 5.20 **C**). It was noteworthy that because Figure 5.20 **C** is a G4 sequence it is not completely denatured and therefore, gives multiple peaks in the HPLC trace. However, only one major peak was observed by mass spectrometry.

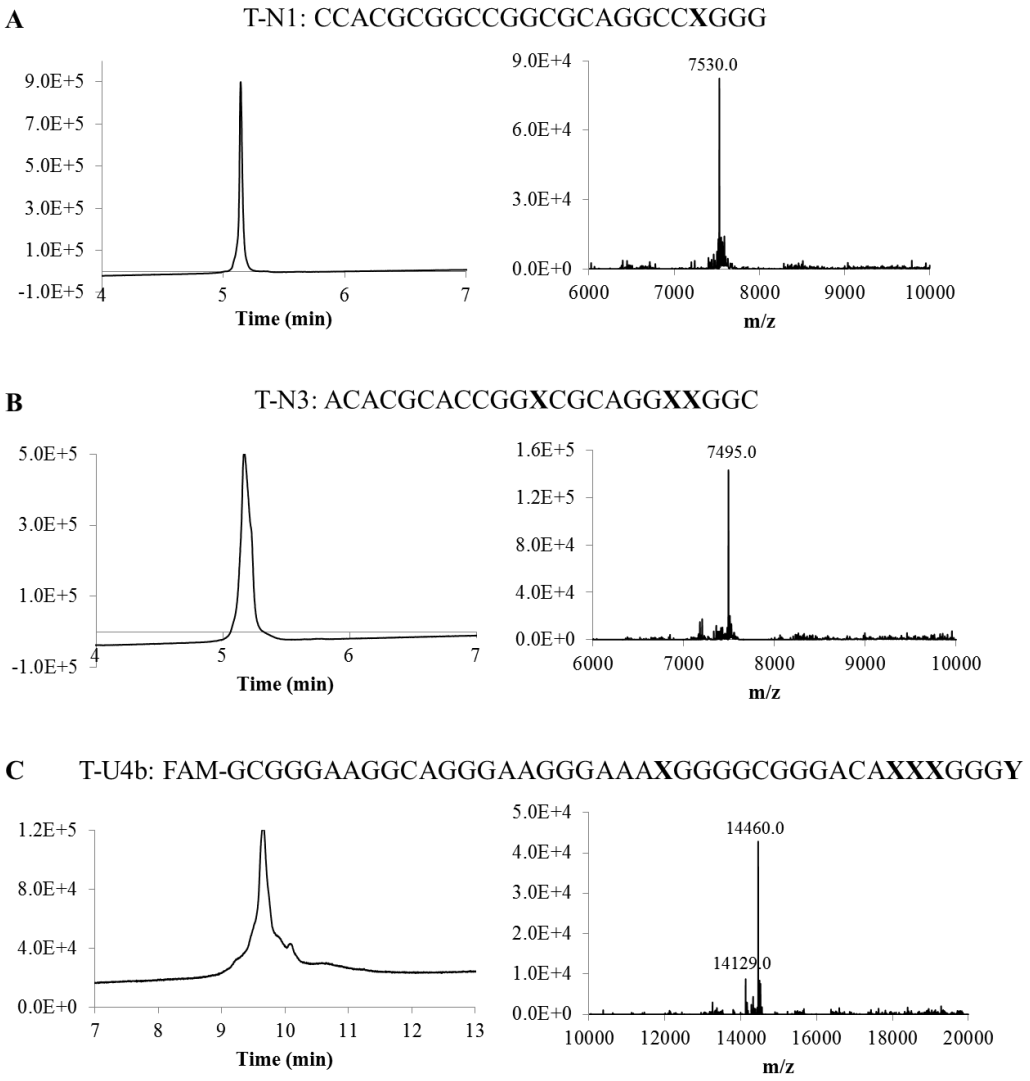


Figure 5.20. HPLC traces and mass spectra after NHS ester coupling reaction. The sequences written as **X** represent 5-PA-dU and **Y** is 3′-propanol. **A**) Labelling reaction using template T-N1 and PTZ-NHS ester. Product strand calc. 7529.9 and found. 7530.0. **B**) Labelling reaction using template T-N3 and PTZ-NHS ester. Product strand calc. 7494.9 and found. 7495.0. **C**) Labelling reaction using template T-U4b and ITZ-NHS ester. Product strand calc. 14456.0 and found. 14460.0.

The mass spec analysis of this G-quadruplex aptamer synthesis using PTZ- or ITZ-NHS ester coupling is shown in Table 5.13 and the mass spec analysis of fragmented aptamer synthesis are shown in Table 5.14.

Table 5.13. **A**) Mass spec results of G-quadruplex aptamer sequences with modifications (AG1 and AU4). **B**) Mass spec results of G-quadruplex aptamer sequences with modifications and 3′-propanol (AG1b and AU4b). P is product and Ac is acetylated product. Representative mass data in appendix 8.14 and 8.15.

A

Modification	AG1		AU4	
	Calc. (Da)	Found (Da)	Calc. (Da)	Found (Da)
dTTP (T)	12926.5	12925.0	13223.7	13223.0
5-PA-dU (PA)	13082.5	13082.0 (P) 13124.0 (P + Ac)	13379.7	13380.6 (P) 13421.0 (P + Ac)
5-PTZ-dU (PTZ)	13519.0	13519.0 (P)	13816.2	13815.5 (P) 13748.3 (P +Ac)
5-ITZ-dU (ITZ)	14022.6	14023.0 (P)	14319.8	14319.0 (P) 14180.8

B

Modification	AG1b		AU4b	
	Calc. (Da)	Found (Da)	Calc. (Da)	Found (Da)
5-PA-dU (PA)	13219.0	13219.0 (P) 13261.0 (P + Ac) 12889.0 (P - dG)	13516.0	13518.0 (P) 13561.0 (P + Ac) 13190.0 (P - dG)
5-PTZ-dU (PTZ)	13655.4	13655.1 (P) 13587.2 (P + Ac) 13325.0 (P - dG)	13952.4	13952.9 (P) 13887.7 (P + Ac) 13623.4 (P - dG)
5-ITZ-dU (ITZ)	14159.0	14159.1 (P) 13829.8 (P - dG)	14456.0	14456.5 (P) 14126.8 (P - dG)

Table 5.14. Mass spec results of AN1 and AN3 containing modifications. Representative mass spec data in appendix 8.16.

Modification	AN1		AN3	
	Calc. (Da)	Found (Da)	Calc. (Da)	Found (Da)
dTTP (T)	7381.8	7382.0	7050.6	7051.0
5-PA-dU (PA)	7420.8	7422.0	7167.6	7169.0
5-PTZ-dU (PTZ)	7529.9	7527.6	7494.9	7493.1
5-ITZ-dU (ITZ)	7655.8	7654.2	7872.6	7872.3

In summary, the post-synthetic labelling via NHS ester coupling reactions was fast and purification was straightforward. This method gave a far higher yield of the product than enzymatic synthesis and uses less of the prototriazole/iodotriazole than the phosphoramidite approach. It also avoids the necessity to make phosphoramidite monomers.

5.6.2 SPR binding analysis

SPR was used to determine the binding affinities of the screened candidate aptamers, AG1-ITZ and AU4-ITZ to streptavidin. Serial dilutions of the potential aptamers (including controls) were exposed to a surface immobilised streptavidin (5,000–6,000 RU). Initially selected aptamers without primer sites (natural dT or 5-PTZ-dU or 5-ITZ-dU variants) were tested (see appendix 8.17). As shown in Figure 5.21, the highest concentration injected for AG1-ITZ (1.76 µM) and AU4-ITZ (1.90 µM) did not show detectable binding to streptavidin.

AG1-ITZ: FAM-GCGGGAAGGCAGGGAAGGGGGA**T**GGGGCGGGACA**TTT**GG

AU4-ITZ: FAM-GCGGGAAGGCAGGGAAGGGAAA**T**GGGGCGGGACA**TTT**GGG

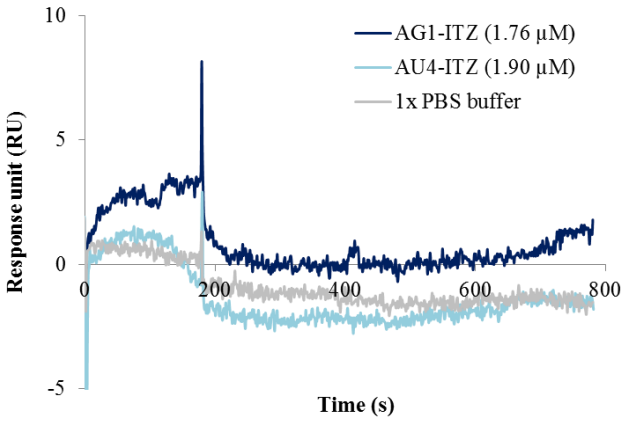


Figure 5.21. Affinity of AG1-ITZ and AU4-ITZ sequences to streptavidin by SPR. The positions of 5-ITZ-dU in the sequences are highlighted in red. The highest concentrations of AG1-ITZ (1.76 µM) and AU4-ITZ (1.90 µM) used are shown with the full titration is shown in appendix 8.17.

Concerned by the negative result, control unmodified streptavidin aptamers from the literature were tested to confirm that the experimental system works. Bing *et al*. reported a canonical DNA aptamer (29 mer, AC1, Table 5.8) which binds with an equilibrium dissociation constant (K_D) of 40 ± 18 nM by static adsorption.[434] In addition, Ruigrok *et al.* found a DNA aptamer (26 mer, AC2, Table 5.8) which binds with a K_D value of 35 ± 1 nM by SPR (1:1 binding model kinetic data).[452]

AC1 was selected in PBS buffer (pH 7.5, 0.02% Tween 20),[434] and AC2 was selected in Tris buffer (20 mM Tris-HCL pH 7.6, 100 mM NaCl, 2 mM $MgCl_2$, 5 mM KCl, 1 mM $CaCl_2$, 0.02% Tween 20).[114] In order to make the experiments consistent with the conditions used for SELEX in this chapter, all samples were tested in PBS buffer. The K_D of AC1 and AC2 were 1.48 ± 0.09 µM and 1.10 ± 0.14 µM, respectively (Figure 5.22 **B**, n=3, independent repeats, the standard deviation for error bars, 1:1 binding model). AC1 and AC2 showed similar binding affinities, but much lower than the reported K_D values in the literature, possibly due to the different experimental settings used (the amount of streptavidin immobilised on the chip, running buffer and the flow-rates). Interestingly, the binding kinetics of two streptavidin aptamers were significantly different. The on-rate and off-rate of AC1 were observed to be more rapid compared to AC2 at identical injection concentrations (Figure 5.22 **A**).

AC1: ATTGACCGCTGTGTGACGCAACACTCAAT

AC2: GGGAACGCACCGATCGCAGGTTTCCC

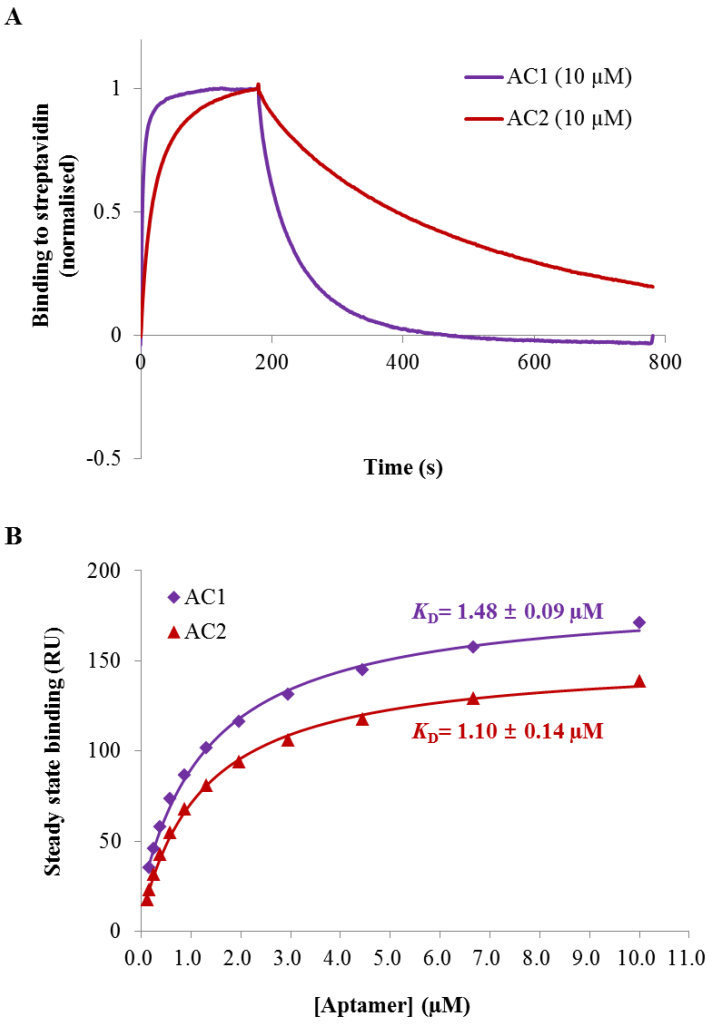


Figure 5.22. **A**) SPR association and dissociation curves for the highest concentration of AC1 and AC2 injected over a streptavidin-coated surface. The full titration curves were shown in appendix 8.18. **B**) The corresponding plot of steady-state binding data from the end of the association phase against aptamer concentration was used to calculate the steady-state affinities (K_D) (n=3, independent repeats, the standard deviation for the error bars, 1:1 binding model).

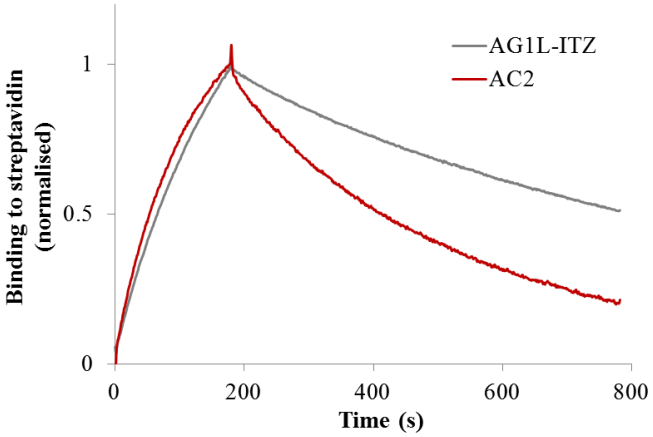
Since aptamer binding was observed for the controls, the failure of the binding assays of AG1-ITZ and AU4-ITZ (Figure 5.21) could be due to the excision of the primer binding sites which could play an important role in the aptamer's structure. For example, the 5′-end of the primer site may keep the FAM label away from the G-quadruplex structure

and the 3′-end contains an extra five 5-ITZ-dU sites which could enhance target interactions. Therefore, the primer binding sites should be attached to AG1-ITZ to test the effects of primer binding regions in its secondary structure conformation. The attachment of primer sites at the 5′- and 3′-ends of the AG1-ITZ makes the aptamer 79 bases long and the resultant oligonucleotides contains nine 5-ITZ-dUs (AG1L-ITZ, Table 5.7). When using the sequence, binding to streptavidin was observed with a steady-state K_D value of 1.70 ± 0.50 µM (Figure 5.23 **B**, n=3, independent repeats, the standard deviation for error bars, 1:1 binding model). In order to validate the effects of iodotriazole, two controls which contain dT or 5-PTZ-dU in place of 5-ITZ-dU (AG1L-T or AG1L-PTZ respectively) were also tested for their affinities (see appendix 8.19). AG1L-T showed the K_D of 14.90 ± 0.20 µM, which is almost a 9-fold decrease in affinity compared to AG1L-ITZ (Figure 5.23 **B**, n=2, independent repeats, a min-max method for error bars, 1:1 binding model). The binding constant for AG1L-PTZ was observed to be 10.88 ± 0.11 µM, which is again almost a 6.5-fold decrease in affinity for the target compared to AG1L-ITZ (Figure 5.23 **B**, n=2, independent repeats, a min-max method for error bars, 1:1 binding model).

AC1 and AC2 have bulge and hairpin structures whereas AG1L-ITZ has a G-quadruplex structure. The binding affinities measured by SPR showed a similar binding affinity to AG1L-ITZ. However, it was noteworthy that AG1L-ITZ has a slower dissociation rate (Figure 5.23 **A**) compared to AC2 indicating that once it is bound, it adheres more tightly to immobilised streptavidin. AG1L-PTZ and AG1L-T also showed slower off-rate which is similar to AG1L-ITZ (see appendix 8.19) and it could be due to highly stable G-quadruplex structures which fit into the binding pocket more tightly than bulge-hairpin shape.

AG1L-ITZ: 5′-FAM-AGCAGCACAGAGGTCAGATGGCGGGAAGGCAGGGA-AGGGGGA**T**GGGGCGGGACA**TTT**GGCC**T**A**T**GCG**T**GC**T**ACCG**T**GAA-3′

A



B

	Binding affinity (K_D in μM)
AG1L-ITZ	1.70 ± 0.50
AG1L-T	14.90 ± 0.20
AG1L-PTZ	10.88 ± 0.11

Figure 5.23. **A**) SPR association and dissociation curves for AG1L-ITZ and AC2 (≈0.35 μM) injected over a streptavidin-coated surface. The positions of 5-ITZ-dU in the sequences are highlighted in red. **B**) A table of the steady-state binding affinities (K_D) of AG1L-ITZ (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model), AG1L-T (n=2, independent repeats, a min-max method for error bars, 1:1 binding model) and AG1L-PTZ (n=2, independent repeats, a min-max method for error bars, 1:1 binding model) calculated from the end of association phases against aptamer concentrations. The full titration curves of AG1L-ITZ, AG1L-T and AG1L-PTZ were shown in appendix 8.19.

Attempts to fit AG1L-ITZ and its controls to a 1:1 association and dissociation kinetics model were made. However, the parameters cannot be fully trusted due to poor fitting of the data as shown in Figure 5.24. The shape of the curves is suggestive of the kinetics but further studies are required. For example, a large amount of streptavidin on the surface may cause re-binding effects that complicate the kinetics. To avoid the mass transport limitation in the kinetic analysis, lower surface densities of target and higher flow rates are recommended.[473] The binding of the aptamer to streptavidin could be more complex than a 1:1 binding mode. In addition, AG1L-ITZ exhibits G4 characteristics and therefore one sequence can form many G-quadruplex structures, possibly causing complex kinetics.

AG1L-ITZ: 5′-FAM-AGCAGCACAGAGGTCAGATGGCGGGAAGGCAGGGA-AGGGGGATGGGGCGGGACATTTGGCCTATGCGTGCTACCGTGAA-3′

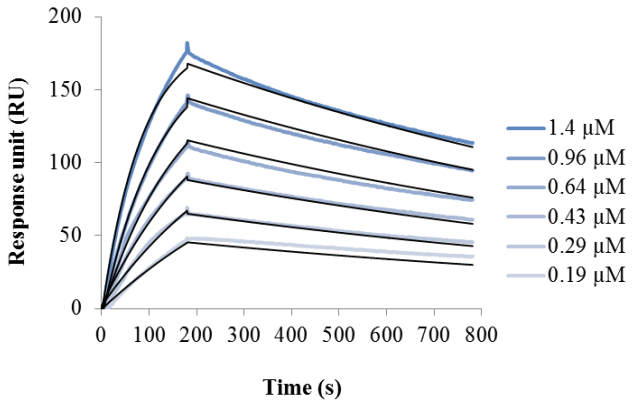


Figure 5.24. Kinetic analysis of AG1L-ITZ binding to streptavidin using SPR. The positions of 5-ITZ-dU in the sequences are highlighted in red. The experimental data (coloured curves) were fitted using the 1:1 binding model (black lines) to determine the kinetic parameters.

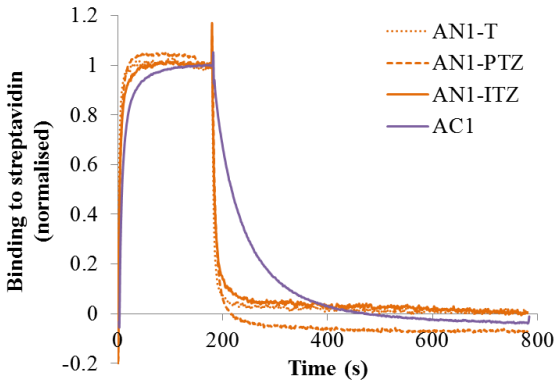
Although these results suggest halogen bonding is important, it cannot be assumed as the different sizes of the proto- and iodo-triazoles might influence G-quadruplex folding and hydrophobic interactions. It is not certain that halogen bonding contibutes to higher binding affinity, but it is interesting that the streptavidin aptamer selected from the 5-ITZ-dU containing libraries forms G-quadruplex secondary conformations which differ from the streptavidin aptamers found in the literature.

In addition, the best enriched NMer fragment sequences (non-G4) were tested for their binding affinities. In Figure 5.25 **A** the binding curves of AN1-ITZ and its two controls which contain dT or 5-PTZ-dU in place of 5-ITZ-dU (AN1-T or AN1-PTZ, respectively) were compared with AC1. The binding kinetics between the modified (ITZ and PTZ) and canonical DNA showed no difference. They bound to the streptavidin and dissociated immediately and therefore it was not possible to fit the data to obtain kinetic information. AN1-ITZ and its controls showed much faster on- and off-rate compared to AC1. The K_D value of AN1-ITZ was 1.21 ± 0.54 µM (Figure 5.25 **B**, n=3, independent repeats, the standard deviation for error bars, 1:1 binding model) and this is similar to

the binding affinities of AC1 (Figure 5.22 **B**). AN1-T and AN1-PTZ showed the K_D value of 0.73 ± 0.06 µM (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model) and 0.82 ± 0.02 µM (n=2, independent repeats, a min-max method for error bars, 1:1 binding model) respectively (Figure 5.25 **B**). These results suggest iodo-containing AN1-ITZ binds more weakly than AN1-T and AN1-PTZ, but the difference in K_D is low (only two-fold). It is noteworthy that this non-G4 forming sequence showed rapid kinetics, suggesting sufficient screening of the SELEX libraries can obtain aptamers with very different properties.

AN1-ITZ: CCACGCGGCCGGCGCAGGCCTGGG

A



B

	Binding affinity (K_D in µM)
AN1-ITZ	1.21 ± 0.54
AN1-T	0.73 ± 0.06
AN1-PTZ	0.82 ± 0.02

Figure 5.25. **A**) SPR association and dissociation curves for AN1-ITZ, AN1-T, AN1-PTZ and AC1 (2.96 µM) injected over a streptavidin-coated surface. The position of 5-ITZ-dU in the sequences is highlighted in red. **B**) A table of the steady-state binding affinities (K_D) of AN1-ITZ (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model), AN1-T (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model) and AN1-PTZ (n=2, independent repeats, a min-max method for error bars, 1:1 binding model) calculated from the end of association phases against aptamer concentration. The full titration curves of AN1-ITZ, AN1-T and AN1-PTZ were shown in appendix 8.20.

Next, AN3-ITZ and its two controls which contain dT or 5-PTZ-dU in place of 5-ITZ-dU (AN3-T or AN3-PTZ, respectively) were compared with AC2 (Figure 5.26 **A**). The association and dissociation rates for AN3-T were more rapid than AN3-ITZ and AN3-

PTZ (Figure 5.26 **A**). However, compared to AC2, they all showed the relatively fast on- and off-rates (Figure 5.26 **A**). There was a slight difference in binding kinetics between them but the binding affinities were similar. K_D values obtained for AN3-ITZ was 1.38 ± 0.03 µM (Figure 5.26 **B**, n=3, independent repeats, the standard deviation or error bars, 1:1 binding model) whereas the obtained K_D of AN3-T and AN3-PTZ were 1.92 ± 0.35 µM (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model) and 1.27 ± 0.25 µM (n=2, independent repeats, a min-max method for error bars, 1:1 binding model) respectively (Figure 5.26 **B**). The K_D values are too close with the values taking errors into account suggesting the K_D values overlap and hence, the modification has a negligible effect for this aptamer candidate.

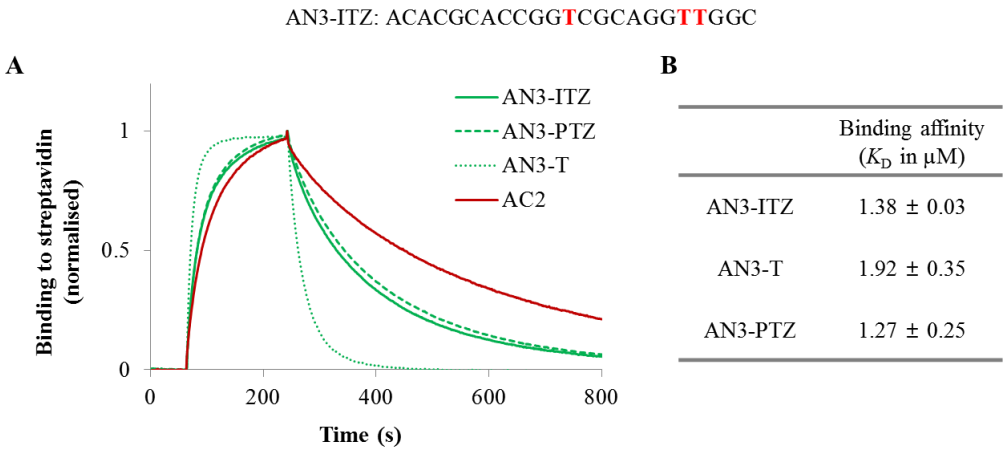


Figure 5.26. **A**) SPR association and dissociation curves for AN3-ITZ, AN3-T, AN3-PTZ and AC2 (≈2.96 µM) injected over a streptavidin-coated surface. The positions of 5-ITZ-dU in the sequences are highlighted in red. **B**) A table of the steady-state binding affinities (K_D) of AN3-ITZ (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model), AN3-T (n=3, independent repeats, the standard deviation for error bars, 1:1 binding model) and AN3-PTZ (n=2, independent repeats, a min-max method for error bars, 1:1 binding model) calculated from the end of association phases against aptamer concentrations. The full titration curves of AN3-ITZ, AN3-dT and AN3-PTZ were shown in appendix 8.21.

5.6.3 EMSA analysis

During SPR binding analysis, the steady-state binding affinity of AG1L-ITZ was 6.5- to 9- fold higher than those of AG1L-PTZ and AG1L-T, respectively. In order to support these results, EMSA was performed. A serial of dilution of streptavidin (SA, from 0 to 100 µM) was incubated with 1 µM aptamer for 30 min before the non-denaturing separation of the SA-aptamer complexes from the unbound aptamer. Gels were visualised using aptamer FAM fluorescence and quantified using Fiji software.

The SA-aptamer complex was formed in a concentration-dependent manner and migrated slower than the corresponding free nucleic acid due to the high molecular weight of the complexes. However, the complex bands were not sharp compared to the unbound DNA bands, possibly due to dissociation of the complex during electrophoretic separation. Therefore, it was assumed any decrease in band intensity of unbound aptamer to the initial aptamer band intensity is proportional to the SA-aptamer complexes formed. The results were fitted to a 1:1 model[474] by Dr Arun Shivalingam using MatLab.

Firstly, AG1L-ITZ was tested. The SA-aptamer complexes remained unclear, but a concentration-dependent decrease in the unbound aptamer band intensity was observed as streptavidin concentration was increased. An additional band above the unbound aptamer band was also observed, possibly due to the secondary structure or dimerisation of the aptamer (Figure 5.27 **A**). The affinity graph which was drawn using the unbound DNA showed the estimated binding constant for AG1L-ITZ is about 1.40 ± 0.31 µM (Figure 5.27 **B**) which is very close to K_D value measured using steady-state binding via SPR (1.70 ± 0.50 µM).

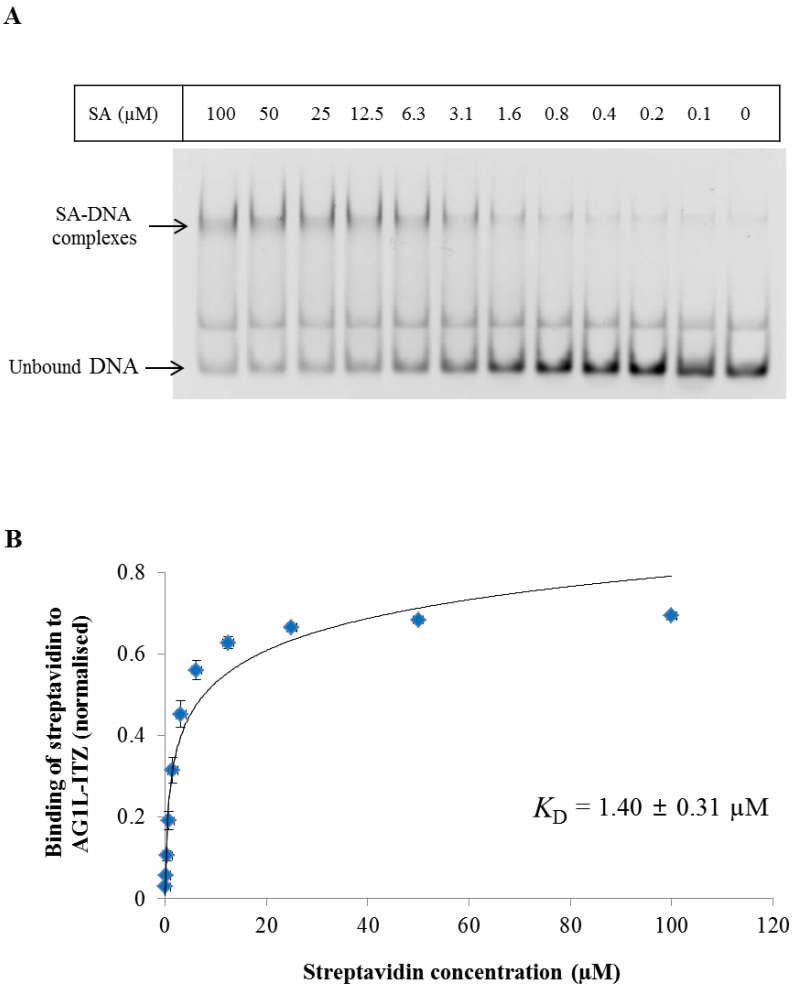


Figure 5.27. Titration of streptavidin protein (SA) incubated with 1 μM AG1L-ITZ. The complex consists of AG1L-ITZ bound to the according concentration of SA. **A**) Electrophoresis was performed with an 8% native PAGE gel running in the 1× Tris-Borate-EDTA buffer at 4 ℃. **B**) The plot shows the fraction of AG1L-ITZ bound as SA was titrated. The value for the K_D obtained from the 1:1 binding model was 1.40 ± 0.31 μM (n=3, independent repeat, the standard deviation for error bars).

The results of an EMSA to measure the affinity of the interaction between AG1L-T and SA are shown in Figure 5.28 **A**. Approximately, 50% of the DNA was in a complex with SA in reactions containing 3.1–6.3 nM SA. A 1:1 binding fitted K_D value of AG1L-T was 4.55 µM (Figure 5.28 **B**), which is lower than the K_D value obtained from the steady-state affinity by SPR (14.90 ± 0.20 µM).

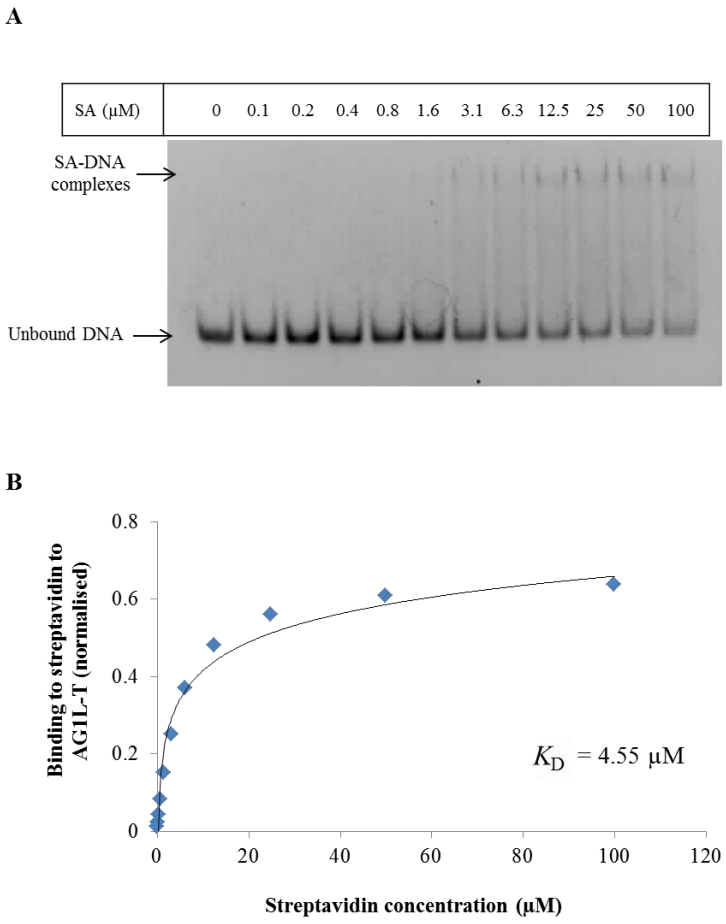


Figure 5.28. Titration of streptavidin protein (SA) incubated with 1 µM AG1L-T. The complex consists of AG1L-T bound to the according concentration of SA. **A**) Electrophoresis was performed with an 8% native PAGE gel running in the 1× Tris-Borate-EDTA buffer at 4 ℃. **B**) The plot shows the fraction of AG1L-T bound as SA was titrated. The value for the K_D obtained from the 1:1 binding model was 4.55 µM.

For AG1L-PTZ, similar gel migration trends to AG1L-ITZ were observed (Figure 5.29 **A**). The second bands appeared above the unbound DNA bands and this could be due to the various G-quadruplex conformations of AG1L-PTZ. A 1:1 binding fitted K_D value of AG1L-PTZ was 3.55 ± 1.93 µM (Figure 5.29 **B**), which is lower than the dissociation constant obtained from the steady-state affinity by SPR (10.77 ± 0.11 µM).

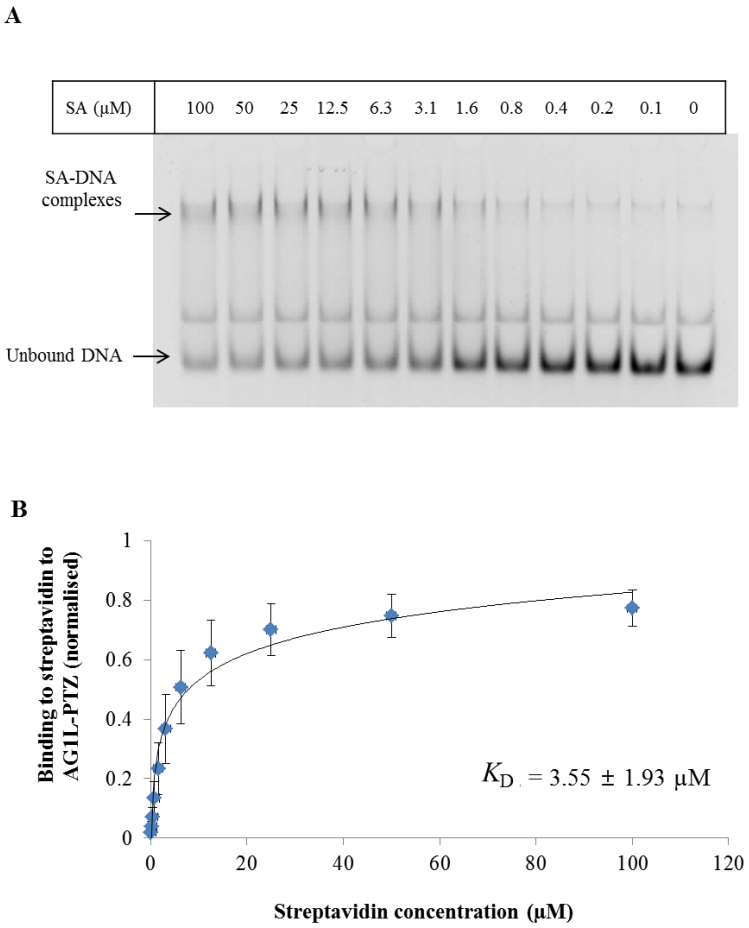


Figure 5.29. Titration of streptavidin protein (SA) incubated with 1 µM AG1L-PTZ. The complex consists of AG1L-PTZ bound to the according concentration of SA. **A**) Electrophoresis was performed with an 8% native PAGE gel running in the 1× Tris-Borate-EDTA buffer at 4 ℃. **B**) The plot shows the fraction of AG1L-PTZ bound as SA was titrated. The value for the K_D obtained from the 1:1 binding model was 3.55 ± 1.93 µM (n=3, independent repeat, the standard deviation for error bars)

In summary, the results of binding affinity trends were consistent with trends observed from SPR experiments. The K_D of AG1L-ITZ was the lowest and the highest K_D value was observed for AG1L-T. Expect AG1L-ITZ, AG1L-T and AG1L-PTZ exhibited much higher binding affinities compared to values obtained from SPR measurement. In order to accurately determine the binding affinities of aptamers to streptavidin, the bound complexes are supposed to be used as determining factors. Here we performed EMSA to qualitatively confirm the binding affinities initially obtained by SPR. The major limitation of EMSA is that the system is not at chemical equilibrium during electrophoresis. Rapid dissociation of streptavidin and DNA during electrophoresis can prevent detection of complexes, and this could be the main reasons for having smeared and streaky complex bands while free DNA migrates as sharp bands.[293]

5.7 Conclusions

SELEX techniques established in the previous chapter were applied to modified ssDNA libraries in this chapter. The goal was to expand aptamer structural diversity and thereby increase the binding affinities of aptamers to their target.

SELEX was performed using 5-ITZ-dUTP in place of dTTP to investigate the effects of halogen bonding interactions between aptamer and streptavidin. In total, 10 rounds of selection were done with negative selection performed prior to positive selection to eliminate the selection of sequences that bound to the solid support. Sequence enrichment was monitored by Q-PCR and found to be substantially higher by round 8. The enriched pools were analysed by NGS and the various data analysis methods confirmed that the most enriched sequences formed G-quadruplex structures and that the modified base was not selected against during SELEX. G-quadruplex aptamers and

two non-G4 motifs were enzymatically and chemically synthesised using various methods to find the highest yielding method. Due to the strong intermolecular interactions between individual strands of G-quadruplex aptamers, primer extension reaction followed by denaturing gel electrophoresis gave low yields of the desired strands. However, the use of a basic buffer (Tris-Caps buffer pH 9.6) in gel electrophoresis showed an improvement in ssDNA production after the primer extension reaction. Still, post-synthetic labelling of oligonucleotides via the NHS esters of the modifications was the most efficient and higher yielding method.

The binding affinities of our aptamers and control unmodified literature aptamers were then determined by SPR. G-quadruplex aptamers without primer binding sites (AG1-ITZ) completely lost their ability to bind suggesting that the primer binding sites are involved in secondary/tertiary structure formation or binding to streptavidin. The G-quadruplex aptamer containing iodotriazole with primer binding sites (AG1L-ITZ) had a binding affinity of 1.70 ± 0.50 µM, much higher than its control sequences (10.88 ± 0.11 µM for AG1L-PTZ and 14.90 ± 0.20 µM AG1L-T). The same binding trends were observed by EMSA. SPR experiments using canonical streptavidin-binding aptamer (AC1 and AC2) showed similar K_D values to AG1L-ITZ, but AG1L-ITZ appears to have a slower dissociation rate; more work is required to confirm this.

These results alone are not sufficient to tell if AG1L-ITZ's enhanced binding to streptavidin is due to halogen bonding. For example, aptamer folding could be affected by the iodotriazole modification or the hydrophobicity of the iodo group may contribute to binding. In order to further investigate halogen bonding interaction between this aptamer and streptavidin, structural studies such as X-ray crystallography need to be performed in future.

Chapter 6

Conclusion and future work

6 Conclusion and future work

There has been significant growth in the discovery and application of aptamers in the past quarter-century. However, in contrast to antibodies, aptamers still remain under-utilised in practical applications. The main bottlenecks limiting the development of aptamers are their rapid degradation *in vivo* and the relatively limited range of intermolecular interactions they can form with target molecules. Hence, next-generation aptamers with greater biological functions and pharmacokinetic are highly anticipated. Given the importance of chemical modifications in designing next-generation aptamers, we chemically synthesised various modified nucleotide triphosphates and examined their capabilities during enzymatic DNA synthesis. At the end of the study, a DNA polymerase-compatible iodo-substituted dU triphosphate was used in SELEX to discover aptamers exhibiting high binding affinity for streptavidin.

In the early stage of the study, a modified DNA template containing 2′-aminoethoxy-T (2′-AmE-T, Figure 6.1 **A**) was used in enzymatic DNA synthesis to examine the ability of DNA polymerases to read-through 2′-AmE-T. It was successfully read-through and dATP was incorporated opposite it in the antisense strand by KOD XL DNA polymerase with supplementary $MnCl_2$. To demonstrate sugar modified triphosphates incorporation into DNA by polymerases, 2′-azidoethoxy-TTP (2′-AzE-T, Figure 6.1 **B**) was synthesised by adaptation of the Yoshikawa method in 16% yield. It was then put instead of dTTP used in primer extension reactions with a canonical DNA template. Even using KOD XL DNA polymerase, an enzyme that is very tolerant of modified triphosphate, 2′-AzE-TTP and 2′-ethynyl-UTP (Figure 6.1 **C**) could not be successfully incorporated into DNA.

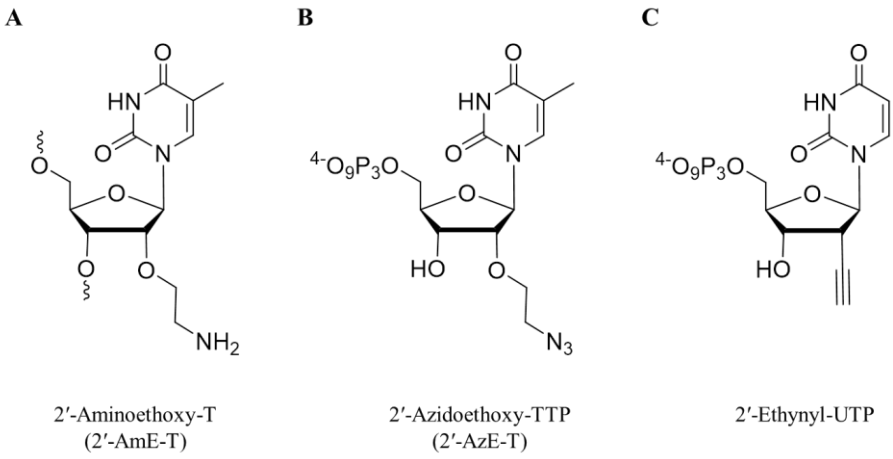


Figure 6.1. Chemical structures of 2′-aminoethoxy-T (**A**), 2′-azidoethoxy-TTP (**B**) and 2′-ethynyl-UTP (**C**).

