
# The compatibility of modified nucleic acids with enzymatic copying and ligand selection

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

Doctor of Philosophy

Yutong Zhang

St. Cross College, University of Oxford

Hilary Term 2022

**The compatibility of modified nucleic acids with enzymatic copying and ligand selection**

Yutong Zhang

(Doctor of Philosophy, St. Cross College, Hilary Term 2022)

# Abstract

Modified nucleic acids have important applications in molecular biology, genetics and genomics. Here I report a series of modified 5-aminomethyl-deoxyuridine (5-AM-dU) analogues with different functional groups on the 5-position of the pyrimidine ring. These modified 5-AM-dU nucleotides were evaluated in primer extension and polymerase chain reaction (PCR) amplification assays and were shown to be compatible with the polymerase enzymes. In total, six 5-AM-dU analogues with either stabilising or destabilising functional groups were prepared and efficiently incorporated by polymerases. Another modified nucleoside, the fluorescent quadracyclic adenine analogue qAN1 (qAN1) is incorporated with low efficiency in PCR. I was able to show that when this highly duplex stabilising modification is paired with the complementary duplex-destabilising 5-AM-dU, PCR amplification becomes possible. This is because the stabilising effects of one analogue offsets the destabilising properties of the other. Two destabilising thymidine analogues with hydrophobic functional groups at the 5′ position of the nucleobase enabled the use of qAN1 in place of deoxyadenosine at full substitution. In combination, these artificial qAN1-5-AM-dU base pairs allow the preparation chemically modified DNA in which all the A and T nucleobases are

modified. These chemically modified 5-AM-dU analogues were used into SELEX to generate a series of aptamers through *in vitro* selection. In parallel, similar approaches were used to construct modified nucleic acid libraries with aminoethyl glycine (AEG) substituents on thymidine.

# *Declaration*

I, Yutong Zhang, declare that the thesis entitled 'The compatibility of modified nucleic acids with enzymatic copying and ligand selection' and the work presented in this thesis are entirely my own.

I confirm that:

- This work was done wholly while in candidature for a research degree at this university;
- No part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution;
- 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 clearly given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help.

Yutong Zhang

September 2022

# *Acknowledgements*

For my memorable DPhil studies, I appreciate the help and support I received from many people and organisations. Without them, my DPhil would not have been achievable.

First of all, I would like to express my sincere gratitude to my supervisor, Professor Tom Brown, for offering me a DPhil position and the continuous support throughout my DPhil research. I am also very thankful for his immense knowledge, research vision, patience, and academic style which all guided me during my studies.

I am also grateful to all the current and previous members of Professor Tom Brown's group. Special thanks are attributed to Professor Afaf El-Sagheer for all the guidance, supervision, and especially, encouragement she provided during my DPhil. She is an ideal academic tutor who was always very amiable and stringent in oversight of my research. Many thanks to all the Brownies, including Arun, Yssy, Sara, Pawan, Cheng, Nicolo, Jagannath, Xiaomei, Jinfeng, Pui, Piotr, Jack H, Ellia, Sarah, Ashley, Phoebe, Brendan, Andrei, Cameron, Sven, Thanishta, Gerald, Lillian, Siqi, and Chunsen. I enjoyed working with them for these past five years and I appreciate the comfortable group environment they created and all the good memories we shared during these years.

I would also like to acknowledge the funding I received from China Scholarship

Council (CSC) and I express my profound thanks to them for their timely financial support. I never imagined I would be so lucky to receive such generosity.

Last, but not least, I would like to show all my gratitude and respect to my parents, grandma, brother, uncles and aunts from both sides of my family. As we always said "Every victory in the frontline is inseparable from the rear." It is their backup in every aspect that has guided me to each and every success I have attained so far in my career.

# *Abbreviations*

| | |
|---|---|
| ACN/MeCN | Acetonitrile |
| ADDP | *1,1'-(Azodicarbonyl)dipiperidine* |
| AEG | Aminoethyl Gylcine |
| aq | Aqueous |
| Boc | *tert*-Butyloxycarbonyl |
| bp | Base pairs |
| °C | Degrees Celsius |
| $CDCl_3$ | Deuterated chloroform |
| Conc. | Concentration |
| CuAAC | Copper(I)-catalysed azide alkyne cycloaddition |
| dA | 2′-Deoxyadenosine |
| DBU | 1,8-Diazabicyclo[5.4.0]undec-7-ene |
| DCC | Dicyclohexylcarbodiimide |
| DCM | Dichloromethane |
| DEL | DNA-Encoded Library |
| DEAD | Diethylazodicarboxylate |
| DIAD | Diisopropylazodicarboxylate |
| DIPEA | *N,N*-diisopropylethylamine |
| DMF | Dimethylformamide |
| DMSO | Dimethyl sulfoxide |
| DMTr | 4,4′-Dimethoxytrityl |

| | |
|---|---|
| DNA | Deoxyribonucleic acid |
| dsDNA | Double-stranded DNA |
| dNTP | Deoxynucleoside triphosphate |
| dT | Thymidine |
| dU | 2′-Deoxyuridine |
| ds | Double-stranded |
| Equiv./eq. | Molar equivalent |
| EDC | 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide |
| EDTA | Ethylenediaminetetraacetic acid |
| $Et_3N$ | Triethylamine |
| EtOH | Ethanol |
| EtOAc | Ethyl acetate |
| ESI | Electrospray ionisation |
| FAM | 6-Carboxyfluorescein |
| Fmoc | Fluorenylmethyloxycarbonyl |
| FRET | Fluorescence resonance energy transfer |
| g | Gram(s) |
| HATU | 1-[*bis*(dimethylamino)methylene]-1H-1,2,3-triazolo [4,5-b] pyridinium 3-oxidhexafluorophosphate |
| HBTU | 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate |
| Her2 | Human epidermal growth factor receptor 2 |
| HOBt | 1-Hydroxybenzotriazole |
| HPLC | High-performance liquid chromatography |

| | |
|---|---|
| hr. | Hour(s) |
| HTS | High throughput sequencing |
| IR | Infrared spectroscopy |
| $K_a$ | Association rate constant |
| $K_d$ | Dissociation rate constant |
| $K_D$ | Binding dissociation constant |
| MeOH | Methanol |
| m.p. | Melting point |
| mmol | Millimole |
| mL | Millilitre |
| min | Minute(s) |
| MW | Molecular weight |
| MS | Mass spectrometry |
| NGS | Next-generation sequencing |
| NMR | Nuclear Magnetic Resonance spectroscopy |
| NHS | *N*-Hydroxysuccinimide |
| μL | Microlitre |
| μM | Micromolar concentration |
| O.D. | Optical density |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |
| PEX | Primer extension |
| Ph | Phenyl |

| | |
|---|---|
| pH | Potential of hydrogen |
| PNA | Peptide Nucleic Acid |
| $PPh_3$/TPP | Triphenylphosphine |
| $PPh_3O$ | Triphenylphosphine oxide |
| ppm | Parts per million |
| PyBOP | Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate |
| RNA | Ribonucleic acid |
| r.t. | Room temperature |
| RT | Real-time |
| $R_f$ | Retention factor |
| SPPS | Solid-phase peptide synthesis |
| sat. | Saturated |
| SABL | Spontaneously Assembling Bivalent Ligand |
| SELEX | Systematic evolution of ligands by exponential enrichment |
| SOMAmer | Slow off-rate modified aptamer |
| ssDNA | single-stranded DNA |
| TBE | Tri-Borate-EDTA |
| TBP | Tributyl phosphate |
| t-Bu | Tertiary Butyl |
| TEAB | Triethylammonium bicarbonate |
| TFA | Trifluoroacetic |
| THF | Tetrahydrofuran |
| Tm | Melting temperature |

| | |
|---|---|
| TMAD | Tetramethylazodicarboxamide |
| TLC | Thin-layer chromatography |
| UV | Ultraviolet |
| V | Volt(s) |
| W | Watt(s) |
| ε | Extinction coefficient |
| δ | Chemical shift |
| Φ | Fluorescence quantum yield |
| λex/λem | Excitation/emission wavelength |