These results are consistent with the literature for 2′-sugar modified nucleotides (2′-F-, 2′-NH_2- and 2′-OMe-dNTPs).[249,325] DNA polymerases, therefore, appear to have a very low tolerance for 2′-sugar modifications due to the presence of a steric gate within DNA polymerase. Therefore, enzymatic synthesis using wild-type or engineered T7 RNA polymerases could be evaluated for RNA aptamer discovery. Alternatively, the nuclease resistance of aptamers could be increased by post-SELEX modifications such as F-, NH_2- or OMe- introduced at C2′ position of sugar ring. The pharmacokinetic properties of aptamers for *in vivo* application could also be optimised using 3′-end capping with inverted deoxythymidine to prevent aptamer against 3′-exonucleases or 5′-capping with 40 kDa PEG to prevent renal clearance. However, it is noteworthy that the secondary structure and binding characteristics of aptamers are not altered by these modifications.

To expand the range of interactions an aptamer can make with its target, chemical modifications at the C5 position of deoxyuridine triphosphate were explored. In order to orient the modification away from the hydrogen bonding face of the bases, 5-propargylamino-dUTP (5-PA-dUTP, Figure 6.2 **A**) was used as a linker. The

propargylamino group of 5-PA-dUTP, introduced via a Sonogashira coupling reaction, was conjugated to various functional groups via amide coupling reactions. Hydrophobic groups and halogen substituted modifications were attached to 5-PA-dUTP and used in the enzymatic synthesis to examine their level of acceptance by DNA polymerase. Base-modified nucleotides were good substrates of KOD XL DNA polymerase as they were efficiently incorporated into DNA, except those with bulky modifications such as 5-propargylamino-1-pyrenebutyl-dUTP (5-PAPBA-dUTP, Figure 6.2 **B**) and 5-propargylamino-2,6-difluoro-4-iodobenzyl-dUTP (5-FFI-dUTP, Figure 6.2 **C**). Among them, 5-propargylamino-iodotriazole-dUTP (5-ITZ-dUTP, Figure 6.2 **D**) was selected for modified DNA aptamer selection via SELEX.

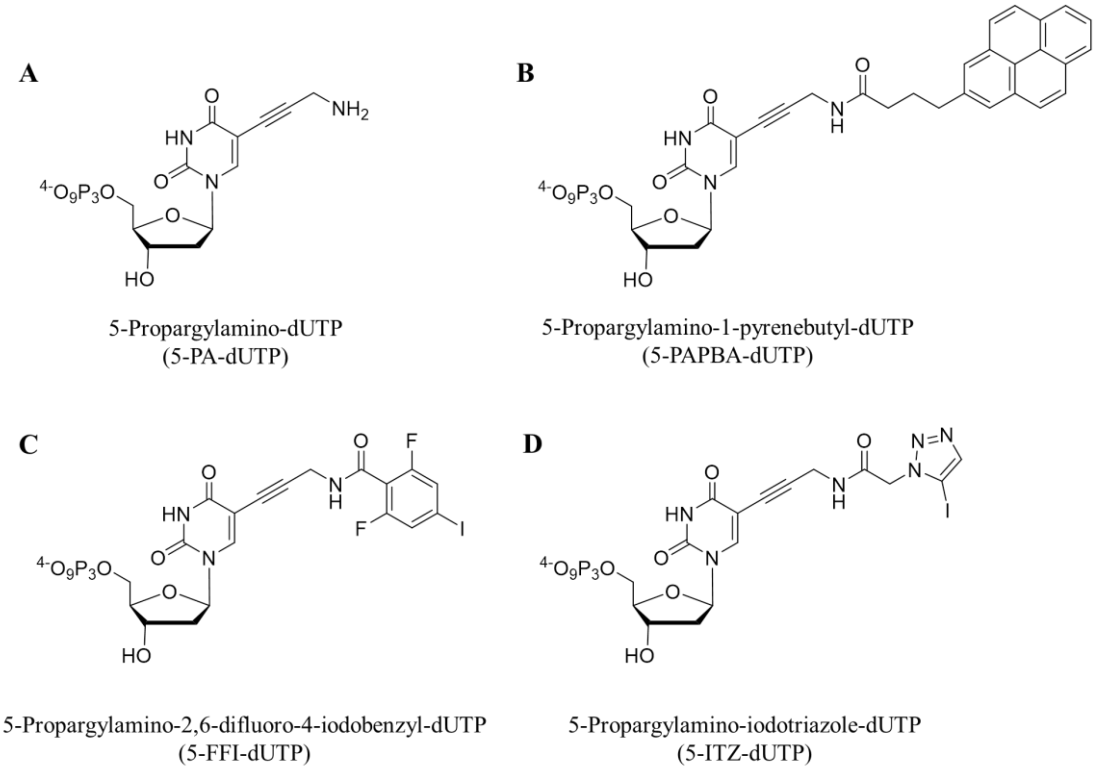


Figure 6.2. Chemical structures of 5-propargylamino-dUTP (**A**), 5-propargylamino-1-pyrenebutyl-dUTP (**B**) 5-propargylamino-2,6-difluoro-4-iodobenzyl-dUTP (**C**) and 5-propargylamino-iodotriazole-dUTP (**D**).

Prior to performing SELEX using 5-ITZ-dUTP, a canonical DNA library was firstly used to establish the SELEX methodology using two different epitopes found in the HER2 extracellular domain as targets. Due to the small sizes of the targets, this was unsuccessful and an aptamer was selected to the streptavidin immobilised on the solid-support (negative target) instead of the desired target. Also, it was hypothesised that the enrichment of aptamers towards the negative target was due to a lack of the negative selection in the early stages of SELEX. Taken this as a model system and to study the effect of halogen bonding in the binding affinity, SELEX using 5-ITZ-dUTP was then performed using streptavidin as the target. The enriched aptamer sequences were discovered after 10 rounds of negative and positive selection. The most enriched sequences were G-rich when analysed by next-generation sequencing (NGS) methods and therefore they could potentially form G-quadruplexes which were distinct from the bulge-hairpin structure of the streptavidin aptamers found in the literature.[409] NGS analysis also confirmed that 5-ITZ-dUTP was tolerant with DNA polymerase during PCR indicating that there was no undesirable selection pressure. G-quadruplex aptamer candidates were synthesised for binding affinity validation using SPR. The importance of primer binding sites in aptamer secondary structure formation was demonstrated, with only the full-length aptamer (AG1L-ITZ) showing good affinity for the streptavidin target (1.70 ± 0.50 μM). This K_D value was 6.5-fold and 9-fold higher than its appropriate control sequences (AG1L-PTZ and AG1L-T, respectively) which contains 5-propargylamino-prototriazole-dU or dT instead of 5-ITZ-dU. The binding results were verified by EMSA which showed similar trends.

The binding affinities of canonical streptavidin-binding aptamer in the literature[409,427] (K_D =1.48 ± 0.09 μM for AC1 and 1.10 ± 0.14 μM for AC2) showed similar K_D values to AG1L-ITZ when analysed by SPR, but AG1L-ITZ appears to have a slower

dissociation rate. AG1L-ITZ contains nine 5-ITZ-dU modifications, and a detailed structural study would show contribution of halogen bonding to streptavidin binding. X-Ray crystallography should be possible as a structure of streptavidin is already available in the literature. Streptavidin can be pre-saturated with biotin to investigate the binding sites of aptamers and also binding analysis using counter targets such as avidin can be performed to test the specificity of the chosen aptamer. In addition, iodotriazole can be changed to bromo- or chlorotriazole. By varying the halogen atom, the conformational changes in the aptamer, and the effects of binding specificity and affinity could be investigated. Alternatively, known aptamers can be site-specifically modified by the incorporation of halogen substituted groups to demonstrate the effects of halogen bonding on aptamer-target binding. Given our continuous interest in immunotherapeutic agent development, attempts to generate aptamers targeting HER2 extracellular peptides can be resumed. In order to increase the success rate of SELEX, the HER2 peptides can be immobilised directly on magnetic beads by the reaction of their –COOH and $–NH_2$ groups. In this case, small peptides will not be shielded by the support which is larger than its target (e.g. streptavidin beads). Throughout the studies, we found several base-modified nucleotides that are efficiently accepted by DNA polymerase during enzymatic DNA synthesis. Therefore, SELEX using two differently chemically modified nucleotides can be introduced in a ssDNA pool to enhance target binding specificity and affinity.

Chapter 7

Experimental

7 Experimental

7.1 General methods

All reagents were purchased from Sigma-Aldrich, Carbosynth ltd, Alfa Aesar, Acros Organics, Merck Millipore, Lonza, Invitrogen and Thermo Fisher Scientific and used without further purification. The exceptions were dichloromethane (DCM), triethylamine (TEA), pyridine, *N*,*N*-diisopropylethylamine (DIEA) which were purified by distillation or dried over activated molecular sieves overnight. Deoxygenated solvents were prepared by bubbling argon through the solvent for 15 min. All air/moisture sensitive reactions were performed under argon in flame-dried glassware. Column chromatography was carried out using Merck Geduran 60 Å (particle size 40–63 micron) silica.

All the molecular sieves used in the experiments were 3 Å (beads, 8–12 mesh, Sigma-Aldrich). Thin layer chromatography (TLC) was performed using Merck TLC silica gel 60 F_{254} plates (0.22 mm thickness, aluminium coated) and the compounds were visualised by staining with *p*-anisaldehyde or potassium permanganate ($KMnO_4$) (10% aq.) unless otherwise stated and by ultraviolet irradiation at 254/365 nm.

^{1}H, ^{13}C, and ^{31}P NMR spectra were recorded on a Bruker AVIIIHD 400 nanobay or AVII 500 with ^{13}C cryoprobe spectrometer. Chemical shifts are given in ppm and *J* coupling values are quoted in Hz. ^{1}H NMR spectra were internally referenced to the appropriate residual undeuterated solvent signal; ^{13}C NMR spectra were internally referenced to the appropriate deuterated solvent signal; assignment of the compounds was aided by COSY (^{1}H–^{1}H), HSQC (^{1}H–^{13}C) and HMBC experiments.

Low-resolution mass spectra (LRMS) were recorded using a Waters ZMD quadrupole mass spectrometer for +/- electrospray ionisation (ESI) in HPLC grade acetonitrile or methanol. High-resolution electrospray mass spectra (HRMS) of crude or purified deoxynucleoside triphosphates and oligonucleotides were characterised in deionised water. MS analysis of the oligonucleotides was performed in the negative mode using Bruker Daltonics micro-TOF mass spectrometer or XEVO G2-QTOF mass spectrometer equipped with an Acquity UPLC BEH C18 column (2.1 × 50 mm, 1.7 µm, 130 Å, Waters).

For oligonucleotide mass, a gradient of acetonitrile in triethylamine (TEA) and hexafluoroisopropanol (HFIP) was employed, increasing from 0–70% buffer B over 7.5 min, with a flow rate of 0.2 mL/min. For triphosphate mass, a gradient of 0 % buffer B for 3 min and then increased to 70% buffer B over 5 min, with a flow rate of 0.2 mL/min (buffer A: 8.6 mM TEA, 200 mM HFIP in 5% methanol/water; buffer B: 20% buffer A in methanol). Raw data was processed and deconvoluted using the Data Analytics function of the Bruker Daltonics Compass TM1.3 software package or the Waters' MassLynx Mass spectrometry software.

7.2 A general method for oligonucleotide synthesis, purification and analysis

Standard DNA phosphoramidites, solid supports and additional reagents were purchased from Link Technologies and Applied Biosystems Ltd. Oligonucleotides were synthesised using an Applied Biosystems 394 automated DNA/RNA synthesiser using a standard 0.2 µmol or 1.0 µmol phosphoramidite cycle of acid-catalysed detritylation, coupling, capping and iodine oxidation. Stepwise coupling efficiencies and overall

yields were determined by an automated trityl cation conductivity monitoring facility and in all cases were >98%. 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 35 sec and the coupling time for the FAM monomer was 360 sec. Cleavage of the oligonucleotides from the solid support and deprotection was achieved by exposure to concentrated aqueous ammonia solution (1 h, 25 ℃) followed by heating in a sealed tube (5 h, 55 ℃). Occasionally, fast-deprotecting monomers (acetyl dC and DMF dG) were used to reduce the transamination side products and all protecting groups can be cleaved in aqueous ammonia solution with heating at 55 ℃ for 1 h. After evaporation of the aqueous ammonia *in vacuo*, the fully deprotected oligonucleotides were purified by reverse-phase HPLC (RP-HPLC) on a Gilson system using a Phenomenex column Luna 10 µm C8 100 Å pore 10 × 250 mm. A gradient of acetonitrile in triethylammonium bicarbonate (TEAB) (buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 10–30% buffer B over 20 min, flow rate: 4 mL/min) was used. Elution was monitored by UV absorption at 280 nm. After RP-HPLC purification, oligonucleotides were desalted using Sephadex illustra NAP columns (GE Healthcare).

7.2.1 Oligonucleotides and triphosphates quantification

Oligonucleotides were quantified based on their absorption at 260 nm using the nearest neighbour molar extinction coefficient. Triphosphates were quantified based on their absorption at 260 nm for natural triphosphates or 290 nm for modified triphosphates using the nearest neighbour molar extinction coefficient.

The concentration of oligonucleotides or triphosphates were calculated by Lambert-Beer equation A = OD (optical density) = εcl (ε = extinction coefficient, c = concentration and l = 1 cm) at 25 ℃ in deionised water.

7.2.2 Denaturing polyacrylamide gel electrophoresis (PAGE)

Denaturing polyacrylamide gel electrophoresis (18 (W) x 0.2 (D) x 24.4 (H) cm) was used to analyse the reaction progress and also the bands of interest were extracted for further analysis such as mass spectrometry analysis. A mixture containing 7 M urea, acrylamide (19:1 acrylamide/bis ratio, 40%) and 10× TBE buffer were prepared in deionised water (70 mL), according to the desired percentage of the gel. To polymerise the gel, 10 % (w/v) ammonium persulfate (APS, 580 µL) and *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED, 58 µL) were added, then immediately the gel plates were filled with the acrylamide mix and the well comb was inserted. After the gel was polymerised, it was placed in a gel running apparatus and 1× TBE buffer (running buffer) was added to the upper and lower reservoirs. The gel was pre-run for 30 min at 20 W at 25 ℃ before the sample was loaded. Oligonucleotides (20 µL) were mixed with formamide (20 µL) and loaded onto each well and the gel was run for 2–4 h at 20 W depending on the percentage of gels or band migration.

The desired DNA bands were excised, crushed and incubated with 15 mL deionised water in a shaking incubator overnight at 37 ℃. After filtration, the solution was concentrated and excess salts were removed using a Sephadex illustra NAP column.

7.3 General buffers

7.3.1 RP-HPLC buffer

Triethylammonium bicarbonate (TEAB) buffer: buffer A, 0.1 M TEAB, pH 7.5; buffer B, 0.1 M TEAB, pH 7.5 with 50% acetonitrile.

Triethylammonium acetate (TEAA) buffer: buffer A, 0.1 M TEAA, pH 7.5; buffer B, 0.1 M TEAA, pH 7.5 with 50% acetonitrile.

Ammonium acetate buffer: buffer A, 0.1 M ammonium acetate, pH 7.0; buffer B, 0.1 M ammonium acetate, pH 7.0 with 50% acetonitrile.

Hexylammonium acetate (HAA) buffer: buffer A, 0.1 M HAA, pH 7.0; buffer B, 0.1 M HAA, pH 7.0 with 50% acetonitrile.

7.3.2 Anion exchange buffer

Triethylammonium bicarbonate (TEAB) buffer: buffer A, 10 mM TEAB, pH 7.5; buffer B, 1 M TEAB, pH 7.5.

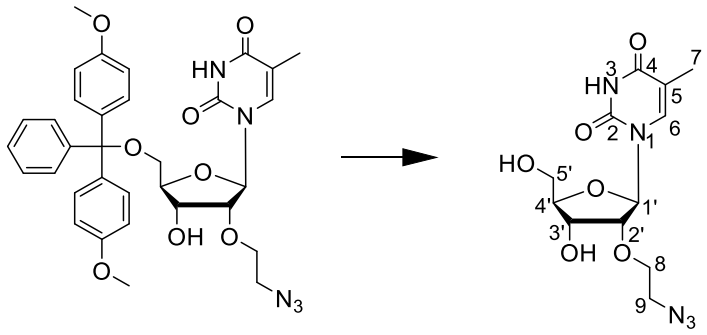
7.3.3 Gel electrophoresis buffer

10× Tris-Borate-EDTA (TBE) buffer, pH 8.3: Tris base (108 g), boric acid (55 g) and ethylenediaminetetraacetic acid (9.3 g) dissolved in deionised water (total = 1 L).

10× Tris-CAPS (TC) buffer, pH 9.6: Tris base (72.7 g) and 3-(cyclohexylamino)-1-propanesulfonic acid (88.5 g) dissolved in deionised water (total = 1 L).

7.4 Chemical Synthesis

7.4.1 2′-Azidoethoxy-thymidine (2′-AzE-T)[475] (1)



Purified 5′-*O*-(4,4′-Dimethoxytrityl)-2′-azidoethoxy-thymidine[476] was provided by Dr Anna Dysko (Brown group). Prior to use in detritylation, a Dowex resin (50W × 2, 30 mL) was equilibrated with 0.1 M aqueous hydrochloric acid (HCl, 90 mL). The column was then washed with deionised water (90 mL) until the pH of the dripping solution became pH 7.0. The wet resin was used for detritylation. 5′-*O*-(4,4′-Dimethoxytrityl)-2′-azidoethoxy-thymidine (0.54 g, 0.86 mmol, 1 eq) was dissolved in methanol (1.5 mL) and loaded onto the resin. After 10 min, deionised water (60 mL) was added to the column to elute the detritylated compound. Then, the eluent was washed twice with dichloromethane (2 × 100 mL) followed by petroleum ether (2 × 100 mL). The water layer was dried *in vacuo*. 2′-AzE-T **1** was obtained as a white solid. Yield: 30% (85 mg, 0.26 mmol). M_w = 327.3 g/mol (Chemical Formula: $C_{12}H_{17}N_5O_6$)

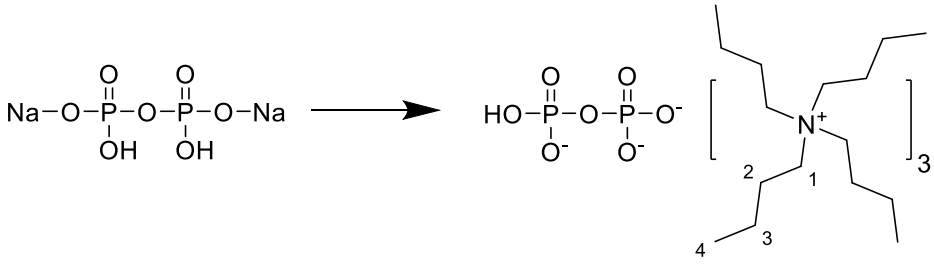
R_f: 0.18 [100 % EtOAc]

^{1}H NMR (400 MHz, CD_3OD) δ 7.95 (d, *J* = 1.1 Hz, 1H, H-6), 5.96 (d, *J* = 3.9 Hz, 1H, H-1′), 4.27 (t, *J* = 5.5 Hz, 1H, H-3′), 4.04 (dd, *J* = 5.5, 3.9 Hz, 1H, H-2′), 4.01 (dt, *J* = 5.5, 2.7 Hz, 1H, H-4′), 3.90 (dd, *J* = 12.2, 2.7 Hz, 1H, H-5′a), 3.82 – 3.89 (m, 2H, H-8), 3.76 (dd, *J* = 12.2, 2.7 Hz, 1H, H-5′b), 3.50 (ddd, *J* = 13.3, 6.5, 4.4 Hz, 1H, H-9a), 3.33 - 3.41 (m, 1H, H-9b), 1.87 (d, *J* = 1.1 Hz, 3H, H-7)

^{13}C NMR (101 MHz, CD_3OD) δ 138.3 (C6), 89.0 (C1′), 86.2 (C4′), 83.9 (C2′), 71.0 (C8), 70.0 (C3′), 61.8 (C5′), 51.9 (C9), 12.6 (C7)

LRMS: [ESI^+, MeOH] calculated mass: 327.1; found *m/z*: 328.1 ($[M+H]^+$, 56) and 350.1 ($[M+Na]^+$, 100)

7.4.2 Tris(tetrabutylammonium) hydrogen pyrophosphate[477] (2)



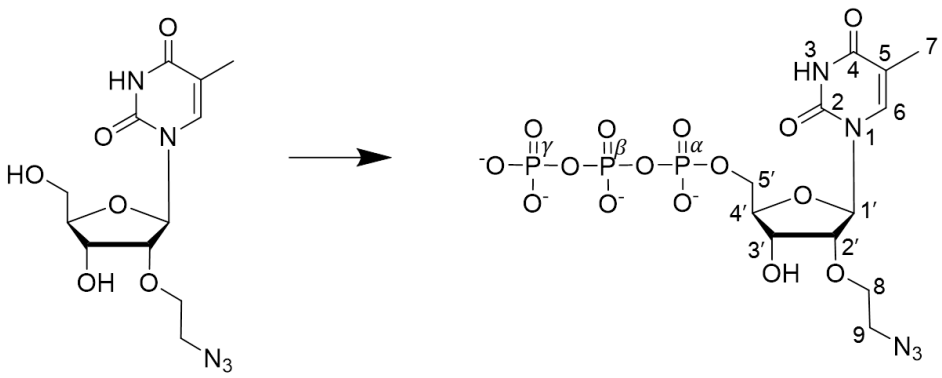
Dibasic sodium pyrophosphate (2.22 g, 10 mmol) was dissolved in deionised water (40 mL) and passed through Dowex resins (50W × 2, 40 mL) that had been equilibrated with 0.1 M aqueous hydrochloric acid (120 mL). The column was then washed with deionised water (120 mL) until the pH of the dripping solution became pH 7.0. The eluent was collected in a flask containing 0.2 M tetra-butylammonium hydroxide (90 mL) that was cooled to -15 ℃. The solution was then titrated with 1.0 M tetra-butylammonium hydroxide (5 mL) until it was neutralised (pH 7.0). The mixture was then evaporated to dryness under high vacuum and left *in vacuo* overnight. Tris hydrogen pyrophosphate **2** was obtained as a white solid. It was aliquoted in a glove box with molecular sieves due to its high hygroscopic nature and stored at 4 ℃. M_w = 902.4 g/mol (Chemical Formula: $C_{48}H_{109}N_3O_7P_2$).

^{1}H NMR (400 MHz, D_2O) δ 3.12 (t, *J* = 8.6, 7.9 Hz, 8H, H-1), 1.57 (m, 8H, H-2), 1.31 (tq, *J* = 7.4 Hz, 8H, H-3), 0.90 (t, *J* = 7.4 Hz, 12H, H-4)

^{13}C NMR (101 MHz, D_2O) δ 58.0 (CH_2) 23.1 (CH_2), 19.2 (CH_2), 13.0 (CH_3)

^{31}P NMR (162 MHz, D_2O) δ -8.4

7.4.3 2′-Azidoethoxy-thymidine-5′-*O*-triphosphate (2′-AzE-TTP)[236] (3)



Before commencing the procedure, the 2′-AzE-T **1**, 1,8-bis(dimethylamino)naphthalene (proton-sponge) and tris(tetrabutylammonium) hydrogen pyrophosphate **2** were dried over potassium hydroxide under high vacuum and trimethyl phosphate was dried over molecular sieves overnight. Trimethyl phosphate (2.13 mL) was added with proton-sponge (84 mg, 0.39 mmol, 1.5 eq) to 2′-AzE-T **1** (85 mg, 0.26 mmol) under argon at -15 ℃. Phosphorus oxychloride (29 µL, 0.31 mmol, 1.2 eq) was then added dropwise with stirring. After 30 min, tris (tetrabutylammonium) hydrogen pyrophosphate **2** (1.17 g, 1.3 mmol, 5 eq) in *N*,*N*-dimethylformamide (1.17 mL) was added with tributylamine (0.23 mL). The reaction mixture was then stirred at -15 ℃ for a further 3 h before the reaction was quenched with 1 M triethylammonium bicarbonate buffer (TEAB, 5 mL, pH 7.5) and the solvent removed *in vacuo*. The crude reaction mixture was dissolved in deionised water (3 mL) and mixed with 3 M sodium chloride solution (300 µL) and ethanol (9 mL) to precipitate the triphosphate. In order to increase product yield, this precipitation procedure was repeated three times on the supernatant (the solvent of the supernatant was first removed *in vacuo* and re-dissolved in 3mL deionised water every time). After performing the precipitation procedure three times, no further precipitate was observed. The resulting precipitated pellets were dissolved in deionised water (1

mL). Finally, the crude product was purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 3.5–35% buffer B over 35 min; flow rate: 8 mL/min). UV absorbance of 2′-AzE-TTP **3** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 16% (420 OD_{267}, 42.4 µmol, ε_{267} = 9,900 $M^{-1}cm^{-1}$). M_w = 567.2 g/mol (Chemical Formula: $C_{12}H_{20}N_5O_{15}P_3$).

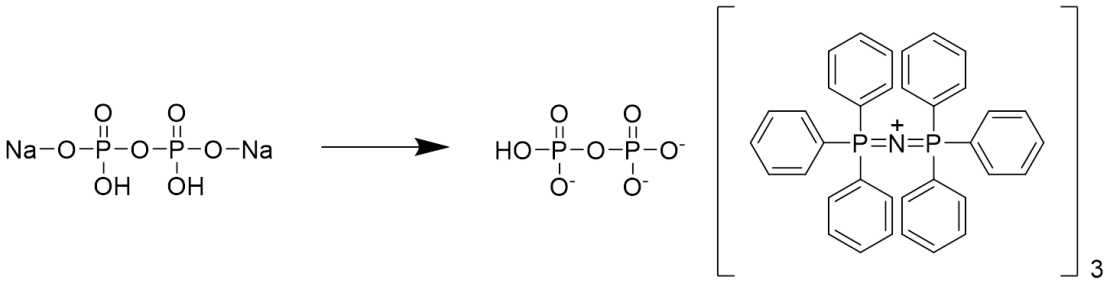
^{1}H NMR (500 MHz, D_2O) δ 7.82 (s, 1H, H-6), 6.08 (app dd, *J* = 11.4, 5.7 Hz, 1H, H-1′), 4.58 (app t, *J* = 4.7 Hz, 1H, H-3′), 4.34 - 4.20 (m, 4H, H-2′, H-4′ and H-5′), 3.92 (ddd, *J* = 11.0, 5.7, 3.2 Hz, 1H, H-8a), 3.85 (ddd, *J* = 11.0, 7.1, 3.0 Hz, 1H, H-8b), 3.50 (ddd, *J*=13.9, 7.1, 3.2 Hz, 1H, H-9a), 3.43 (ddd, *J* = 13.9, 5.7, 3.0 Hz, 1H, H-9b), 3.11 (q, *J* = 7.3 Hz, 24H, $Et_2NH^+C\boldsymbol{H_2}CH_3$), 1.94 (s, 3H, H-7), 1.19 (t, *J* = 7.3 Hz, 36H, $Et_2NH^+CH_2C\boldsymbol{H_3}$)

^{13}C NMR (126 MHz, D_2O) δ 137.0 (C6), 86.4 (C1′), 83.7 (C2′ or C4′), 81.1 (C2′ or C4′), 69.4 (C8 or C9), 68.1 (C3′), 64.8 (C5′), 50.2 (C8 or C9), 46.6 ($Et_2NH^+\boldsymbol{C}H_2CH_3$), 11.6 (C7), 8.2 ($Et_2NH^+CH_2\boldsymbol{C}H_3$)

^{31}P NMR (202 MHz, D_2O) δ -10.5 (d, *J*=18.5 Hz, P_γ), -10.7 (d, *J*=18.5 Hz, P_α), -21.8 (t, *J*=18.5 Hz, P_β)

HRMS: [ESI^-, H_2O] Calculated mass: 567.0169; found *m/z*: 566.0056 ($[M-H]^-$)

7.4.4 Tris(bis(triphenylphosphoranylidene)ammonium) pyrophosphate (PPN pyrophosphate)[386] (4)



Tetrasodium pyrophosphate decahydrate (0.79 g, 1.77 mmol, 1eq) was dissolved in deionised water (25 mL) at 50 ℃ (in an oil bath) and the pH of the solution was adjusted to pH 5.0 with 1 M aqueous hydrochloric acid. This solution was added dropwise, with mixing, to a solution of *bis*(triphenylphosphoranylidene)ammonium chloride (2.00 g, 3.48 mmol, 2 eq) in deionised water (260 mL) at 50 ℃. The mixture was transferred to six 50 mL centrifuge tubes and allowed to cool to 25 ℃, resulting in the formation of a crystalline solid and solution then became crystalised. Samples were centrifuged at 8,500 rpm for 30 min at 18 ℃. The supernatant was removed and the solid residues were combined, re-crystallised from distilled water (45 mL) at 80 ℃ and collected by centrifugation at 8,500 rpm for 30 min. The supernatant was discarded and the pellets were dried, ground to a powder and further dried in a vacuum desiccator over phosphorus pentoxide for 2 days. PPN pyrophosphate **4** was obtained as a white powder. Yield: 78% (2.48 g, 1.38 mmol). M_w = 1790.7 g/mol. (Chemical Formula: $C_{108}H_{91}N_3O_7P_8$)

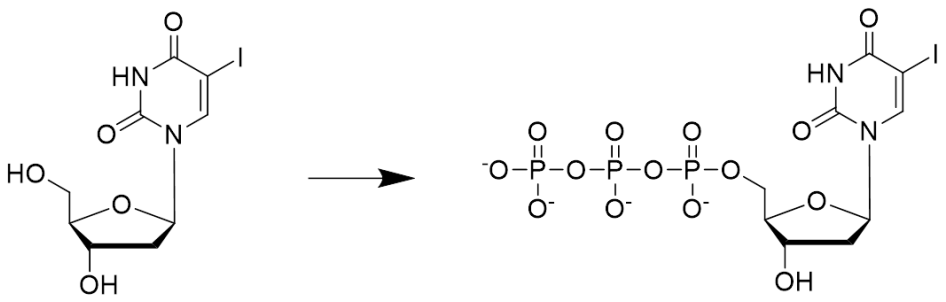
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.74 – 7.51 (m, 30H, H-Ar).

^{13}C NMR (101 MHz, DMSO-d_6) δ 133.7, 132.3 – 131.8 (m), 129.8 – 129.3 (m), 127.3, 126.3.

^{31}P NMR (162 MHz, DMSO-d_6) δ 20.6 (s, 6P, PPh_3), -5.5 (s, 2P,PPi).

LRMS: [ESI^+, MeOH] calculated mass: 1790.7; found *m/z*: 538.2 ($[PPN]^+$, 100) and 539.2 ($[PPN+H]^+$, 39)

7.4.5 5-Iodo-2′-deoxyuridine-5′-*O*-triphosphate (5-I-dUTP)[386] (5.1)

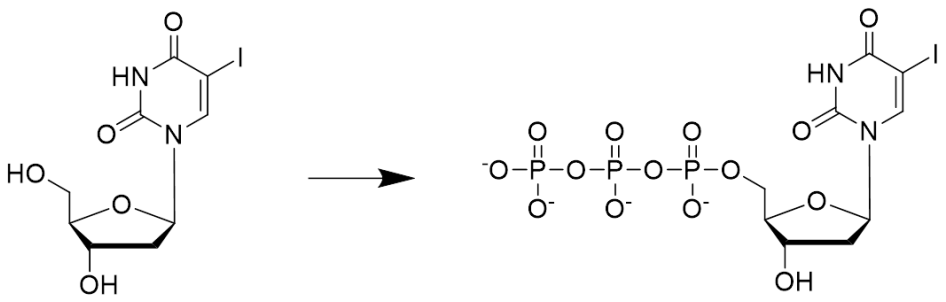


Before commencing the procedure, the 5-Iodo-2′-deoxyuridine, 1,8-bis(dimethylamino)naphthalene (proton-sponge) and PPN pyrophosphate **4** were dried over potassium hydroxide under high vacuum and triethylphosphate was dried over molecular sieves overnight. Triethylphosphate (1.4 mL) was then then added with proton-sponge (90 mg, 0.42 mmol, 1.5 eq) to 5-Iodo-2′-deoxyuridine (100 mg, 0.28 mmol, 1eq) under argon at -10 ℃. Phosphorus oxychloride (31.8 µL, 0.34 mmol, 1.2 eq) was then added dropwise with stirring. After 1 h, PPN pyrophosphate **4** (2.51 g, 1.4 mmol, 5 eq) in anhydrous acetonitrile (3.5 mL) was added with dry n-trioctylamine (0.25 mL, 0.56 mmol, 2 eq). The reaction mixture was then stirred at 0 ℃ for a further 4 h before the reaction was quenched with 1 M TEAB buffer (5 mL, pH 7.5) and the solvent removed *in vacuo*. The crude reaction mixture was dissolved in deionised water (3 mL) and mixed with 3 M sodium chloride solution (300 µL) and ethanol (9 mL) to precipitate the triphosphate at -20 ℃. In order to increase product yield, this precipitation procedure was repeated three times on the supernatant (the solvent of the supernatant was first removed *in vacuo* and re-dissolved in 3 mL deionised water every time). After performing the precipitation three times, no further precipitate was observed. The resulting precipitated pellets were dissolved in deionised water (1 mL). Finally, the crude product was purified by an anion exchange chromatography.

(Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 5–60% buffer B over 25 min; flow rate: 6 mL/min) and further purification by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 3.5–35% buffer B over 35 min; flow rate: 8 mL/min). UV absorbance of 5-I-dUTP **5.1** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 10% (231 OD_{287}, 28 µmol, ε_{287} = 8,275 $M^{-1}cm^{-1}$). M_w = 594.0 g/mol (Chemical Formula: $C_9H_{14}IN_2O_{14}P_3$)

HRMS: [ESI^-, H_2O] Calculated mass: 593.8703 *m/z*: 592.8554 ($[M-H]^-$)

7.4.6 5-Iodo-2′-deoxyuridine-5′-*O*-triphosphate (5-I-dUTP)[241] (5.2)

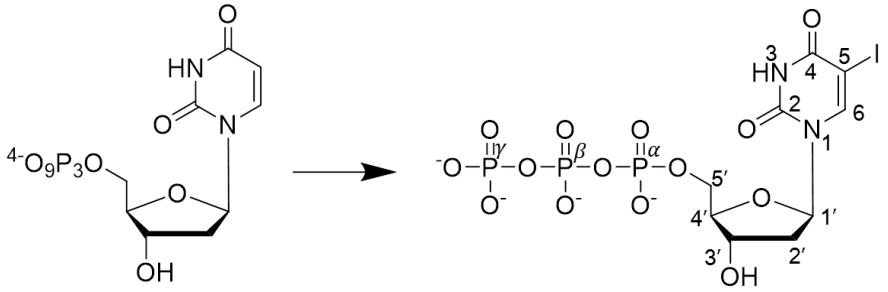


5-Iodo-2′-deoxyuridine (100 mg, 0.28 mmol, 1 eq) was dried over potassium hydroxide under high vacuum and dissolved in anhydrous *N*,*N*-dimethylformamide (4 mL) under argon. 2-Chloro-1,3,2-benzodioxaphosphorin-4-one (85 mg, 0.42 mmol, 1.5 eq) in anhydrous dioxane (1 mL) was injected into the well-stirred solution. After 1 h at 25 ℃, PPN pyrophosphate **4** (2.51 g, 1.40 mmol, 5 eq) in anhydrous *N*,*N*-dimethylformamide (0.81 mL) and n-tributylamine (0.27 mL, 1.12 mmol, 4 eq) were added to the reaction mixture. After 3 h at 25 ℃, a solution of 1% (w/v) iodine in pyridine/water (10 mL, 98:2, v/v) was added. After stirring of the sample for 20 min, excess iodine was quenched by adding a 5% (v/v) aq. $NaHSO_3$ solution (3 mL) and the reaction solution was evaporated to dryness. The crude reaction mixture was dissolved in deionised water (3 mL) and mixed with 3 M sodium chloride solution (300 µL) and ethanol (9 mL) to precipitate the triphosphate at -20 ℃. In order to increase product yield, this precipitation procedure was repeated three times on the supernatant (the solvent of the supernatant was first removed *in vacuo* and re-dissolved in 3 mL deionised water every time). After performing the precipitation three times, no further precipitate was observed. The resulting precipitated pellets were dissolved in deionised water (1 mL). Finally, the crude product was purified by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1

M TEAB, pH 7.5; gradient: 5–60% buffer B over 25 min; flow rate: 6 mL/min) and further purification by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 3.5–35% buffer B over 35 min; flow rate: 8 mL/min). UV absorbance of 5-I-dUTP **5.2** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 23% (533 OD_{287}, 64.4 µmol, ε_{287} = 8,275 $M^{-1}cm^{-1}$). M_w = 594.0 g/mol (Chemical Formula: $C_9H_{14}IN_2O_{14}P_3$)

HRMS: [ESI^-, H_2O] Calculated mass: 593.8703; found *m/z*: 592.8654 ($[M-H]^-$)

7.4.7 5-Iodo-2′-deoxyuridine-5′-*O*-triphosphate (5-I-dUTP)[478] (5.3)



To 2′-Deoxyuridine-5′-*O*-triphosphate (100 mg, 214 µmol, 1 eq) in deionised water (1 mL), *N*-iodosuccinimide (144 mg, 0.64 mmol, 3 eq), and sodium azide (83 mg, 1.28 mmol, 6 eq) were added and the reaction mixture was stirred at 25 ℃ for 15 h. 20 mL of deionised water was then added and the mixture was centrifuged at 9,500 rpm for 30 min. The supernatant was collected, adjusted to pH 6.5 and subjected to purification by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 5–60% buffer B over 25 min; flow rate: 6 mL/min). UV absorbance of 5-I-dUTP **5.3** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 70% (1248 OD_{287}, 150.8 µmol, ε_{287} = 8,275 $M^{-1}cm^{-1}$). M_w = 594.0 g/mol (Chemical Formula: $C_9H_{14}IN_2O_{14}P_3$)

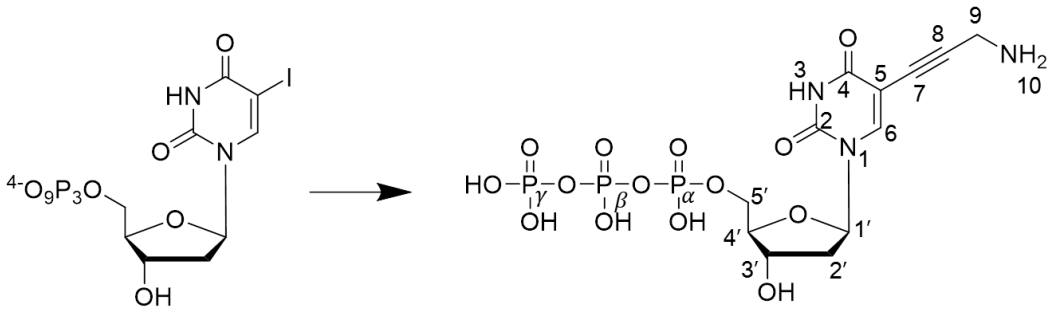
^{1}H NMR (500 MHz, D_2O) δ 8.27 (s, 1H, H-6), 6.28 (app t, *J* = 6.9 Hz, 1H, H-1′), 4.66 (m, 1H, H-3′), 4.27 – 4.15 (m, 3H, H-5′ and H-4′), 3.20 (q, *J* = 7.4 Hz, 24H, $Et_2NH^+C\boldsymbol{H_2}CH_3$), 2.39 (app dd, *J* = 6.9, 4.9 Hz , 2H, H-2′), 1.28 (t, *J* = 7.4 Hz, 36H, $Et_2NH^+CH_2C\boldsymbol{H_3}$).

^{13}C NMR (126 MHz, D_2O) δ 151.5 (C=O), 146.2 (C6), 85.7 (C4′), 85.5 (C1′), 70.6 (C3′), 65.4 (C5′), 46.6 ($Et_2NH^+\boldsymbol{C}H_2CH_3$), 38.7 (C2′), 7.4 ($Et_2NH^+CH_2\boldsymbol{C}H_3$).

^{31}P NMR (202 MHz, D_2O) δ -9.3 (m, P_γ), -11.6 (d, *J* = 19.7 Hz, P_α), -22.8.(m, P_β)

HRMS: [ESI$^-$, H_2O] Calculated mass: 593.8703; found *m/z*: 592.8604 ([M-H]$^-$)

7.4.8 5-(3-Aminoprop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PA-dUTP)[478] (6)



The freeze-dried 5-I-dUTP **5.3** (1248 OD_{287}, 150.8 µmol, 1 eq) was dissolved in dry *N*,*N*-dimethylformamide (3 mL). Palladium tetrakis triphenylphosphine (17 mg, 15 µmol, 0.1 eq), copper iodide (6 mg, 33 µmol, 0.22 eq) and propargylamine (25 mg, 29 µL, 452 µmol, 3 eq) were added. The mixture was degassed. Separately degassed triethylamine (76 mg, 105 µL, 754 µmol, 5 eq) was then added and the reaction mixture was stirred at 25 ℃ for 6 h under argon. 20 mL of water was added to the mixture and stirred for 1 h. The reaction mixture was then centrifuged at 8,500 rpm for 30 min and the supernatant was filtered through a 0.45 µM filter. The filtrate was collected, adjusted to pH 6.5 and purified by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 0–60% buffer B over 25 min; flow rate: 6 mL/min). UV absorbance of 5-PA-dUTP **6** at 289 nm was measured. 5-PA-dUTP **6** was then aliquoted into small vial for labelling reaction and obtained as a TEA salt after freeze-drying. Yield: 82% (1311 OD_{289}, 123.7 µmol, ε_{289} = 10,600 $M^{-1}cm^{-1}$). M_w = 521.20 g/mol (Chemical Formula: $C_{12}H_{18}N_3O_{14}P_3$)

R_f: 0.1 [n-propanol/ammonium hydroxide/H_2O 3:1:1 *v/v/v*]

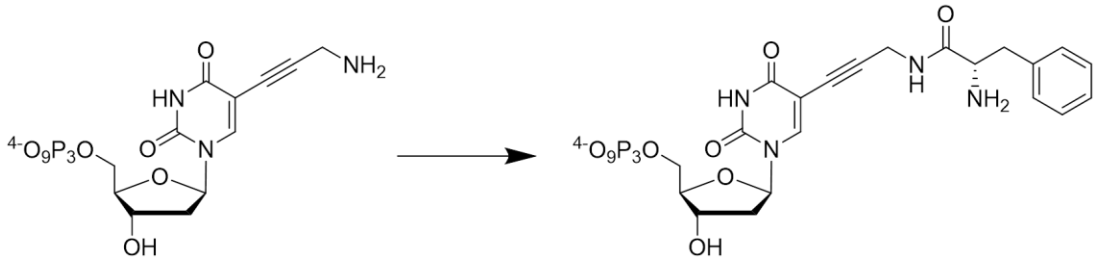
^{1}H NMR (400 MHz, D_2O) δ 8.35 (s, 1H, H-6), 6.29 (app t, *J* = 6.4 Hz, 1H, H-1′), 4.65 (app dt, *J* = 6.4, 5.5, 4.4 Hz, 1H, H-3′), 4.25 (m, 2H, H-5′), 4.14 (app dt, *J* = 4.4, 2.5 Hz, 1H, H-4′), 3.92 (s, 2H, H-9), 3.17 (q, *J* = 7.4 Hz, 24H, $Et_2NH^+C\boldsymbol{H_2}CH_3$), 2.45 – 2.29 (m, 2H, H-2′), 1.26 (t, *J* = 7.4 Hz, 36H, $Et_2NH^+CH_2C\boldsymbol{H_3}$).