# Contents

*But to the degree that large industry develops, the creation of real wealth comes to depend less on labour time and on the amount of labour employed than on the power of the agencies set in motion during labour time, whose 'powerful effectiveness' is itself in turn out of all proportion to the direct labour time spent on their production, but depends rather on the general state of science and on the progress of technology, or the application of this science to production.*

*Karl Heinrich Marx*

*Fundamentals of a Critique of Political Economy*

# *Chapter 1.*

# *General Introduction*

# 1.1 General structure of nucleic acids (DNA/RNA)

## 1.1.1 Primary structure of nucleic acids

Nucleic acids are important biopolymers which store the genetic information of all known forms of life. There are two types of naturally occurring nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), which share substantive similarities in terms of their chemical structure. They are comprised of nucleotide monomers which are in turn composed of a phosphate group, a pentose (five-carbon) sugar (2′-deoxy-D-ribose in DNA and D-ribose in RNA) and a nitrogenous heterocyclic base (also known as nucleobase). Four nucleobases constitute the DNA code, namely adenine (A), guanine (G), cytosine (C) and thymine (T). In RNA, thymine (T) is replaced by uracil (U), although the other 3 nucleobases (adenine, guanine, and cytosine) are conserved. Among these nucleobases, adenine and guanine are purines, while cytosine, thymine and uracil are pyrimidines.[1]

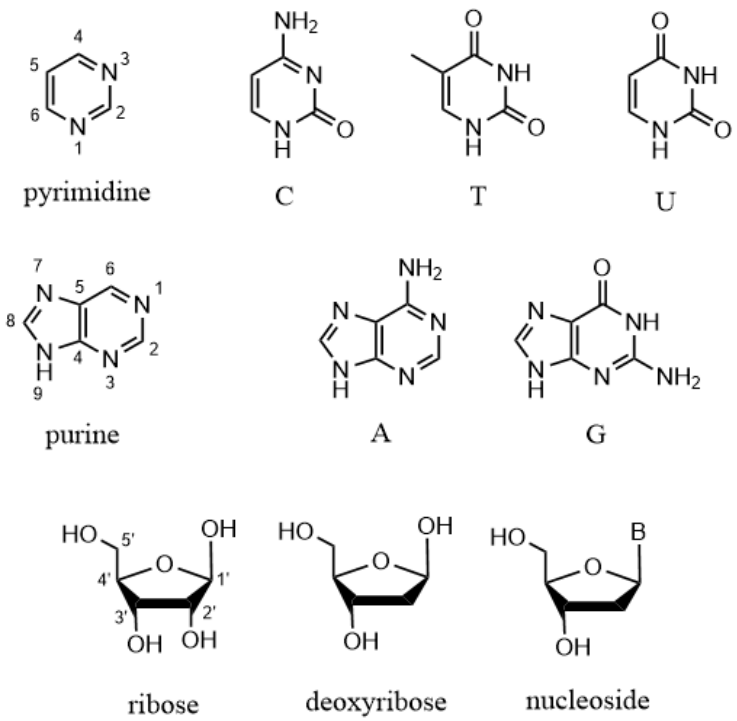



**Figure 1.1** Structures of the heterocyclic bases (B) present in nucleic acids and chemical structures of a ribose, deoxyribose and nucleoside. Cytosine (C), thymine (T), uracil (U) are pyrimidines; while adenine (A) and guanine (G) are purines.

Nucleobases are attached to the pentose sugar (in furanose form) via an N-glycosidic bond connecting the C-1′ of the pentose sugar ring and the N-9 (in purines), and the N-1 (in pyrimidines). These structures are known as nucleosides. There are four types of deoxyribonucleosides in DNA, namely deoxyadenosine (dA), deoxyguanosine (dG), deoxycytidine (dC) and thymidine (dT); and four types of ribonucleosides present in RNA, specifically adenosine (A), guanosine (G), cytidine (C) and uridine (U). The C-1′ position of the pentose sugar is the anomeric centre. If the nucleobase attached to the 1′-carbon lies on the same face of the sugar ring as the 5′-hydroxyl group, it is known as a β-anomer. On the other hand, if the nucleobase is on the opposite side of the sugar

ring to the 5′-hydroxyl group it is termed an α-anomer. Naturally occurring DNA and RNA nucleosides adopt the β-configuration.[2]

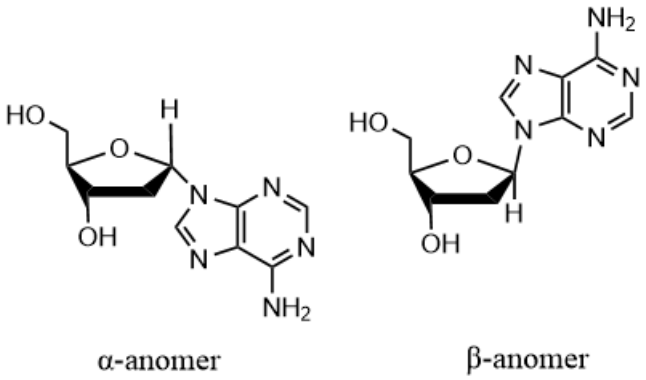



**Figure 1.2** Structures of α- and β-anomers of deoxyadenosine.

Nucleobases in DNA and RNA can rotate around the N-glycosidic bond giving rise to syn- and anti-conformations. However, the rotation about this bond is restricted due to steric interference from the sugar moiety. Hence, the anti-conformation is generally favoured (although there are still some exceptions to this rule; e.g. guanosine monophosphate, in which the guanine base readily adopts the *syn*-conformation about the glycosidic bond).[2]

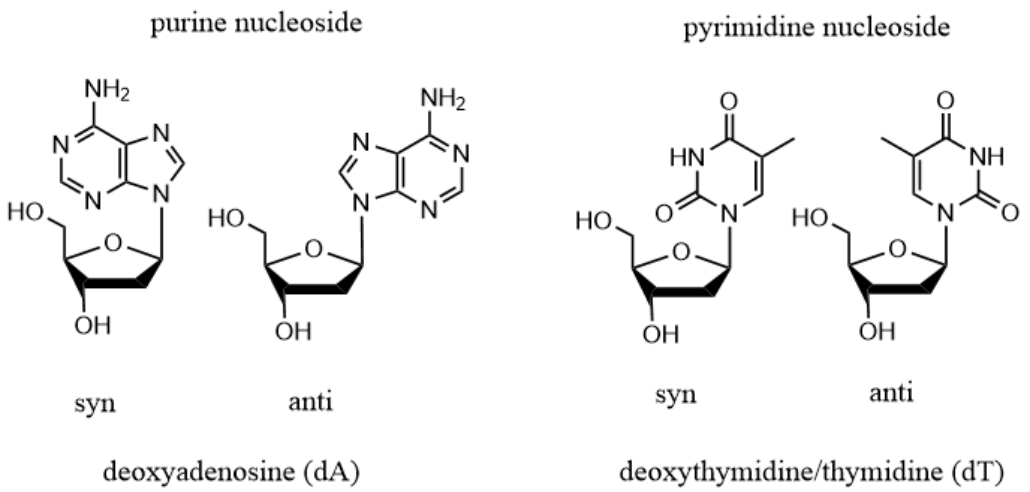



**Figure 1.3** Structures of *syn* and *anti*-nucleoside conformations of purines and pyrimidines. Deoxythymidine is commonly referred to as thymidine.

The phosphate group can be found either at the 5′- or 3′-position of the sugar depending on the method used to break down DNA to yield the individual nucleotides. A dinucleotide (dimer) of DNA/RNA is formed by covalently linking the 5′-phosphate group of one nucleotide to the 3′-hydroxyl group of another to form a *phosphodiester bond*. A single-stranded oligonucleotide (oligomer) is formed when several of these phosphodiester bonds are formed between nucleotides, resulting in a linear, high molecular weight molecule. As each phosphodiester group is anionic, oligonucleotides are highly charged polyanionic molecules (hence the term *nucleic acid*). The resulting oligonucleotide strand contains a 5′-hydroxyl group at one end and a 3′-hydroxyl group at the other. Conventionally, the oligonucleotide sequence is written in the 5′- to 3′-direction.