^{13}C NMR (101 MHz, D_2O) δ 145.6 (C6), 98.4 (C≡C), 85.5 (C4′), 85.5 (C1′), 69.4 (C3′), 64.6 (C5′), 46.6 ($Et_2NH^+\boldsymbol{C}H_2CH_3$), 39.5 (C2′), 29.80 (C9), 8.2 ($Et_2NH^+CH_2\boldsymbol{C}H_3$).

^{31}P NMR (162 MHz, D_2O) δ -6.31 (d, *J* = 20 Hz, P_γ), -11.37 (d, *J* = 20 Hz, P_α), -22.20 (app t, *J* = 20 Hz, P_β).

HRMS: [ESI^-, H_2O] Calculated mass: 521.0002; found *m/z*: 519.9929 ($[M-H]^-$)

7.4.9 -(3-((2-Amino-3-phenyl-N-)propanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PAF-dUTP)[391] (7)

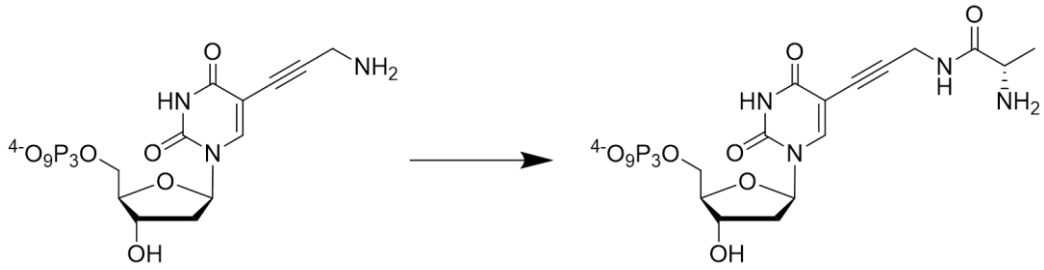


The freeze-dried 5-PA-dUTP **6** (15.9 OD_{298}, 1.5 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). This solution was then added to a solution of 1-hydroxybenzotriazole hydrate (1 mg, 7.5 µmol, 5 eq) and Fmoc-L-phenylalanine pentafluorophenyl ester (4.2 mg, 7.5 µmol, 5 eq) in dimethylformamide (100 µL). The reaction mixture was shaken at 45 ℃ for 3 h. Deionised water (1 mL) was then added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate the excess reagent. The supernatant solution was freeze-dried and the crude product was dissolved in deionised water and purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 35 min; flow rate: 8 mL/min) to afford Fmoc-protected 5-PAF-dUTP. The Fmoc protecting group was cleaved by treatment with concentrated aqueous ammonia solution at 55 ℃ for 2 h. Ammonium hydroxide was then removed *in vacuo*. The residue was dissolved in deionised water (1 mL) and purified by RP-HPLC (Phenomenex column Luna 10µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–80% buffer B over 20 min; flow rate: 4 mL/min). UV absorbance of 5-PAF-dUTP **7** was measured and this compound was freeze-dried

and obtained as a TEA salt. Yield: 53% (8.6 OD_{289}, 0.8 µmol, ε_{289} = 10,800). M_w = 668.38 g/mol (Chemical Formula: $C_{21}H_{27}N_4O_{15}P_3$).

HRMS: [ESI^-, H_2O] Calculated mass: 668.0686; found *m/z*: 667.0383 ($[M-H]^-$)

7.4.10 5-(3-(2-Aminopropanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PAA-dUTP) (8)

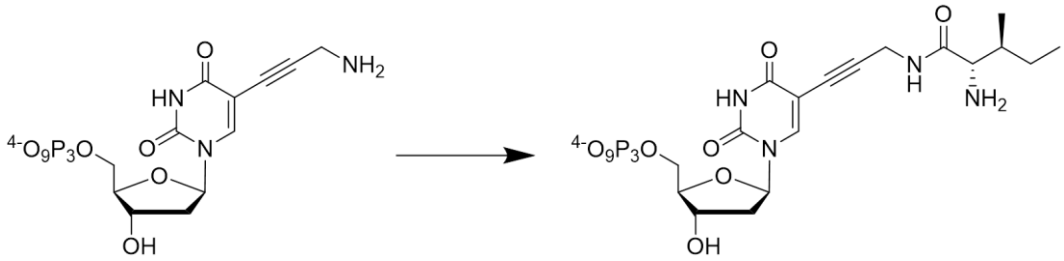


The freeze-dried 5-PA-dUTP **6** (15.9 OD_{298}, 1.5 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). This solution was then added to a solution of 1-hydroxybenzotriazole hydrate (1 mg, 7.5 µmol, 5 eq) and Fmoc-L-alanine pentafluorophenyl ester (4 mg, 7.5 µmol, 5 eq) dissolved in dimethylformamide (100 µL). The reaction mixture was shaken at 45 ℃ for 3 h. Deionised water (1 mL) was then added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate the excess reagent. The supernatant solution was freeze-dried and the crude product was dissolved in deionised water and purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 35 min; flow rate: 8 mL/min) to afford Fmoc protected 5-PAA-dUTP. The Fmoc group was cleaved by treatment with concentrated aqueous ammonia solution at 55 ℃ for 2 h. Ammonium hydroxide was then removed *in vacuo*. The residue was dissolved in deionised water (1 mL) and purified by RP-HPLC (Phenomenex column Luna 10µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–80% buffer B over 20 min; flow rate: 4 mL/min). UV absorbance of 5-PAA-dUTP **8** was measured and this compound was freeze-dried

and obtained as a TEA salt. Yield: 56% (8.9 OD_{289}, 0.84 µmol). M_w = 592.3 g/mol (Chemical Formula: $C_{15}H_{23}N_4O_{15}P_3$).

HRMS: [ESI^-, H_2O] Calculated mass: 592.0373; found *m/z*: 591.0639 ($[M-H]^-$)

7.4.11 5-(3-((2-Amino-3-methyl)pentanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PAI-dUTP) (9)

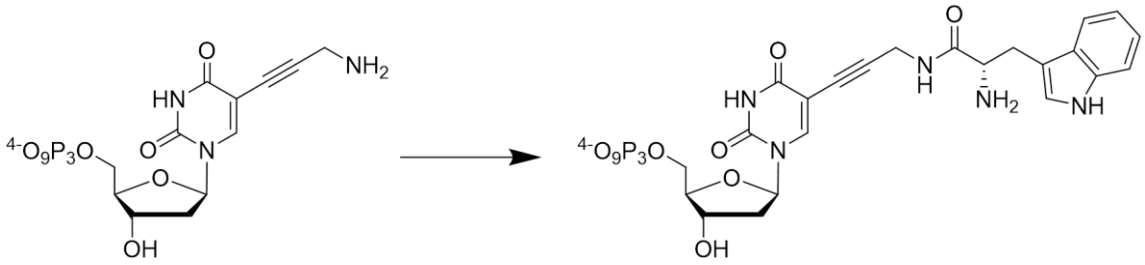


The freeze-dried 5-PA-dUTP **6** (15.9 OD_{298}, 1.5 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). This solution was then added to a solution of 1-hydroxybenzotriazole hydrate (1 mg, 7.5 µmol, 5 eq) and Fmoc-L-isoleucine pentafluorophenyl ester (4 mg, 7.5 µmol, 5 eq) in dimethylformamide (100 µL). The reaction mixture was shaken at 45 ℃ for 3 h. Deionised water (1 mL) was then added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate the excess reagent. The supernatant solution was freeze-dried and the crude product was dissolved in deionised water and purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 35 min; flow rate: 8 mL/min) to afford Fmoc protected 5-PAI-dUTP. The Fmoc group was cleaved by treatment with concentrated aqueous ammonia solution at 55 ℃ for 2 h. Ammonium hydroxide was then removed *in vacuo*. The residue was dissolved in deionised water (1 mL) and purified by RP-HPLC (Phenomenex column Luna 10µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–80% buffer B over 20 min; flow rate: 4 mL/min). UV absorbance of 5-PAI-dUTP **9** was measured and this compound was freeze-dried

and obtained as a TEA salt. Yield: 42% (6.7 OD_{289}, 0.63 µmol). M_w = 634.4 g/mol (Chemical Formula: $C_{18}H_{29}N_4O_{15}P_3$).

HRMS: [ESI$^-$, H_2O] Calculated mass: 634.0842; found *m/z*: 633.1164 ([M-H]$^-$)

7.4.12 5-(3-((2-Amino-3-1*H*-indole-N-)propanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PAW-dUTP) (10)

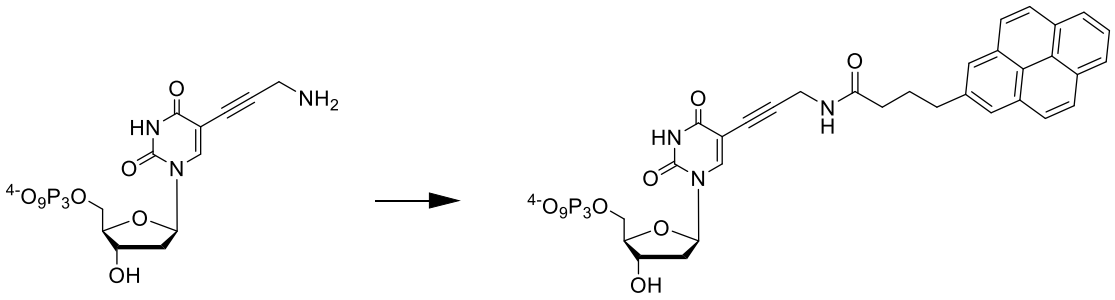


The freeze-dried 5-PA-dUTP **6** (15.9 OD_{298}, 1.5 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). This solution was then added to a solution of 1-hydroxybenzotriazole hydrate (1 mg, 7.5 µmol, 5 eq) and Fmoc-L-tryptophan pentafluorophenyl ester (4 mg, 7.5 µmol, 5 eq) in dimethylformamide (100 µL). The reaction mixture was shaken at 45 ℃ for 3 h. Deionised water (1 mL) was then added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate the excess reagent. The supernatant solution was freeze-dried and the crude product was dissolved in deionised water and purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 35 min; flow rate: 8 mL/min) to afford Fmoc protected 5-PAW-dUTP. The Fmoc group was cleaved by treatment with concentrated aqueous ammonia solution at 55 ℃ for 2 h. Ammonium hydroxide was then removed *in vacuo*. The residue was dissolved in deionised water (1 mL) and purified by RP-HPLC (Phenomenex column Luna 10µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–80% buffer B over 20 min; flow rate: 4 mL/min). UV absorbance of 5-PAW-dUTP **10** was measured and this compound was freeze-dried

and obtained as a TEA salt. Yield: 55% (13.4 OD_{289}, 0.83 µmol, ε_{289} = 16,100). M_w = 707.4 g/mol (Chemical Formula: $C_{23}H_{28}N_5O_{15}P_3$).

HRMS: [ESI$^-$, H_2O] calculated mass: 707.0795; found *m/z*: 706.1179 ([M-H]$^-$)

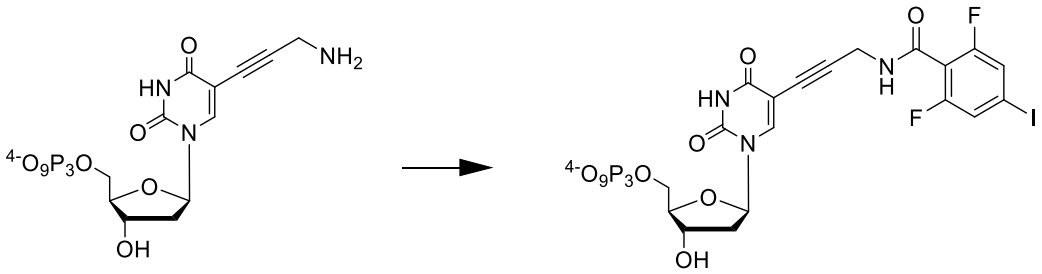
7.4.13 5-(3-(1-Pyrenylbutanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PAPBA-dUTP) (11)



The freeze-dried 5-PA-dUTP **6** (9.5 OD_{298}, 0.9 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). A solution of HATU (35 mg, 90 µmol, 100 eq), dry *N*,*N*-diisopropylethylamine (35 mg, 47 µL, 0.27 mmol, 300 eq) and pyrenebutyric acid (13 mg, 45 µmol, 50 eq) in dimethylformamide (1 mL) was pre-activated at 37 ℃ for 15 min. The pre-activated solution (200 µL) was then added to the 5-PA-dUTP in buffer and the reaction mixture was stirred at 45 ℃ for 3 h. Deionised water (1 mL) was then added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate the excess reagent. The supernatant solution was freeze-dried and the crude product was dissolved in deionised water and purified by RP-HPLC (Phenomenex column Luna 10µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 20 min; flow rate: 4 mL/min). UV absorbance of 5-PAPBA-dUTP **11** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 78% (7.4 OD_{289}, 0.70 µmol). M_w = 791.5 g/mol (Chemical Formula: $C_{32}H_{32}N_3O_{15}P_3$).

HRMS: [ESI$^-$, H_2O] Calculated mass: 791.1046; found *m/z*: 790.1417 ([M-H]$^-$)

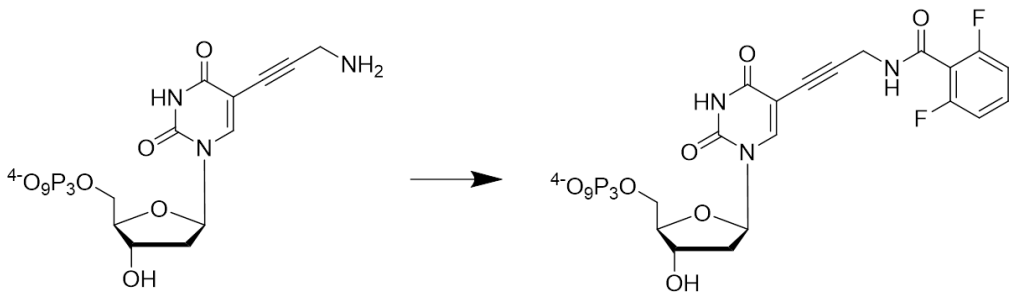
7.4.14 5-(3-((2,6-Difluoro-4-iodobenzyl)acetamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-FFI-dUTP) (12)



The freeze-dried 5-PA-dUTP **6** (19 OD_{289}, 1.8 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). This was then added to a solution of HATU (7 mg, 18 µmol, 10 eq), *N*,*N*-diisopropylethylamine (7 mg, 9.4 µL, 54 µmol, 30 eq) and 2,6-difluoro-4-iodobenzoic acid (5.1 mg, 18 µmol, 10 eq) dissolved in dimethylformamide (450 µL). The reaction mixture was shaken at 25 ℃ for 15 h. Deionised water (1 mL) was added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate excess reagents. The supernatant solution was freeze-dried in excess of deionised water and the crude product was dissolved in deionised water and purified by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 0–80% buffer B over 25 min; flow rate: 6 mL/min). Then, the product was further purified by RP-HPLC (Phenomenex column Luna 10 µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 20 min; flow rate: 4 mL/min). After freeze-drying, 5-FFI-dUTP **12** was obtained as a TEA salt. Yield: 48% (9.2 OD_{289}, 0.86 µmol). M_w = 787.2 g/mol (Chemical Formula: $C_{19}H_{19}F_2IN_3O_{15}P_3$).

HRMS: [ESI$^-$, H_2O] Calculated mass: 786.9042; found *m/z*: 785.9017 ($[M-H]^-$)

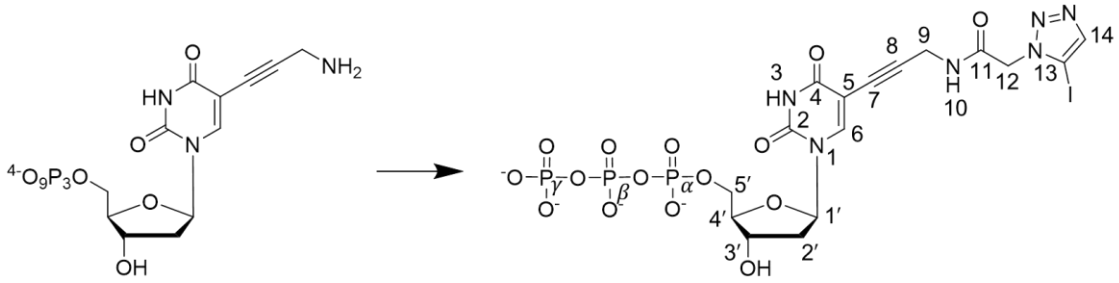
7.4.15 5-(3-((2,6-Difluorobenzyl)acetamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-FF-dUTP) (13)



The freeze-dried 5-PA-dUTP **6** (19 OD_{289}, 1.8 µmol, 1 eq) was dissolved in 1 M TEAB buffer (50 µL, pH 8.5). This was then added to a solution of HATU (7 mg, 18 µmol, 10 eq), *N*,*N*-diisopropylethylamine (7 mg, 9.4 µL, 54 µmol, 30 eq) and 2,6-difluorobenzoic acid (3 mg, 18 µmol, 10 eq) dissolved in dimethylformamide (450 µL). The reaction mixture was shaken at 25 ℃ for 15 h. Deionised water (1 mL) was added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate excess reagents. The supernatant solution was freeze-dried in excess of deionised water and the crude product was dissolved in deionised water and purified by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 0–80% buffer B over 25 min; flow rate: 6 mL/min). Then, the product was further purified by RP-HPLC (Phenomenex column Luna 10µm C8 100Å , 10 × 250 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 20 min; flow rate: 4 mL/min). After freeze-drying, 5-FF-dUTP **13** was obtained as a TEA salt. Yield: 68% (12.7 OD_{289}, 1.2 µmol). M_w = 661.3 g/mol (Chemical Formula: $C_{19}H_{20}F_2N_3O_{15}P_3$).

HRMS: [ESI^-, H_2O] Calculated mass: 661.0075; found *m/z*: 660.0548 ($[M-H]^-$) and 698.0040 ($[M+K-2H]^-$)

7.4.16 5-(3-((2-(5-Iodo-1*H*-1,2,3-triazole))ethanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-ITZ-dUTP) (14)



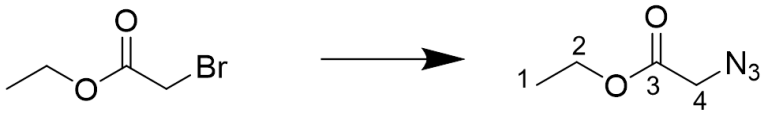
The freeze-dried 5-PA-dUTP **6** (742 OD_{289}, 70 µmol, 1 eq) was dissolved in 1 M TEAB buffer (1 mL, pH 8.5). This was then added to a solution of HATU (266 mg, 0.7 mmol, 10 eq), *N*,*N*-diisopropylethylamine (DIEA) (271 mg, 366 µL, 2.1 mmol, 30 eq) and ITZ-acid **17** (177 mg, 0.7 mmol, 10 eq) dissolved in dimethylformamide (5 mL). The reaction mixture was stirred at 25 ℃ for 15 h. Deionised water (20 mL) was then added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate excess reagents. The supernatant solution was freeze-dried in excess of deionised water and the crude product was dissolved in deionised water and purified by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 0–80% buffer B over 25 min; flow rate: 6 mL/min). Then, the product was further purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–100% buffer B over 35 min; flow rate: 8 mL/min). UV absorbance of 5-ITZ-dUTP **14** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 68% (511 OD_{289}, 47.6 µmol, ε_{289} = 10,730). M_w = 756.2 g/mol (Chemical Formula: $C_{16}H_{20}IN_6O_{15}P_3$).

^{1}H NMR (500 MHz, D_2O) δ 8.15 (s, 1H, H-6), 7.83 (s, 1H, H-14), 6.21 (app t, *J* = 6.7 Hz, 1H, H-1′), 5.32 (s, 2H, H-12), 4.57 (app dt, *J* = 6.8, 3.7 Hz, 1H, H-3′), 4.22 – 4.09 (m, 5H, H-5′, H-4′ and H-9), 3.11 (q, *J* = 7.4 Hz, 18H, $Et_2NH^+C\boldsymbol{H_2}CH_3$), 2.38 – 2.31 (m, 1H, H-2′a), 2.34 – 2.27 (m, 1H, H-2′b), 1.19 (t, *J* = 7.4 Hz, 27H, $Et_2NH^+CH_2C\boldsymbol{H_3}$).

^{13}C NMR (126 MHz, D_2O) δ 167.2 (C11), 164.6 (C4), 150.5 (C2), 145.1 (C6), 141.7 (C14), 99.0 (C≡C), 89.0 (C5), 85.8 (C4′), 85.7 (C1′), 82.0 (C13), 74.1 (C≡C), 70.4 (C3′), 65.20 (C5′), 52.4 (C12), 46.6 ($Et_2NH^+\boldsymbol{C}H_2CH_3$), 39.00 (C2′), 30.0 (C9), 8.2 ($Et_2NH^+CH_2\boldsymbol{C}H_3$).

HRMS: [ESI$^-$, H_2O] Calculated mass: 755.9244; found *m/z*: 754.9156 ($[M-H]^-$)

7.4.17 Ethyl 2-azidoacetate[479] (15)



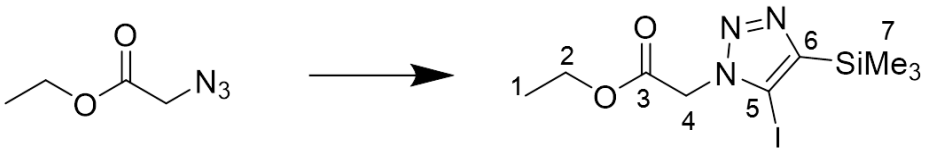
Ethyl 2-bromoacetate (0.60 g, 0.40 mL, 3.59 mmol, 1 eq) was dissolved in dimethyl sulfoxide (2 mL) and sodium azide (0.47 g, 7.18 mmol, 2 eq) was slowly added. The round bound flask was wrapped with aluminum foil to protect the contents from light and the reaction mixture was stirred at 25 ℃ for 72 h under argon. After 72 h, deionsied water (10 mL) was added and product was extracted with diethyl ether (4 × 15 mL). The combined organic extracts were washed twice with brine (20 mL), dried over magnesium sulfate for 20 min, filtered and the filtrate was concentrated on a rotary evaporator at 25 ℃. Ethyl 2-azidoacetate **15** was obtained as brown oil. Yield: 80% (0.37 g, 2.87 mmol). M_w = 129.1 g/mol (Chemical formula: $C_4H_7N_3O_2$)

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.26 (q, *J* = 7.2 Hz, 2H, H-2), 3.86 (s, 2H, H-4), 1.31 (t, *J* = 7.2 Hz, 3H, H-1).

^{13}C NMR (101 MHz, $CDCl_3$) δ 168.4 (C3), 62.0 (C2), 50.5 (C4), 14.3 (C1).

LRMS: [ESI$^+$, MeCN] Calculated mass: 129.1; found *m/z*: 259 ([2M+H]$^+$, 100)

7.4.18 5-Iodo-4-trimethylsilyl-1*H*-1,2,3-triazole-1-acetic acid ethyl ester (ITZ-$SiMe_3$)[394] (16)



Ethyl 2-azidoacetate **15** (0.37 g, 2.87 mmol, 1 eq) was dissolved in dry tetrahydrofuran (9 mL). Sodium iodide (1.72 g, 11.48 mmol, 4 eq) and copper (II) perchlorate hexahydrate (2.13 g, 5.74 mmol, 2 eq) were added. The solution became a dark brown colour. Tris[(1-benzyl-1*H*-1,2,3-triazole-4-yl)methyl]amine (TBTA) (154 mg, 0.29 mmol, 0.1 eq) was then added and the mixture was stirred at 25 ℃ for 5 min. 1,8-Diaabicyclo[5.4.0]undec-7 (DBU) (437 mg, 0.43 mL, 2.87 mmol, 1 eq) and trimethylsilylacetylene (311 mg, 0.44 mL, 3.16 mmol, 1.1 eq) were added. The flask was wrapped with aluminum foil to protect the contents from light and stirred at 25 ℃ for 48 h. The reaction mixture was diluted with dichloromethane (60 mL) and filtered. The filtrate was washed twice with saturated ammonium chloride solution (40 mL) and twice with brine (50 mL). The organic layer was dried over magnesium sulfate for 20 min, filtered and the filtrate was concentrated on a rotary evaporator. The residue was purified by column chromatography on silica (5–30% Ethylacetate/Hexane). ITZ-$SiMe_3$ **16** was obtained as a white solid. Yield: 61% (618 mg, 1.75 mmol). M_w = 353.2 g/mol (Chemical formula: $C_9H_{16}IN_3O_2Si$)

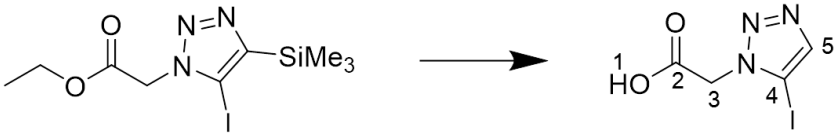
R_f: 0.51 [EtOAc/Hexane 3:7 *v/v*], stain: potassium permanganate

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.20 (s, 2H, H-4), 4.27 (q, *J* = 7.2 Hz, 2H, H-2), 1.29 (t, *J* = 7.2 Hz, 3H, H-1), 0.41 (s, 9H, H-7).

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.8 (C3), 151.5 (C=C), 86.2 (C=C), 62.52 (C2), 51.0 (C4), 14.2 (C1), -1.2 (C7).

LRMS: [ESI^+, MeCN] Calculated mass: 353.0; found *m/z*: 354.2 ($[M+H]^+$, 100) and 376.0 ($[M+Na]^+$, 40)

7.4.19 5-Iodo-1*H*-1,2,3-triazole-1-acetic acid (ITZ-acid) (17)



ITZ-$SiMe_3$ **16** (100 mg, 0.28 mmol) was dissolved in methanol (3 mL) and a solution of 0.7 M sodium hydroxide solution (2 mL) was added. The reaction was stirred at 25 ℃ for 72 h. After 72 h, the solvent was removed on a rotary evaporator and the residual solid was dissolved water (10 mL). The pH of this solution was adjusted to pH 4.0 by careful addition of 1 M aqueous hydrochloric acid. Then, the product was extracted four times with ethyl acetate (15 mL) and the organic layer was washed twice with brine, dried over magnesium sulfate for 20 min, filtered and concentrated on a rotary evaporator. ITZ-acid **17** was obtained as a white salt. Yield: 86% (60.7 mg, 0.24 mmol). M_w = 253.0 g/mol (Chemical Formula: $C_4H_4IN_3O_2$)

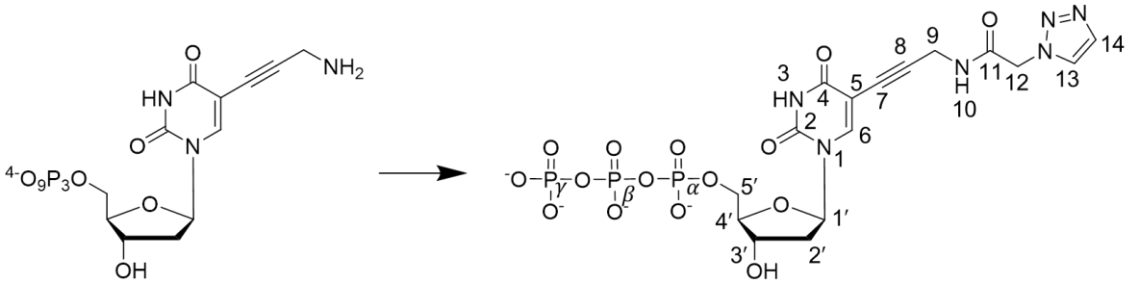
R_f: 0.37 [DCM/MeOH/Acetic acid 10:2:1 *v/v*], stain: anisaldehyde

^{1}H NMR (400 MHz, CD_3OD) δ 7.82 (s, 1H, H-5), 5.29 (s, 2H, H-3).

^{13}C NMR (101 MHz, CD_3OD) δ 169.1 (C2), 142.2 (C=C), 82.5 (C=C), 52.1 (C3).

LRMS: [ESI$^-$, MeOH] Calculated mass: 252.9; found *m/z*: 251.9 ([M-H]$^-$, 100) and 504.9 ([2M-H]$^-$, 50)

7.4.20 5-(3-((2-(1*H*-1,2,3-triazole))ethanamido)prop-1-yn-1-yl)-2′-deoxyuridine-5′-*O*-triphosphate (5-PTZ-dUTP) (18)



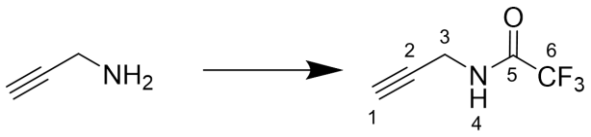
The freeze-dried 5-PA-dUTP **6** (742 OD_{289}, 70 µmol, 1 eq) was dissolved in 1 M TEAB buffer (1 mL, pH 8.5). This was then added to a solution of HATU (266 mg, 0.7 mmol, 10 eq), *N*,*N*-diisopropylethylamine (271 mg, 365.8 µL, 2.1 mmol, 30 eq) and 1*H*-1,2,3-Triazole-1-acetic acid (89 mg, 0.7 mmol, 10 eq) dissolved in dimethylformamide (5 mL). The reaction mixture was stirred at 25 ℃ for 15 h. Deionised water (20 mL) was added and the mixture was vortexed and centrifuged (17,000 rpm, 4 ℃, 30 min) to precipitate excess reagents. The supernatant solution was freeze-dried in excess of deionised water and the crude product was dissolved in deionised water and purified by an anion exchange chromatography (Resource Q; 6 ml column, 16 × 30 mm; buffer A: 10 mM TEAB, pH 7.5; buffer B: 1 M TEAB, pH 7.5; gradient: 0–80% buffer B over 25 min; flow rate: 6 mL/min). Then, the product was further purified by RP-HPLC (Phenomenex column Gemini-NX 10 µm C18 110A AXIA, 250 × 21.2 mm; buffer A: 0.1 M TEAB, pH 7.5; buffer B: 0.1 M TEAB, pH 7.5 with 50% acetonitrile; gradient: 20–80% buffer B over 35 min; flow rate: 8 mL/min). UV absorbance of 5-PTZ-dUTP **18** was measured and this compound was freeze-dried and obtained as a TEA salt. Yield: 57% (430 OD_{289}, 39.9 µmol, ε_{289} = 10,771). M_w = 630.3 g/mol (Chemical Formula: $C_{16}H_{21}N_6O_{15}P_3$).

^{1}H NMR (500 MHz, D_2O) δ 8.17 (s, 1H, H-6), 7.99 (s, 1H, H-14), 7.76 (s, 1H, H-13), 6.21 (app t, *J* = 6.6 Hz, 1H, H-1′), 5.29 (s, 2H, H-12), 4.57 (m, 1H, H-3′), 4.19 – 4.12 (m, 5H, H-5′, H-4′ and H-9), 3.12 (q, *J* = 7.4 Hz, 18H, $Et_2NH^+C\boldsymbol{H_2}CH_3$), 2.38 – 2.27 (m, 2H, H-2′), 1.19 (t, *J* = 7.4 Hz, 27H, $Et_2NH^+CH_2C\boldsymbol{H_3}$).

^{13}C NMR (126 MHz, D_2O) δ 167.8 (C11), 164.6 (C4), 150.5 (C2), 145.0 (C6), 134.0 (C14), 127.1 (C13), 99.0 (C≡C), 89.0 (C5), 85.8 (C4′), 85.7 (C1′), 74.1 (C≡C), 70.4 (C3′), 65.2 (C5′), 51.8 (C12), 46.6 ($Et_2NH^+\boldsymbol{C}H_2CH_3$), 39.1 (C2′), 29.9 (C9), 8.2 ($Et_2NH^+CH_2\boldsymbol{C}H_3$).

HRMS: [ESI$^-$, H_2O] Calculated mass: 630.0278; found *m/z*: 629.0183 ($[M-H]^-$)

7.4.21 2,2,2-Trifluoro-*N*-2-propyn-1-yl-acetamide (TFA-propargylamine)[480] (19)



Ethyl trifluoroacetate (17.9 g, 15 mL, 126 mmol, 1.3 eq) was dissolved in anhydrous methanol (30 mL) and the solution was cooled to 0 ℃ under argon. Propargylamine (5.3 g, 6.2 mL, 96.2 mmol, 1.0 eq) in anhydrous methanol was then added dropwise. The reaction was stirred at 0 ℃ for 3 h. Solvent was removed *in vacuo* and the residue was purified by column chromatography on silica (100% DCM). TFA-propargylamine **19** was obtained as colourless oil. Yield: 63% (12.0 g, 79.4 mmol). M_w = 151.1 g/mol (Chemical Formula: $C_5H_4F_3NO$)

R_f: 0.7 [DCM/MeOH 95:5 *v/v*], stain: ninhydrin and potassium permanganate (the product is volatile, no strong heat).

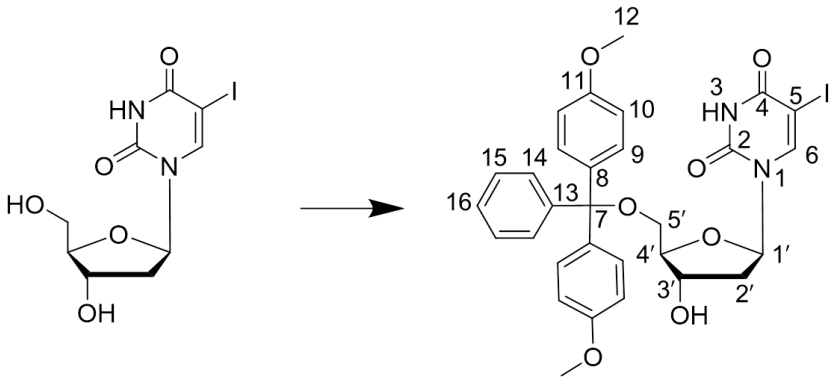
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.24 (br. s, 1H, H-4), 4.12 (dd, *J* = 5.5, 2.6 Hz, 2H, H-3), 2.30 (t, *J* = 2.6 Hz, 1H, H-1).

^{13}C NMR (101 MHz, $CDCl_3$) δ 157.4 (q, $^2J_{C\text{-}F}$= 37.7 Hz, C5), 115.7 (q, $^1J_{C\text{-}F}$ = 287.3 Hz, C6), 77.2 (C2), 73.0 (C1), 29.7 (C3).

^{19}F NMR (376 MHz, $CDCl_3$) δ -76.1 (s)

LRMS: [ESI$^-$, MeCN] Calculated mass: 151.0; found *m/z*: 150.1 ([M-H]$^-$, 100) and 185.9 ([M+Cl]$^-$, 33).

7.4.22 5′-*O*-(4,4′-Dimethoxytrityl)-5-iodo-2′-deoxyuridine (DMT-5-I-dU)[481] (20)



5-Iodo-2′-deoxyuridine (8.7 g, 24.6 mmol, 1 eq) was co-evaporated three times with dry pyridine (3 × 25 mL) and then dissolved in dry pyridine (40 mL). 4-dimethylaminopyridine (79 mg, 0.65 mmol, 0.026 eq) was then added followed by slow addition of 4,4′-dimethoxytritylchloride (9.16 g, 27.1 mmol, 1.1 eq) under argon. The reaction was stirred at 25 ℃ under argon for 3 h. The volume of the reaction mixture was reduced after TLC showed the reaction was complete. The residue was dissolved in dichloromethane (60 mL) and washed with saturated sodium bicarbonate solution (50 mL), dried over sodium sulphate, filtered and the filtrate was evaporated to dryness. The crude product was purified by column chromatography on pre-equilibrated silica gel (90% DCM/EtOAc with 5% toluene and 0.5% TEA, v/v), eluting with 10–50% EtOAc in DCM. DMT-5-I-dU **20** was obtained as a white solid. Yield: 35% (5.61 g, 8.55 mmol). M_w = 656.5 g/mol (Chemical Formula: $C_{30}H_{29}IN_2O_7$)

R_{Prod}: 0.61, R_{IdU}: 0.1, R_{DMT}: 0.94 [DCM/EtOAc 2:3 *v/v*], stain: anisaldehyde

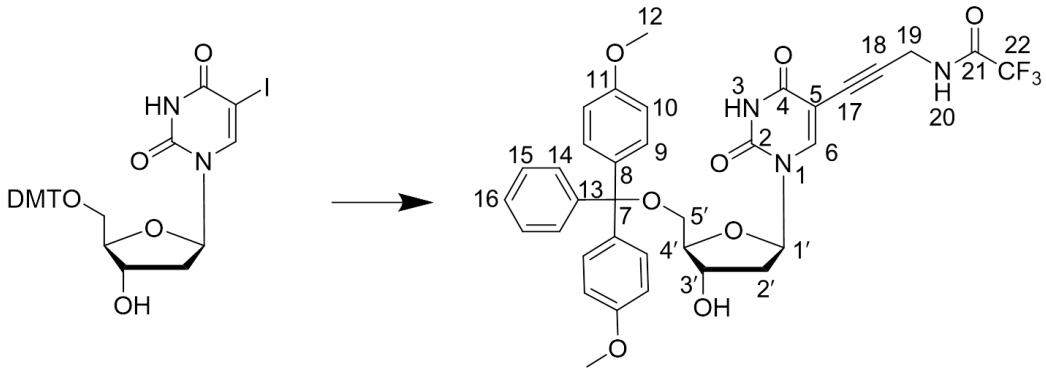
^{1}H NMR (400 MHz, DMSO-d_6) δ 11.73 (s, 1H, H-3), 8.01 (s, 1H, H-6), 7.44 – 7.36 (m, 2H, H-Ar), 7.35 – 7.24 (m, 7H, H-Ar), 6.90 (d, *J* = 8.9 Hz, 4H, H-10), 6.10 (app t, *J* = 7.0, 6.6 Hz, 1H, H-1′), 5.31 (d, *J* = 4.5 Hz, 1H, OH), 4.22 (app dt, *J* = 7.4, 3.5 Hz, 1H, H-3′), 3.90 (app dt, *J* = 4.8, 3.5 Hz, 1H, H-4′), 3.74 (s, 6H, H-12), 3.24 – 3.12 (m, 2H, H-5′), 2.29 – 2.12 (m, 2H, H-2′).