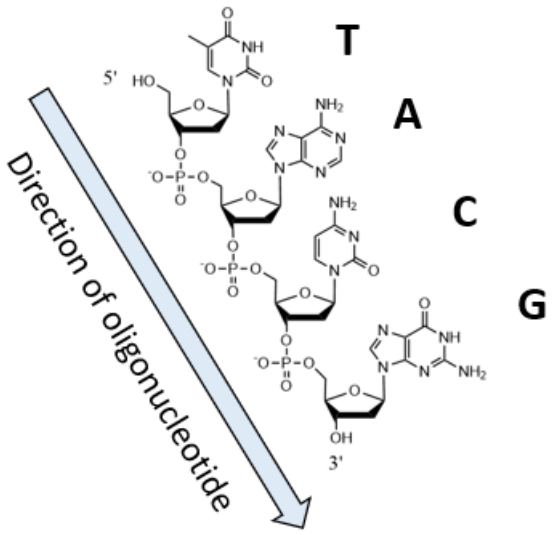



**Figure 1.4** Oligonucleotide sequence 5′-TACG-3’ commonly written as TACG.

## 1.1.2 Secondary structure of nucleic acids

In 1949, Erwin Chargaff discovered that the molar quantity of adenine in DNA was always equivalent to that of thymine (i.e. the number of moles of A = number of moles of T) and the same also held true for guanine and cytosine (number of moles of G = number of moles of C).[3] This fundamental discovery was known as Chargaff’s rule. In 1952, Rosalind Franklin and Maurice Wilkins proposed that the secondary structure of DNA is helical, based on X-ray DNA fibre diffraction data.[4,5] Taking advantage of these findings, James Watson and Francis Crick first published the double-helical model of DNA in 1953, proposing that the two strands of DNA are held together by hydrogen bonds between individual bases on opposing strands.[6] The double-helical model clearly elucidates the mechanism of base pairing and genetic information heritage wherein the purine base A always pairs with an opposing pyrimidine T through 2 hydrogen bonds,

and the purine base G always pairs with a pyrimidine C via 3 hydrogen bonds in double-stranded DNA.

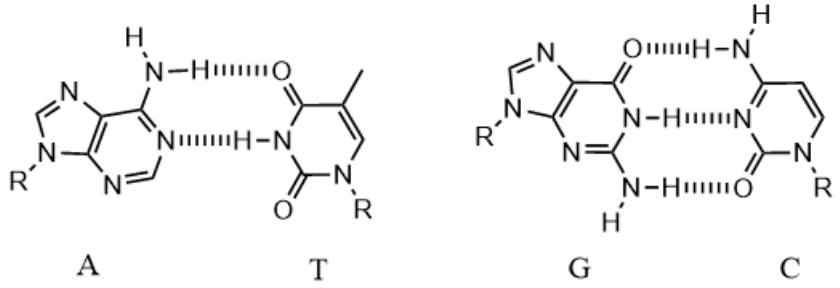



**Figure 1.5** Hydrogen bonding within A·T and C·G Watson-Crick DNA base pairs.

The two DNA strands run antiparallel in opposing directions, one in the 5′ - 3′ direction and the other in the 3′ - 5′ direction. The sequence of these two DNA strands is thus reverse- complementary due to the obligatory base pairing (with the 5′-end of one strand being adjacent to the 3′-end of the other). Hence, the sequence of one strand of DNA precisely defines the sequence of the other one. Two hybridised DNA strands direct the hydrophobic base pairs into the centre of the duplex surrounded by a hydrophilic external backbone constructed of sugars and negatively charged phosphate moieties. There are three major possible conformations of the DNA double helix structure, designated A-, B- and Z- forms. B-DNA is the most common form and is the form predominately found in living cells. B-DNA is a right-handed helix (right-handed stands for a clockwise screwing motion move away from the observer with the line of sight along the helical axis), with a wide major groove and a narrow minor groove. Each 36 Å helix pitch (a complete turn about the axis) contains 10.5 base pairs and the resulting helix diameter is 20 Å. The inter-base distance (distance between two

successive base pairs) in B-DNA is around 3.4 Å.[7,8]

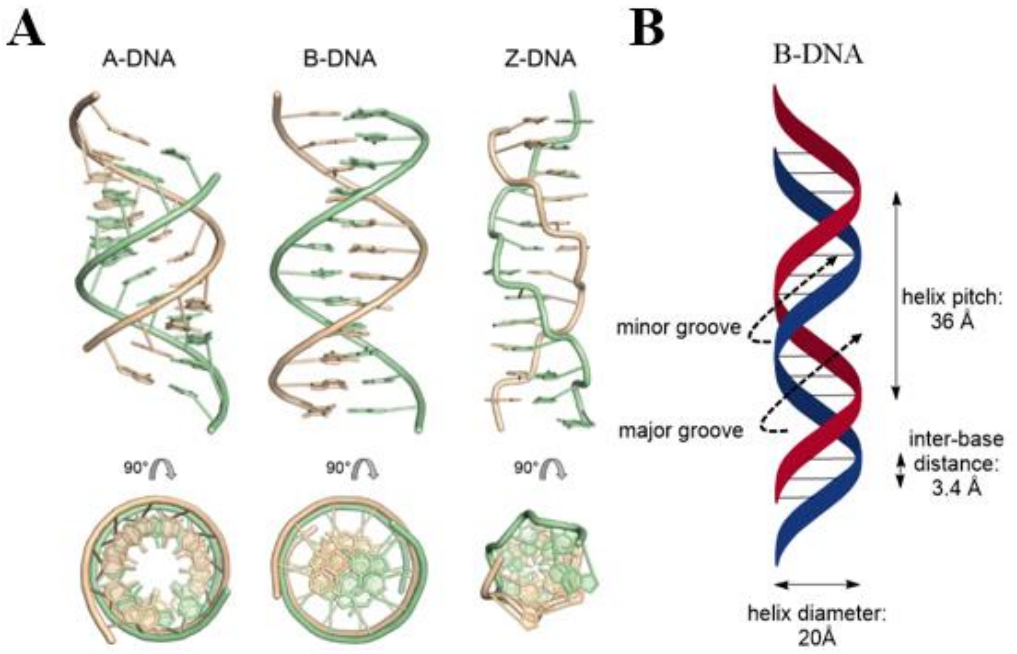



**Figure 1.6 (A)** Three-dimensional structures of A-DNA, B-DNA and Z-DNA. Figure was adapted with permission from [9]; **(B)** Schematic representation of the B-DNA.

### 1.1.3 Stability of the DNA duplex

The stability of the DNA double helix depends on a balance of several electrostatic interactions, including hydrogen bonding between base pairs, nucleobases and their surrounding water molecules, and base-stacking interactions between adjacent bases. The heterocyclic bases of single-stranded DNA may present polar amido, amidino, guanidino and carbonyl groups, which together form a complex network of hydrogen bonds with surrounding water molecules. During the formation of a DNA duplex, some of these bonds may be broken along with the formation of inter-base hydrogen bonds.