** Extra signals at 7.05–7.20 ppm and 2.30 ppm were also observed owing to the presence of a small amount of residual toluene.*

^{13}C NMR (101 MHz, DMSO-d_6) δ 160.5 (C2), 158.1 (C4), 150.0 (C-Ar), 144.7 (C-Ar), 144.2 (C6), 135.5 (C-Ar), 135.4 (C-Ar), 128.2 (CH-Ar), 127.9 (CH-Ar), 127.6 (C16), 125.3 (C-Ar), 113.3 (C10), 85.8 (C4′), 84.9 (C1′), 70.5 (C3′), 69.8 (C5), 63.7 (C5′), 55.1 (C12)

** Extra signals at 137.3 ppm, 129.7 ppm, 128.9 ppm, 126.7 ppm and 21.1 ppm were also observed owing to the presence of a small amount of residual toluene.*

7.4.23 5′-*O*-(4,4′-Dimethoxytrityl)-5-(3-(2,2,2-trifluoroacetamido)prop-1-yn-1-yl)-2′-deoxyuridine (DMT-5-PA-TFA-dU)[257,357,481] (21)



DMT-5-I-dU **20** (5.10 g, 7.77 mmol, 1 eq), TFA-propargylamine **19** (0.34 g, 2.28 mmol, 3 eq), palladium tetrakis triphenylphosphine (88 mg, 0.08 mmol, 0.1 eq) and copper (II) iodide (31.80 mg, 0.17 mmol, 0.22 eq) were dissolved in dry *N*,*N*-dimethylformamide (2.5 mL). The reaction mixture was degassed under vacuum. Separately degassed trimethylamine (0.36 g, 0.53 mL, 3.8 mmol, 5 eq) was added under argon and the reaction mixture was stirred at 25 ℃ for 6 h at which point TLC analysis confirmed that the reaction was complete. The reaction was reduced in volume and diluted in dichloromethane (50 mL). This solution was washed with 5% $Na_2EDTA_{(aq)}$ and then 5% $Na_2HSO_{3(aq)}$, dried over sodium sulphate, filtered and the filtrate was evaporated to dryness to give a brown oil. The crude product was purified by column chromatography on pre-equilibrated silica gel (DCM with 0.5% TEA), eluting with 0–5% MeOH/DCM. DMT-5-PA-TFA-dU **21** was obtained as a pale yellow solid. Yield: 66% (3.5 g, 5.15 mmol). M_w = 679.6 g/mol (Chemical Formula: $C_{35}H_{32}F_3N_3O_8$)

R_{Prod}: 0.24, R_{SM}: 0.29 [DCM/MeOH 95:5 *v/v*], stain: potassium permanganate

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.11 (s, 1H, H-6), 7.38 – 7.34 (m, 2H, H-Ar), 7.33 – 7.11 (m, 7H, H-Ar), 6.77 (d, *J* = 9.0 Hz, 4H, H-10), 6.30 (app t, *J* = 6.3 Hz, 1H, H-1′), 4.50 (app dt, *J* = 6.3, 3.1 Hz, 1H, H-3′), 4.05 (app q, *J* = 3.1 Hz, 1H, H-4′), 3.87 (dd, *J* = 25.2, 17.8 Hz, 2H, H-19), 3.72 (s, 6H, H-12), 3.36 – 3.24 (m, 2H, H-5′), 2.51 – 2.40 (ddd, *J* = 13.6, 6.3, 3.1 Hz, 1H, H-2′a), 2.23 (dt, *J* = 13.6, 6.3 Hz, 1H, H-2′b)

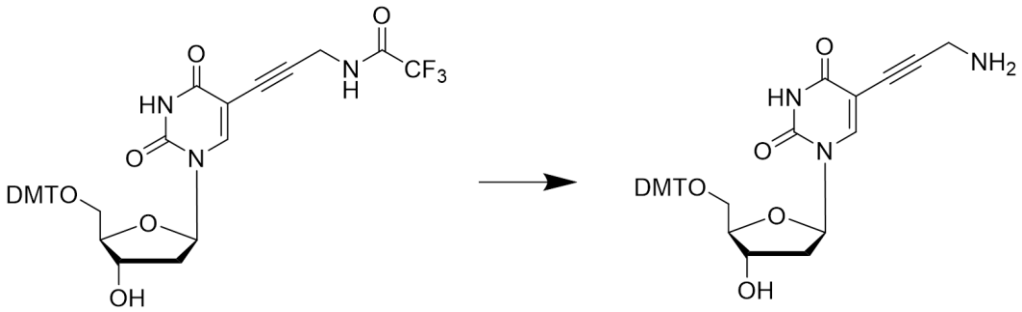
** The spectrum also showed two multiplet peaks at 2.61 ppm and 1.06–0.98 ppm owing to the presence of a triethylammonium salt contaminant.*

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.4 (C=O), 158.7 (C=O), 158.7 (C11), 151.4(C=O), 144.7 (C-Ar), 143.3 (C6), 135.7 (C8), 130.1 (CH-Ar), 130.1 (CH-Ar), 128.2 (CH-Ar), 128.1 (CH-Ar), 127.0 (CH-Ar), 114.3 (C22), 113.4 (C10), 98.9 (C5), 87.0 (C7), 86.7 (C4′), 86.6 (C18), 85.8 (C1′), 76.7 (C≡C), 71.9 (C3′), 63.8 (C5′), 55.3 (C12), 41.8 (C2′), 30.6 (C19)

** Additional peaks at 45.7 ppm and 10.6 ppm were observed, corresponding to a triethylammonium salt contaminant.*

LRMS: [ESI$^-$, MeCN] Calculated mass: 679.2; found *m/z*: 678.2 ([M-H]$^-$, 30) and 717.0 ([M+K-2H]$^-$, 100)

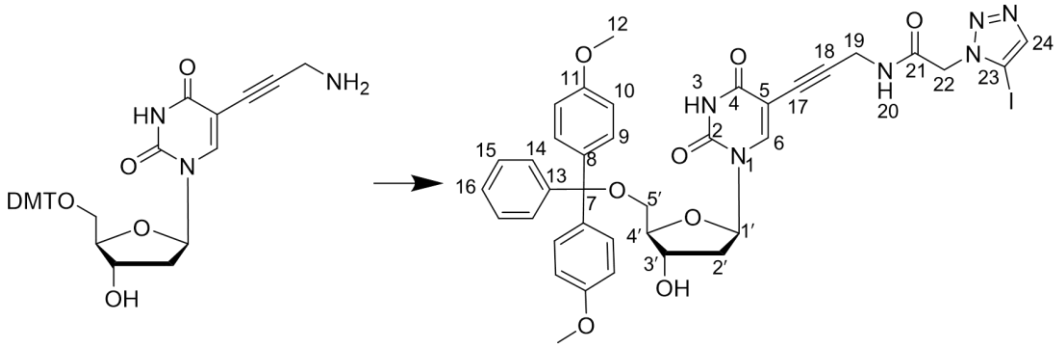
7.4.24 5′-*O*-(4,4′-Dimethoxytrityl)-5-(3-aminoprop-1-yn-1-yl)-2′-deoxyuridine (DMT-5-PA-dU)[357] (22)



DMT-5-PA-TFA-dU **21** (800 mg, 1.18 mmol, 1 eq) was dissolved in anhydrous methanol (15 mL). Concentrated ammonia solution (15 mL) was added and the reaction mixture was stirred at 25 ℃. The progress of the reaction was monitored by TLC, which indicated that deprotection was complete after 3 h. The mixture was then evaporated to dryness. DMT-5-PA-dU **22** was obtained as a pale yellow solid. Yield: 99.9% (690 mg, 1.18 mmol). M_w = 583.6 g/mol (Chemical Formula: $C_{33}H_{33}N_3O_7$)

R_{Prod}: 0.05, R_{SM}: 0.22 [EtOAc/MeOH/$NH_{3(aq)}$ 5:1:1 v/*v*/*v*]

7.4.25 5′-*O*-(4,4′-Dimethoxytrityl)-5-(3-((2-(5-iodo-1*H*-1,2,3-triazole))ethanamido)prop-1-yn-1-yl)-2′-deoxyuridine (DMT-5-ITZ-dU) (23)



ITZ-acid **17** (300 mg, 1.18 mmol, 1 eq) and PyBOP (614 mg, 1.18 mmol, 1 eq) were dried over potassium hydroxide pellets overnight and dissolved in dry dimethylformamide (20 mL). 4-methylmorpholine (NMM) (418 mg, 455 µL, 4.13 mmol, 3.5 eq) was added. DMT-5-PA-dU **22** (690 mg, 1.18 mmol, 1 eq) was dissolved in dry *N*,*N*-dimethylformamide (10 mL) was added to the reaction mixture and stirred at 0 ℃ for 4 h under argon at which point TLC analysis confirmed that the reaction had reached completion. The reaction mixture was reduced in volume and diluted in ethyl acetate (50 mL). Then it was washed with saturated sodium bicarbonate (40 mL) and brine (40 mL), dried over magnesium sulphate, filtered and the filtrate was evaporated to dryness. The crude product was purified by column chromatography on pre-equilibrated silica gel ($CHCl_3$ with 0.5% TEA), eluting with 0–5% MeOH/$CHCl_3$. DMT-5-ITZ-dU **23** was obtained as a pale yellow solid. Yield: 75% (720 mg, 0.88 mmol). M_w = 818.6 g/mol (Chemical Formula: $C_{37}H_{35}IN_6O_8$)

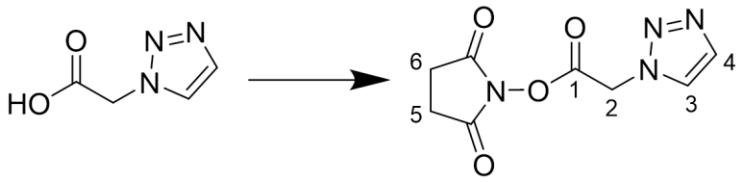
R_{Prod}: 0.21, R_{SM}: 0.14 [EtOAc/MeOH/$NH_{3(aq)}$ 5:1:1 *v/v/v*]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.06 (s, 1H, H-6), 7.63 (s, 1H, H-24), 7.39 – 7.20 (m, 8H, H-Ar), 7.17 – 7.11 (m, 1H, H-Ar), 6.77 (d, *J* = 8.5 Hz, 4H, H-10), 6.54 (t, *J* = 4.7 Hz, 1H, H-1′), 6.26 (m, H-3′), 4.86 (br.s, 2H, H-22), 4.01 (app q, *J* = 3.3 Hz, 1H, H-4′), 3.89 – 3.82 (br.s, 2H, H-19), 3.70 (s, 6H, H-12), 3.31 (app dd, *J* = 10.8, 3.3 Hz, 1H, H-5′a), 3.26 (app dd, *J* = 10.8, 3.3 Hz, 1H, H-5′b), 2.48 – 2.39 (m, 1H, H-2′a), 2.21 (dt, *J* = 13.6, 6.6, 4.7 Hz, 1H, H-2′b)

** The spectrum also showed two multiplet peaks at 2.54 ppm and 0.98 ppm owing to the presence of a triethylammonium salt contaminant.*

LRMS: [ESI$^-$, MeCN] Calculated mass: 818.2; found *m/z*: 817.0 ($[M-H]^-$, 40) and 853.0 ($[M+Cl]^-$, 100)

7.4.26 1*H*-1,2,3-Triazole-1-acetic acid *N*-hydroxysuccinimide ester (PTZ-NHS) (24)

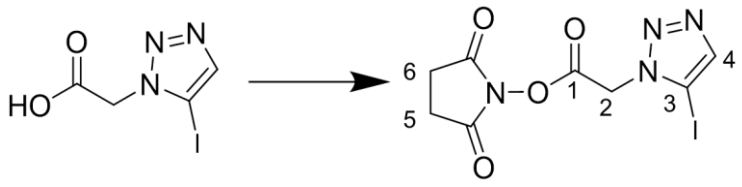


1*H*-1,2,3-Triazole-1-acetic acid (50 mg, 0.39 mmol, 1 eq), purchased from Alfa Chemistry was dissolved in dry dichloromethane (3 mL). DCC (89 mg, 0.43 mmol, 1.1 eq) and *N*-hydroxysuccinimide (68 mg, 0.59 mmol, 1.5 eq) were added to the reaction mixture and stirred at 25 ℃ for 15 h under argon. The reaction mixture was filtered and the precipitate was washed with dry dichloromethane (3 mL). The combined filtrate and washings were evaporated to dryness on a rotary evaporator. The crude PTZ-NHS **24** was obtained as a white crystal solid. The crude product **24** was obtained in quantitative yield and used in the next step without further purification. M_w = 224.2 g/mol (Chemical Formula: $C_8H_8N_4O_4$).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.79 (m, 2H, H-3 and H-4), 5.58 (s, 2H, H-2), 2.96 – 2.77 (m, 4H, H-5 and H-6).

LRMS: [ESI$^+$, MeCN] Calculated mass: 224.1; found *m/z*: 225.2 ([M+H]$^+$, 40)

7.4.27 5-Iodo-1*H*-1,2,3-triazole-1-acetic acid *N*-hydroxysuccinimide ester (ITZ-NHS) (25)



ITZ-acid **17** (40 mg, 0.158 mmol, 1 eq) was dissolved in dry dichloromethane (3 mL). *N*,*N*′-dicyclohexylcarbodiimide (DCC) (36 mg, 0.17 mmol, 1.1 eq) and *N*-hydroxysuccinimide (27 mg, 0.24 mmol, 1.5 eq) were added to the reaction mixture and stirred at 25 ℃ for 15 h under argon. The reaction mixture was filtered and the precipitates were washed with dry dichloromethane (3 mL). The combined filtrate and washings were evaporated to dryness on a rotary evaporator. The crude ITZ-NHS **25** was obtained as a white crystal solid. The crude product **25** was obtained in quantitative yield and used in the next step without further purification. M_w = 350.1 g/mol (Chemical Formula: $C_8H_7IN_4O_4$).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (s, 1H, H-4), 5.57 (s, 2H, H-2), 2.90 – 2.84 (m, 4H, H-5 and H-6).

LRMS: [ESI$^+$, MeCN] Calculated mass: 350.0; found *m/z*: 351.0 ([M+H]$^+$, 20)

7.5 Primer extension experiment protocol

Primer extension assays were performed using 66 pmol of 5′-FAM labelled DNA primer, 132 pmol of either modified or unmodified DNA template, 3.2 nmol of each dATP, dGTP and dCTP, 2 unit (U) of DNA polymerase (Gotaq, Klenow, KOD and KOD XL DNA polymerase) and 1× buffer (supplied with the enzymes) in a 20 μL total reaction volume. KOD XL DNA polymerase was supplemented with additional 1 mM $MnCl_2$. To each reaction, 3.2 nmol of dTTP or modified TTP/dUTP was added as required. Extensions were performed at 65 ℃ for thermostable enzymes (except for Klenow fragment DNA polymerase which is functional at 37 ℃). Various time course experiments were tested. Primer extension reactions were conducted in a BIO-RAD T100^{TM} Thermal Cycler. Samples were then analysed by denaturing PAGE as described in the section 7.2.2.

KOD and KOD XL DNA polymerases were purchased from Merck Millipore; DNA Polymerase I-Large (Klenow) Fragment was purchased from New England Biolabs; Gotaq polymerase was purchased from Promega.

7.5.1 Primer extension results analysis by mass spectrometry

For the mass spectrometry analysis, 4 × 20 μL primer extension reactions were performed. They were then combined (80 μL in total) and DNA was isolated firstly by phenol/chloroform extraction and followed by ethanol precipitation.

Phenol/chloroform extraction: 80 μL of UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v), purchased from Invitrogen was added to 80 μL of the primer

extended products. Enzymes in the organic layer were discarded and only nucleic acids in the aqueous layer were collected.

Ethanol precipitation: 3 M sodium acetate (8 µL, pH 5.3) was added to the reaction mixture followed by ethanol (240 µL). This mixture was then left on dry ice for 10 min, kept at -20 ℃ for a further 20 min and centrifuged (17,000 rpm, 4 ℃, 40 min). The supernatant was then removed and the pelleted sample was dissolved in deionised water (10 µL) and analysed on a Bruker micrOTOF^{TM} II focus ESI$^-$ TOF MS instrument in ES$^-$ mode or XEVO G2-QTOF MS instrument.

7.6 Polymerase chain reaction (PCR) protocol

Polymerase chain reactions (PCR) were performed using 10 pmol of forward and reverse primers, 2 fmol of PCR template, 5 nmol of each dATP, dGTP and dCTP, 0.5 U KOD XL DNA polymerase or Vent exo- DNA polymerase (New England Biolabs) and 1× buffer (supplied with the enzymes, pH 8.0) in a 20 µL total reaction volume. 1 mM $MnCl_2$ was supplemented for KOD XL DNA polymerase. To each reaction, 5 nmol of dTTP or modified TTP/dUTP was added as required. SYBR Green (Sigma, 0.1×) or Evagreen dye (Biotium, 0.8×) was added if the PCR amplification was monitored in real-time or when the PCR products were subsequently used for melting temperature experiments.

Amplification was performed on a real-time PCR instrument (Bio-Rad CFX96) using the following procedure unless otherwise stated. Initial denaturation was performed at 94 ℃ for 3 min. 25 cycles of PCR consist of DNA duplex denaturation step at 94 ℃ for

20 sec, followed by a primer annealing step at 52 ℃ for 20 sec and an extension step at 65 ℃ for 1 min were performed with a final extension at 65 ℃ for 10 min.

The samples were then analysed by 8% denaturing PAGE gel or 2% agarose gel with 20 µL of ethidium bromide (Sigma, 1 µg/mL) or SYBR gold (Thermo Fisher Scientific, 0.5×). The gel was run at constant voltage (126 V) in 1× TBE buffer.

7.6.1 Melting temperature (T_m) experiments protocol

Bio-Rad CFX96: T_m of the PCR product was determined by heating the sample to 95 ℃ for 2 min, cooling to 30 ℃, then increasing the temperature by 0.1 ℃ per sec to 95 ℃. Samples were then cooled and held at 4 ℃.

Note that melting temperature for primer annealing temperature optimisation was prepared by mixing PCR reagents described above except the polymerase.

Lightcycler 1.5: After PCR, T_m curve was acquired by heating the sample to 95 ℃ for 2 min and slowly cooling it at 0.05 ℃ per sec to 30 ℃.

7.7 Single-stranded DNA generation

7.7.1 Lambda(λ) exonuclease digestion protocol

Modified duplex products containing one strand with 5′-phosphate were produced via PCR. The reaction mixtures were extracted using phenol/chloroform and precipitated with ethanol to remove the DNA polymerase. The DNA precipitates were re-dissolved in 1× λ-exonuclease reaction buffer (20 μL) with 5 U of λ-exonuclease and SYBR green (0.1×). The reactions were left at 37 ℃ for 2 h in a real-time PCR instrument (Bio-Rad CFX96) followed by T_m measurement to check the presence of remaining duplex products. If un-digested duplexes existed, an additional 5 U of λ-exonuclease was added and incubated at 37 ℃ for another 1 h. When the digestion was complete, the melting temperature of the digested products was conducted by heating at 95 ℃ for 30 sec, cooling down to 30 ℃ and then increased to 95 ℃ by 1℃ per sec.

7.7.2 Magnetic separation with streptavidin-coated beads

Modified duplex products containing one strand with 5′-biotin were produced via PCR. Dynabeads M-280 Streptavidin (Thermo Fisher Scientific, 40 μL) was washed three times with 2× TE buffer (40 μL, 10 mM Tris HCl, pH 7.5, 1 mM EDTA, 2 M NaCl, according to the manufacturer's recommendation). The PCR products containing 5′-biotin modified strand in buffer (40 μL) were mixed with Dynabeads in 2× TE buffer (40 μL) and gently shaken at 25 ℃ for 30 min. The supernatant was removed (sample a) and the beads were washed three times with 2× TE buffer (40 μL, the third wash was kept as sample b) The beads were re-suspended in 20 mM NaOH (40 μL) and gently shaken at 25 ℃ for 30 min. The supernatant was collected in 40 mM aqueous HCl (30 μL) with 100× TE buffer (1 μL) (sample c). The beads were washed three times with

2× TE buffer (40 µL) and re-suspended in 2× TE buffer (40 µL). The beads were re-suspended in formamide (40 µL) and incubated at 65 ℃ for 5 min (sample d). Samples (a–d) were analysed by 12% native PAGE gel at constant voltage (126 V) in 1× TBE buffer, followed by the gel staining with SYBR Gold (1×) .

7.7.3 Size separation method

For size separation derived from unequal length primers (a reverse primer comprised of an HEG spacer as a terminator and an extension of PolyA or PolyT (20 mer) in conjunction with an FAM-labelled forward primer), PCR products with stands of unequal length (3 mL) were concentrated to 1 mL using 10 kDa molecular weight cut-off (MWCO) Amicon ultracentrifugal filter (Merck Millipore) and analysed by 10% denaturing PAGE at 20 W in 1× TBE buffer for 2 h. The bands of interest were excised from the gel and incubated in deionised water (15 mL) at 37 ℃ for overnight. The extracted DNA was then filtered and the filtrate was desalted using 10 kDa MWCO ultracentrifugal filter and Sephadex illustra NAP columns (GE healthcare).

7.8 SELEX techniques

7.8.1 FAM-labelled ssDNA quantification

FAM-labelled oligonucleotides in each round of selection process were quantified by comparison to calibration curve of FAM-labelled primer (1 nM to 200 nM) in 1× PBS buffer. Fluorescence spectroscopy measurements were performed using a Hellma® fluorescence cuvette (ultra Micro, Sigma-Aldrich) on Luminescence Spectrometer (LB50B, Perkin Elmer) using excitation wavelength of 485 nm and emission wavelength of 535 nm.

7.8.2 Fluorescamine assay for peptide quantification

Each peptide (200 µg) was dissolved in B/W buffer (1 mL, 40 mM HEPE pH 7.5, 125 mM NaCl, 5 mM KCl, 1mM $MgCl_2$ and 1 mM $CaCl_2$).[184] The peptide (150 µL) was then mixed with fluorescamine (50 µL of 4 mg fluorescamine dissolved in 1.5 mL acetone) and the fluorescence intensity measured using a Hellma® fluorescence on a Luminescence Spectrometer with an excitation wavelength of 365 nm and emission wavelength of 470 nm. Unknown samples were compared to a standard curve of peptide **1**. All measurements were replicated triplicates.

7.8.3 Peptide labelling with biotin

Lyophilised peptide (Peptide Protein Reseach ltd, 200 µg) was added to EZ-Link™ Sulfo-NHS-LC-Biotin (Thermo Fisher Scientific, 5 mg) in 1× PBS buffer (Phosphate buffered saline tablet, Thermo Fisher Scientific, 200 µL). The reaction was left at 25 ℃ for 40 min and the biotinylated peptide was purified using a Pierce™ C18 spin column (Thermo Fisher Scientific).

The peptides were analysed by mass spectrometry using a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% TFA. The purified peptide (2 µg in 1 µL 1 × PBS buffer) was mixed with maxtrix solution (1 µL). The mixture (1.6 µL) was then deposited on the sample plate and air-dried. The peptides were analysed using MALDI micro MX - TOF Mass Spectrometer (Waters corporation).

7.8.4 Non-templated control PCR test

Non-templated Q-PCR was performed in a real-time PCR instrument (Bio-Rad CFX96) with a reaction mixture (20 µL) containing Evagreen dye (0.8×), 10 pmol of each primer, 5 nmol of each dATP, dCTP, dGTP and dTTP or 5-propargylamino-iodotriazole-dUTP (5-ITZ-dUTP), 0.5 U DNA polymerase (Vent exo-, Q5 or KOD XL) and 1× buffer (supplied with the enzyme). Additional 2 mM $MgSO_4$ for Vent exo- DNA polymerase mixture or 1 mM $MnCl_2$ for KOD XL DNA polymerase mixture was supplemented.

Non-templated Q-PCR consisted of: 1) initial denaturation at 95 ℃ for 3 min; 2) 30 cycles of denaturation at 95 ℃ for 20 sec, annealing at 58 ℃ for 30 sec and extension at 72 ℃ for 30 sec; 3) final extension at 72 ℃ for 1 min.

7.8.5 PCR amplification and purification for each selection cycle

5′-FAM labelled forward primer and 5′-(dT or dA)n-HEG appended reverse primer and ssDNA that was bound to the SELEX target was used for PCR. 1.5 nmol of each primer, 0.75 µmol of each dATP, dGTP, dCTP and dTTP or 5-ITZ-dUTP, 60 U DNA polymerase (Q5 or KOD XL) and 1× buffer (supplied with the enzyme) in a 3 mL total reaction volume and 0.05–0.2 pmol of ssDNA (70 µL) eluted from the target in each

selection was added to 3 mL PCR mixture. 1mM $MnCl_2$ was supplemented to KOD XL polymerase mixture just before commencing PCR experiments.

PCR amplification for canonical ssDNA pool consisted of: 1) initial denaturation at 98 ℃ for 30 sec; 2) 25 cycles consisting of denaturation at 98 ℃ for 10 sec, annealing at 65 ℃ for 20 sec and extension at 72 ℃ for 30 sec; 3) final extension at 72 ℃ for 2 min.

PCR amplification for ssDNA pool containing 5-ITZ-dUTP consisted of: 1) initial denaturation at 98 ℃ for 30 sec; 2) 30 cycles consisting of denaturation at 98 ℃ for 30 sec, annealing at 65 ℃ for 20 sec and extension at 70 ℃ for 1 min; 3) final extension at 70 ℃ for 5 min.

After PCR, samples were combined and concentrated, loaded onto 10% denaturing PAGE gel and run at 20 W for 2 h in 1× TBE buffer. The band of interest was excised and extracted from the gel. Extracted ssDNA was concentrated and washed with deionised water using 10 kDa MWCO Amicon ultracentrifugal filter for 10 times and finally desalted by a Sephadex illustra NAP-10 column. Eluted DNA was freeze-dried and re-dissovled in 1× PBS before use in subsequent cycles of SELEX.

7.8.6 Q-PCR quantification

After each round of selection, Q-PCR assay was established for quantification of eluted ssDNA using a real-time PCR instrument (Bio-Rad CFX96). Q-PCR was performed in a reaction mixture (20 μL) containing Evagreen dye (0.8×), 10 pmol of each primer, 5 nmol of each dATP, dGTP, dCTP and dTTP or 5-ITZ-dUTP, 0.5 U DNA polymerase (Q5 or KOD XL DNA polymerase), 1× buffer (supplied with the enzyme) and ssDNA

eluted from each selection cycle (1 µL). 1 mM $MnCl_2$ was supplemented for KOD XL DNA polymerase mixture.

A standard sample, a 10-fold serial dilution of FAM-labelled ssDNA library (1 µL, 1.6 nM – 0.016 pM) was also prepared to determine the copy numbers ssDNA bound to target in individual samples (copy numbers were calculated for both dilutions individually and then averaged). In order to make the result accurate, standard samples were measured in triplicates.

Q-PCR consisted of: 1) initial denaturation at 98 ℃ for 30 sec; 2) 30 cycles of denaturation at 98 ℃ for 10 sec, annealing at 65 ℃ for 20 sec and extension at 72 ℃ for 20 sec; 3) final extension at 72 ℃ for 20 sec, concluded by measuring T_m, starting from 60 ℃ increasing the temperature by 0.1 ℃ per sec to 95 ℃.

7.8.7 Enzymatic synthesis of aptamer sequences for target binding validation

Primer extension assays were performed using 30 nmol of 5′-FAM labelled primer and template, 1.5 µmol of each dATP, dGTP and dCTP or modified dUTP, 400 U Vent exo- DNA polymerase (New England Biolabs) and 1× Thermopol buffer (supplied with the enzymes) in a 8 mL total reaction volume. The reaction mixture was aliquoted to each well in the 96-well PCR plate (Thermo Fisher Scientific, 80 µL). The reactions were conducted in a BIO-RAD T100 Thermal Cycler at 65 ℃ for 2 h. The reaction mixtures were then combined and concentrated to 1 mL using 10 kDa MWCO Amicon ultracentrifugal filters (Merck Millipore). Formamide (1 mL) and 3 M NaOH (7 µL) were added to the reaction mixture and heat denatured at 95 ℃ for 15 min. Samples were then loaded onto 8% denaturing PAGE gel (Table 7.1) and run at 20 W in 1 × TBE or TC buffer for 2 h. The bands of interest were excised from the gel and incubated

in deionised water (15 mL) at 37 ℃ for overnight. The gels were filtered and the filtrate containing DNA was washed with deionsed water using 10 kDa MWCO Amicon ultracentrifugal filters (Merck Millipore) for 10 times and finally desalted by a Sephadex illustra NAP-10 column (GE healthcare).

Table 7.1. Composition of 8% denaturing PAGE gel.

Composition of denaturing PAGE gel (8%)	
10 × TBE or TC buffer	7 mL
UREA	29.4 g
40% (w/v) acrylamide-biscrylamide (19:1)	14 mL
H_2O	26 mL
10% (w/v) ammonium persulfate	580 µL
TEMED	58 µL

7.8.8 SPR binding analysis

7.8.8.1 Running conditions

All SPR experiments were performed using CM5 dextran sensor chip (GE Healthcare) with a Biacore X100 instrument at 25 ℃. The SPR running buffer was 1× PBS buffer (Thermo Fisher Scientific, pH 7.4) for streptavidin-binding aptamer and 1× HBS-EP buffer (GE healthcare, 0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) for peptide work. Both are at a flow rate of 10 µL/min.

7.8.8.2 Target immobilisation

Flow-cell 1 (FC1) and flow-cell 2 (FC2) were activated by injection of a freshly prepared mixture of 0.4 M EDC (100 µL) and 0.1 M NHS (100 µL) over the sensor chip for 6 min.

Streptavidin (New England Biolabs, 1 mg/mL in 140 mM NaCl, 8 mM sodium phosphate, 2 mM potassium phosphate and 10 mM KCl, pH 7.4) was buffer exchanged to 10 mM sodium acetate (pH 5.0) and concentrated to using a 10 kDa MWCO Amicon ultracentrifugal filter (Merck Millipore) to give a final concentration of 2 mg/mL. Alternatively, lyophilised streptavidin (Sigma Aldrich, 1 mg) was dissolved in 10 mM sodium acetate (500 µL, pH 5.0).

For streptavidin aptamer analysis, streptavidin was immobilised on the activated surface of FC2 using the Biacore wizard immobilisation programme with a target response of 6,000 RU. The remaining active sites in both FC1 and FC2 were quenched by the injection of 1 M ethanolamine-HCL (100 µL, pH 8.5) for 6 min.

For peptide immobilisation studies, 2060 RU and 1960 RU of streptavidin were immobilised to FC1 and FC2 respectively. 10 µM BP1 (100 µL) was then injected onto FC1 for 180 sec and 10 µM BP2 (100 µL) was injected onto FC2 for 180 sec. The response unit for BP1 and BP2 were 104 RU and 72 RU respectively.

For peptide aptamer analysis, peptide (Peptide Protein Reseach ltd, 200 µg) in 50 mM sodium borate buffer (100 µL, pH 8.4) was injected for 6 min and manually immobilised on the activated surface of FC2. The remaining active sites in both FC1 and FC2 were quenched by the injection of 1 M ethanolamine-HCL (100 µL, pH 8.5)

for 6 min. 410 RU of peptide **1** (Figure 7.1 **A**) and 232.4 RU of peptide **2** (Figure 7.1 **B**) were immobilised onto two separate CM5 chips.

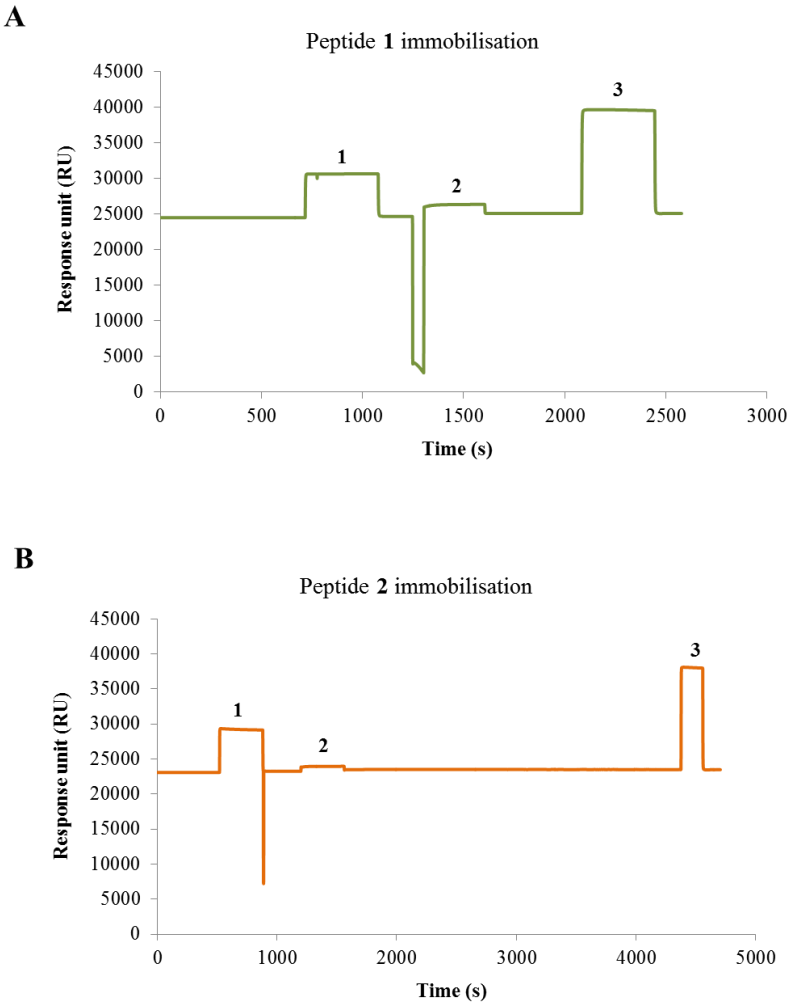


Figure 7.1. **A**) Peptide **1** and **B**) peptide **2** immobilisation on the flow-cell 2 (FC2) on CM5 chips respectively. **Step 1**, EDC/NHS activation; **step 2**, peptide injection; **step 3**, 1 M ethanolamine-HCL (pH 8.5) deactivation.

7.8.8.3 Analyte preparation and testing

Freeze-dried oligonucleotides were dissolved in the appropriate 1× running buffer. The samples were then denatured at 95 ℃ for 15 min, frozen in liquid nitrogen and allowed to thaw to room temperature in order to facilitate unimolecular folding. This solution was then serially diluted in buffer to give the desired concentration range and injected at 25 ℃ at a flow rate of 10 µL/min for 180 sec (association phase) followed by a dissociation phase of 600 sec. 50 mM NaOH (pH 12.7) was then injected for 20 sec to regenerate the surface. The binding assay was performed in duplicate or triplicate at different time points within each experiment.

7.8.8.4 Data analysis

Data processing and analysis were performed using the Biacore X100 Evaluation Software v2.0 (GE healthcare). A double-referencing method was performed to process all datasets. In brief, data from the reaction FC2 was referenced first by subtracting data from the reference FC1 to correct for bulk refractive index changes, matrix effects, non-specific binding, injection noise and baseline drift. The reference-subtracted data (FC2-1) was referenced again by a blank injection (running buffer) to account for any systematic drift over the course of the injection. The double-referenced data was fit to the default steady-state affinity model for thermodynamic analysis.

7.8.9 Electrophoretic Mobility Shift Assay

Gels of size 18 (W) × 0.2 (D) × 24.4 (H) cm were prepared (Table 7.2). The gel was pre-run for 30 min at 100 V in a cold room.

Table 7.2. Composition of 8% native gel used in EMSA experiments.

Composition of native gel (8%)	
10× TBE buffer	7 mL
40% (w/v) acrylamide-biscrylamide (19:1)	14 mL
H_2O	56 mL
10% (w/v) ammonium persulfate	580 µL
TEMED	58 µL

1 µM aptamer (2 µL) was denatured at 95 ℃ for 15 min, frozen in liquid nitrogen and allowed to thaw to room temperature in order to facilitate unimolecular folding. Once the aptamer was thawed, the concentrated streptavidin in 1× PBS was added to the ssDNA according to various concentration. Total binding volume was 20 µL and the complexes were incubated at 25 ℃ for 30 min at 900 rpm. After 30 min, 4% (v/v) glycerol was added with loading dye (bromophenol blue) before loaded on to the 8% native gel. The gel was run at 100 V for 5 h in a cold room. The value of the K_D can be estimated from looking at the gel. ImageJ software can be used to quantitate the fluorescent signal in each DNA band on the gel. Background signals from blank regions of the gel were subtracted from the signal intensities obtained from bands. The fraction of DNA bound was determined from the background-subtracted signal intensities using the expression: $\frac{\text{Bound}}{\text{Bound+Unbound}}$.

References

References

1. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737 (1953).

2. Chargaff, E. & Vischer, E. The composition of the desoxypentose nucleic acids of thymus and spleen. *J. Biol. Chem.* **177,** 405–416 (1949).

3. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10,** 654–659 (1951).

4. Chargaff, E., Lipshitz, R. & Green, C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.* **195,** 155–160 (1952).

5. Donohue, J. Fourier analysis and the structure of DNA. *Science* **165,** 1091–1096 (1969).

6. Dickerson, R. E. *et al.* The anatomy of A-, B-, and Z-DNA. *Science* **216,** 475–485 (1982).

7. Nakamoto, K., Tsuboi, M. & Strahan, G. D. Drug-DNA interactions: structures and spectra. *Methods Biochem. Anal.* **51,** 1–366 (2008).

8. Neidle, S. DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.* **18,** 291–309 (2001).

9. Hamilton, P. L. & Arya, D. P. Natural product DNA major groove binders. *Nat. Prod. Rep.* **29,** 134–143 (2012).

10. Cai, X., Gray, P. J. J. & Von Hoff, D. D. DNA minor groove binders: back in the groove. *Cancer Treat. Rev.* **35,** 437–450 (2009).

11. Tse, W. C. & Boger, D. L. Sequence-Selective DNA Recognition: Natural Products and Nature's Lessons. *Chem. Biol.* **11,** 1607–1617 (2004).

12. Khan, G. S., Shah, A., Zia-Ur-Rehman & Barker, D. Chemistry of DNA minor groove binding agents. *J. Photochem. Photobiol. B Biol.* **115,** 105–118 (2012).

13. Kulkarni, M. & Mukherjee, A. Understanding B-DNA to A-DNA transition in the right-handed DNA helix: Perspective from a local to global transition. *Prog. Biophys. Mol. Biol.* **128,** 63–73 (2017).

14. Vargason, J. M., Henderson, K. & Ho, P. S. A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U. S. A* **98,** 7265 LP-7270 (2001).

15. Perera, R. T. *et al.* Unzipping of A-Form DNA-RNA, A-Form DNA-PNA, and B-Form DNA-DNA in the α-Hemolysin Nanopore. *Biophys. J.* **110,** 306–314 (2016).

16. Gao, Y. G., Robinson, H. & Wang, A. H. High-resolution A-DNA crystal structures of d(AGGGGCCCCT). An A-DNA model of poly(dG) x poly(dC). *Eur. J. Biochem.* **261,** 413–420 (1999).

17. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2179–2183 (1981).

18. Medhi, C., Mitchell, J. B., Price, S. L. & Tabor, A. B. Electrostatic factors in DNA intercalation. *Biopolymers* **52,** 84–93 (1999).

19. Li, S., Cooper, V. R., Thonhauser, T., Lundqvist, B. I. & Langreth, D. C. Stacking Interactions and DNA Intercalation. *J. Phys. Chem. B* **113,** 11166–11172 (2009).

20. Johnson, C. A. *et al.* Effect of intercalator substituent and nucleotide sequence on the stability of DNA- and RNA-naphthalimide complexes. *Bioorg. Med. Chem.* **23,** 3586–3591 (2015).

21. Neto, B. A. D. & Lapis, A. A. M. Recent developments in the chemistry of deoxyribonucleic acid (DNA) intercalators: principles, design, synthesis, applications and trends. *Molecules* **14,** 1725–1746 (2009).

22. Ihmels, H., Faulhaber, K., Vedaldi, D., Dall'Acqua, F. & Viola, G. Intercalation of Organic Dye Molecules into Double-stranded DNA. Part 2: The Annelated Quinolizinium Ion as a Structural Motif in DNA Intercalators. *Photochem. Photobiol.* **81,** 1107–1115 (2005).

23. Gudnason, H., Dufva, M., Bang, D. D. & Wolff, A. Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic Acids Res.* **35,** e127–e127 (2007).

24. Nafee, N. *et al.* Aptamers as Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **57,** 61–79 (2017).

25. Nimjee, S. M., Rusconi, C. P. & Sullenger, B. A. Aptamers: An Emerging Class of Therapeutics. *Annu. Rev. Med.* **56,** 555–583 (2005).

26. Sun, H. *et al.* Oligonucleotide aptamers: New tools for targeted cancer therapy. *Mol. Ther. - Nucleic Acids* **3,** (2014).

27. Tang, Z., Parekh, P., Turner, P., Moyer, R. W. & Tan, W. Generating Aptamers for Recognition of Virus-Infected Cells. *Clin. Chem.* **55,** 813 LP-822 (2009).

28. Shangguan, D. *et al.* Aptamers evolved from live cells as effective molecular probes for cancer study. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11838–11843 (2006).

29. Zhou, J. & Rossi, J. Aptamers as targeted therapeutics: Current potential and challenges. *Nat. Rev. Drug Discov.* **16,** 181–202 (2017).

30. Platella, C., Riccardi, C., Montesarchio, D., Roviello, G. N. & Musumeci, D. G-quadruplex-based aptamers against protein targets in therapy and diagnostics. *Biochim. Biophys. acta. Gen. Subj.* **1861,** 1429–1447 (2017).

31. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34,** 5402–5415 (2006).

32. Georgiades, S. N., Abd Karim, N. H., Suntharalingam, K. & Vilar, R. Interaction of Metal Complexes with G-Quadruplex DNA. *Angew. Chemie Int. Ed.* **49,** 4020–4034 (2010).

33. Monchaud, D. & Teulade-Fichou, M.-P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **6,** 627–636 (2008).

34. Mergny, J.-L., De Cian, A., Ghelab, A., Sacca, B. & Lacroix, L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.* **33,** 81–94 (2005).

35. Yang, D. & Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.* **2,** 619–646 (2010).

36. Neidle, S. & Parkinson, G. Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discov.* **1,** 383–393 (2002).

37. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90,** 1149–1171 (2008).

38. Buket, O., Clement, L. & DanZhou, Y. DNA G-quadruplex and its potential as anticancer drug target. *Sci. China. Chem.* **57,** 1605–1614 (2014).

39. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43,** 8627–8637 (2015).

40. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33,** 2908–2916 (2005).

41. Wong, H. M., Stegle, O., Rodgers, S. & Huppert, J. L. A toolbox for predicting g-quadruplex formation and stability. *J. Nucleic Acids* **2010,** 564946 (2010).

42. Ma, D.-L. *et al.* Recent Developments in G-Quadruplex Probes. *Cell Chem. Biol.* **22,** 812–828 (2015).

43. Konig, S. L. B., Evans, A. C. & Huppert, J. L. Seven essential questions on G-quadruplexes. *Biomol. Concepts* **1,** 197–213 (2010).

44. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1,** 263–282 (1993).

45. Risitano, A. & Fox, K. R. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **32,** 2598–2606 (2004).

46. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36,** 5482–5515 (2008).

47. Gatto, B., Palumbo, M. & Sissi, C. Nucleic acid aptamers based on the G-quadruplex structure: therapeutic and diagnostic potential. *Curr. Med. Chem.* **16,** 1248–1265 (2009).

48. Tucker, W. O., Shum, K. T. & Tanner, J. A. G-quadruplex DNA aptamers and their ligands: structure, function and application. *Curr. Pharm. Des.* **18,** 2014–2026 (2012).

49. Nagatoishi, S., Isono, N., Tsumoto, K. & Sugimoto, N. Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie* **93,** 1231–1238 (2011).

50. Olsen, C. M., Lee, H.-T. & Marky, L. A. Unfolding thermodynamics of intramolecular G-quadruplexes: base sequence contributions of the loops. *J. Phys. Chem. B* **113,** 2587–2595 (2009).

51. Smirnov, I. & Shafer, R. H. Effect of Loop Sequence and Size on DNA Aptamer Stability. *Biochemistry* **39,** 1462–1468 (2000).

52. Ou, T. *et al.* G-quadruplexes: targets in anticancer drug design. *ChemMedChem* **3,** 690–713 (2008).

53. Leroy, J. L., Gehring, K., Kettani, A. & Gueron, M. Acid multimers of oligodeoxycytidine strands: Stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry* **32,** 6019–6031 (1993).

54. Gehring, K., Leroy, J. L. & Gueron, M. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* **363,** 561–565 (1993).

55. Reilly, S. M., Morgan, R. K., Brooks, T. A. & Wadkins, R. M. Effect of interior loop length on the thermal stability and pK(a) of i-motif DNA. *Biochemistry* **54,** 1364–1370 (2015).

56. Gurung, S. P., Schwarz, C., Hall, J. P., Cardin, C. J. & Brazier, J. A. The importance of loop length on the stability of i-motif structures. *Chem. Commun.* **51,** 5630–5632 (2015).

57. Webb, B. A., Chimenti, M., Jacobson, M. P. & Barber, D. L. Dysregulated pH: a perfect storm for cancer progression. *Nature reviews. Cancer* **11,** 671–677 (2011).

58. Huang, J. *et al.* I-motif-based nano-flares for sensing pH changes in live cells. *Chem. Commun.* **50,** 15768–15771 (2014).

59. Modi, S. *et al.* A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.* **4,** 325–330 (2009).

60. Ng, E. W. M. & Adamis, A. P. Anti-VEGF aptamer (pegaptanib) therapy for ocular vascular diseases. *Ann. N. Y. Acad. Sci.* **1082,** 151–171 (2006).

61. Ng, E. W. M. *et al.* Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug Discov.* **5,** 123–132 (2006).

62. Rusconi, C. P. *et al.* RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature* **419,** 90–94 (2002).

63. Vavalle, J. P. & Cohen, M. G. The REG1 anticoagulation system: A novel actively controlled factor IX inhibitor using RNA aptamer technology for treatment of acute coronary syndrome. *Future Cardiol.* **8,** 371–382 (2012).

64. McNamara, J. O. *et al.* Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. *Nat. Biotechnol.* **24,** 1005–1015 (2006).

65. Lai, W. Y. *et al.* Synergistic inhibition of lung cancer cell invasion, tumor growth and angiogenesis using aptamer-siRNA chimeras. *Biomaterials* **35,** 2905–2914 (2014).

66. Li, L. *et al.* Nucleolin-targeting liposomes guided by aptamer AS1411 for the delivery of siRNA for the treatment of malignant melanomas. *Biomaterials* **35,** 3840–3850 (2014).

67. Herrmann, A. *et al.* CTLA4 aptamer delivers STAT3 siRNA to tumor-associated and malignant T cells. *J. Clin. Invest.* **124,** 2977–2987 (2014).

68. Santulli-Marotto, S., Nair, S. K., Rusconi, C., Sullenger, B. & Gilboa, E. Multivalent RNA aptamers that inhibit CTLA-4 and enhance tumor immunity. *Cancer Res.* **63,** 7483–7489 (2003).

69. Reyes-Reyes, E. M., Teng, Y. & Bates, P. J. A new paradigm for aptamer therapeutic AS1411 action: uptake by macropinocytosis and its stimulation by a nucleolin-dependent mechanism. *Cancer Res.* **70,** 8617–8629 (2010).

70. Bates, P. J., Laber, D. A., Miller, D. M., Thomas, S. D. & Trent, J. O. Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Exp. Mol. Pathol.* **86,** 151–164 (2009).

71. Ruckman, J. *et al.* 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.* **273,** 20556–20567 (1998).

72. Trujillo, C. A., Nery, A. A., Alves, J. M., Martins, A. H. & Ulrich, H. Development of the anti-VEGF aptamer to a therapeutic agent for clinical ophthalmology. *Clin. Ophthalmol.* **1,** 393–402 (2007).

73. Burmeister, P. E. *et al.* Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. *Chem. Biol.* **12,** 25–33 (2005).

74. Jellinek, D. *et al.* Potent 2'-amino-2'-deoxypyrimidine RNA inhibitors of basic fibroblast growth factor. *Biochemistry* **34,** 11363–11372 (1995).

75. Green, L. S. *et al.* Inhibitory DNA ligands to platelet-derived growth factor B-chain. *Biochemistry* **35,** 14413–14424 (1996).

76. Boyer, D. A Phase 2b Study of Fovista^{TM}, a Platelet Derived Growth Factor (PDGF) inhibitor in combination with a Vascular Endothelial Growth Factor (VEGF) inhibitor for Neovascular Age-Related Macular Degeneration (AMD). *Invest. Ophthalmol. Vis. Sci.* **54,** 2175 (2013).

77. Boyer, D. S. & Group, O. A.-P. in A. M. D. S. Combined Inhibition of Platelet Derived (PDGF) and Vascular Endothelial (VEGF) Growth Factors for the Treatment of Neovascular Age-Related Macular Degeneration (NV-AMD) - Results of a Phase 1 Study. *Invest. Ophthalmol. Vis. Sci.* **50,** 1260 (2009).

78. Cohen, M. G. *et al.* First clinical application of an actively reversible direct factor IXa inhibitor as an anticoagulation strategy in patients undergoing percutaneous coronary intervention. *Circulation* **122,** 614–622 (2010).

79. Povsic, T. J. *et al.* Use of the REG1 anticoagulation system in patients with acute coronary syndromes undergoing percutaneous coronary intervention: results from the phase II RADAR-PCI study. *EuroIntervention J. Eur. Collab. with Work. Gr. Interv. Cardiol. Eur. Soc. Cardiol.* **10,** 431–438 (2014).

80. Morita, Y., Leslie, M., Kameyama, H., Volk, D. E. & Tanaka, T. Aptamer Therapeutics in Cancer: Current and Future. *Cancers (Basel).* **10,** 80 (2018).

81. Soundararajan, S. *et al.* Plasma membrane nucleolin is a receptor for the anticancer aptamer AS1411 in MV4-11 leukemia cells. *Mol. Pharmacol.* **76,** 984–991 (2009).

82. Marasca, R. & Maffei, R. NOX-A12: mobilizing CLL away from home. *Blood* **123,** 952 LP-953 (2014).

83. Hoellenriegel, J. *et al.* The Spiegelmer NOX-A12, a novel CXCL12 inhibitor, interferes with chronic lymphocytic leukemia cell motility and causes chemosensitization. *Blood* **123,** 1032 LP-1039 (2014).

84. Ismail, S. I. & Alshaer, W. Therapeutic aptamers in discovery, preclinical and clinical stages. *Adv. Drug Deliv. Rev.* **134,** 51–64 (2018).

85. Salamanca, H. H., Antonyak, M. A., Cerione, R. A., Shi, H. & Lis, J. T. Inhibiting Heat Shock Factor 1 in Human Cancer Cells with a Potent RNA Aptamer. *PLoS One* **9,** e96330 (2014).

86. Seiwert, S. D., Stines Nahreini, T., Aigner, S., Ahn, N. G. & Uhlenbeck, O. C. RNA aptamers as pathway-specific MAP kinase inhibitors. *Chem. Biol.* **7,** 833–843 (2000).

87. Vaish, N. K., Larralde, R., Fraley, A. W., Szostak, J. W. & McLaughlin, L. W. A novel, modification-dependent ATP-binding aptamer selected from an RNA library incorporating a cationic functionality. *Biochemistry* **42,** 8842–8851 (2003).

88. Zamay, T. N. *et al.* DNA-aptamer targeting vimentin for tumor therapy in vivo. *Nucleic Acid Ther.* **24,** 160–170 (2014).

89. Maasch, C. *et al.* Polyetheylenimine-polyplexes of Spiegelmer NOX-A50 directed against intracellular high mobility group protein A1 (HMGA1) reduce tumor growth in vivo. *J. Biol. Chem.* **285,** 40012–40018 (2010).

90. Ryou, S.-M. *et al.* Gold nanoparticle-assisted delivery of small, highly structured RNA into the nuclei of human cells. *Biochem. Biophys. Res. Commun.* **416,** 178–183 (2011).

91. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818 (1990).

92. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–510 (1990).

93. Pfeiffer, F. *et al.* Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat. Protoc.* **13,** 1153 (2018).

94. Jijakli, K. *et al.* The in vitro selection world. *Methods* **106,** 3–13 (2016).

95. Tolle, F., Wilke, J., Wengel, J. & Mayer, G. By-product formation in repetitive PCR amplification of DNA libraries during SELEX. *PLoS One* **9,** 1–12 (2014).

96. Kanagawa, T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* **96,** 317–323 (2003).

97. Musheev, M. U. & Krylov, S. N. Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue. *Anal. Chim. Acta* **564,** 91–96 (2006).

98. Ouellet, E., Foley, J. H., Conway, E. M. & Haynes, C. Hi-Fi SELEX: A High-Fidelity Digital-PCR Based Therapeutic Aptamer Discovery Platform. *Biotechnol. Bioeng.* **112,** 1506–1522 (2015).

99. Takahashi, M. *et al.* High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Sci. Rep.* **6,** 33697 (2016).

100. Shao, K. *et al.* Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection. *PLoS One* **6,** e24910–e24910 (2011).

101. Liang, C. *et al.* Comparison of the methods for generating single-stranded DNA in SELEX. *Analyst* **140,** 3439–3444 (2015).

102. Wilson, R. Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads. *Nucleic Acid Ther.* **21,** 437–440 (2011).

103. Paul, A., Avci-Adali, M., Ziemer, G. & Wendel, H. P. Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell-SELEX. *Oligonucleotides* **19,** 243–254 (2009).

104. Gyllensten, U. B. & Erlich, H. A. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 7652–7656 (1988).

105. Avci-Adali, M., Paul, A., Wilhelm, N., Ziemer, G. & Wendel, H. P. Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation. *Molecules* **15,** 1–11 (2009).

106. Citartan, M., Tang, T.-H., Tan, S.-C. & Gopinath, S. C. B. Conditions optimized for the preparation of single-stranded DNA (ssDNA) employing lambda exonuclease digestion in generating DNA aptamer. *World J. Microbiol. Biotechnol.* **27,** 1167–1173 (2011).

107. Higuchi, R. G. & Ochman, H. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucleic Acids Res.* **17,** 5865 (1989).

108. Williams, K. P. & Bartel, D. P. PCR product with strands of unequal length. *Nucleic Acids Res.* **23,** 4220–4221 (1995).

109. Cao, X. *et al.* Combining use of a panel of ssDNA aptamers in the detection of Staphylococcus aureus. *Nucleic Acids Res.* **37,** 4621–4628 (2009).

110. Marimuthu, C., Tang, T.-H., Tominaga, J., Tan, S.-C. & Gopinath, S. C. B. Single-stranded DNA (ssDNA) production in DNA aptamer generation. *Analyst* **137,** 1307–1315 (2012).

111. Zhuo, Z. *et al.* Recent Advances in SELEX Technology and Aptamer Applications in Biomedicine. *Int. J. Mol. Sci.* **18,** (2017).

112. Ellington, A. D. & Szostak, J. W. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **355,** 850–852 (1992).

113. Bruno, J. G. In vitro selection of DNA to chloroaromatics using magnetic microbead-based affinity separation and fluorescence detection. *Biochem. Biophys. Res. Commun.* **234,** 117–120 (1997).

114. Stoltenburg, R., Reinemann, C. & Strehlitz, B. FluMag-SELEX as an advantageous method for DNA aptamer selection. *Anal. Bioanal. Chem.* **383,** 83–91 (2005).

115. Wang, C., Yang, G., Luo, Z. & Ding, H. In vitro selection of high-affinity DNA aptamers for streptavidin. *Acta Biochim. Biophys. Sin. (Shanghai).* **41,** 335–340 (2009).

116. Xu, S. *et al.* Selection of DNA aptamers against polychlorinated biphenyls as potential biorecognition elements for environmental analysis. *Anal. Biochem.* **423,** 195–201 (2012).

117. Kim, Y. S., Hyun, C. J., Kim, I. A. & Gu, M. B. Isolation and characterization of enantioselective DNA aptamers for ibuprofen. *Bioorg. Med. Chem.* **18,** 3467–3473 (2010).

118. Stoltenburg, R., Nikolaus, N. & Strehlitz, B. Capture-SELEX: Selection of DNA Aptamers for Aminoglycoside Antibiotics. *J. Anal. Methods Chem.* **2012,** 415697 (2012).

119. Khati, M. *et al.* Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2'F-RNA aptamers. *J. Virol.* **77,** 12692–12698 (2003).

120. Dausse, E. *et al.* Aptamer selection by direct microfluidic recovery and surface plasmon resonance evaluation. *Biosens. Bioelectron.* **80,** 418–425 (2016).

121. Mendonsa, S. D. & Bowser, M. T. In Vitro Evolution of Functional DNA Using Capillary Electrophoresis. *J. Am. Chem. Soc.* **126,** 20–21 (2004).

122. Mendonsa, S. D. & Bowser, M. T. In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. *Anal. Chem.* **76,** 5387–5392 (2004).

123. Mosing, R. K., Mendonsa, S. D. & Bowser, M. T. Capillary Electrophoresis-SELEX Selection of Aptamers with Affinity for HIV-1 Reverse Transcriptase. *Anal. Chem.* **77,** 6107–6112 (2005).

124. Mendonsa, S. D. & Bowser, M. T. In Vitro Selection of Aptamers with Affinity for Neuropeptide Y Using Capillary Electrophoresis. *J. Am. Chem. Soc.* **127,** 9382–9383 (2005).

125. Tang, J., Xie, J., Shao, N. & Yan, Y. The DNA aptamers that specifically recognize ricin toxin are selected by two in vitro selection methods. *Electrophoresis* **27,** 1303–1311 (2006).

126. Mallikaratchy, P., Stahelin, R. V, Cao, Z., Cho, W. & Tan, W. Selection of DNA ligands for protein kinase C-δ. *Chem. Commun.* 3229–3231 (2006). doi:10.1039/B604778E

127. Cella, L. N. *et al.* Nano Aptasensor for Protective Antigen Toxin of Anthrax. *Anal. Chem.* **82,** 2042–2047 (2010).

128. Berezovski, M. V, Musheev, M. U., Drabovich, A. P., Jitkova, J. V & Krylov, S. N. Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides. *Nat. Protoc.* **1,** 1359–1369 (2006).

129. Ashley, J., Ji, K. & Li, S. Selection of bovine catalase aptamers using non-SELEX. *Electrophoresis* **33,** (2012).

130. Chen, M. *et al.* Development of Cell-SELEX Technology and Its Application in Cancer Diagnosis and Therapy. *Int. J. Mol. Sci.* **17,** 2079 (2016).

131. Mercier, M.-C., Dontenwill, M. & Choulier, L. Selection of Nucleic Acid Aptamers Targeting Tumor Cell-Surface Protein Biomarkers. *Cancers (Basel).* **9,** (2017).

132. Shangguan, D., Bing, T. & Zhang, N. Cell-SELEX: Aptamer Selection Against Whole Cells. In: Tan W., Fang X. (eds) Aptamers Selected by Cell-SELEX for Theranostics. Springer, Berlin, Heidelberg (2015)

133. Catuogno, S. & Esposito, C. L. Aptamer Cell-Based Selection: Overview and Advances. *Biomedicines* **5,** (2017).

134. Daniels, D. A., Chen, H., Hicke, B. J., Swiderek, K. M. & Gold, L. A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15416–15421 (2003).

135. Ninomiya, K. *et al.* Cell-SELEX based selection and characterization of DNA aptamer recognizing human hepatocarcinoma. *Bioorg. Med. Chem. Lett.* **23,** 1797–1802 (2013).

136. Ara, M. N., Hyodo, M., Ohga, N., Hida, K. & Harashima, H. Development of a Novel DNA Aptamer Ligand Targeting to Primary Cultured Tumor Endothelial Cells by a Cell-Based SELEX Method. *PLoS One* **7,** e50174 (2012).

137. Rong, Y. *et al.* Identification of an aptamer through whole cell-SELEX for targeting high metastatic liver cancers. *Oncotarget* **7,** 8282–8294 (2016).

138. Bayrac, A. T. *et al.* In Vitro Selection of DNA Aptamers to Glioblastoma Multiforme. *ACS Chem. Neurosci.* **2,** 175–181 (2011).

139. Dastjerdi, K., Tabar, G. H., Dehghani, H. & Haghparast, A. Generation of an enriched pool of DNA aptamers for an HER2-overexpressing cell line selected by Cell SELEX. *Biotechnol. Appl. Biochem.* **58,** 226–230 (2011).

140. Kunii, T., Ogura, S., Mie, M. & Kobatake, E. Selection of DNA aptamers recognizing small cell lung cancer using living cell-SELEX. *Analyst* **136,** 1310–1312 (2011).

141. Jiménez, E. *et al.* Generation of Lung Adenocarcinoma DNA Aptamers for Cancer Studies. *PLoS One* **7,** e46222 (2012).

142. Wang, Y. *et al.* DNA aptamer evolved by cell-SELEX for recognition of prostate cancer. *PLoS One* **9,** e100243 (2014).

143. Li, W.-M. *et al.* Cell-SELEX-based selection of aptamers that recognize distinct targets on metastatic colorectal cancer cells. *Biomaterials* **35,** 6998–7007 (2014).

144. Raddatz, M.-S. L. *et al.* Enrichment of cell-targeting and population-specific aptamers by fluorescence-activated cell sorting. *Angew. Chem. Int. Ed. Engl.* **47,** 5190–5193 (2008).

145. Winkler, J. Oligonucleotide conjugates for therapeutic applications. *Ther. Deliv.* **4,** 791–809 (2013).

146. Juliano, R. L., Ming, X. & Nakagawa, O. The Chemistry and Biology of Oligonucleotide Conjugates. *Acc. Chem. Res.* **45,** 1067–1076 (2012).

147. Pagano, B., Martino, L., Randazzo, A. & Giancola, C. Stability and Binding Properties of a Modified Thrombin Binding Aptamer. *Biophys. J.* **94,** 562–569 (2008).

148. Perrin, D. M., Garestier, T. & Hélène, C. Expanding the Catalytic Repertoire of Nucleic Acid Catalysts: Simultaneous Incorporation of Two Modified Deoxyribonucleoside Triphosphates Bearing Ammonium and Imidazolyl Functionalities. *Nucleosides and Nucleotides* **18,** 377–391 (1999).

149. Keefe, A. D. & Cload, S. T. SELEX with modified nucleotides. *Curr. Opin. Chem. Biol.* **12,** 448–456 (2008).

150. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709 (1993).

151. Lauridsen, L. H., Rothnagel, J. A. & Veedu, R. N. Enzymatic recognition of 2'-modified ribonucleoside 5'-triphosphates: towards the evolution of versatile aptamers. *Chembiochem* **13,** 19–25 (2012).

152. Diafa, S. & Hollenstein, M. Generation of Aptamers with an Expanded Chemical Repertoire. *Molecules* **20,** 16643–16671 (2015).

153. Shigdar, S. *et al.* Aptamers as theranostic agents: modifications, serum stability and functionalisation. *Sensors (Basel).* **13,** 13624–13637 (2013).

154. Davydova, A. *et al.* Development and characterization of novel 2′-F-RNA aptamers specific to human total and glycated hemoglobins. *Anal. Biochem.* **570,** 43–50 (2019).

155. Rhie, A. *et al.* Characterization of 2'-fluoro-RNA aptamers that bind preferentially to disease-associated conformations of prion protein and inhibit conversion. *J. Biol. Chem.* **278,** 39697–39705 (2003).

156. Tahiri-Alaoui, A. *et al.* High affinity nucleic acid aptamers for streptavidin incorporated into bi-specific capture ligands. *Nucleic Acids Res.* **30,** e45–e45 (2002).

157. Yan, X., Gao, X. & Zhang, Z. Isolation and characterization of 2'-amino-modified RNA aptamers for human TNFalpha. *Genomics. Proteomics Bioinformatics* **2,** 32–42 (2004).

158. Manoharan, M. 2'-carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. *Biochim. Biophys. Acta* **1489,** 117–130 (1999).

159. Li, N., Nguyen, H. H., Byrom, M. & Ellington, A. D. Inhibition of cell proliferation by an anti-EGFR aptamer. *PLoS One* **6,** e20299 (2011).

160. Svobodova, M., Bunka, D. H. J., Nadal, P., Stockley, P. G. & O'Sullivan, C. K. Selection of 2'F-modified RNA aptamers against prostate-specific antigen and their evaluation for diagnostic and therapeutic applications. *Anal. Bioanal. Chem.* **405,** 9149–9157 (2013).

161. Padilla, R. & Sousa, R. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.* **30,** e138 (2002).

162. Padilla, R. & Sousa, R. Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutantT7 RNA polymerase (RNAP). *Nucleic Acids Res.* **27,** 1561–1563 (1999).

163. Chelliserrykattil, J. & Ellington, A. D. Evolution of a T7 RNA polymerase variant that transcribes 2'-O-methyl RNA. *Nature biotechnology* **22,** 1155–1160 (2004).

164. Siegmund, V., Santner, T., Micura, R. & Marx, A. Screening mutant libraries of T7 RNA polymerase for candidates with increased acceptance of 2'-modified nucleotides. *Chem. Commun. (Camb).* **48,** 9870–9872 (2012).

165. Kuwahara, M. *et al.* Systematic analysis of enzymatic DNA polymerization using oligo-DNA templates and triphosphate analogs involving 2',4'-bridged nucleosides. *Nucleic Acids Res.* **36,** 4257–4265 (2008).

166. Pavey, J. B. J., Lawrence, A. J., O'Neil, I. A., Vortler, S. & Cosstick, R. Synthesis and transcription studies on 5'-triphosphates derived from 2'-C-branched-uridines: 2'-homouridine-5'-triphosphate is a substrate for T7 RNA polymerase. *Org. Biomol. Chem.* **2,** 869–875 (2004).

167. Kato, Y. *et al.* New NTP analogs: the synthesis of 4'-thioUTP and 4'-thioCTP and their utility for SELEX. *Nucleic Acids Res.* **33,** 2942–2951 (2005).

168. Inoue, N. *et al.* Amplification of 4'-ThioDNA in the Presence of 4'-Thio-dTTP and 4'-Thio-dCTP, and 4'-ThioDNA-Directed Transcription in Vitro and in Mammalian Cells. *J. Am. Chem. Soc.* **129,** 15424–15425 (2007).

169. Kojima, T. *et al.* PCR Amplification of 4′-ThioDNA Using 2′-Deoxy-4′-thionucleoside 5′-Triphosphates. *ACS Synth. Biol.* **2,** 529–536 (2013).

170. Virno, A. *et al.* A novel thrombin binding aptamer containing a G-LNA residue. *Bioorg. Med. Chem.* **15,** 5710–5718 (2007).

171. Hernandez, F. J., Kalra, N., Wengel, J. & Vester, B. Aptamers as a model for functional evaluation of LNA and 2'-amino LNA. *Bioorg. Med. Chem. Lett.* **19,** 6585–6587 (2009).

172. Veedu, R. N., Vester, B. & Wengel, J. Efficient enzymatic synthesis of LNA-modified DNA duplexes using KOD DNA polymerase. *Org. Biomol. Chem.* **7,** 1404–1409 (2009).

173. Højland, T., Veedu, R. N., Vester, B. & Wengel, J. Enzymatic synthesis of DNA strands containing α-L-LNA (α-L-configured locked nucleic acid) thymine nucleotides. *Artif. DNA PNA XNA* **3,** 14–21 (2012).

174. Elle, I. C. *et al.* Selection of LNA-containing DNA aptamers against recombinant human CD73. *Mol. Biosyst.* **11,** 1260–1270 (2015).

175. Weisbrod, S. H. & Marx, A. Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem. Commun.* 5675–5685 (2008). doi:10.1039/B809528K

176. Hottin, A. & Marx, A. Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Acc. Chem. Res.* **49,** 418–427 (2016).

177. Hollenstein, M. Nucleoside Triphosphates - Building Blocks for the Modification of Nucleic Acids. *Molecules* **17,** 13569–13591 (2012).

178. Hocek, M. Synthesis of Base-Modified 2′-Deoxyribonucleoside Triphosphates and Their Use in Enzymatic Synthesis of Modified DNA for Applications in Bioanalysis and Chemical Biology. *J. Org. Chem.* **79,** 9914–9921 (2014).

179. Vaught, J. D. *et al.* Expanding the Chemistry of DNA for in Vitro Selection. *J. Am. Chem. Soc.* **132,** 4141–4151 (2010).

180. Kuwahara, M. *et al.* Direct PCR amplification of various modified DNAs having amino acids: Convenient preparation of DNA libraries with high-potential activities for in vitro selection. *Bioorg. Med. Chem.* **14,** 2518–2526 (2006).

181. Kuwahara, M. *et al.* Substrate properties of C5-substituted pyrimidine 2'-deoxynucleoside 5'-triphosphates for thermostable DNA polymerases during PCR. *Bioorg. Med. Chem. Lett.* **13,** 3735–3738 (2003).

182. Sawai, H. *et al.* Differences in substrate specificity of C(5)-substituted or C(5)-unsubstituted pyrimidine nucleotides by DNA polymerases from thermophilic bacteria, archaea, and phages. *Chem. Biodivers.* **4,** 1979–1995 (2007).

183. Ohbayashi, T. *et al.* Expansion of repertoire of modified DNAs prepared by PCR using KOD Dash DNA polymerase. *Org. Biomol. Chem.* **3,** 2463–2468 (2005).

184. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5,** (2010).

185. Davies, D. R. *et al.* Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19971 LP-19976 (2012).

186. Rohloff, J. C. *et al.* Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol. Ther. - Nucleic Acids* **3,** e201 (2014).

187. Gawande, B. N. *et al.* Selection of DNA aptamers with two modified bases. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 2898 LP-2903 (2017).

188. Tolle, F., Brandle, G. M., Matzner, D. & Mayer, G. A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew. Chem. Int. Ed. Engl.* **54,** 10971–10974 (2015).

189. Yamashige, R. *et al.* Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **40,** 2793–2806 (2012).

190. Lee, K. H., Hamashima, K., Kimoto, M. & Hirao, I. Genetic alphabet expansion biotechnology by creating unnatural base pairs. *Curr. Opin. Biotechnol.* **51,** 8–15 (2018).

191. Kimoto, M., Matsunaga, K. & Hirao, I. DNA Aptamer Generation by Genetic Alphabet Expansion SELEX (ExSELEX) Using an Unnatural Base Pair System. *Methods Mol. Biol.* **1380,** 47–60 (2016).

192. Georgiadis, M. M. *et al.* Structural basis for a six nucleotide genetic alphabet. *J. Am. Chem. Soc.* **137,** 6947–6955 (2015).

193. Zhang, L. *et al.* Evolution of Functional Six-Nucleotide DNA. *J. Am. Chem. Soc.* **137,** 6734–6737 (2015).

194. Brukner, I. & Tremblay, G. A. Cellular proteins prevent antisense phosphorothioate oligonucleotide (SdT18) to target sense RNA (rA18): development of a new in vitro assay. *Biochemistry* **39,** 11463–11466 (2000).

195. Putney, S. D., Benkovic, S. J. & Schimmel, P. R. A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 7350–7354 (1981).

196. Volk, D. E. & Lokesh, G. L. R. Development of Phosphorothioate DNA and DNA Thioaptamers. *Biomedicines* **5,** (2017).

197. Abeydeera, N. D. *et al.* Evoking picomolar binding in RNA by a single phosphorodithioate linkage. *Nucleic Acids Res.* **44,** 8052–8064 (2016).

198. Jhaveri, S., Olwin, B. & Ellington, A. D. In vitro selection of phosphorothiolated aptamers. *Bioorg. Med. Chem. Lett.* **8,** 2285–2290 (1998).

199. King, D. J. *et al.* Combinatorial Selection and Binding of Phosphorothioate Aptamers Targeting Human NF-κB RelA(p65) and p50. *Biochemistry* **41,** 9696–9706 (2002).

200. Yang, X. *et al.* Aptamers containing thymidine 3'-O-phosphorodithioates: synthesis and binding to nuclear factor-kappaB. *Bioorg. Med. Chem. Lett.* **9,** 3357—3362 (1999).

201. Cummins, L., Graff, D., Beaton, G., Marshall, W. S. & Caruthers, M. H. Biochemical and physicochemical properties of phosphorodithioate DNA. *Biochemistry* **35,** 8734–8741 (1996).

202. King, D. J., Ventura, D. A., Brasier, A. R. & Gorenstein, D. G. Novel combinatorial selection of phosphorothioate oligonucleotide aptamers. *Biochemistry* **37,** 16489–16493 (1998).

203. Deleavey, G. F. & Damha, M. J. Designing chemically modified oligonucleotides for targeted gene silencing. *Chem. Biol.* **19,** 937–954 (2012).

204. Wan, W. B. *et al.* Synthesis, biophysical properties and biological activity of second generation antisense oligonucleotides containing chiral phosphorothioate linkages. *Nucleic Acids Res.* **42,** 13456–13468 (2014).

205. Oka, N. & Wada, T. Stereocontrolled synthesis of oligonucleotide analogs containing chiral internucleotidic phosphorus atoms. *Chem. Soc. Rev.* **40,** 5829–5843 (2011).

206. Kang, J., Lee, M. S., Copland, J. A. 3rd, Luxon, B. A. & Gorenstein, D. G. Combinatorial selection of a single stranded DNA thioaptamer targeting TGF-beta1 protein. *Bioorg. Med. Chem. Lett.* **18,** 1835–1839 (2008).

207. Kang, J., Lee, M. S., Watowich, S. J. & Gorenstein, D. G. Combinatorial selection of a RNA thioaptamer that binds to Venezuelan equine encephalitis virus capsid protein. *FEBS Lett.* **581,** 2497–2502 (2007).

208. Zaitseva, M. *et al.* Conformation and thermostability of oligonucleotide d(GGTTGGTGTGGTTGG) containing thiophosphoryl internucleotide bonds at different positions. *Biophys. Chem.* **146,** 1–6 (2010).

209. Samani, T. D., Jolles, B. & Laigle, A. Best minimally modified antisense oligonucleotides according to cell nuclease activity. *Antisense Nucleic Acid Drug Dev.* **11,** 129–136 (2001).

210. Somasunderam, A. *et al.* Combinatorial selection, inhibition, and antiviral activity of DNA thioaptamers targeting the RNase H domain of HIV-1 reverse transcriptase. *Biochemistry* **44,** 10388–10395 (2005).

211. Varizhuk, A. M. *et al.* Synthesis, characterization and in vitro activity of thrombin-binding DNA aptamers with triazole internucleotide linkages. *Eur. J. Med. Chem.* **67,** 90–97 (2013).

212. Lato, S. M. *et al.* Boron-containing aptamers to ATP. *Nucleic Acids Res.* **30,** 1401–1407 (2002).

213. Varizhuk, A. M. *et al.* Synthesis of triazole-linked oligonucleotides with high affinity to DNA complements and an analysis of their compatibility with biosystems. *J. Org. Chem.* **78,** 5964–5969 (2013).

214. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467 (1977).

215. Paolillo, C., Londin, E. & Fortina, P. Next generation sequencing in cancer: opportunities and challenges for precision cancer medicine. *Scand. J. Clin. Lab. Invest. Suppl.* **245,** S84-91 (2016).

216. Zhao, W. *et al.* Biomarker identification from next-generation sequencing data for pathogen bacteria characterization and surveillance. *Biomark. Med.* **9,** 1253–1264 (2015).

217. Lopez, J. P. *et al.* Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med. Genomics* **8,** 35 (2015).

218. Schütze, T. *et al.* Probing the SELEX Process with Next-Generation Sequencing. *PLoS One* **6,** e29604 (2011).

219. Nguyen Quang, N., Perret, G. & Ducongé, F. Applications of High-Throughput Sequencing for In Vitro Selection and Characterization of Aptamers. *Pharmaceuticals* **9,** (2016).

220. Yuce, M., Ullah, N. & Budak, H. Trends in aptamer selection methods and applications. *Analyst* **140,** 5379–5399 (2015).

221. Hoon, S., Zhou, B., Janda, K. D., Brenner, S. & Scolnick, J. Aptamer selection by high-throughput sequencing and informatic analysis. *Biotechniques* **51,** 413–416 (2011).

222. Cho, M. *et al.* Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15373–15378 (2010).

223. Cho, M. *et al.* Quantitative selection and parallel characterization of aptamers. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18460–18465 (2013).

224. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome Biol.* **14,** R51 (2013).

225. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).

226. Illumina. An introduction to Next-Generation Sequencing Technology (2007). at <https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf>

227. Caruthers, M. H. Chemical synthesis of DNA and DNA analogs. *Acc. Chem. Res.* **24,** 278–284 (1991).

228. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22,** 1859–1862 (1981).

229. Singh, Y., Murat, P. & Defrancq, E. Recent developments in oligonucleotide conjugation. *Chem. Soc. Rev.* **39,** 2054–2070 (2010).

230. Iwasawa, T., Wash, P., Gibson, C. & Rebek, J. Reaction of an introverted carboxylic acid with carbodiimide. *Tetrahedron* **63,** 6506–6511 (2007).

231. Konig, W. & Geiger, R. [A new method for synthesis of peptides: activation of the carboxyl group with dicyclohexylcarbodiimide using 1-hydroxybenzotriazoles as additives]. *Chem. Ber.* **103,** 788–798 (1970).

232. Valeur, E. & Bradley, M. Amide bond formation: beyond the myth of coupling reagents. *Chem. Soc. Rev.* **38,** 606–631 (2009).

233. Al-Warhi, T. I., Al-Hazimi, H. M. A. & El-Faham, A. Recent development in peptide coupling reagents. *J. Saudi Chem. Soc.* **16,** 97–116 (2012).

234. Carter, J. D. & Labean, T. H. Coupling strategies for the synthesis of Peptide-oligonucleotide conjugates for patterned synthetic biomineralization. *J. Nucleic Acids* **2011,** 926595 (2011).

235. Burgess, K. & Cook, D. Syntheses of Nucleoside Triphosphates. *Chem. Rev.* **100,** 2047–2060 (2000).

236. Yoshikawa, M., Kato, T. & Takenishi, T. A novel method for phosphorylation of nucleosides to 5′-nucleotides. *Tetrahedron Lett.* **8,** 5065–5068 (1967).

237. Mishra, N. C. & Broom, A. D. A novel synthesis of nucleoside 5′-triphosphates. *J. Chem. Soc. Chem. Commun.* 1276–1277 (1991). doi:10.1039/C39910001276

238. Roy, B., Depaix, A., Périgaud, C. & Peyrottes, S. Recent Trends in Nucleotide Synthesis. *Chem. Rev.* **116,** 7854–7897 (2016).

239. Kovács, T. & Ötvös, L. Simple synthesis of 5-vinyl- and5-ethynyl-2′-deoxyuridine-5′-triphosphates. *Tetrahedron Lett.* **29,** 4525–4528 (1988).

240. Gillerman, I. & Fischer, B. An Improved One-Pot Synthesis of Nucleoside 5′-Triphosphate Analogues. *Nucleosides, Nucleotides and Nucleic Acids* **29,** 245–256 (2010).

241. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54,** 631–635 (1989).

242. Hollenstein, M. Synthesis of Deoxynucleoside Triphosphates that Include Proline, Urea, or Sulfonamide Groups and Their Polymerase Incorporation into DNA. *Chem. – A Eur. J.* **18,** 13320–13330 (2012).

243. Hollenstein, M., Wojciechowski, F. & Leumann, C. J. Polymerase incorporation of pyrene-nucleoside triphosphates. *Bioorg. Med. Chem. Lett.* **22,** 4428–4430 (2012).

244. Caton-Williams, J., Lin, L., Smith, M. & Huang, Z. Convenient synthesis of nucleoside 5′-triphosphates for RNA transcription. *Chem. Commun.* **47,** 8142–8144 (2011).

245. Caton-Williams, J., Smith, M., Carrasco, N. & Huang, Z. Protection-free one-pot synthesis of 2'-deoxynucleoside 5'-triphosphates and DNA polymerization. *Org. Lett.* **13,** 4156–4159 (2011).

246. Vashishtha, A. K., Wang, J. & Konigsberg, W. H. Different Divalent Cations Alter the Kinetics and Fidelity of DNA Polymerases. *J. Biol. Chem.* **291,** 20869–20875 (2016).

247. Xia, S., Wang, M., Blaha, G., Konigsberg, W. H. & Wang, J. Structural insights into complete metal ion coordination from ternary complexes of B family RB69 DNA polymerase. *Biochemistry* **50,** 9114–9124 (2011).

248. Filee, J., Forterre, P., Sen-Lin, T. & Laurent, J. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J. Mol. Evol.* **54,** 763–773 (2002).

249. Vashishtha, A. K. & Kuchta, R. D. Polymerase and exonuclease activities in herpes simplex virus type 1 DNA polymerase are not highly coordinated. *Biochemistry* **54,** 240–249 (2015).

250. Rothwell, P. J. & Waksman, G. Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71,** 401–440 (2005).

251. Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3,** 31–38 (1993).

252. Beese, L. S. & Steitz, T. A. Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10,** 25–33 (1991).

253. Xia, S., Wang, J. & Konigsberg, W. H. DNA mismatch synthesis complexes provide insights into base selectivity of a B family DNA polymerase. *J. Am. Chem. Soc.* **135,** 193–202 (2013).

254. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H. & Kraut, J. A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta. *Biochemistry* **35,** 12762–12777 (1996).

255. Wang, J. *et al.* Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell* **89,** 1087–1099 (1997).

256. Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8,** 54–63 (1998).

257. Jäger, S. *et al.* A Versatile Toolbox for Variable DNA Functionalization at High Density. *J. Am. Chem. Soc.* **127,** 15071–15082 (2005).

258. Kuwahara, M. *et al.* Systematic characterization of 2′-deoxynucleoside- 5′-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. *Nucleic Acids Res.* **34,** 5383–5394 (2006).

259. Chen, T. & Romesberg, F. E. Directed polymerase evolution. *FEBS Lett.* **588,** 219–229 (2014).

260. Takagi, M. *et al.* Characterization of DNA polymerase from Pyrococcus sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* **63,** 4504–4510 (1997).

261. Kuroita, T. *et al.* Structural Mechanism for Coordination of Proofreading and Polymerase Activities in Archaeal DNA Polymerases. *J. Mol. Biol.* **351,** 291–298 (2005).

262. Kuwahara, M. *et al.* Study on suitability of KOD DNA polymerase for enzymatic production of artificial nucleic acids using base/sugar modified nucleoside triphosphates. *Molecules* **15,** 8229–8240 (2010).

263. Gardner, A. F. & Jack, W. E. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* **27,** 2545–2553 (1999).

264. Liu, E., Lam, C. H. & Perrin, D. M. Synthesis and Enzymatic Incorporation of Modified Deoxyuridine Triphosphates. *Molecules* **20,** 13591–13602 (2015).

265. Kong, H., Kucera, R. B. & Jack, W. E. Characterization of a DNA polymerase from the hyperthermophile archaea Thermococcus litoralis. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J. Biol. Chem.* **268,** 1965–1975 (1993).

266. Southworth, M. W. *et al.* Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on Thermococcus sp. 9 degrees N-7 and mutations affecting 3’-5’ exonuclease activity. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 5281–5285 (1996).

267. Fa, M., Radeghieri, A., Henry, A. A. & Romesberg, F. E. Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* **126,** 1748–1754 (2004).

268. Ong, J. L., Loakes, D., Jaroslawski, S., Too, K. & Holliger, P. Directed Evolution of DNA Polymerase, RNA Polymerase and Reverse Transcriptase Activity in a Single Polypeptide. *J. Mol. Biol.* **361,** 537–550 (2006).

269. Shinkai, A., Patel, P. H. & Loeb, L. A. The conserved active site motif A of Escherichia coli DNA polymerase I is highly mutable. *J. Biol. Chem.* **276,** 18836–18842 (2001).

270. Brown, J. A. & Suo, Z. Unlocking the sugar 'steric gate' of DNA polymerases. *Biochemistry* **50,** 1135–1142 (2011).

271. Ono, T., Scalf, M. & Smith, L. M. 2'-Fluoro modified nucleic acids: polymerase-directed synthesis, properties and stability to analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res.* **25,** 4581–4588 (1997).

272. Richardson, F. C., Kuchta, R. D., Mazurkiewicz, A. & Richardson, K. A. Polymerization of 2'-fluoro- and 2'-O-methyl-dNTPs by human DNA polymerase alpha, polymerase gamma, and primase. *Biochem. Pharmacol.* **59,** 1045–1052 (2000).

273. Tjian, R. The biochemistry of transcription in eukaryotes: a paradigm for multisubunit regulatory complexes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **351,** 491–499 (1996).

274. Tracy, R. L. & Stern, D. B. Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr. Genet.* **28,** 205–216 (1995).

275. von Hippel, P. H., Bear, D. G., Morgan, W. D. & McSwiggen, J. A. Protein-Nucleic Acid Interactions in Transcription: A Molecular Analysis. *Annu. Rev. Biochem.* **53,** 389–446 (1984).

276. Ivanov, S. A., Welz, R., Gottikh, M. B. & Müller, S. RNA Synthesis by T7 RNA Polymerase Supported Primer Extension. *Mol. Biol.* **38,** 674–679 (2004).

277. McAllister, W. T. & Raskin, C. A. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol. Microbiol.* **10,** 1–6 (1993).

278. Sousa, R. Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem. Sci.* **21,** 186–190 (1996).

279. Cermakian, N., Ikeda, T. M., Cedergren, R. & Gray, M. W. Sequences Homologous to Yeast Mitochondrial and Bacteriophage T3 and T7 RNA Polymerases Are Widespread Throughout the Eukaryotic Lineage. *Nucleic Acids Res.* **24,** 648–654 (1996).

280. Tunitskaya, V. L. & Kochetkov, S. N. Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochemistry. (Mosc).* **67,** 1124–1135 (2002).

281. Kochetkov, S. N., Rusakova, E. E. & Tunitskaya, V. L. Recent studies of T7 RNA polymerase mechanism. *FEBS Lett.* **440,** 264–267 (1998).

282. Aurup, H., Williams, D. M. & Eckstein, F. 2'-Fluoro- and 2'-amino-2'-deoxynucleoside 5'-triphosphates as substrates for T7 RNA polymerase. *Biochemistry* **31,** 9636–9641 (1992).

283. Conrad, F., Hanne, A., Gaur, R. K. & Krupp, G. Enzymatic synthesis of 2'-modified nucleic acids: identification of important phosphate and ribose moieties in RNase P substrates. *Nucleic Acids Res.* **23,** 1845–1853 (1995).

284. Raines, K. & Gottlieb, P. A. Enzymatic incorporation of 2'-thio-CTP into the HDV ribozyme. *RNA* **4,** 340–345 (1998).

285. Pavlov, A. R., Pavlova, N. V, Kozyavkin, S. A. & Slesarev, A. I. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol.* **22,** 253–260 (2004).

286. Beckman, R. A., Mildvan, A. S. & Loeb, L. A. On the fidelity of DNA replication: manganese mutagenesis in vitro. *Biochemistry* **24,** 5810–5817 (1985).

287. Mullis, K. B. *et al.* Process for amplifying, detecting, and/or-cloning nucleic acid sequences. (1987).

288. Lapa, S. A., Chudinov, A. V & Timofeev, E. N. The Toolbox for Modified Aptamers. *Mol. Biotechnol.* **58,** 79–92 (2016).

289. Sokoloski, J. E., Dombrowski, S. E. & Bevilacqua, P. C. Thermodynamics of ligand binding to a heterogeneous RNA population in the malachite green aptamer. *Biochemistry* **51,** 565–572 (2012).

290. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077,** 101–113 (2014).

291. Quang, N. N., Miodek, A., Cibiel, A. & Duconge, F. Selection of Aptamers Against Whole Living Cells: From Cell-SELEX to Identification of Biomarkers. *Methods Mol. Biol.* **1575,** 253–272 (2017).

292. Patching, S. G. Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery. *Biochim. Biophys. Acta - Biomembr.* **1838,** 43–55 (2014).

293. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2,** 1849–1861 (2007).

294. Tang, Y., Zeng, X. & Liang, J. Surface Plasmon Resonance: An Introduction to a Surface Spectroscopy Technique. *J. Chem. Educ.* **87,** 742–746 (2010).

295. Cooper, M. A. Optical biosensors in drug discovery. *Nat. Rev. Drug Discov.* **1,** 515–528 (2002).

296. Raghavan, M. & Bjorkman, P. J. BIAcore: a microchip-based system for analyzing the formation of macromolecular complexes. *Structure* **3,** 331–333 (1995).

297. Karlsson, R. SPR for molecular interaction analysis: a review of emerging application areas. *J. Mol. Recognit.* **17,** 151–161 (2004).

298. Homola, J. Surface Plasmon Resonance Sensors for Detection of Chemical and Biological Species. *Chem. Rev.* **108,** 462–493 (2008).