Hence the net change in enthalpy upon duplex formation is partially due to $\Delta H$ (H-bonds formed) − $\Delta H$ (H-bonds broken).[10] Most of the duplex formation process is exothermic and entropically unfavourable (with negative $\Delta S$). Another important factor for duplex stability, other than inter-strand hydrogen bonding, is the base-stacking effect. Base-stacking interactions depend on the aromaticity of the bases and their dipole moments are strongly hydrophobic and electrostatically favourable.[11] The degree of stability affected by the base-stacking effect varies among different DNA sequences with different combinations of base pairs. Generally, nearest neighbour base-stacking interactions are important determinants of duplex stability.[12] Salt concentration is another important factor that affects the degree of base-stacking interactions. Since high salt concentrations in solution (e.g. $Mg^{2+}$ and $Na^{+}$) reduce the destabilising charge repulsion arising between two negatively charged phosphodiester backbones, base-stacking interactions tend to increase with increasing salt concentration.[12,13]

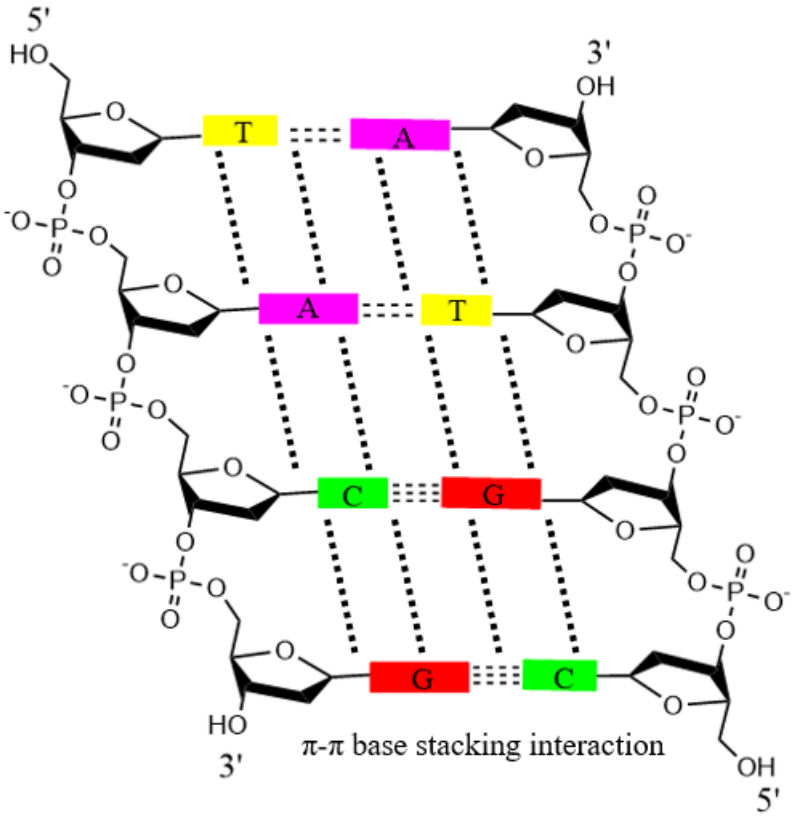



**Figure 1.7** DNA duplex, schematic representation of intermolecular π-π stacking interactions between internal aromatic rings.

The thermal denaturation (unwinding) of double-stranded DNA through heating disrupts base-stacking and causes hydrogen bonds to break. The whole process can be monitored by measuring the change in UV absorbance of the nucleobases at 260 nm.[14] The melting temperature ($T_m$) is defined as the temperature at which half of the DNA strands are in random coil form, in a single-stranded state. At the melting temperature ($T_m$), the equilibrium is at the mid-point, at which the ratio of the dsDNA to the single-stranded DNA becomes equal. $T_m$ is a very important index of the thermal stability of double stranded DNA and is dependent on several conditions including the DNA base sequence, presence of mismatched base pairs, DNA concentration, solvent, nature and

concentration of salts, and pH.

To measure the thermal stability of DNA duplexes, the intrinsic properties of nucleic acids are monitored. DNA nucleobases are naturally UV active. Therefore, monitoring by UV signal is commonly used to acquire the $T_m$ of a DNA duplex. This is because the UV signal of a dsDNA duplex is lower than for ssDNA due to its stronger π-π base-stacking interactions. At low concentrations, the UV signal of DNA nucleobases is too weak to be detected, and a more sensitive method than UV detection is required. Fluorescence melting using DNA intercalators can be used to determine the $T_m$ of DNA duplexes at nanomolar (nM) concentrations. Fluorescence DNA intercalators such as SYBR Green or EvaGreen exhibit a greater than 10-fold increase in fluorescence after intercalation into DNA duplexes (through the insertion of a planar aromatic substituent between the two neighbouring nucleobase pairs of dsDNA) compared to the single-stranded form.[15] These intercalators have been widely used to monitor the progression of real-time PCR and to investigate the thermal stability of PCR products (the higher the amounts of duplex product generated during amplification, the stronger the signal, as more intercalator molecules are bound).[16,17]

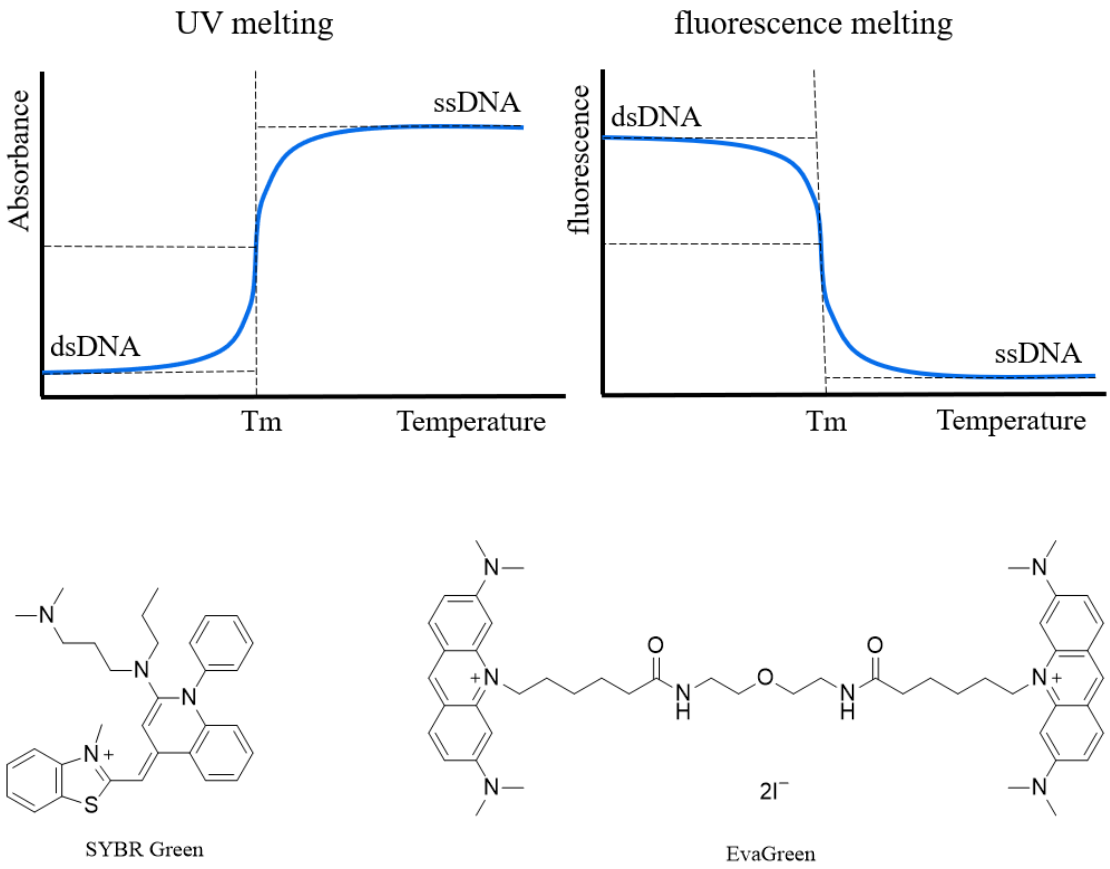



**Figure 1.8** A schematic of a DNA UV melting curve, fluorescence melting curve and chemical structures of two fluorescent DNA intercalators: SYBR Green and EvaGreen.