299. Garner, M. M. & Revzin, A. The use of gel electrophoresis to detect and study nucleic acid— protein interactions. *Trends Biochem. Sci.* **11,** 395–396 (1986).

300. Fried, M. G. Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* **10,** 366–376 (1989).

301. Lane, D., Prentki, P. & Chandler, M. Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiol. Rev.* **56,** 509–528 (1992).

302. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 560–564 (1977).

303. Rye, H. S., Drees, B. L., Nelson, H. C. & Glazer, A. N. Stable fluorescent dye-DNA complexes in high sensitivity detection of protein-DNA interactions. Application to heat shock transcription factor. *J. Biol. Chem.* **268,** 25229–25238 (1993).

304. Forwood, J. K. & Jans, D. A. Quantitative analysis of DNA-protein interactions using double-labeled native gel electrophoresis and fluorescence-based imaging. *Electrophoresis* **27,** 3166–3170 (2006).

305. Kang, J., Lee, M. S. & Gorenstein, D. G. Quantitative analysis of chemiluminescence signals using a cooled charge-coupled device camera. *Anal. Biochem.* **345,** 66–71 (2005).

306. Li, Y., Jiang, Z., Chen, H. & Ma, W.-J. A modified quantitative EMSA and its application in the study of RNA--protein interactions. *J. Biochem. Biophys. Methods* **60,** 85–96 (2004).

307. Fried, M. G. & Daugherty, M. A. Electrophoretic analysis of multiple protein-DNA interactions. *Electrophoresis* **19,** 1247–1253 (2005).

308. Rasimas, J. J., Kar, S. R., Pegg, A. E. & Fried, M. G. Interactions of human O6-alkylguanine-DNA alkyltransferase (AGT) with short single-stranded DNAs. *J. Biol. Chem.* **282,** 3357–3366 (2007).

309. Musso, M., Bianchi-Scarrà, G. & Van Dyke, M. W. The yeast CDP1 gene encodes a triple-helical DNA-binding protein. *Nucleic Acids Res.* **28,** 4090–4096 (2000).

310. Tolstonog, G. V, Li, G., Shoeman, R. L. & Traub, P. Interaction in vitro of type III intermediate filament proteins with higher order structures of single-stranded DNA, particularly with G-quadruplex DNA. *DNA Cell Biol.* **24,** 85–110 (2005).

311. Nordheim, A. & Meese, K. Topoisomer gel retardation: detection of anti-Z-DNA antibodies bound to Z-DNA within supercoiled DNA minicircles. *Nucleic Acids Res.* **16,** 21–37 (1988).

312. Holden, N. S. & Tacon, C. E. Principles and problems of the electrophoretic mobility shift assay. *J. Pharmacol. Toxicol. Methods* **63,** 7–14 (2011).

313. Alves, C. & Cunha, C. in *Gel Electrophoresis-Advanced techniques* (ed. Magdeldin, S.) 205–228 (InTech, 2012). doi:10.5772/37619

314. Fried, M. G. & Liu, G. Molecular sequestration stabilizes CAP-DNA complexes during polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **22,** 5054–5059 (1994).

315. Fried, M. G. & Bromberg, J. L. Factors that affect the stability of protein-DNA complexes during gel electrophoresis. *Electrophoresis* **18,** 6–11 (1997).

316. Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9,** 6505–6525 (1981).

317. Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* **9,** 3047–3060 (1981).

318. Berman, J., Eisenberg, S. & Tye, B. K. An agarose gel electrophoresis assay for the detection of DNA-binding activities in yeast cell extracts. *Methods Enzymol.* **155,** 528–537 (1987).

319. Thirunavukarasu, D., Chen, T., Liu, Z., Hongdilokkul, N. & Romesberg, F. E. Selection of 2′-Fluoro-Modified Aptamers with Optimized Properties. *J. Am. Chem. Soc.* **139,** 2892–2895 (2017).

320. Wilson, C. & Keefe, A. D. Building oligonucleotide therapeutics using non-natural chemistries. *Curr. Opin. Chem. Biol.* **10,** 607–614 (2006).

321. Lin, Y., Qiu, Q., Gill, S. C. & Jayasena, S. D. Modified RNA sequence pools for in vitro selection. *Nucleic Acids Res.* **22,** 5229–5234 (1994).

322. Dube, D. K. & Loeb, L. A. Manganese as a mutagenic agent during in vitro DNA synthesis. *Biochem. Biophys. Res. Commun.* **67,** 1041–1046 (1975).

323. Bock, C. W., Katz, A. K., Markham, G. D. & Glusker, J. P. Manganese as a Replacement for Magnesium and Zinc: Functional Comparison of the Divalent Ions. *J. Am. Chem. Soc.* **121,** 7360–7372 (1999).

324. Miyaki, M., Murata, I., Osabe, M. & Ono, T. Effect of metal cations on misincorporation by E. coli DNA polymerases. *Biochem. Biophys. Res. Commun.* **77,** 854–860 (1977).

325. Goodman, M. F., Keener, S., Guidotti, S. & Branscomb, E. W. On the enzymatic basis for mutagenesis by manganese. *J. Biol. Chem.* **258,** 3469–3475 (1983).

326. Vashishtha, A. K. & Konigsberg, W. H. Effect of Different Divalent Cations on the Kinetics and Fidelity of RB69 DNA Polymerase. *Biochemistry* **55,** 2661–2670 (2016).

327. Wang, W., Hellinga, H. W. & Beese, L. S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17644–17648 (2011).

328. Kunkel, T. A. & Soni, A. Exonucleolytic proofreading enhances the fidelity of DNA synthesis by chick embryo DNA polymerase-gamma. *J. Biol. Chem.* **263,** 4450–4459 (1988).

329. Doetsch, P. W., Chan, G. L. & Haseltine, W. A. T4 DNA polymerase (3'-5') exonuclease, an enzyme for the detection and quantitation of stable DNA lesions: the ultraviolet light example. *Nucleic Acids Res.* **13,** 3285–3304 (1985).

330. Hori, K., Mark, D. F. & Richardson, C. C. Deoxyribonucleic acid polymerase of bacteriophage T7. Characterization of the exonuclease activities of the gene 5 protein and the reconstituted polymerase. *J. Biol. Chem.* **254,** 11598–11604 (1979).

331. Joyce, C. M. How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **264,** 10858–10866 (1989).

332. Eremeeva, E., Abramov, M., Margamuljana, L. & Herdewijn, P. Base-Modified Nucleic Acids as a Powerful Tool for Synthetic Biology and Biotechnology. *Chemistry* **23,** 9560–9576 (2017).

333. Veedu, R. N., Vester, B. & Wengel, J. Enzymatic Incorporation of LNA Nucleotides into DNA Strands. *ChemBioChem* **8,** 490–492 (2007).

334. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chem. Soc. Rev.* **39,** 1388–1405 (2010).

335. Shivalingam, A. & Brown, T. Synthesis of chemically modified DNA. *Biochem. Soc. Trans.* **44,** 709–715 (2016).

336. 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 LP-11343 (2011).

337. Glebs, A., Stencel, E. & Knoche, K. The 'Go-To' Choice: GoTaq™ DNA Polymerase. *Promega Corporation* Promega Notes 83:21–24 (2003).

338. Joyce, C. M. & Grindley, N. D. Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 1830–1834 (1983).

339. Rychlik, W., Spencer, W. J. & Rhoads, R. E. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **18,** 6409–6412 (1990).

340. Zhao, G., Tang, S., Li, J., Hu, T. & Guan, Y. Effects of cations on small fragment of DNA polymerase I using a novel FRET assay. *Acta Biochim. Biophys. Sin. (Shanghai).* **46,** 659–667 (2014).

341. Yang, W., Lee, J. Y. & Nowotny, M. Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. *Mol. Cell* **22,** 5–13 (2006).

342. Fei, L. *et al.* Structural and Biochemical Studies of a Moderately Thermophilic Exonuclease I from Methylocaldum szegediense. *PLoS One* **10,** e0117470 (2015).

343. Summer, H., Grämer, R. & Dröge, P. Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *J. Vis. Exp.* 1485 (2009). doi:10.3791/1485

344. Clark, J. M., Joyce, C. M. & Beardsley, G. P. Novel blunt-end addition reactions catalyzed by DNA polymerase I of Escherichia coli. *J. Mol. Biol.* **198,** 123–127 (1987).

345. Clark, J. M. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16,** 9677–9686 (1988).

346. Güixens-Gallardo, P., Hocek, M. & Perlíková, P. Inhibition of non-templated nucleotide addition by DNA polymerases in primer extension using twisted intercalating nucleic acid modified templates. *Bioorg. Med. Chem. Lett.* **26,** 288–291 (2016).

347. Obayashi, T. *et al.* Enzymatic synthesis of labeled DNA by PCR using new fluorescent thymidine nucleotide analogue and superthermophilic KOD dash DNA polymerase. *Bioorg. Med. Chem. Lett.* **12,** 1167–1170 (2002).

348. Berg, J. M., Tymoczko, J. L. & Stryer L. *Biochemistry*. 5th edition. New York: W H Freeman (2002) Section 8.4, The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22430/

349. Langer, A. *et al.* Polymerase/DNA interactions and enzymatic activity: multi-parameter analysis with electro-switchable biosurfaces. *Sci. Rep.* **5,** 12066 (2015).

350. Tabor, S. & Richardson, C. C. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 4076–4080 (1989).

351. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8067–8072 (2012).

352. Cox, W. G. & Singer, V. L. Fluorescent DNA hybridization probe preparation using amine modification and reactive dye coupling. *Biotechniques* **36,** 114–122 (2004).

353. Dellafiore, M. A., Montserrat, J. M. & Iribarren, A. M. Modified Nucleoside Triphosphates for In-vitro Selection Techniques. *Front. Chem.* **4,** 18 (2016).

354. Gupta, S. *et al.* Pharmacokinetic Properties of DNA Aptamers with Base Modifications. *Nucleic Acid Ther.* **27,** 345–353 (2017).

355. Lee, S. E. *et al.* Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Res.* **29,** 1565–1573 (2001).

356. Wagner, R. W. *et al.* Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science* **260,** 1510–1513 (1993).

357. Diezmann, F., Eberhard, H. & Seitz, O. Native chemical ligation in the synthesis of internally modified oligonucleotide-peptide conjugates. *Biopolymers* **94,** 397–404 (2010).

358. Borsenberger, V., Kukwikila, M. & Howorka, S. Synthesis and enzymatic incorporation of modified deoxyuridine triphosphates. *Org. Biomol. Chem.* **7,** 3826–3835 (2009).

359. Ren, X., Gerowska, M., El-Sagheer, A. H. & Brown, T. Enzymatic incorporation and fluorescent labelling of cyclooctyne-modified deoxyuridine triphosphates in DNA. *Bioorg. Med. Chem.* **22,** 4384–4390 (2014).

360. Obeid, S., Baccaro, A., Welte, W., Diederichs, K. & Marx, A. Structural basis for the synthesis of nucleobase modified DNA by Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21327 LP-21331 (2010).

361. Janeway, C.A Jr; Travers, P; Walport, M; Schlomchik, M. in (Garland Science, 2001).

362. Peng, H.-P., Lee, K. H., Jian, J.-W. & Yang, A.-S. Origins of specificity and affinity in antibody-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E2656–E2665 (2014).

363. Desiraju, G. R. *et al.* Definition of the halogen bond (IUPAC Recommendations 2013). *Pure and Applied Chemistry* **85,** 1711 (2013).

364. Metrangolo, P. & Resnati, G. Halogen bonding: a paradigm in supramolecular chemistry. *Chemistry* **7,** 2511–2519 (2001).

365. Eskandari, K. & Lesani, M. Does Fluorine Participate in Halogen Bonding? *Chem. – A Eur. J.* **21,** 4739–4746 (2015).

366. Wilcken, R., Zimmermann, M. O., Lange, A., Joerger, A. C. & Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **56,** 1363–1388 (2013).

367. Ding, X., Tuikka, M. & Haukka, M (2012). Halogen Bonding in Crystal Engineering, Recent Advances in Crystallography, Jason B. Benedict, IntechOpen, DOI: 10.5772/48592. Available from: https://www.intechopen.com/books/recent-advances-in-crystallography/halogen-bonding-in-crystal-engineering

368. Njoroge, F. G. *et al.* Enhanced FTase activity achieved via piperazine interaction with catalytic zinc. *Bioorg. Med. Chem. Lett.* **16,** 984–988 (2006).

369. Maignan, S. *et al.* Molecular structures of human factor Xa complexed with ketopiperazine inhibitors: preference for a neutral group in the S1 pocket. *J. Med. Chem.* **46,** 685–690 (2003).

370. Warmus, J. S. *et al.* 2-Alkylamino- and alkoxy-substituted 2-amino-1,3,4-oxadiazoles-O-Alkyl benzohydroxamate esters replacements retain the desired

inhibition and selectivity against MEK (MAP ERK kinase). *Bioorg. Med. Chem. Lett.* **18,** 6171–6174 (2008).

371. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

372. Gilday, L. C. *et al.* Halogen Bonding in Supramolecular Chemistry. *Chem. Rev.* **115,** 7118–7195 (2015).

373. Scholfield, M. R., Zanden, C. M. Vander, Carter, M. & Ho, P. S. Halogen bonding (X-bonding): a biological perspective. *Protein Sci.* **22,** 139–152 (2013).

374. Riley, K. E. & Hobza, P. Strength and Character of Halogen Bonds in Protein–Ligand Complexes. *Cryst. Growth Des.* **11,** 4272–4278 (2011).

375. Sirimulla, S., Bailey, J. B., Vegesna, R. & Narayan, M. Halogen Interactions in Protein–Ligand Complexes: Implications of Halogen Bonding for Rational Drug Design. *J. Chem. Inf. Model.* **53,** 2781–2791 (2013).

376. Kolář, M. H. & Tabarrini, O. Halogen Bonding in Nucleic Acid Complexes. *J. Med. Chem.* **60,** 8681–8690 (2017).

377. Hays, F. A., Vargason, J. M. & Ho, P. S. Effect of Sequence on the Conformation of DNA Holliday Junctions. *Biochemistry* **42,** 9586–9597 (2003).

378. Voth, A. R., Hays, F. A. & Ho, P. S. Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6188–6193 (2007).

379. Carter, M. & Ho, P. S. Assaying the Energies of Biological Halogen Bonds. *Cryst. Growth Des.* **11,** 5087–5095 (2011).

380. Carter, M. *et al.* Enthalpy–Entropy Compensation in Biomolecular Halogen Bonds Measured in DNA Junctions. *Biochemistry* **52,** 4891–4903 (2013).

381. Vander Zanden, C. M., Carter, M. & Ho, P. S. Determining thermodynamic properties of molecular interactions from single crystal studies. *Methods* **64,** 12–18 (2013).

382. Robertson, C. C., Perutz, R. N., Brammer, L. & Hunter, C. A. A solvent-resistant halogen bond. *Chem. Sci.* **5,** 4179–4183 (2014).

383. Robertson, C. C. *et al.* Hydrogen bonding vs. halogen bonding: the solvent decides. *Chem. Sci.* **8,** 5392–5398 (2017).

384. Langton, M. J., Robinson, S. W., Marques, I., Félix, V. & Beer, P. D. Halogen bonding in water results in enhanced anion recognition in acyclic and rotaxane hosts. *Nat. Chem.* **6,** 1039 (2014).

385. Chinchilla, R. & Nájera, C. Recent advances in Sonogashira reactions. *Chem. Soc. Rev.* **40,** 5084–5121 (2011).

386. Korhonen, H. J., Bolt, H. L., Vicente-Gines, L., Perks, D. C. & Hodgson, D. R. W. PPN Pyrophosphate: A New Reagent for the Preparation of Nucleoside Triphosphates. *Phosphorus, Sulfur Silicon Relat. Elem.* **190,** 758–762 (2015).

387. Kore, A. R., Senthilvelan, A. & Shanmugasundaram, M. Highly Regioselective C-5 Iodination of Pyrimidine Nucleotides and Subsequent Chemoselective Sonogashira Coupling with Propargylamine. *Nucleosides, Nucleotides and Nucleic Acids* **34,** 92–102 (2015).

388. Kumar, R., Wiebe, L. I. & Knaus, E. E. A mild and efficient methodology for the synthesis of 5-halogeno uracil nucleosides that occurs via a 5-halogeno-6-azido-5,6-dihydro intermediate. *Can. J. Chem.* **72,** 2005–2010 (1994).

389. Kore, A. R., Senthilvelan, A. & Shanmugasundaram, M. Highly chemoselective palladium-catalyzed Sonogashira coupling of 5-iodouridine-5′-triphosphates with propargylamine: A new efficient method for the synthesis of 5-aminopropargyl-uridine-5′-triphosphates. *Tetrahedron Lett.* **53,** 3070–3072 (2012).

390. Sonogashira, K. Development of Pd–Cu catalyzed cross-coupling of terminal acetylenes with sp2-carbon halides. *J. Organomet. Chem.* **653,** 46–49 (2002).

391. Hojo, H. *et al.* Efficient preparation of Fmoc-aminoacyl-N-ethylcysteine unit, a key device for the synthesis of peptide thioesters. *Org. Biomol. Chem.* **9,** 6807–6813 (2011).

392. Windridge, G. & Jorgensen, E. C. 1-Hydroxybenzotriazole as a racemization-suppressing reagent for the incorporation of im-benzyl-L-histidine into peptides. *J. Am. Chem. Soc.* **93,** 6318–6319 (1971).

393. Li, Y. *et al.* Optimized Reaction Conditions for Amide Bond Formation in DNA-Encoded Combinatorial Libraries. *ACS Comb. Sci.* **18,** 438–443 (2016).

394. Barsoum, D. N. *et al.* Synthesis of 5-Iodo-1,2,3-triazoles from organic azides and terminal alkynes: Ligand acceleration effect, substrate scope, and mechanistic insights. *Synth.* **45,** 2372–2386 (2013).

395. Cherkasov, D., Biet, T., Bäuml, E., Traut, W. & Lohoff, M. New Nucleotide Analogues with Enhanced Signal Properties. *Bioconjug. Chem.* **21,** 122–129 (2010).

396. Schechter, A. L. *et al.* The neu gene: an erbB-homologous gene distinct from and unlinked to the gene encoding the EGF receptor. *Science* **229,** 976–978 (1985).

397. Cho, H.-S. *et al.* Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421,** 756 (2003).

398. Wolf-Yadlin, A. *et al.* Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* **2,** 54 (2006).

399. Rubin, I. & Yarden, Y. The basic biology of HER2. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **12 Suppl 1,** S3-8 (2001).

400. Yarden, Y. Biology of HER2 and Its Importance in Breast Cancer. *Oncology* **61 Suppl 2,** 1–13 (2001).

401. Serra, V. *et al.* PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer. *Oncogene* **30,** 2547–2557 (2011).

402. Peckys, D. B., Korf, U. & de Jonge, N. Local variations of HER2 dimerization in breast cancer cells discovered by correlative fluorescence and liquid electron microscopy. *Sci. Adv.* **1,** e1500165–e1500165 (2015).

403. Pal, S. K. & Pegram, M. HER2 targeted therapy in breast cancer...beyond Herceptin. *Rev. Endocr. Metab. Disord.* **8,** 269–277 (2007).

404. Frejd, F. Y. & Kim, K.-T. Affibody molecules as engineered protein drugs. *Exp. Mol. Med.* **49,** e306–e306 (2017).

405. Agus, D. B. *et al.* Targeting ligand-activated ErbB2 signaling inhibits breast and prostate tumor growth. *Cancer Cell* **2,** 127–137 (2002).

406. Rockberg, J., Schwenk, J. M. & Uhlen, M. Discovery of epitopes for targeting the human epidermal growth factor receptor 2 (HER2) with antibodies. *Mol. Oncol.* **3,** 238–247 (2009).

407. Franklin, M. C. *et al.* Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* **5,** 317–328 (2004).

408. Lee, S. B. *et al.* Affibody Molecules for In vivo Characterization of HER2-Positive Tumors by Near-Infrared Imaging. *Clin. Cancer Res.* **14,** 3840 LP-3849 (2008).

409. Tolmachev, V. *et al.* Radionuclide Therapy of HER2-Positive Microxenografts using a 177Lu-Labeled HER2-Specific Affibody Molecule. *Cancer Res.* **67,** 2773 LP-2782 (2007).

410. Ekerljung, L. *et al.* Dimeric HER2-specific affibody molecules inhibit proliferation of the SKBR-3 breast cancer cell line. *Biochem. Biophys. Res. Commun.* **377,** 489–494 (2008).

411. Baum, R. P. *et al.* Molecular imaging of HER2-expressing malignant tumors in breast cancer patients using synthetic 111In- or 68Ga-labeled affibody molecules. *J. Nucl. Med.* **51,** 892–897 (2010).

412. Eigenbrot, C., Ultsch, M., Dubnovitsky, A., Abrahmsen, L. & Hard, T. Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15039–15044 (2010).

413. Kang, H., Min Huh, Y., Kim, S. & Lee, D. K. Isolation of RNA aptamers targeting HER-2-overexpressing breast cancer cells using Cell-SELEX. *Bulletin of the Korean Chemical Society* **30,** (2009).

414. Kim, M. Y. & Jeong, S. In vitro selection of RNA aptamer and specific targeting of ErbB2 in breast cancer cells. *Nucleic Acid Ther.* **21,** 173–178 (2011).

415. Thiel, K. W. *et al.* Delivery of chemo-sensitizing siRNAs to HER2+-breast cancer cells using RNA aptamers. *Nucleic Acids Res.* **40,** 6319–6337 (2012).

416. Gupta, S. *et al.* Rapid histochemistry using slow off-rate modified aptamers with anionic competition. *Appl. Immunohistochem. Mol. Morphol. AIMM* **19,** 273–278 (2011).

417. Liu, Z. *et al.* Novel HER2 aptamer selectively delivers cytotoxic drug to HER2-positive breast cancer cells in vitro. *J. Transl. Med.* **10,** 148 (2012).

418. Mahlknecht, G. *et al.* Aptamer to ErbB-2/HER2 enhances degradation of the target and inhibits tumorigenic growth. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 8170–8175 (2013).

419. Chu, M. *et al.* Evaluation of human epidermal growth factor receptor 2 in breast cancer with a novel specific aptamer. *Cell. Mol. Immunol.* **14,** 398 (2015).

420. Vorobyeva, M. A., Davydova, A. S., Vorobjev, P. E. & Venyaminova, A. G. Key Aspects of Nucleic Acid Library Design for in Vitro Selection. *Int. J. Mol. Sci.* **19,** (2018).

421. Cowperthwaite, M. C. & Ellington, A. D. Bioinformatic analysis of the contribution of primer sequences to aptamer structures. *J. Mol. Evol.* **67,** 95–102 (2008).

422. Hall, B. *et al.* Design, synthesis, and amplification of DNA pools for in vitro selection. *Curr. Protoc. Mol. Biol.* **Chapter 24,** Unit 24.2 (2009).

423. Zimmermann, B., Bilusic, I., Lorenz, C. & Schroeder, R. Genomic SELEX: A discovery tool for genomic aptamers. *Methods* **52,** 125–132 (2010).

424. Kilili, G. K., Tilton, L. & Karbiwnyk, C. M. [Letter to the Editor] NaOH concentration and streptavidin bead type are key factors for optimal DNA aptamer strand separation and isolation. *BioTechniques* **61,** 114–116 (2016).

425. Tolle, F., Wilke, J., Wengel, J. & Mayer, G. By-product formation in repetitive PCR amplification of DNA libraries during SELEX. *PLoS One* **9,** e114693–e114693 (2014).

426. Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. & Polz, M. F. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71,** 8966–8969 (2005).

427. Kalle, E., Gulevich, A. & Rensing, C. External and semi-internal controls for PCR amplification of homologous sequences in mixed templates. *J. Microbiol. Methods* **95,** 285–294 (2013).

428. Spill, F. *et al.* Controlling uncertainty in aptamer selection. *Proc. Natl. Acad. Sci. U. S. A.* (2016). doi:10.1073/pnas.1605086113

429. Song, K.-M., Lee, S. & Ban, C. Aptamers and their biological applications. *Sensors (Basel).* **12,** 612–631 (2012).

430. Ruscito, A. & DeRosa, M. C. Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Front. Chem.* **4,** 14 (2016).

431. Nieuwlandt, D. In vitro selection of functional nucleic acid sequences. *Curr. Issues Mol. Biol.* **2,** 9–16 (2000).

432. Varshney, A. *et al.* Identification of an RNA aptamer binding hTERT-derived peptide and inhibiting telomerase activity in MCF7 cells. *Mol. Cell. Biochem.* **427,** 157–167 (2017).

433. Bittker, J. A., Le, B. V & Liu, D. R. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat. Biotechnol.* **20,** 1024 (2002).

434. Bing, T., Yang, X., Mei, H., Cao, Z. & Shangguan, D. Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories. *Bioorg. Med. Chem.* **18,** 1798–1805 (2010).

435. Masud, M. M., Kuwahara, M., Ozaki, H. & Sawai, H. Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorg. Med. Chem.* **12,** 1111–1120 (2004).

436. Shoji, A., Kuwahara, M., Ozaki, H. & Sawai, H. Modified DNA Aptamer That Binds the (R)-Isomer of a Thalidomide Derivative with High Enantioselectivity. *J. Am. Chem. Soc.* **129,** 1456–1464 (2007).

437. Baldrich, E. Aptamer arrays. *Methods Mol. Biol.* **671,** 35–54 (2011).

438. Maio, G. *et al.* Systematic optimization and modification of a DNA aptamer with 2'-O-methyl RNA analogues. *ChemistrySelect* **2,** 2335–2340 (2017).

439. Kettenbach, A. N., Rush, J. & Gerber, S. A. Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides. *Nat. Protoc.* **6,** 175–186 (2011).

440. Noble, J. E. & Bailey, M. J. A. Quantitation of protein. *Methods Enzymol.* **463,** 73–95 (2009).

441. Vemuri, S. Comparison of assays for determination of peptide content for lyophilized thymalfasin. *J. Pept. Res.* **65,** 433–439 (2005).

442. Stern, L. J. & Wiley, D. C. The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. *Cell* **68,** 465–477 (1992).

443. Preston, G. W. & Phillips, D. H. Quantification of a peptide standard using the intrinsic fluorescence of tyrosine. *Anal. Bioanal. Chem.* **408,** 2187–2193 (2016).

444. Hammer, K. D. & Nagel, K. An automated fluorescence assay for subnanogram quantities of protein in the presence of interfering material. *Anal. Biochem.* **155,** 308–314 (1986).

445. Lorenzen, A. & Kennedy, S. W. A fluorescence-based protein assay for use with a microplate reader. *Anal. Biochem.* **214,** 346–348 (1993).

446. Udenfriend, S. *et al.* Fluorescamine: A Reagent for Assay of Amino Acids, Peptides, Proteins, and Primary Amines in the Picomole Range. *Science.* **178,** 871 LP-872 (1972).

447. Asermely, K. E., Broomfield, C. A., Nowakowski, J., Courtney, B. C. & Adler, M. Identification of a recombinant synaptobrevin-thioredoxin fusion protein by capillary zone electrophoresis using laser-induced fluorescence detection. *J. Chromatogr. B. Biomed. Sci. Appl.* **695,** 67–75 (1997).

448. Bantan-Polak, T., Kassai, M. & Grant, K. B. A comparison of fluorescamine and naphthalene-2,3-dicarboxaldehyde fluorogenic reagents for microplate-based detection of amino acids. *Anal. Biochem.* **297,** 128–136 (2001).

449. You, W. W., Haugland, R. P., Ryan, D. K. & Haugland, R. P. 3-(4-Carboxybenzoyl)quinoline-2-carboxaldehyde, a reagent with broad dynamic range for the assay of proteins and lipoproteins in solution. *Anal. Biochem.* **244,** 277–282 (1997).

450. Avci-Adali, M. *et al.* Absolute quantification of cell-bound DNA aptamers during SELEX. *Nucleic Acid Ther.* **23,** 125–130 (2013).

451. Zhang, M., Biswas, S., Deng, W. & Yu, H. The Crystal Structure of Monovalent Streptavidin. *Sci. Rep.* **6,** 35915 (2016).

452. Ruigrok, V. J. B. *et al.* Kinetic and stoichiometric characterisation of streptavidin-binding aptamers. *Chembiochem* **13,** 829–836 (2012).

453. Hyre, D. E. *et al.* Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci.* **15,** 459–467 (2006).

454. Le Trong, I. *et al.* Streptavidin and its biotin complex at atomic resolution. *Acta Crystallogr. Sect. D* **67,** 813–821 (2011).

455. Liu, F., Zhang, J. Z. H. & Mei, Y. The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Sci. Rep.* **6,** 27190 (2016).

456. Laitinen, O. H., Hytonen, V. P., Nordlund, H. R. & Kulomaa, M. S. Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.* **63,** 2992–3017 (2006).

457. Dundas, C. M., Demonte, D. & Park, S. Streptavidin–biotin technology: improvements and innovations in chemical and biological applications. *Appl. Microbiol. Biotechnol.* **97,** 9343–9353 (2013).

458. Howarth, M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3,** 267–273 (2006).

459. O'Sullivan, V. J. *et al.* Development of a tetrameric streptavidin mutein with reversible biotin binding capability: engineering a mobile loop as an exit door for biotin. *PLoS One* **7,** e35203 (2012).

460. Leppek, K. & Stoecklin, G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* **42,** e13 (2014).

461. Srisawat, C. & Engelke, D. R. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7,** 632–641 (2001).

462. Le, T. T., Chumphukam, O. & Cass, A. E. G. Determination of minimal sequence for binding of an aptamer. A comparison of truncation and hybridization inhibition methods. *RSC Adv.* **4,** 47227–47233 (2014).

463. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355,** 564–566 (1992).

464. Guittat, L. *et al.* Quadruplex structures and quadruplex ligands. In Dynamical Genetics, V. Parisi, V.D. Fonzo, and F. Aluffi-Pentini, eds. (Research Signpost) 199–236 (2004).

465. Kumari, R., Nambiar, M., Shanbagh, S. & Raghavan, S. C. Detection of G-Quadruplex DNA Using Primer Extension as a Tool. *PLoS One* **10,** e0119722 (2015).

466. Kaur, H. & Yung, L.-Y. L. Probing High Affinity Sequences of DNA Aptamer against VEGF165. *PLoS One* **7,** e31196 (2012).

467. Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nature reviews. Drug discovery* **9,** 537–550 (2010).

468. Hasegawa, H., Savory, N., Abe, K. & Ikebukuro, K. Methods for Improving Aptamer Binding Affinity. *Molecules* **21,** 421 (2016).

469. McLellan, T. Electrophoresis buffers for polyacrylamide gels at various pH. *Anal. Biochem.* **126,** 94–99 (1982).

470. Paredes, E., Evans, M. & Das, S. R. RNA labeling, conjugation and ligation. *Methods* **54,** 251–259 (2011).

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

472. Anderson, G. W., Zimmerman, J. E. & Callahan, F. M. N-Hydroxysuccinimide Esters in Peptide Synthesis. *J. Am. Chem. Soc.* **85,** 3039 (1963).

473. Hahnefeld, C., Drewianka, S. & Herberg, F. W. Determination of kinetic data using surface plasmon resonance biosensors. *Methods Mol. Med.* **94,** 299–320 (2004).

474. Thordarson, P. Determining association constants from titration experiments in supramolecular chemistry. *Chem. Soc. Rev.* **40,** 1305–1323 (2011).

475. Iyer, R. P., Jiang, Z., Yu, D., Tan, W. & Agrawal, S. Improved Procedure for the Detritylation of DMT-Oligonucleotides: Use of Dowex® . *Synth. Commun.* **25,** 3611–3623 (1995).

476. Richardson, J. A., Gerowska, M., Shelbourne, M., French, D. & Brown, T. Six-Colour HyBeacon Probes for Multiplex Genetic Analysis. *ChemBioChem* **11,** 2530–2533 (2010).

477. Dixit, V. M., Laskovics, F. M., Noall, W. I. & Poulter, C. D. Tris(tetrabutylammonium) hydrogen pyrophosphate. A new reagent for the preparation of allylic pyrophosphate esters. *J. Org. Chem.* **46,** 1967–1969 (1981).

478. Kore, A. R., Senthilvelan, A. & Shanmugasundaram, M. Highly regioselective C-5 iodination of pyrimidine nucleotides and subsequent chemoselective Sonogashira coupling with propargylamine. *Nucleosides. Nucleotides Nucleic Acids* **34,** 92–102 (2015).

479. Yamamoto, K., Bruun, T., Kim, J. Y., Zhang, L. & Lautens, M. A New Multicomponent Multicatalyst Reaction (MC)2R: Chemoselective Cycloaddition and Latent Catalyst Activation for the Synthesis of Fully Substituted 1,2,3-Triazoles. *Org. Lett.* **18,** 2644–2647 (2016).

480. Trybulski, E. J., Zhang, J., Kramss, R. H. & Mangano, R. M. The synthesis and biochemical pharmacology of enantiomerically pure methylated oxotremorine derivatives. *J. Med. Chem.* **36,** 3533–3541 (1993).

481. Shelbourne, M., Brown, T., El-Sagheer, A. H. & Brown, T. Fast and efficient DNA crosslinking and multiple orthogonal labelling by copper-free click chemistry. *Chem. Commun.* **48,** 11184–11186 (2012).

Appendix

8 Appendix

8.1 Authority permission for picture reuse by the journal of biological chemistry.

Journal of biological chemistry

E-JOURNAL

ISSN:	1083-351X	**Language:**	English
Publication year(s):	1905 - present	**Country of publication:**	United States of America
Author/Editor:	AMERICAN SOCIETY FOR BIOCHEMISTRY & MOLECULAR BIOL		
Publisher:	AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BI		
Rightsholder:	AMERICAN SOC FOR BIOCHEMISTRY & MOLECULAR BIOLOGY		

Journal of Biological Chemistry permissions policies: • If you are an author of the content for which you are seeking permission, or if you are not an author but are requesting permission to copy, distribute, transmit and adapt the work for noncommercial purposes (e.g. reproduction of a figure for educational purposes such as schoolwork, or appending a reprinted article to a PhD dissertation), you do not need to seek permission using the options listed below, as long as any reuse includes the credit line in the reuse policies listed above. • Parties who are not authors on the article who wish to reuse content for commercial purposes such as reproducing a figure in a book, journal, or coursepack published by a commercial publisher, do need permission and should request permission by completing the form below. For more information please see Journal of Biological Chemistry: http://www.jbc.org/site/misc/edpolicy.xhtml#copyright

8.2 Authority permission for picture reuse by the Elsevier.

Copyright Clearance Center RightsLink® Home Create Account Help



Title: Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes

Author: Chad A Brautigam,Thomas A Steitz

Publication: Current Opinion in Structural Biology

Publisher: Elsevier

Date: February 1998

Copyright © 1998 Published by Elsevier Ltd.

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials.

Already a RightsLink user or want to learn more?

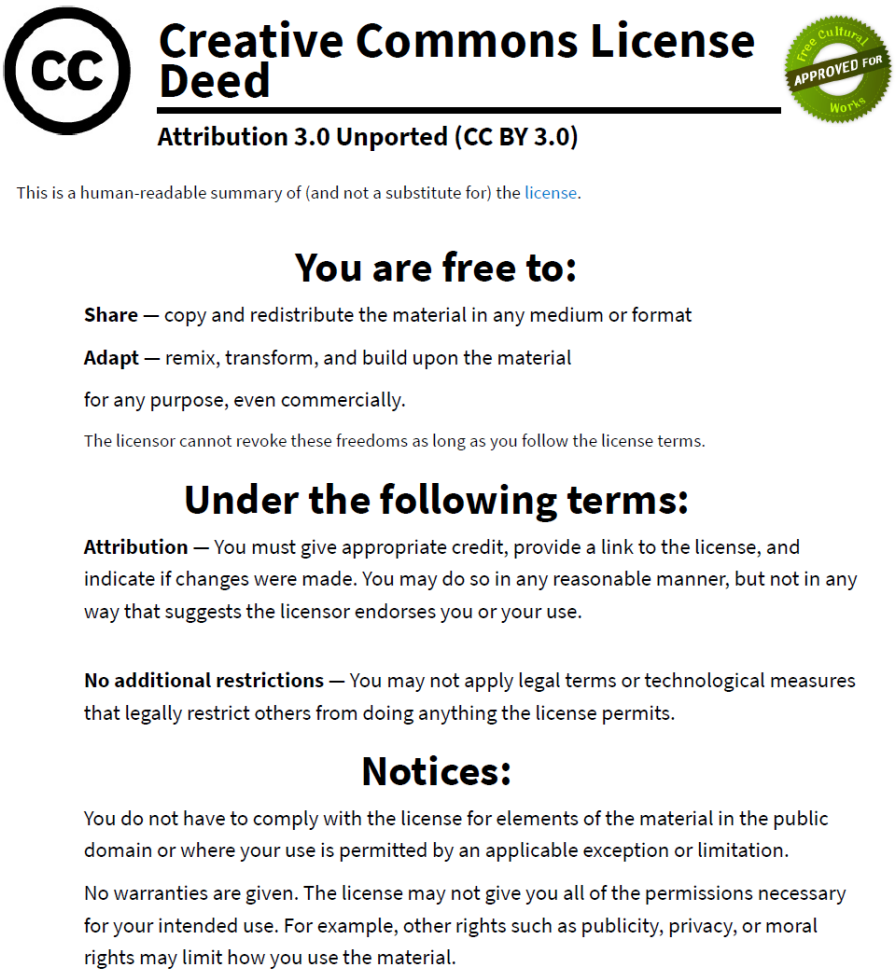
Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Copyright © 2019 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

8.3 Authority permission for picture reuse by the IntechOpen under a license.

This picture is licensed under the Creative Commons Attribution 3.0 (CC BY 3.0) license. To view a copy of this license, visit https://creativecommons.org/licenses/by/3.0/



Creative Commons License Deed

Attribution 3.0 Unported (CC BY 3.0)

This is a human-readable summary of (and not a substitute for) the license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