## 1.2 Solid-Phase oligonucleotide synthesis

The solid-phase synthesis method was invented in the 1950s by Bruce Merrifield. In this method starting molecules are chemically bound to a solid support material and polymers (e.g. peptides or oligonucleotides) are synthesised/assembled in a step-by-step process through exposure to reagents in the solution phase.[18,19] Solid supports (also called resins) are insoluble particles that are typically 50-200 μm in diameter. The most frequently used solid supports in oligonucleotide synthesis are controlled pore glass (CPG) and polystyrene (PS).[20] Compared to normal solution phase chemical synthesis, solid-phase synthesis: (i) increases reaction rates due to large excesses of solution reagents; (ii) eliminates impurities and by-products by a simple washing of the reaction vessel after each step, and (iii) allows the whole process to be automated on computer-controlled solid-phase synthesisers. Today both small and large scale synthesis of oligonucleotides is almost exclusively carried out by the solid-phase method.

Several chemical processes have been developed for use in solid-phase oligonucleotide synthesis throughout. In the early 1950s, H-phosphonate and phosphotriester methods were described by Michelson and Todd.[21,22] Khorana developed phosphodiester approaches in the 1950s wherein 3'-*O*-acetylnucleoside-5'-*O*-phosphate was activated by DCC and Ts-Cl.[23] A reinvestigation of the phosphotriester method and the development of phosphite triester approaches by Reese and Letsinger respectively, was eventually followed by the use of the 2-cyanoethyl P(III) derivatives of nucleosides (3'-*O*-chlorophosphites).[24–26] The phosphoramidite method was pioneered by Marvin

Caruthers in the early 1980s. Since then, this method was quickly adopted and has been used universally for oligonucleotide synthesis.[27] In the phosphoramidite method, the 5′-position of sugar is protected by an acid-labile 4,4′-dimethoxytrityl (DMTr) protecting group. Nucleobases with exocyclic primary amino groups (A, C and G) are nucleophilic and, therefore, also need to be protected through the synthetic processes (**Figure 1.9**).[28,29] The benzoyl protecting group on A and C is cleaved quickly in ammonium hydroxide. However, the isobutyryl protecting group on G is more resistant to hydrolysis, and the rate-determining step in oligonucleotide deprotection is the cleavage of the isobutyryl group from guanine bases. In some cases, using certain chemically modified oligonucleotides, heating in ammonia can lead to degradation, so a more labile guanine protecting group, dimethylformamidine (dmf-dG) is used, which allows oligonucleotide deprotection under milder conditions.

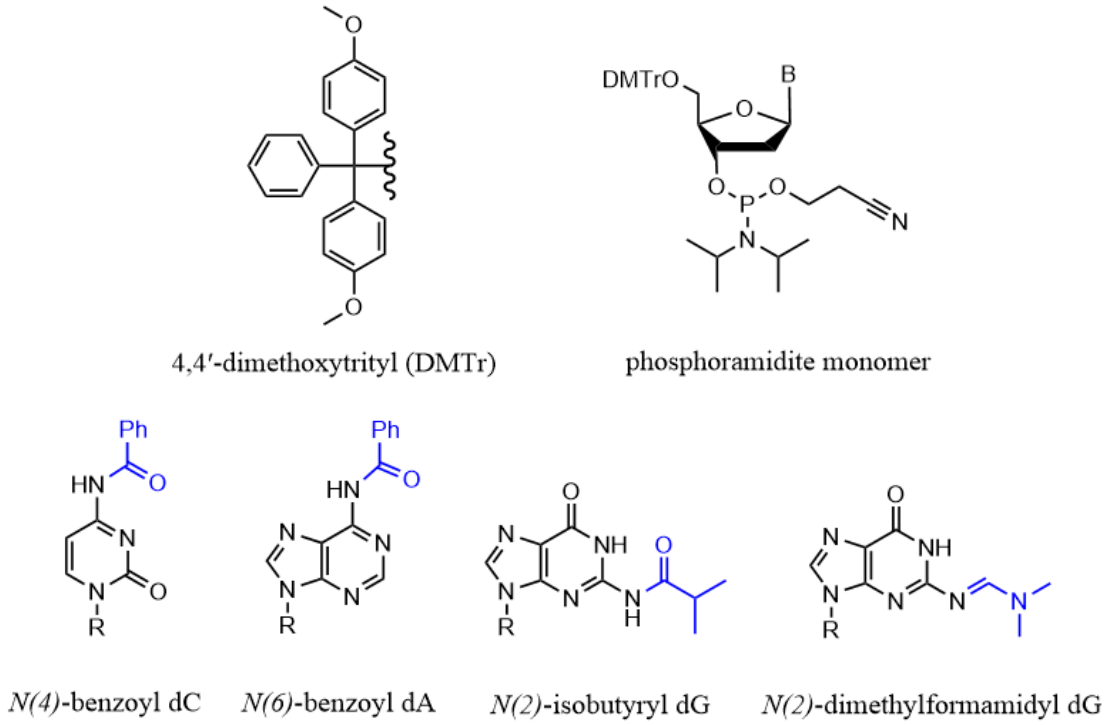



**Figure 1.9** Structure of phosphoramidite monomer with DMTr protecting group and nucleobase protecting groups required in the phosphoramidite method.

In the phosphoramidite method the synthesis of DNA sequences is normally carried out in the 3′- to 5′-direction, which is opposite to the biosynthesis of DNA by replication which takes place in the 5′- to 3′- direction. A solid support is linked to the 3′- nucleoside and each nucleotide is added, step-by-step, at the 5′- position of the support-bound growing DNA strand. There are 4 steps in each synthesis cycle, namely (i) detritylation, (ii) activation and coupling, (iii) capping and, (iv) oxidation.[30] In the detritylation step, the DMTr protection group is removed by washing with 3% trichloroacetic acid (TCA) in dichloromethane or toluene to cleave the expose the 5′-hydroxyl group of the support-bound terminal nucleoside. In the activation and coupling steps, a phosphoramidite monomer is first dissolved in acetonitrile and then activated by mixing with the tetrazole catalyst. Protonation of the phosphoramidite by tetrazole (or a derivative) makes diisopropylamino a good leaving group, which is directly replaced by the 5′-hydroxyl group of the solid support-bound nucleoside, thereby generating a phosphite triester linkage. An alternative mechanism involves displacement of the di-isopropylamino group by tetrazole which is then displaced by the 5′-hydroxyl group. Since it is not possible to achieve a 100% coupling efficiency in this reaction, a capping step is then performed to block any remaining unreacted 5′-hydroxyl groups on the solid support-bound nucleotide. Two capping solutions are used in this step, namely acetic anhydride and *N*-methylimidazole (NMI). This step lowers the number of deletion mutations in the oligonucleotide product. The oxidation step serves to convert the unstable P (III) to a stable P (V) state before moving on to the next TCA detritylation step. Oxidation is achieved by iodine oxidation in the presence of pyridine and water. Once all synthesis cycles are completed, the synthesised oligonucleotide is heated at

55 °C in concentrated aqueous ammonium hydroxide for 5 hours to cleave it from the solid support resin and to remove all the protective groups on nucleobases as well as the cyanoethyl groups present on phosphates ($E_1$cB-elimination). This aqueous solution is then removed by evaporation and the resulting oligonucleotide is ready for further purification.