8.4 Representative mass spectrometry results of modified triphosphates

Figure S 8-1. MS result of 2′-AzE-TTP

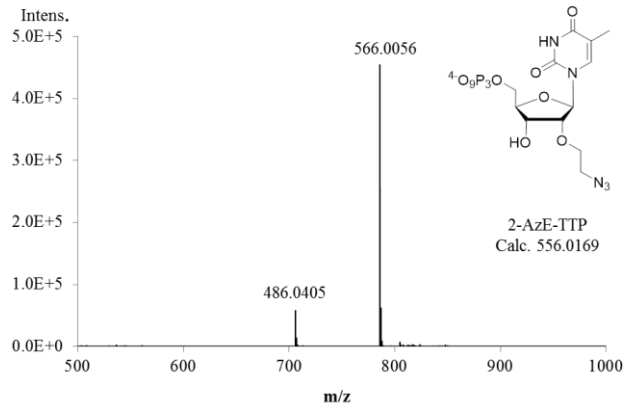


Figure S 8-2. MS result of 5-Iodo-dUTP

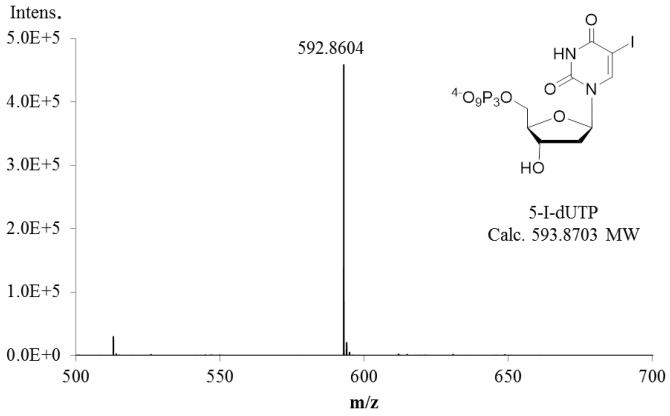


Figure S 8-3. MS result of 5-PA-dUTP

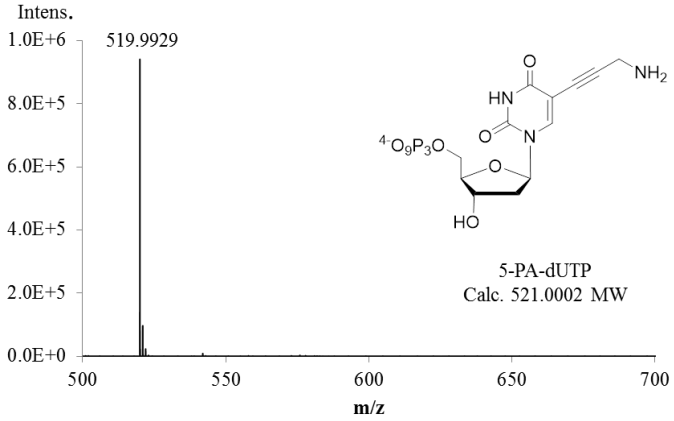


Figure S 8-4. MS result of 5-PAF-dUTP

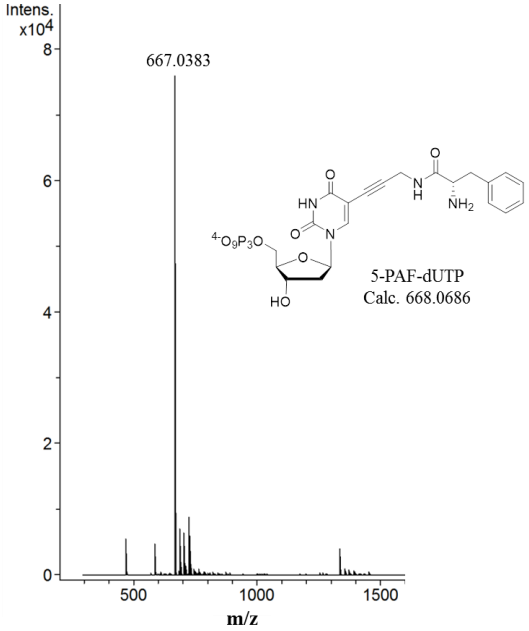


Figure S 8-5. MS result of 5-PAW-dUTP

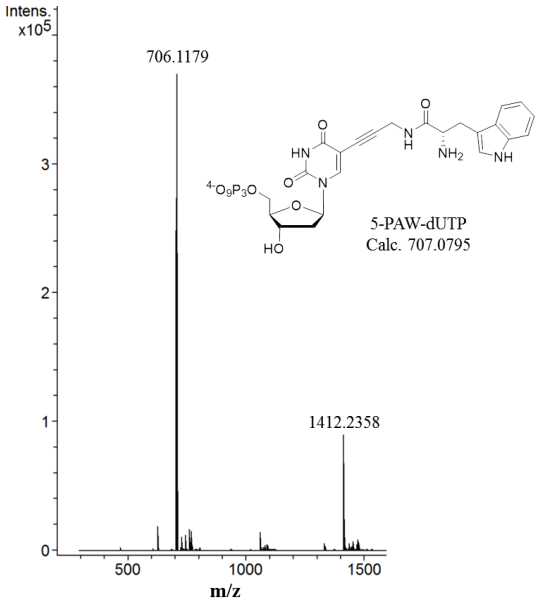


Figure S 8-6. MS result of 5-PAI-dUTP

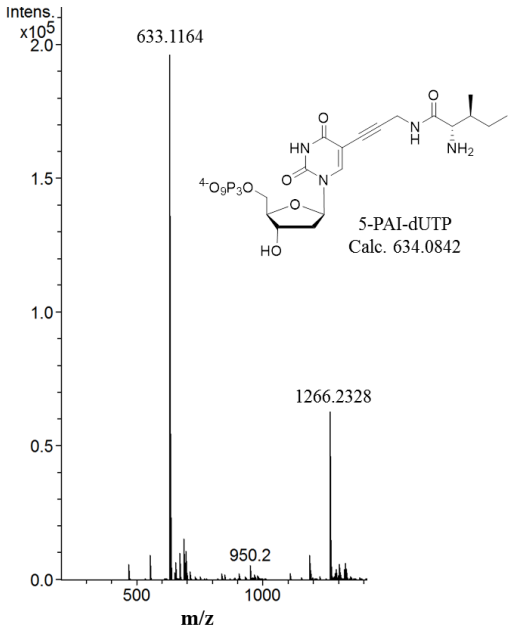


Figure S 8-7. MS result of 5-PAA-dUTP

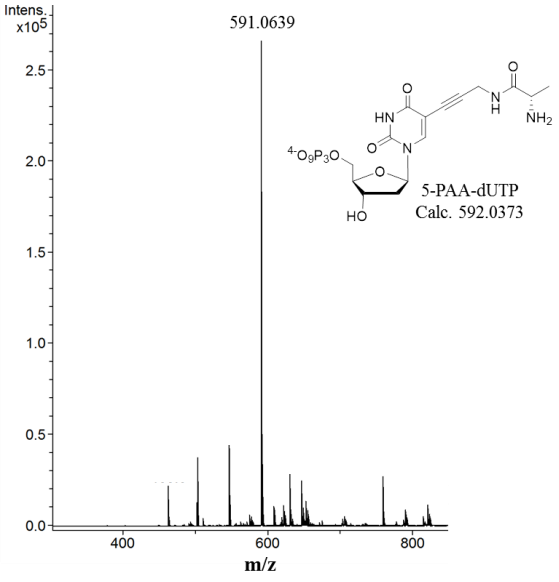


Figure S 8-8. MS result of 5-PAPBA-dUTP

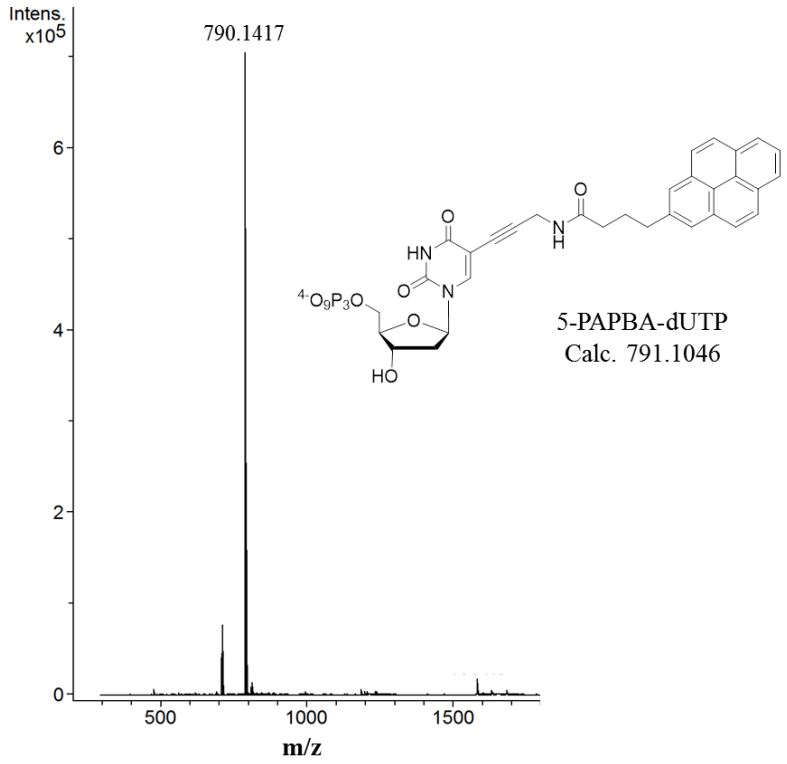


Figure S 8-9. MS result of 5-FF-dUTP

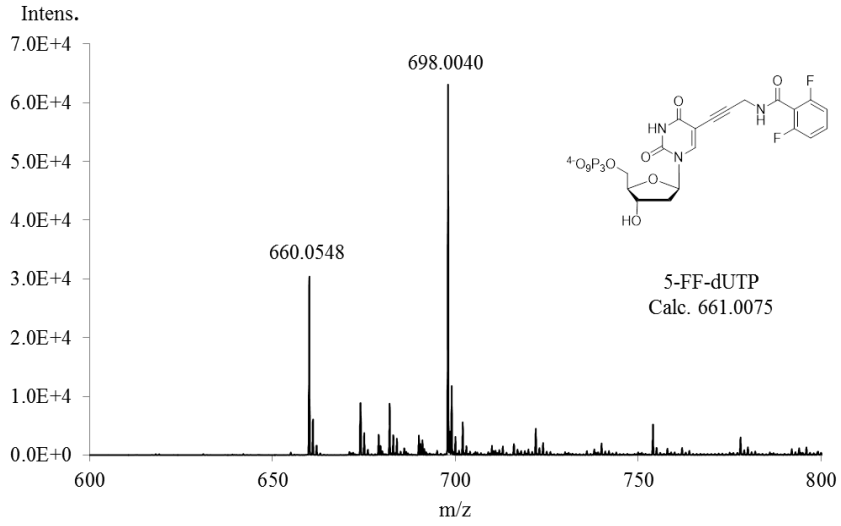


Figure S 8-10. MS result of 5-FFI-dUTP

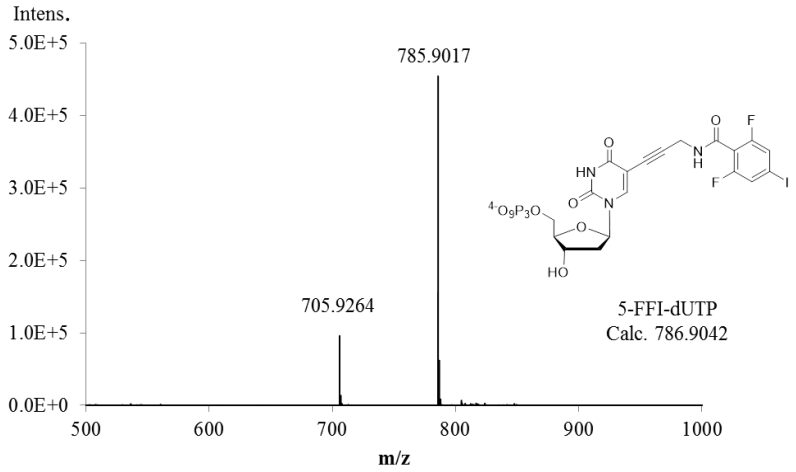


Figure S 8-11. MS result of 5-ITZ-dUTP

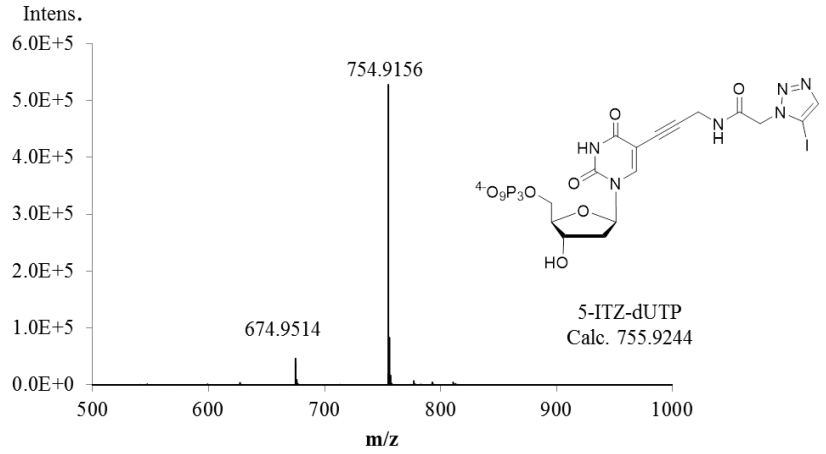
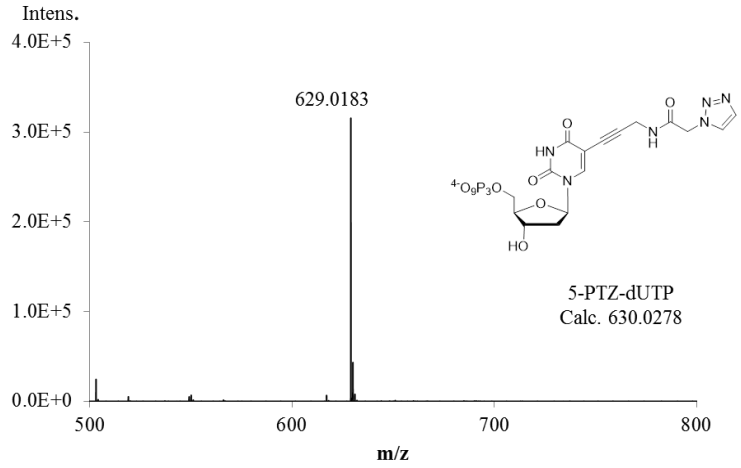
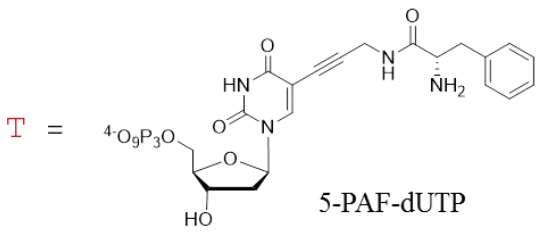


Figure S 8-12. MS result of 5-PTZ-dUTP

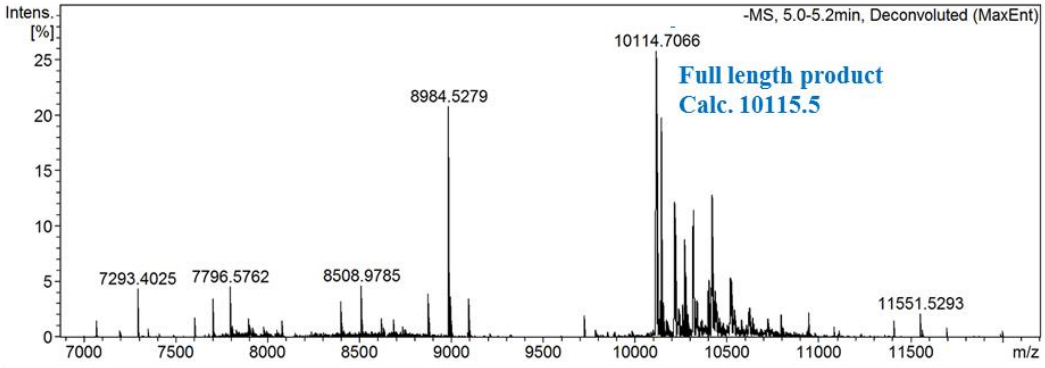


8.5 Representative mass spectrometry results for primer extension products containing modified nucleotides (Hydrophobic amino acid side chains and pyrene substitutied groups)



A

T2: CAGTCACTGTACTGCCGACACACATAACC
GTCAGTGACATGACGGCTGTGTGTATTGG-FAM



B

T3: CAGTCACAAAACTGCCGACACACATAACC
GTCAGTGTTTTGACGGCTGTGTGTATTGG-FAM

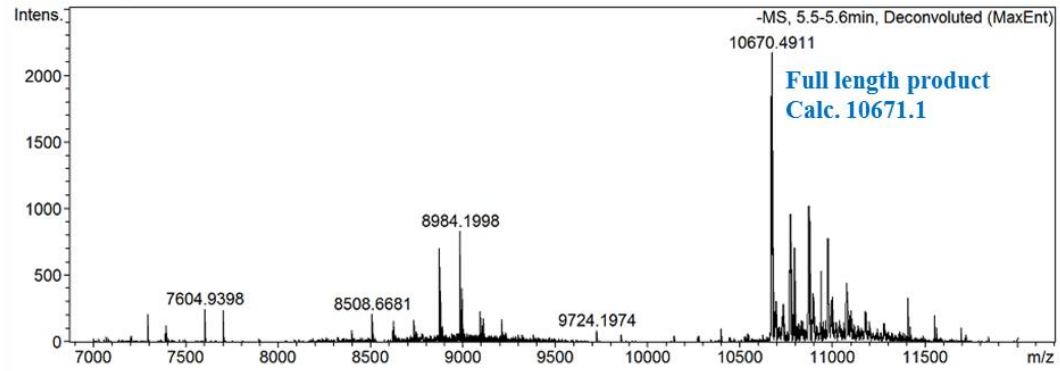
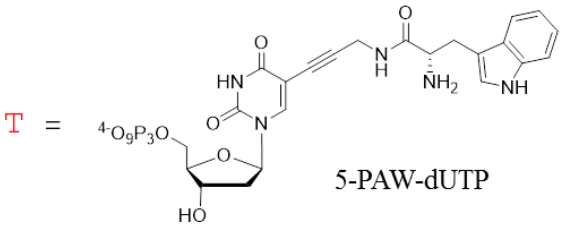
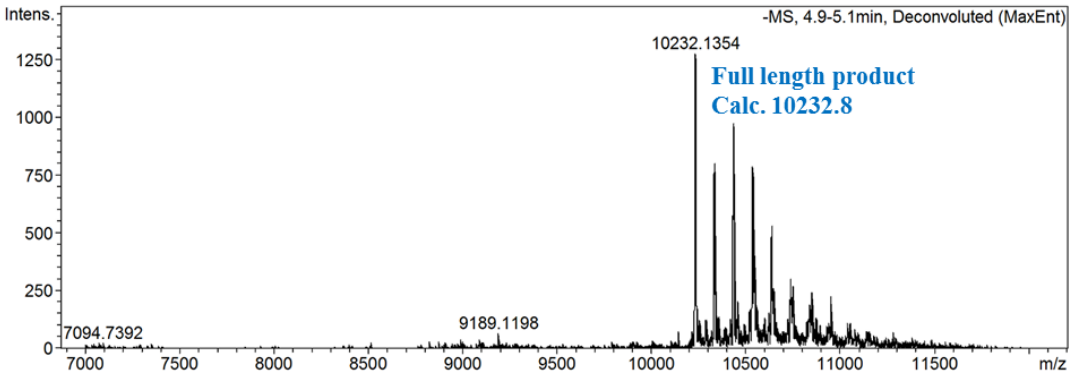


Figure S 8-13. MS-ES$^-$analysis of final primer extension reaction using 5-PAF-dUTP and template T2 (**A**) or T3 (**B**). 5-PAF-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 20 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.



A

T2: CAGTCACTGTACTGCCGACACACATAACC
GTCAGTGACATGACGGCTGTGTGTATTGG-FAM



B

T3: CAGTCACAAAACTGCCGACACACATAACC
GTCAGTGTTTTGACGGCTGTGTGTATTGG-FAM

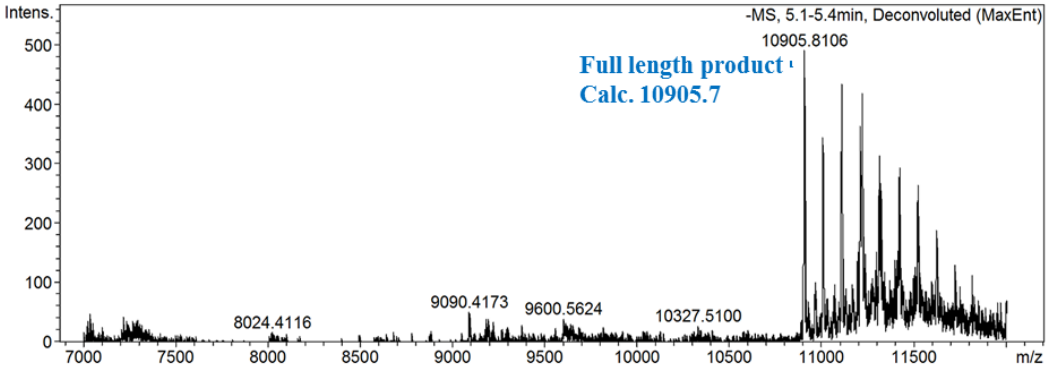
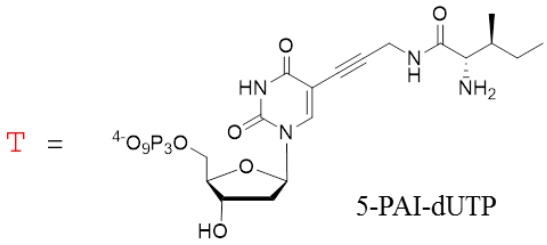
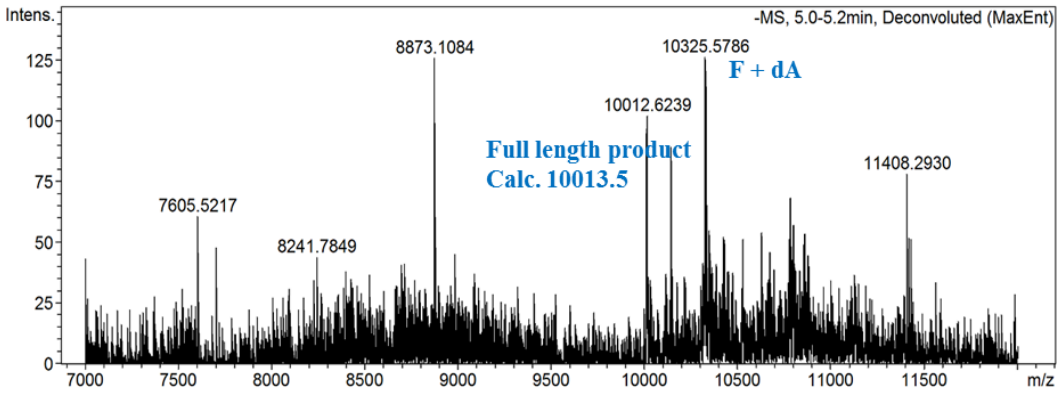


Figure S 8-14. MS-ES$^-$analysis of final primer extension reaction using 5-PAW-dUTP and template T2 (**A**) or T3 (**B**). 5-PAW-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 5 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.



A

T2: CAGTCACTGTACTGCCGACACACATAACC
GTCAGTGACATGACGGCTGTGTGTATTGG-FAM



B

T3: CAGTCACAAAACTGCCGACACACATAACC
GTCAGTGTTTTGACGGCTGTGTGTATTGG-FAM

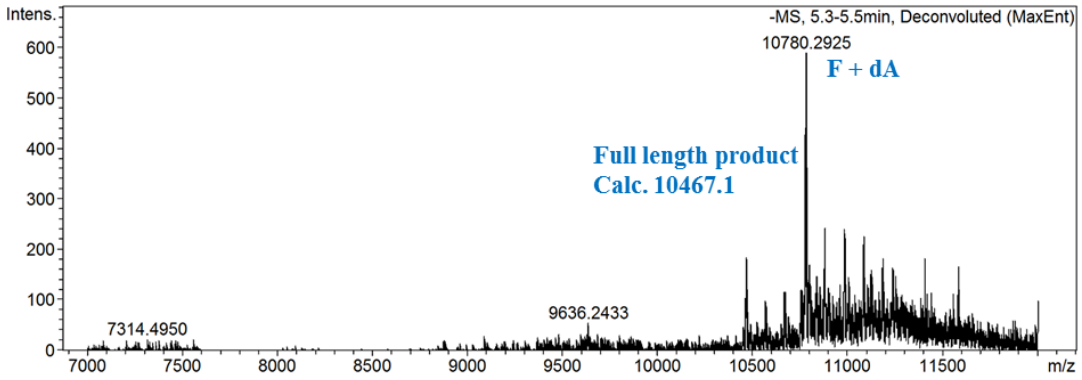
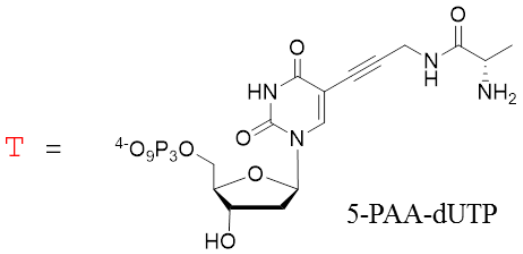
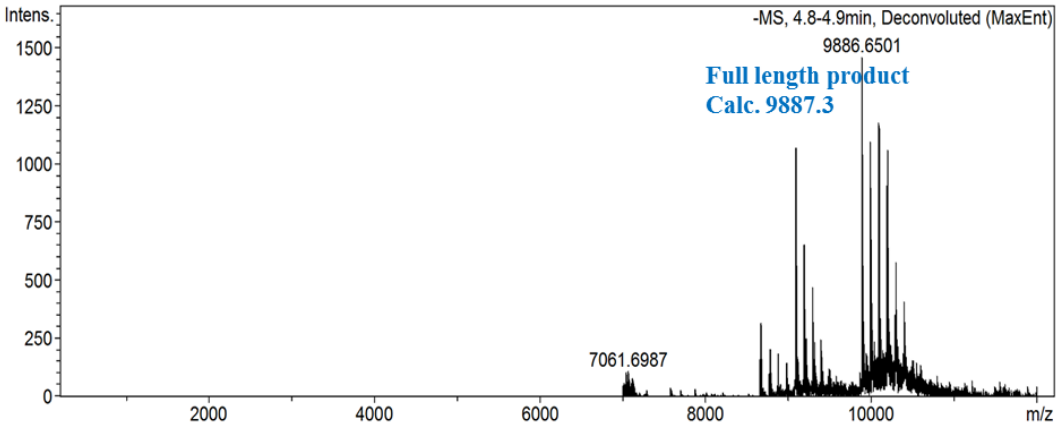


Figure S 8-15. MS-ES$^-$analysis of final primer extension reaction using 5-PAI-dUTP and template T2 (**A**) or T3 (**B**). 5-PAI-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 5 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.



A

T2: CAGTCACTGTACTGCCGACACACATAACC
GTCAGTGACATGACGGCTGTGTGTATTGG-FAM



B

T3: CAGTCACAAAACTGCCGACACACATAACC
GTCAGTGTTTTGACGGCTGTGTGTATTGG-FAM

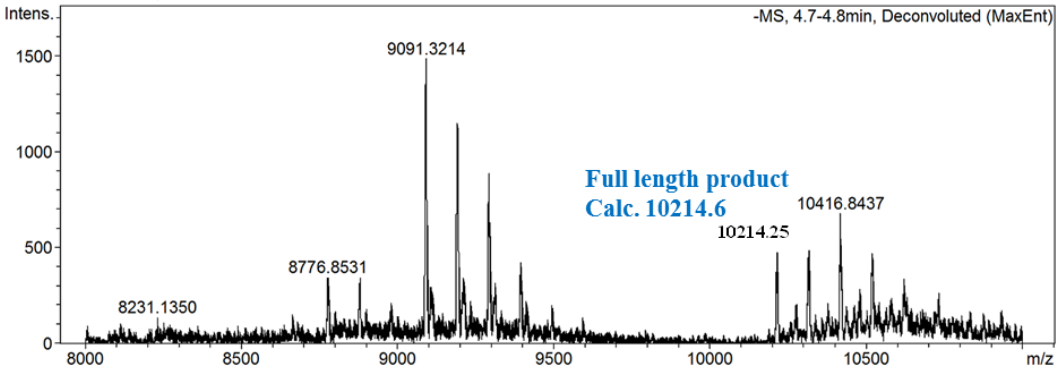
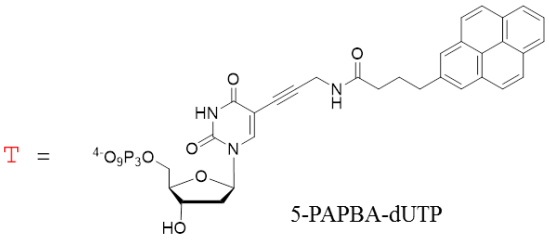
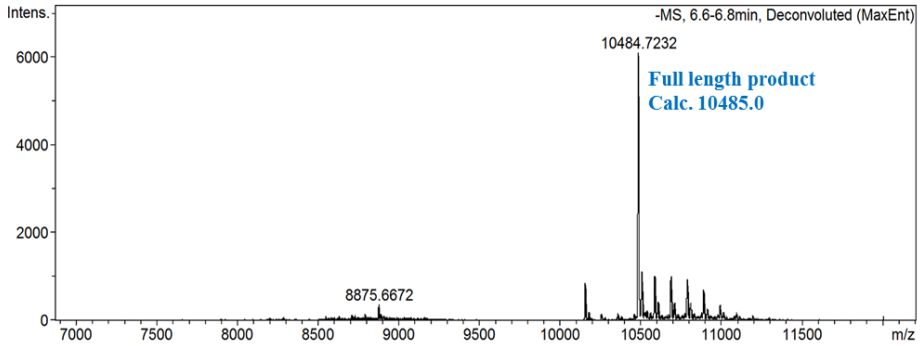


Figure S 8-16. MS-ES$^-$analysis of final primer extension reaction using 5-PAA-dUTP and template T2 (**A**) or T3 (**B**). 5-PAA-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 °C, 15 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.



A

T2: CAGTCACTGTACTGCCGACACACATAACC
GTCAGTGACATGACGGCTGTGTGTATTGG-FAM



B

T3: CAGTCACAAAACTGCCGACACACATAACC
GTCAGTGTTTTTGACGGCTGTGTGTATTGG-FAM

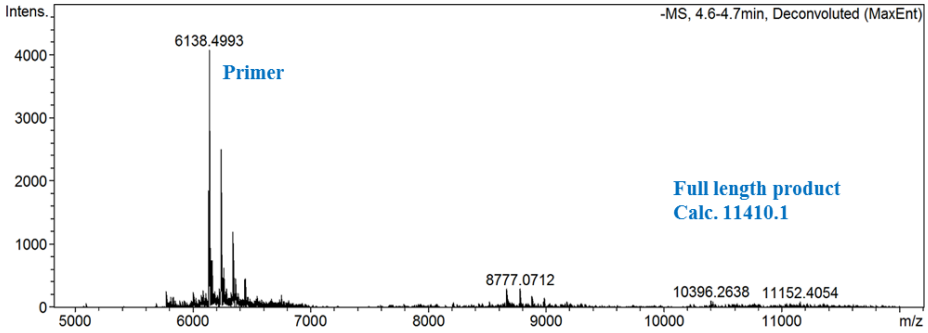
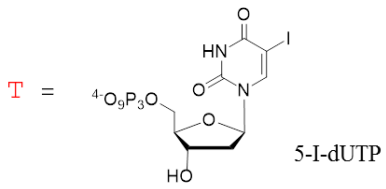


Figure S 8-17. MS-ES¯analysis of final primer extension reaction using 5-PAPBA-dUTP and template T2 (**A**) or T3 (**B**). 5-PAPBA-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 20 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.

8.6 Representative mass spectrometry results for primer extension products containing modified nucleotides (Halogen substitutied groups)



T2: CAGTCACTGTACTGCCGACACACATAACC
GTCAGTGACATGACGGCTGTGTGTATTGG-FAM

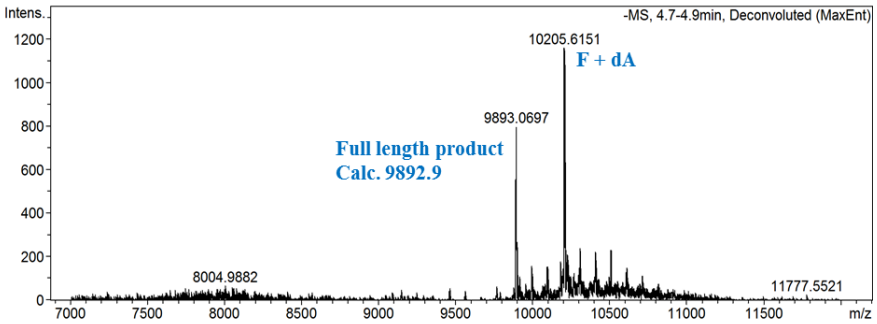


Figure S 8-18. MS-ES$^-$analysis of final primer extension reaction using 5-I-dUTP and template T2. 5-I-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 15 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.

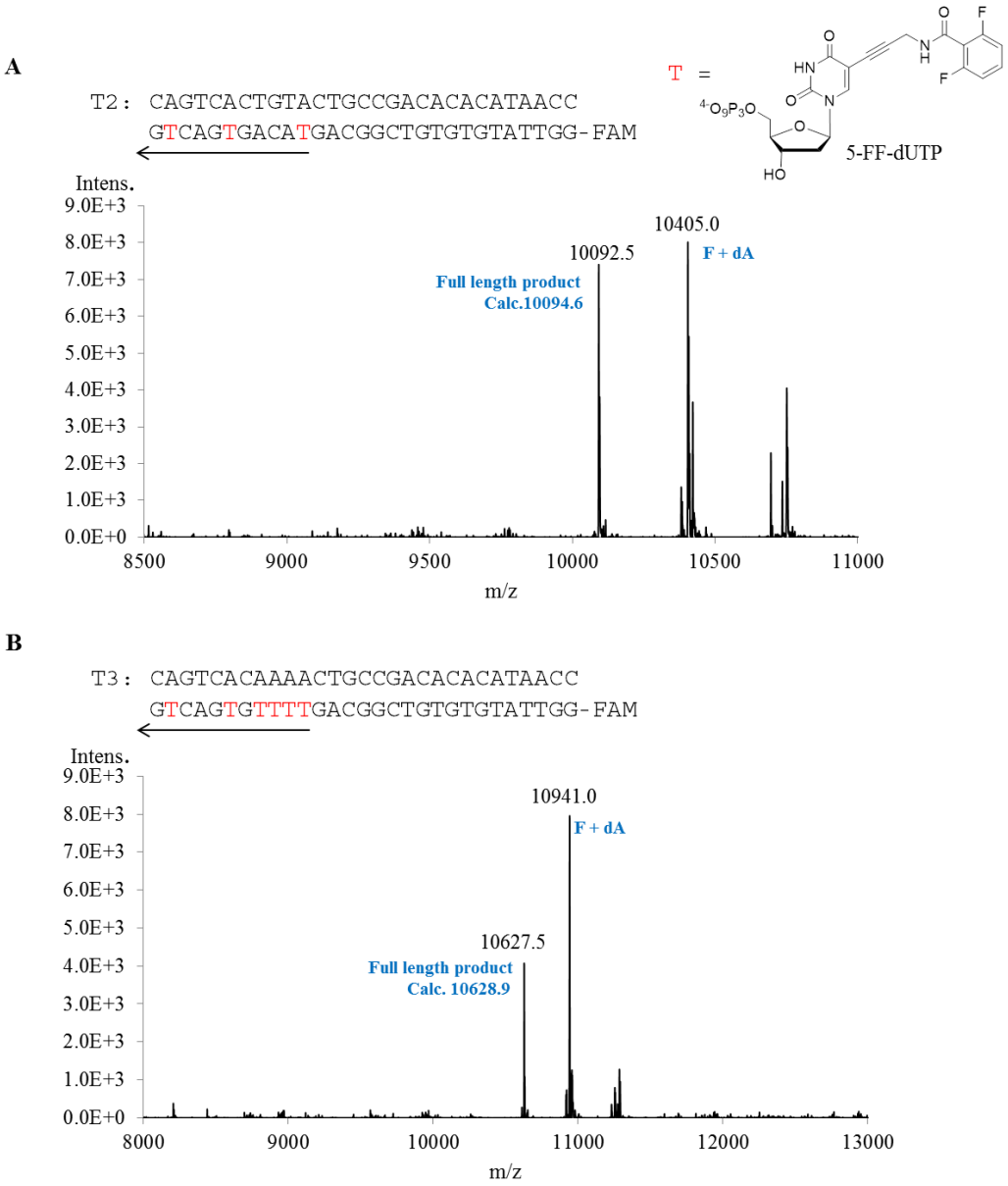


Figure S 8-19. MS-ES$^-$analysis of final primer extension reaction using 5-FF-dUTP and template T2 (**A**) or T3 (**B**). 5-FF-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 ℃, 15 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.

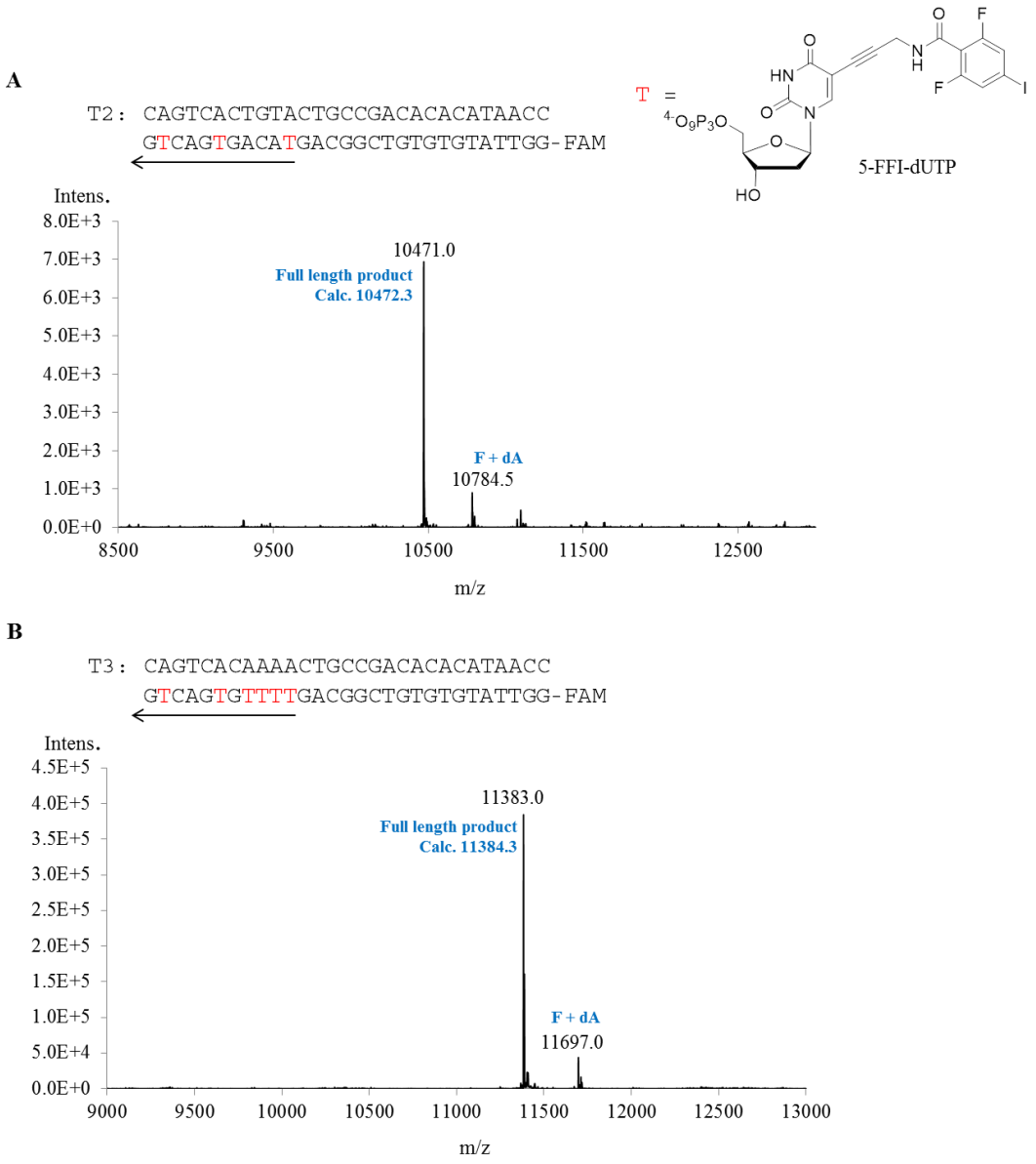


Figure S 8-20. MS-ES$^-$analysis of final primer extension reaction using 5-FFI-dUTP and template T2 (**A**) or T3 (**B**). 5-FFI-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 °C, 15 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.

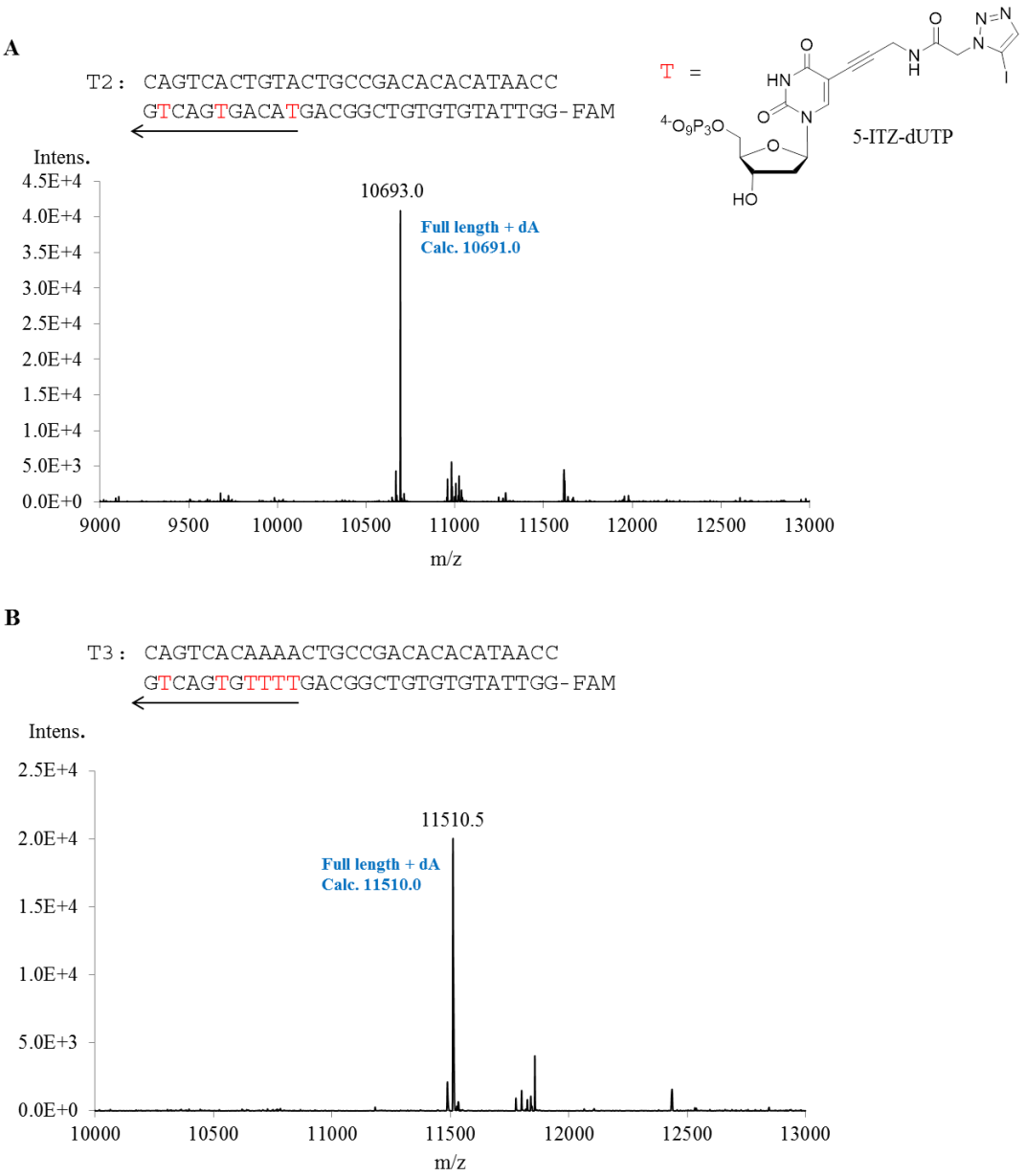


Figure S 8-21. MS-ES$^-$analysis of final primer extension reaction using 5-ITZ-dUTP and template T2 (**A**) or T3 (**B**). 5-ITZ-dUTP incorporation sites are highlighted in red. Reaction was performed using 2 U KOD XL DNA polymerase (65 °C, 15 min), 3.2 nmol of each deoxynucleotide and 1 mM $MnCl_2$.