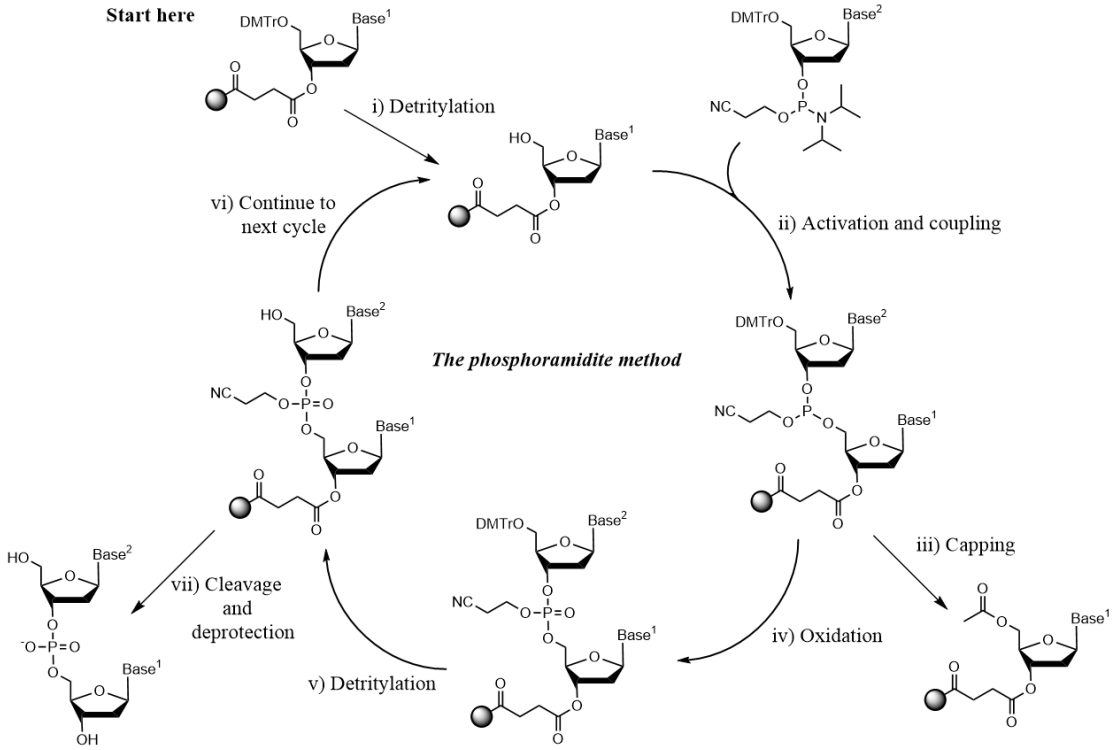



**Figure 1.10** Synthetic cycle of oligonucleotide synthesis by the phosphoramidite method.

# 1.3 Purification of oligonucleotides

## 1.3.1 Gel filtration

Gel filtration, also known as size-exclusion chromatography and molecular-sieve chromatography, is a technique used for the separation of chemical components in a mixture based on molecular size. In this process, separation occurs due to the differential capacities of the molecules in the sample to enter the pores of the gel filtration medium. The column in the stationary phase consists of beads of a hydrate, sponge-like material (for example, Cytiva illustra NAP$^{TM}$ columns packed with Sephadex G-25 DNA grade beads) which act as molecular sieves. When an aqueous solution with molecules of different sizes passes through the column, molecules larger than the pores of the medium move quickly through the column, while smaller molecules enter the pores and diffuse more slowly.[31] Therefore, the molecules are eluted from the column in order of molecular size. Inorganic salts such as $Na^+$ and $K^+$ are small and are retained within the column. Gel filtration is also used as a simple technique to remove salts from oligonucleotides that have been purified by HPLC.

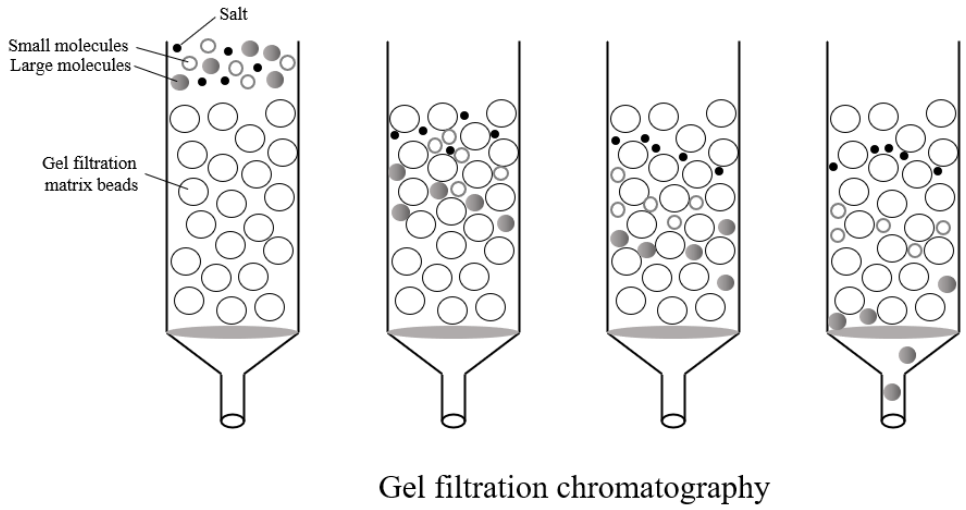



**Figure 1.11** Gel filtration. Schematic representation of gel filtration for the purification of oligonucleotides. Large molecules elute quickly through the column, while smaller molecules are retained in the gel matrix and elute more slowly.

### 1.3.2 High-performance liquid chromatography

High-performance liquid chromatography (HPLC) is an efficient technique for the separation, analysis, and purification of synthetic oligonucleotides. In this process, the dissolved sample mixture is delivered to a column filled with a solid adsorbent material by high-pressure pumps. Each component in the sample interacts with the adsorbent material differently, causing different migration rates through the column, leading to the separation of its constituent components.[32] Two HPLC methods have proven extremely useful in the separation and purification of oligonucleotides, reversed-phase and anion-exchange HPLC. A variation on reversed-phase, ion-pairing, is also a useful method.

Reversed-phase HPLC separates oligonucleotides based on differences in their hydrophobicity. The mixture of oligonucleotides is injected into a column of immobilised hydrocarbon chains (normally $C_8H_{17}$ or $C_{18}H_{37}$) bound to a silica support stationary phase. Molecules with greater hydrophobicity elute more slowly, and a gradient of acetonitrile in ammonium acetate buffer leads to separation of oligonucleotides.[33] In general long oligonucleotides (with more nucleobases) elute slower than shorter impurities.

Anion-exchange HPLC separates oligonucleotides through differences in their anionic charges. Since each nucleic acid building block contains a negative charge within the phosphodiester groups, oligonucleotides are polyanionic and the net charge of the molecule depends on the total number of phosphodiester groups. The anion-exchange column contains a stationary phase composed of an ion exchange resin with positively charged groups such as diethyl-aminoethyl (DEAE) attached. The separation of a mixture of oligonucleotides is accomplished by gradually increasing the ionic strength of the mobile phase to weaken the interactions between the oligonucleotides and the stationary phase. Oligonucleotides with higher charges (longer sequences) elute more slowly, leading to a separation of the desired oligonucleotide from truncated sequences.[34]

### 1.3.3 Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) is a technique used for separating macromolecules (e.g., oligonucleotides of 100 bases or longer) based on their

respective electrophoretic mobility. Electrophoresis is the motion of dispersed particles relative to a fluid under the influence of a uniform electrical field. A polyacrylamide gel is formed through the polymerisation of acrylamide and the size of its pores is determined by the concentration of acrylamide used. When an electric field is applied to the gel containing the sample, the negatively charged oligonucleotides travel toward the anode with different mobilities. Oligonucleotides with higher charges (since each nucleotide possess only one phosphate group with negative charge, oligonucleotide with higher charges could also represent oligonucleotide with longer sequences) move slowly through the electric field than oligonucleotide with lower charges. The difference of moving speed between different size of oligonucleotides through electric field leading to the separation of oligonucleotides from impurities such as failure sequences from the solid-phase synthesis.[35,36]

Denaturing PAGE is a specific technique that eliminates oligonucleotide secondary structures or duplexes to ensure consistent gel migration. This is essential for the analysis of oligonucleotide species based on size differences, as the secondary structures change the mobility of the sample travelling through the gel. In denaturing PAGE, a high concentration of urea is used to disrupt base pairing interactions (by acting as hydrogen bond donor/acceptor) and a mixture of formamide and $H_2O$ are loaded together with the sample to prevent duplex formation.

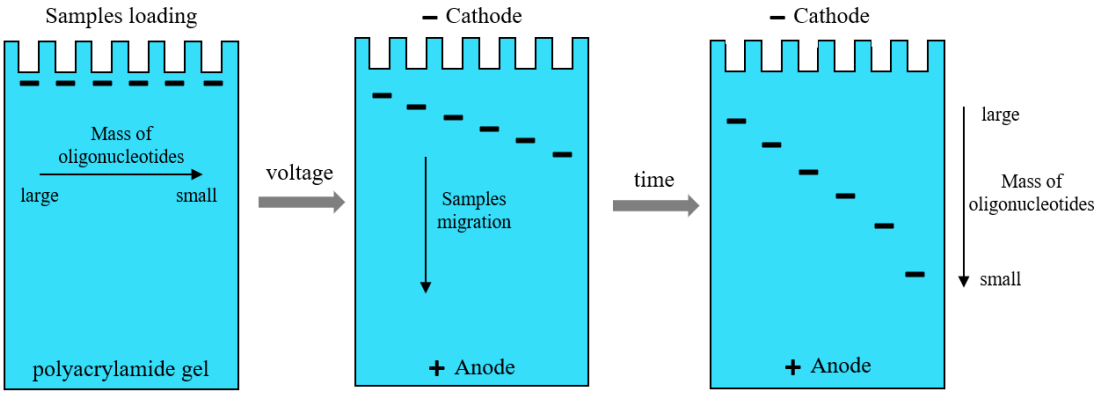



**Figure 1.12** Schematic representation of polyacrylamide gel electrophoresis (PAGE). Samples of oligonucleotides are loaded into the top of the gel and run in a buffer that contains conductive ions.

# 1.4 PCR and real-time PCR

## 1.4.1 PCR

PCR is a technique that is widely used in molecular biology, diagnostics, forensic science, genetics and genomics to amplify a specific region of a very small sample of DNA.[37] PCR was invented by Kary Mullis in 1984.[38] A short region of template DNA is copied exponentially by DNA polymerases in a buffer solution system using deoxynucleotide triphosphates (dNTPs) as building blocks to form new DNA strands. The region of the DNA to be amplified (amplicon) is defined by two short oligonucleotides (primers) which are designed to be complementary to the 3′- end of each target strand.

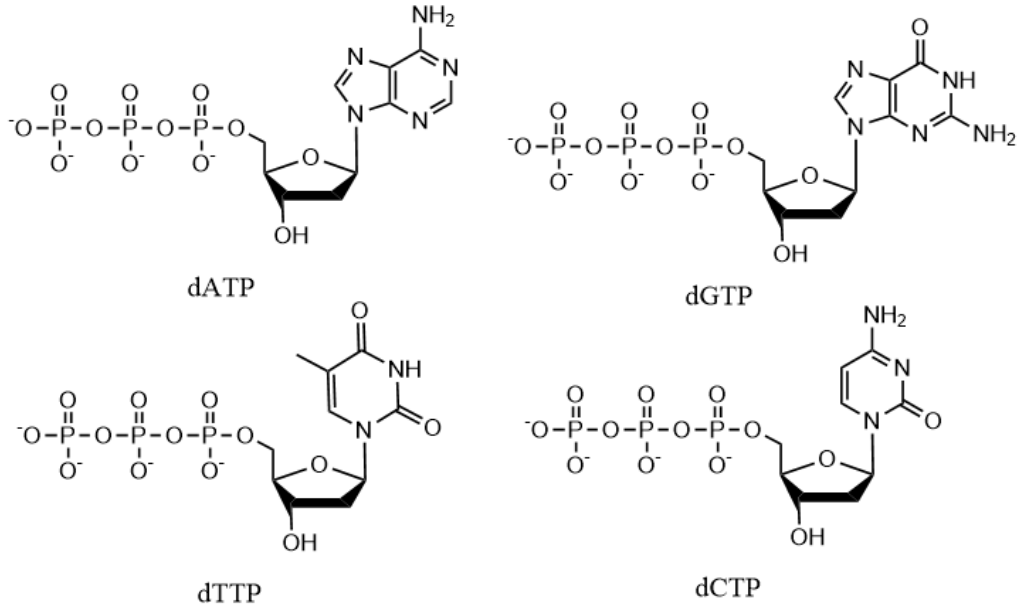



**Figure 1.13** Structures of canonical nucleoside triphosphates (dNTPs).

The DNA strands are first denatured by breaking inter-base hydrogen bonds with heat. The primers with complementary sequences are then annealed to both strands of the single-stranded DNA template, allowing the DNA polymerase to create the complementary sequence. DNA polymerase extends primers along the DNA template by continuously adding dNTPs to the 3′- hydroxyl group the growing DNA strand. The newly formed duplex can be denatured once again (in the next cycle of PCR) by heating to form single-stranded DNA which serves as a template for further DNA replication. By denaturation at 95 °C, then annealing at between 50 °C and 65 °C, followed by extension (elongation) at 72°C for around 30 seconds, a complete thermal cycle of PCR is thus achieved. Typically, a PCR reaction comprises 20 – 40 such thermal cycles and, in theory, a total of $2^n$ DNA copies in the form of PCR products will be produced after n cycles of PCR. A PCR reaction of 30 cycles will result in $2^{30}$, or 1,073,741,824 copies of the original double-stranded DNA target region.[39]

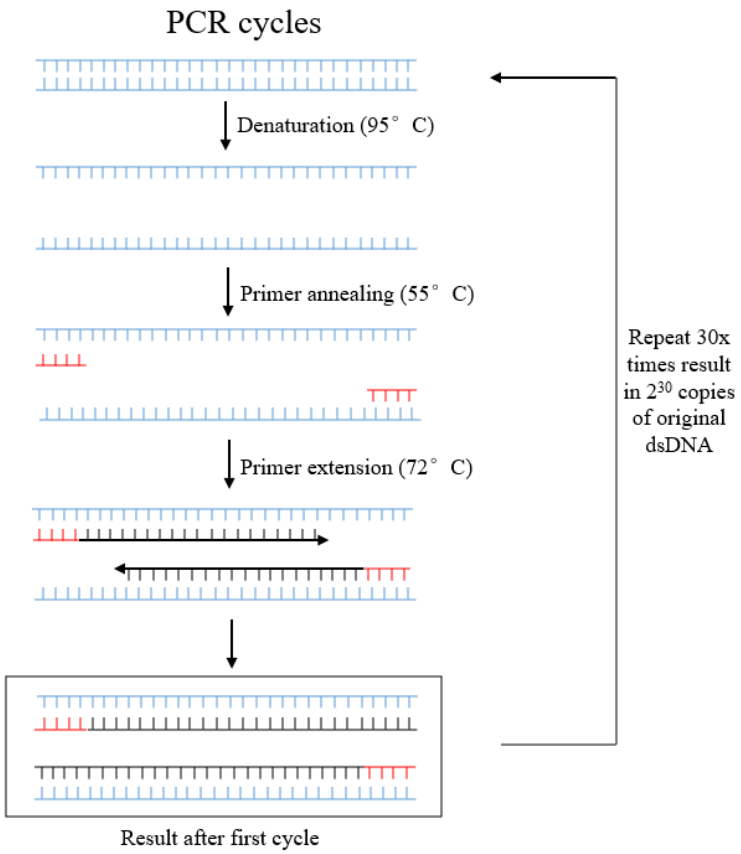



**Figure 1.14** PCR cycles of denaturation, annealing and primer extension/elongation. In the first cycle of PCR the 5'-ends of the amplicons are defined but the 3'ends are variable. After several cycles the synthesised DNA strands are dominated by products with defined 5'- and 3'-ends.