8.7 Denaturing PAGE gels after PCR amplification using 5-FFI-dUTP

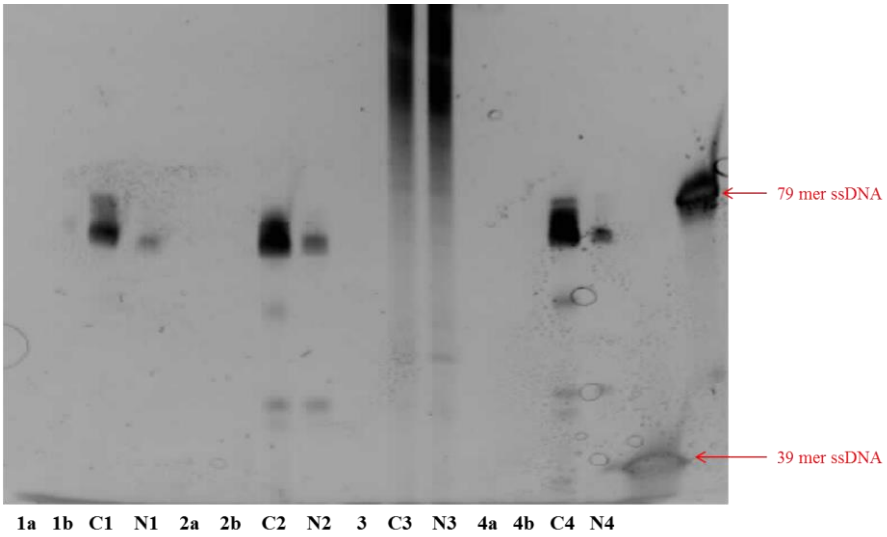


Figure S 8-22. Analysis of PCR products containing 5-FFI-dUTP (see Figure 3.31). Lane **1a**, PCR amplicons using Vent exo- DNA polymerase without additional Mn^{2+}; lane **1b**, PCR amplicons using Vent exo- DNA polymerase with 5 mM $MgSO_4$; lane **2a**, PCR amplicons using Q5 Hot Start DNA polymerase without additional Mn^{2+}; lane **2b**, PCR amplicons using Q5 Hot Start DNA polymerase with additional 5 mM $MgSO_4$; lane **3**, PCR amplicons using KOD XL DNA polymerase with additional 1 mM $MnCl_2$; lane **4a**, PCR amplicons using Phusion HF DNA polymerase without additional Mn^{2+}; lane **4b**, PCR amplicons using Phusion HF DNA polymerase with additional 5 mM $MgSO_4$; lanes **C1-C4**, positive control reaction with natural dNTPs; lanes **N1-N4**, negative control reaction without template T4. (8% Denaturing PAGE)

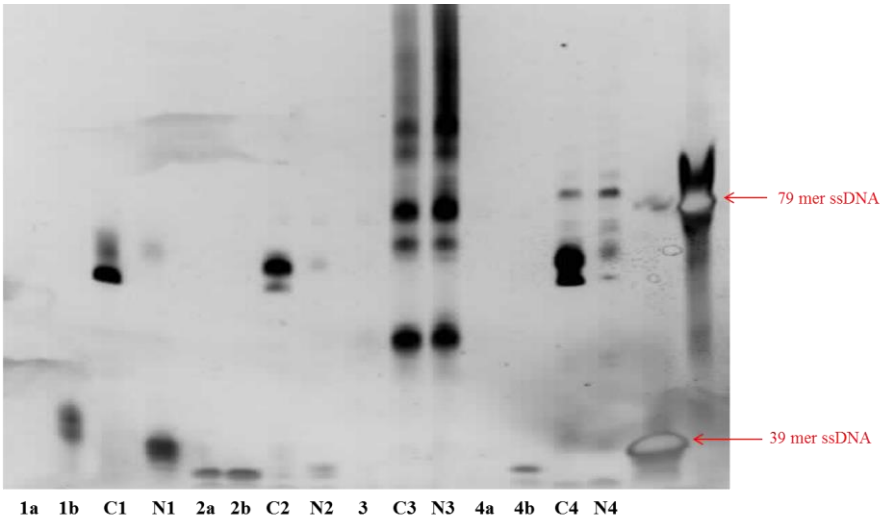


Figure S 8-23. Analysis of PCR products containing 5-FFI-dUTP (see Figure 3.32). Lane **1a**, PCR amplicons using Vent exo- DNA polymerase without additional Mn^{2+}; lane **1b**, PCR amplicons using Vent exo- DNA polymerase with additional 5 mM $MgSO_4$; lane **2a**, PCR amplicons using Q5 Hot Start DNA polymerase without additional Mn^{2+}; lane **2b**, PCR amplicons using Q5 Hot Start DNA polymerase with additional 5 mM $MgSO_4$; lane **3**, PCR amplicons using KOD XL DNA polymerase with additional 1 mM $MnCl_2$; lane **4a**, PCR amplicons using Phusion HF DNA polymerase without additional Mn^{2+}; lane **4b**, PCR amplicons using Phusion HF DNA polymerase with additional 5 mM $MgSO_4$; lanes **C1-C4**, positive control reaction with natural dNTPs; lanes **N1-N4**, negative control reaction without template T5. (8% Denaturing PAGE)

8.8 A table of oligonucleotide recovery rates calculated by Q-PCR after each BP-SELEX selection round (R1-R7)

		R1	R2	R3	R4	R5	R6	R7-1	R7-2
BP1	**Positive target**	3.55E-5	2.54E-4	5.19E-5	1.38E-3	1.27E-3	1.96E-4	2.94E-5	1.91E-5
	Negative target	6.25E-7	-	-	-	1.28E-1	-	2.39E+2	-
BP2	**Positive target**	3.64E-5	2.36E-4	2.62E-5	1.19E-3	3.52E-2	9.77E-6	8.71E-5	1.83E-4
	Negative target	2.45E-7	-	-	-	6.24E-2	8.83E-6	2.91E+2	-

* The negative selection target was streptavidin magnetic beads and the positive selection target was peptide **1** for BP1 and peptide **2** for BP2.

8.9 SPR sensorgrams of enriched ssDNA pools from BP1 SELEX

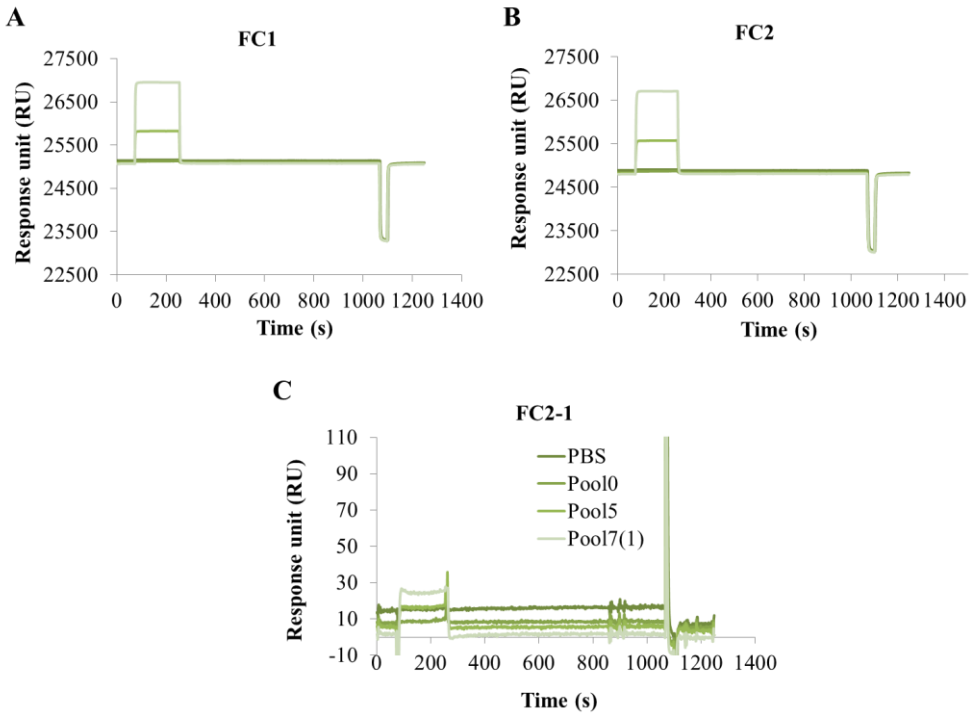


Figure S 8-24. BP1 SELEX starting library, pool5 and pool7 binding to peptide **1**. **A**) ssDNA pools binding to reference (FC1). **B**) ssDNA pools binding to peptide **1** (FC2). **C**) Binding analysis of ssDNA to peptide **1** by subtracting FC1 from FC2 (FC2-1).

8.10 SPR sensorgrams of enriched ssDNA pools from BP2 SELEX

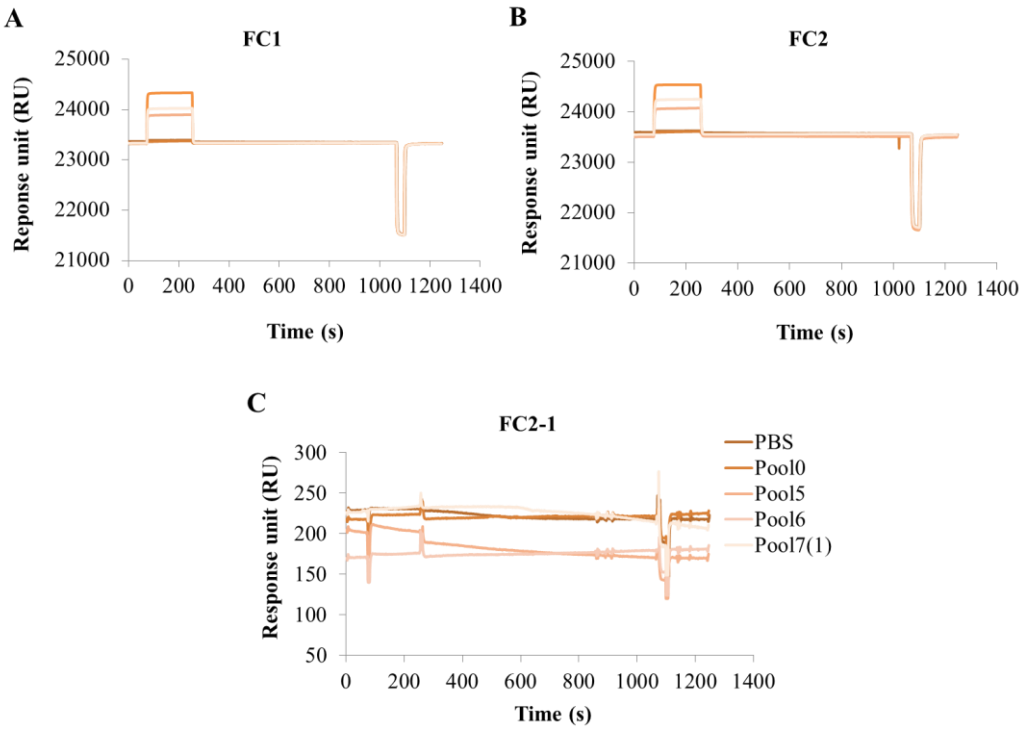


Figure S 8-25. BP2 SELEX starting library, pool5, pool 6 and pool7 binding to peptide **2**. **A**) ssDNA pools binding to reference (FC1). **B**) ssDNA pools binding to peptide **2** (FC2). **C**) Binding analysis of ssDNA to peptide **2** by subtracting FC1 from FC2 (FC2-1).

8.11 Authority permission for picture reuse by the Springer Nature under a license.



SPRINGER NATURE

Title: The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations

Author: Fengjiao Liu, John Z. H. Zhang, Ye Mei

Publication: Scientific Reports

Publisher: Springer Nature

Date: Jun 1, 2016

Copyright © 2016, Springer Nature

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact Springer Nature

To order reprints of this content, please contact Springer Nature by e-mail at reprintswarehouse@springernature.com, and you will be contacted very shortly with a quote.

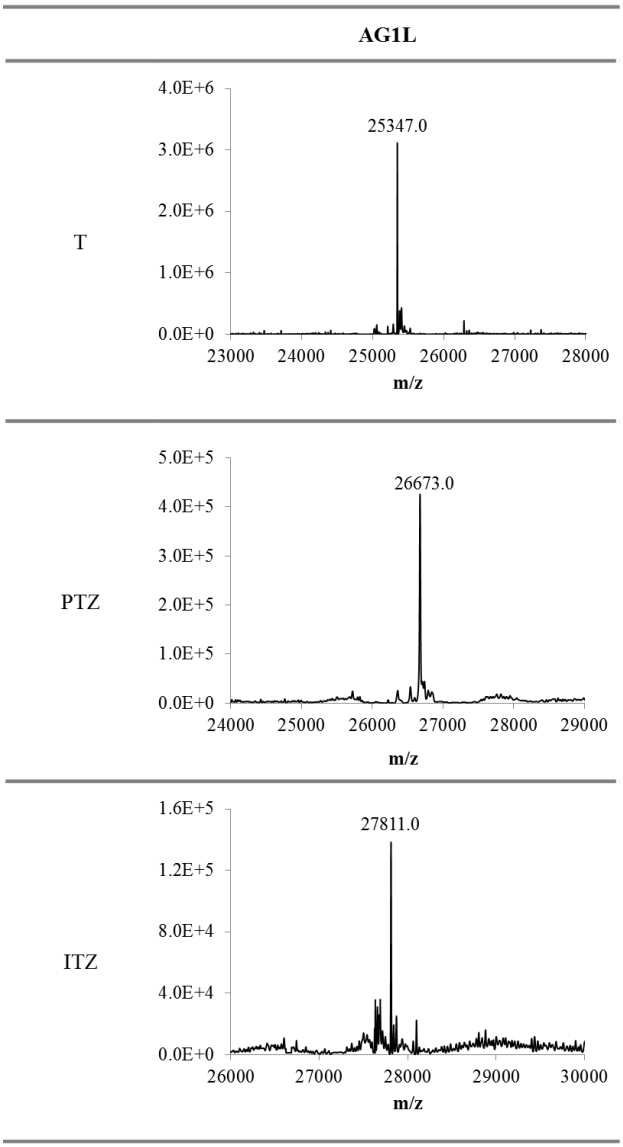
This picture is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

8.12 A table of oligonucleotide recovery rates calculated by Q-PCR after each of streptavidin-SELEX selection round (R1-R9)

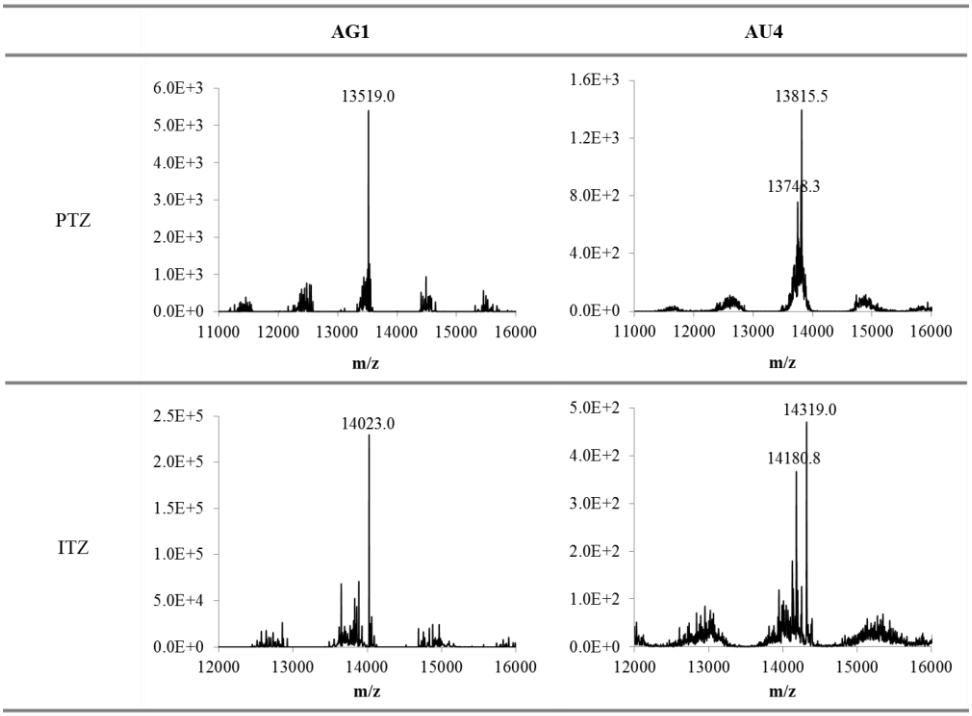
	R1	R2	R3	R4	R5	R6	R7	R8	R9
Recovery rate (Negative target)	1.12E-4	2.94E-4	8.22E-5	3.03E-4	6.47E-7	5.98E-5	7.35E-4	8.56E-4	5.75E-4
Recovery rate (Positive target)	4.90E-4	7.99E-5	2.36E-4	4.74E-4	6.26E-5	7.79E-4	2.31E-3	1.1E-2	1.10E-2

* The negative selection target was carboxylic acid magnetic beads and the positive selection target was streptavidin beads.

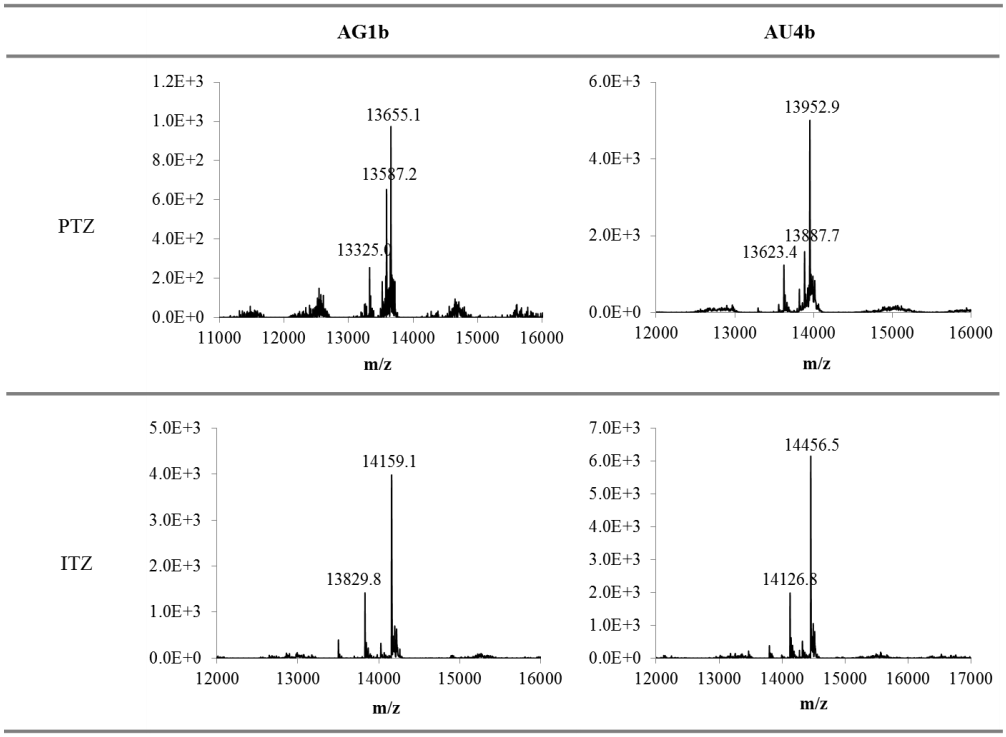
8.13 Representative mass spectrometry results of AG1L and its control sequences



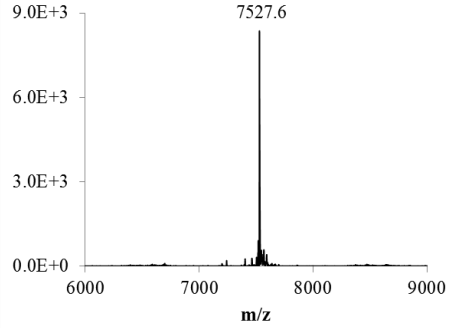
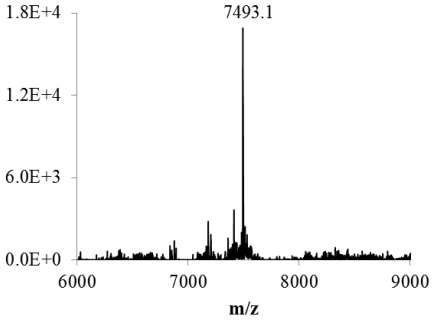
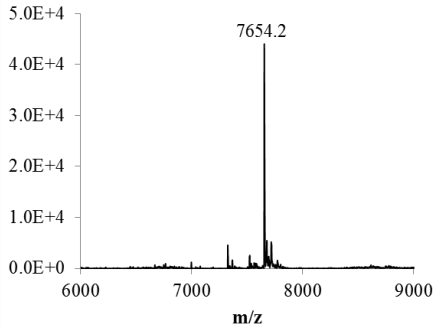
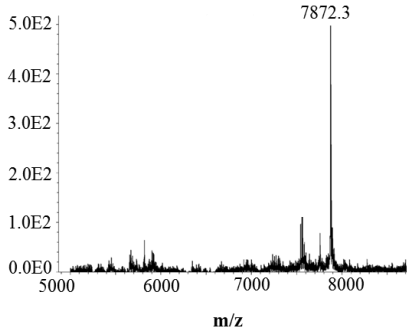
8.14 Representative mass spectrometry results of AG1 and AU4 and corresponding control sequences



8.15 Representative mass spectrometry results of AG1b and AU4b and the corresponding control sequences.



8.16 Representative mass spectrometry results of AN1 and AN3

	AN1	AN3
PTZ		
ITZ		

8.17 SPR sensorgrams of AG1-ITZ and AU4-ITZ and their control sequences binding to streptavidin

AG1-ITZ: FAM-GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTTGG

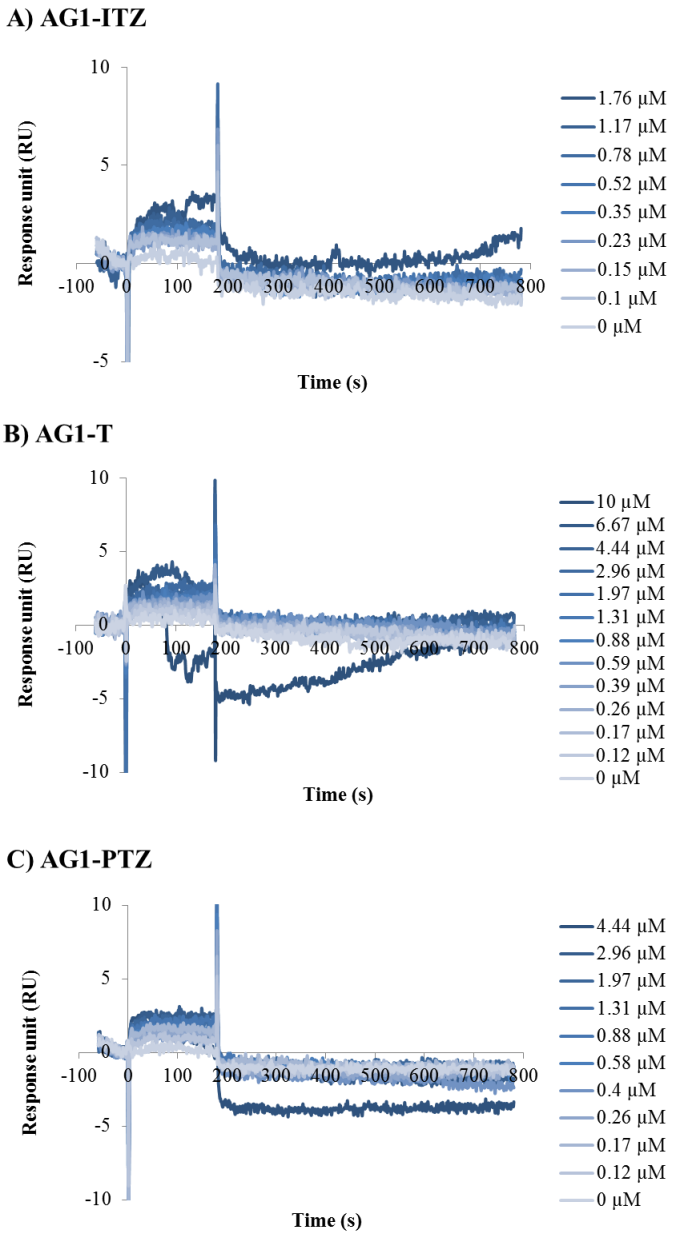


Figure S 8-26. Binding analysis of AG1-ITZ and its control sequences using CM5 chip immobilised with ~5,000 RU of streptavidin. The positions of 5-ITZ-dU are coloured in red. **A**) Reference subtracted binding response of AG1-ITZ. **B**) Reference subtracted binding response of AG1-T. **C**) Reference subtracted binding response of AG1-PTZ.

AU4-ITZ: FAM-GCGGGAAGGCAGGGAAGGGAAA**T**GGGGCGGGACA**TTT**GGG

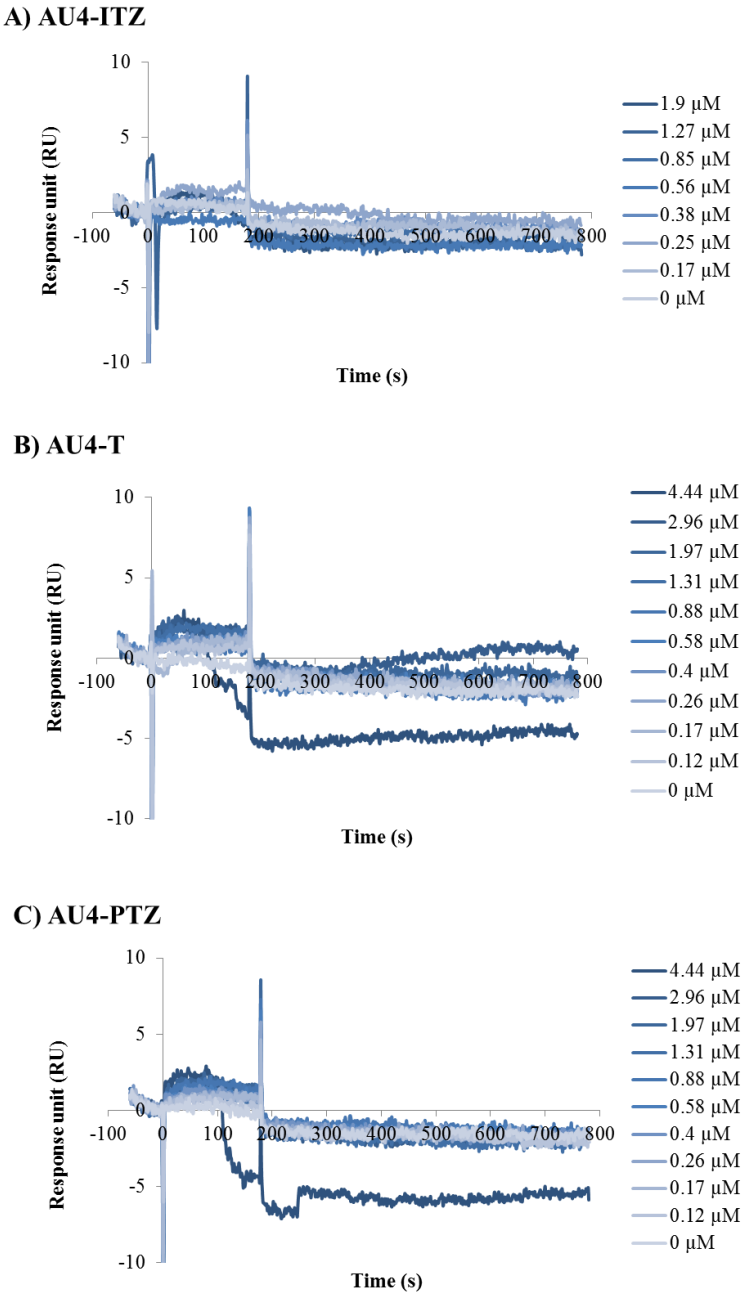


Figure S 8-27. Binding analysis of AU4-ITZ and its control sequences using CM5 chip immobilised with ~5,000 RU of streptavidin. The positions of 5-ITZ-dU are coloured in red. **A**) Reference subtracted binding response of AU4-ITZ. **B**) Reference subtracted binding response of AU4-T. **C**) Reference subtracted binding response of AU4-PTZ.

8.18 SPR sensorgrams of AC1 and AC2 binding to streptavidin

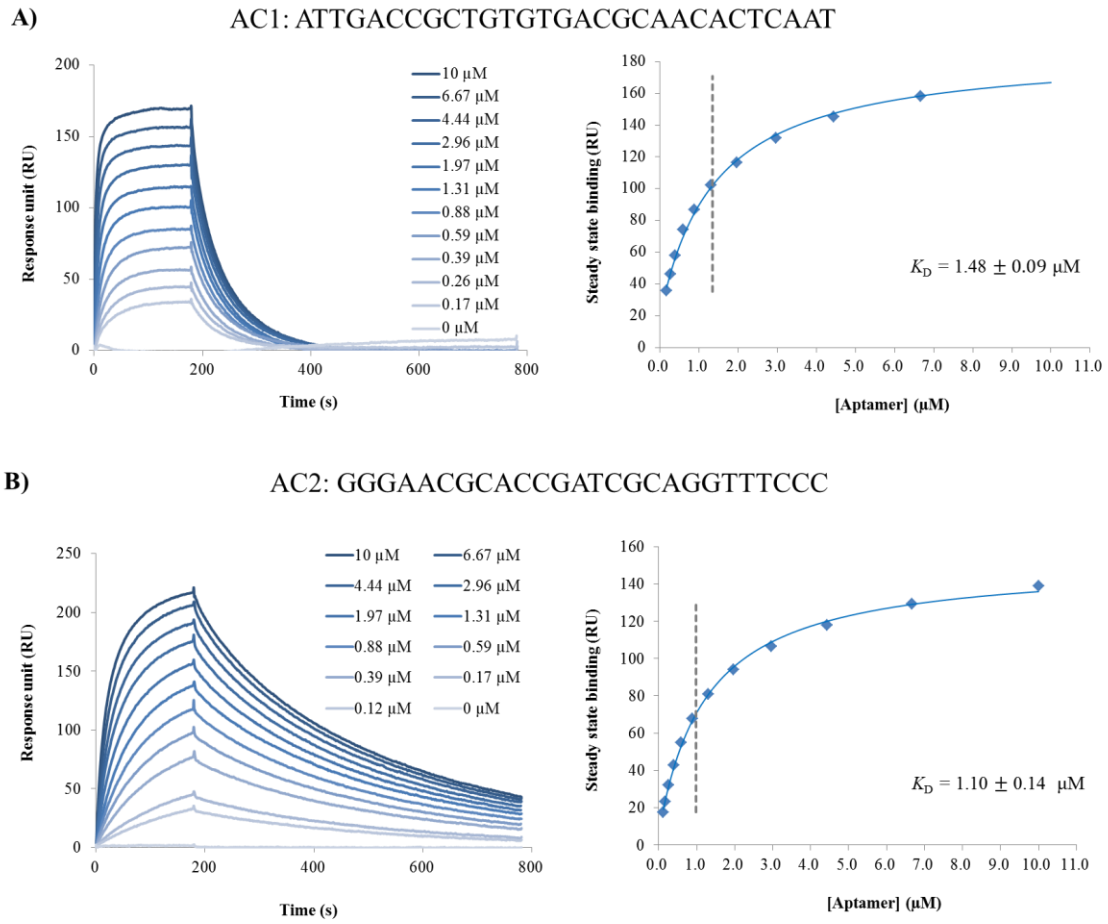


Figure S 8-28. SPR-based interaction analyses of the affinity of AC1 (**A**) and AC2 (**B**) to streptavidin and the corresponding steady-state binding data from the end of the association phases against analyte concentrations. This was used to calculate the steady-state affinities (K_D). The dashed grey line indicates the point at which half the aptamer is bound to the immobilised SA, which is an estimated value of binding constant (K_D).

8.19 SPR sensorgrams of AG1L-ITZ and its control sequences binding to streptavidin

AG1L-ITZ: 5′-FAM-AGCAGCACAGAGGTCAGATGGCGGGAAGGCAGGGA-
AGGGGGATGGGGCGGGACATTTGGCCTATGCGTGCTACCGTGAA-3′

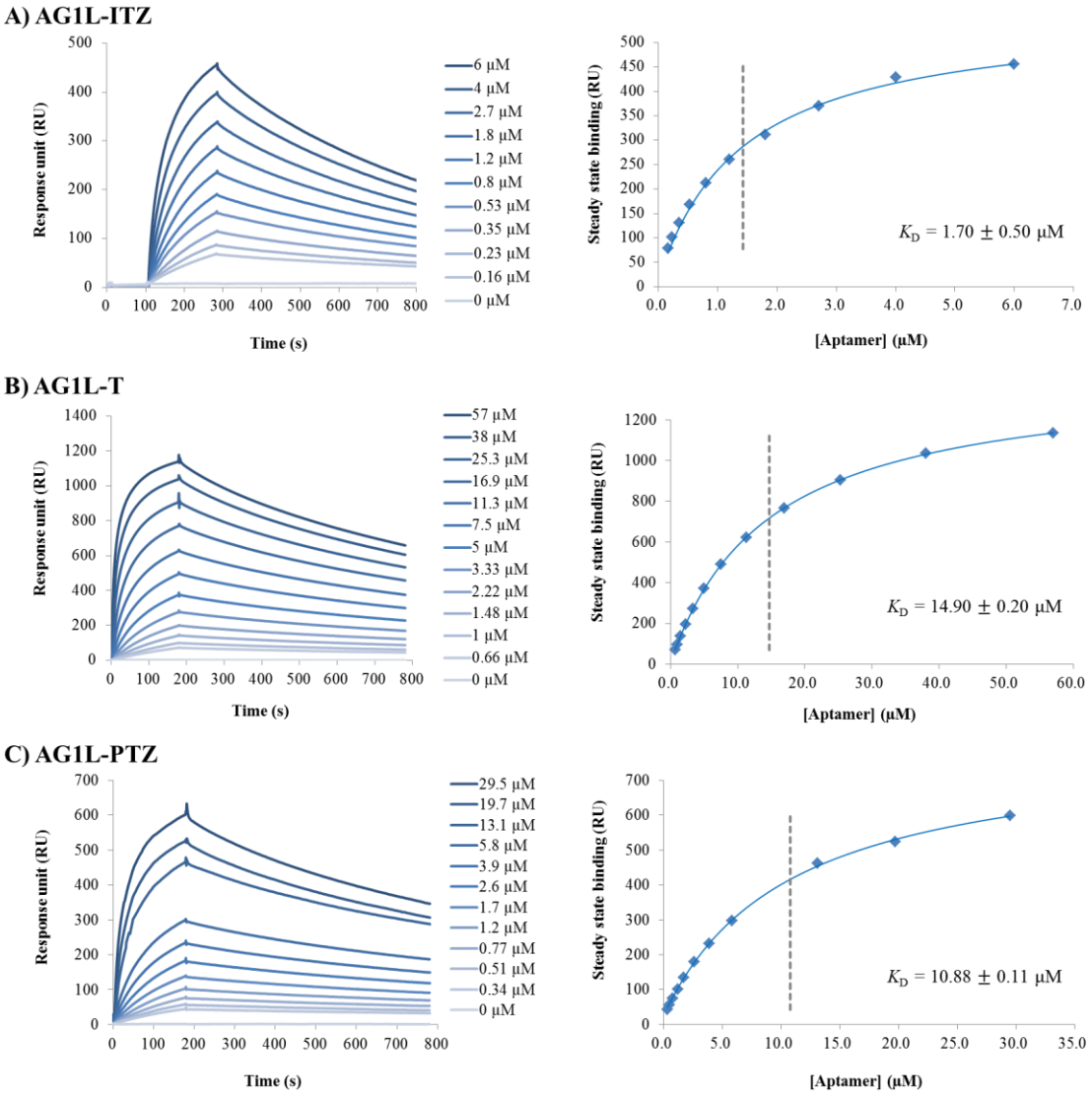


Figure S 8-29. SPR-based interaction analyses of the affinity of AG1L-ITZ and its control sequences to streptavidin (SA) and the corresponding steady-state binding data from the end of the association phases against analyte concentrations. This was used to calculate the steady-state affinities (K_D) of AG1L-ITZ (**A**), AG1L-T (**B**) and AG1L-PTZ (**C**). The positions of 5-ITZ-dU are coloured in red. The dashed grey line indicates the value when half the aptamer is bound to the immobilised SA, which is an estimated value of binding constant (K_D).

8.20 SPR sensorgrams of AN1-ITZ and its control sequences binding to streptavidin

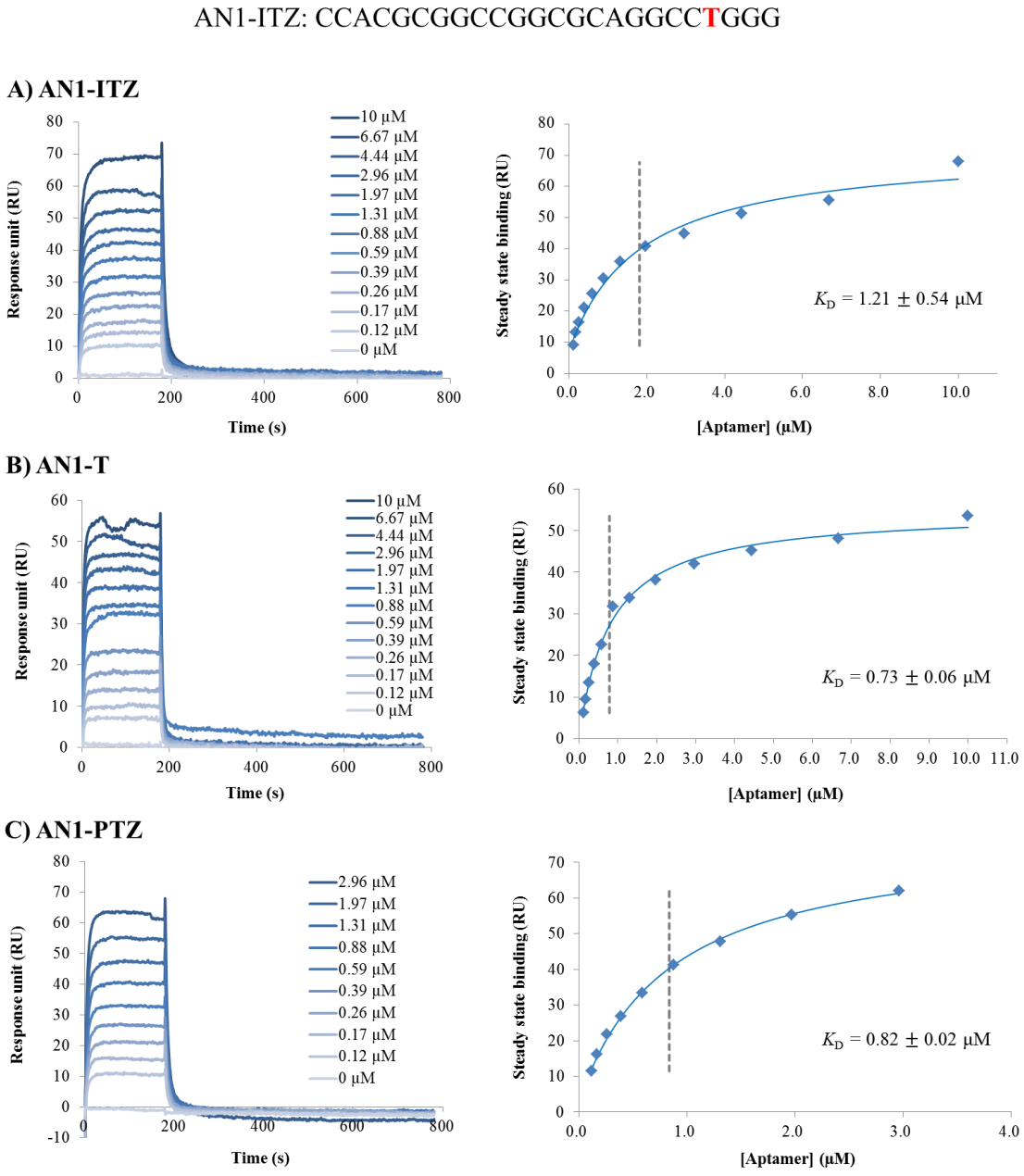


Figure S 8-30. SPR-based interaction analyses of the affinity of AN1-ITZ and its control sequences to streptavidin (SA) and the corresponding steady-state binding data from the end of the association phases against analyte concentrations. This was used to calculate the steady-state affinities (K_D) of AN1-ITZ (**A**), AN1-T (**B**) and AN1-PTZ (**C**). The position of 5-ITZ-dU is coloured in red. The dashed grey line indicates the point at which half the aptamer is bound to the immobilised SA, which is an estimated value of binding constant (K_D).

8.21 SPR sensorgrams of AN3-ITZ and its control sequences binding to streptavidin

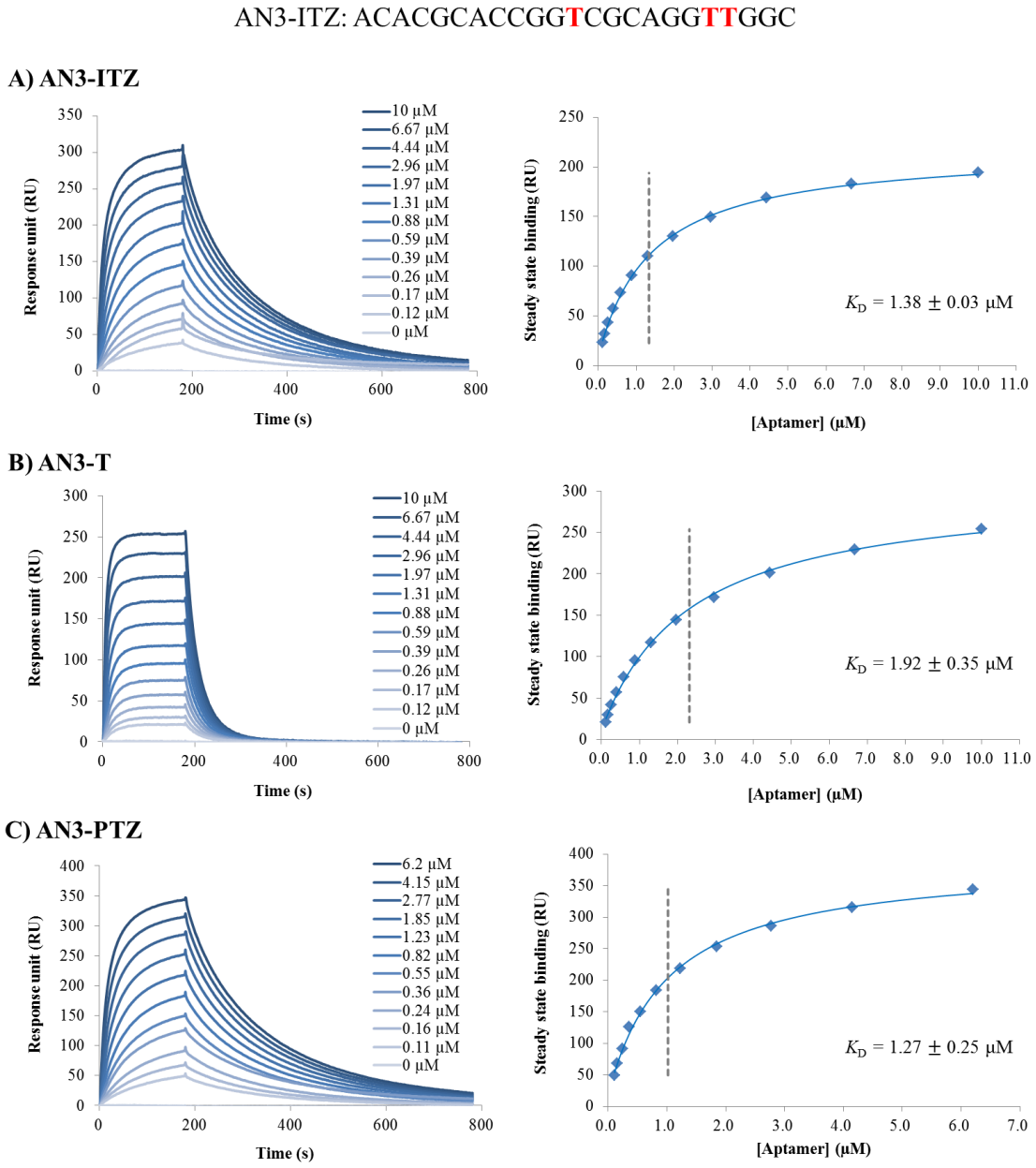


Figure S 8-31. SPR-based interaction analyses of the affinity of AN3-ITZ and its control sequences to streptavidin (SA) and the corresponding steady-state binding data from the end of the association phases against analyte concentrations. This was used to calculate the steady-state affinities (K_D) of AN3-ITZ (**A**), AN3-T (**B**) and AN3-PTZ (**C**). The positions of 5-ITZ-dU are coloured in red. The dashed grey line indicates the point at which half the aptamer is bound to the immobilised SA which is an estimated value of binding constant (K_D).