Though all organisms contain DNA polymerases, the polymerases used in PCR are special as they are highly heat-resistant (stable at 95 °C). A commonly used polymerase in PCR is *Taq* polymerase which was named after the thermophilic bacterium *Thermus aquaticus* which was first discovered in 1969 by Thomas D. Brock. Its polymerase was subsequently isolated in 1976 by Alice Chien.[40,41] *Taq* polymerase can withstand temperatures up to 97.5 °C (for 9 minutes) with an optimum temperature for polymerase

activity in the region of 75-80 °C which makes it an ideal polymerase for PCR.[42] *Taq* polymerase is however not the only option for PCR, many other DNA polymerases (e.g., *Pfu* polymerase, Vent polymerase) were subsequently found (or engineered).

### 1.4.2 real-time PCR/qPCR

Real-time PCR, also known as the quantitative polymerase chain reaction (qPCR), is based on the original PCR method. The major difference between conventional PCR and real-time PCR is that real-time PCR monitors the amplification of targeted DNA oligonucleotides during the PCR process itself, whereas the output of conventional PCR can only be monitored at its conclusion (end-point).[43] The real-time detection of PCR products is achieved either by adding non-specific fluorescent dyes that only intercalate with double-stranded DNA or by using oligonucleotide probes that only produce signals after hybridisation with specific complementary DNA sequences.[44,45] SYBR Green is a typical fluorescent dye that is generally used in real-time PCR as it binds to double-stranded DNA to form a fluorescent complex.[15]

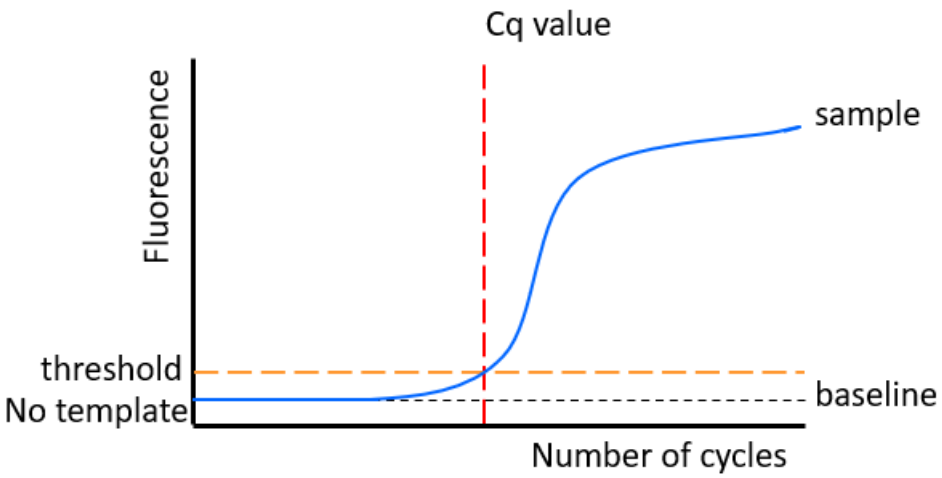



**Figure 1.15** The increase of fluorescence corresponds to the exponential increase in the concentration of PCR product. Quantification of the PCR product is determined from the value of the quantification cycle ($C_q$) or threshold cycle ($C_t$) in each reaction.

Studies of DNA quantification by real-time PCR rely on plotting the fluorescence signal against the number of cycles (**Figure 1.15**). A threshold is usually set for the detection of DNA-based fluorescence that is 3 to 5 times the standard deviation of the background noise signal (this is commonly set automatically by software). The number of cycles at which the fluorescence signal exceeds the threshold is termed the threshold cycle ($C_t$) or, according to the MIQE guidelines, quantification cycle ($C_q$).[46]

The $C_q$ value indicates how many cycles it takes to detect a significant signal from the sample in real-time PCR. Lower $C_q$ values indicate good PCR efficacy and a higher quantity of target sequences, whereas a higher $C_q$ value means a lower amount of target sequence was present and/or lower PCR efficacy.

# 1.5 Aptamers and SELEX approaches

## 1.5.1 Aptamers

Nucleic acid aptamers are single-stranded (ss) DNA or RNA oligonucleotides with unique three-dimensional conformations that bind to specific target molecules including proteins (including receptors)peptides, carbohydrates, small molecules and even live cells.[47–51] The term "Aptamer" was coined by Ellington and is derived from the Latin word "*aptus*" meaning 'to fit' and the Greek term "*meros*" meaning 'part'.[52] The concept of aptamers were first reported in 1990 by Gold and Szostak.[52,53] Aptamers can possess a variety of shapes with a combination of stems, loops, hairpins, bulges, G-quadruplexes and/or pseudoknots.[54–58] Their unique three-dimensional conformations make them extremely versatile in binding to their targets with high selectivity and specificity. Target recognition and binding of aptamers involves aromatic interactions, hydrogen bonding, π-π stacking, hydrophobic interactions, electrostatics, intercalation and three-dimensional shape-dependent interactions.[54,59] Based on these properties, aptamers are also known as "chemical antibodies".[60]

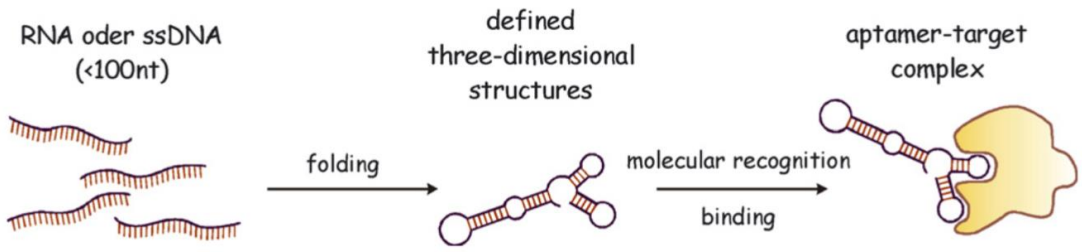



**Figure 1.16** Schematic diagram of the functionality of oligonucleotide aptamers. Figure adapted with permission from [61].

Although the applications of antibodies tend to be more well-developed, aptamers possess several desirable features which antibodies lack, namely:

(i) Large-scale synthesis and easy modification: The process of production of monoclonal antibodies is both time-consuming and expensive, thereby limiting large-scale antibody production.[62] In contrast, aptamers can be automatically synthesised on a large scale with excellent efficiency and reproducibility by chemical methods, in particular solid-phase oligonucleotide synthesis. Aptamers can also be chemically modified at specific positions on their sugar moieties, bases and/or phosphate backbones to enhance their stability and nuclease resistance. For example, the reported sugar substitutions of aptamers include 2′-fluoro, 2′-hydroxymethyl, 2′-azido, 2′-amino and 2′-methoxy modifications[63–66] and DNA backbone modifications can include changing the phosphodiester bond into phosphorothioate or phosphorodithioate.[67,68]

(ii) Higher stability and better storage properties: Antibodies are easily denatured

under non-physiological conditions and are also difficult to re-fold correctly.[69] In contrast, the denaturation of aptamers is reversible; their secondary structure is readily re-formed.[70–72] The higher stability of aptamers enables a wider range of experimental environments. Aptamers can be transported globally without need for special storage conditions.

(iii) Minimal immunogenicity and lower toxicity: Aptamers usually possess a much lower molecular weight when compared to antibodies and are also less toxic. The important aspect is that nucleic acid aptamers cannot be directly recognised by the immune system making them suitable in the development of anti-cancer drugs.[60]

(iv) Wider variety of targets: As substitutes for antibodies, aptamers can bind to targets that cannot be recognised by antibodies, including toxins and those molecules that do not elicit strong immune responses, or ligands such as ions and small molecules.[72–75] Therefore, employing aptamers as recognition units could significantly broaden the applications of biosensors.[72]

However, despite these advantages, and recent progress in chemical modification, aptamers have some fundamental limitations. Aptamers are essentially comprised of DNA or RNA strands and their libraries thus lack true diversity. For instance, aptamers are polyanionic, largely confined to A/T/G/C nucleotides (or their close analogues), and some motifs (e.g. G-quadruplex) are overrepresented.[57,58] Currently, the affinities of aptamers for their respective substrates are, in most cases, lower than for antibodies.[49,56,68]

### 1.5.2 SELEX (**S**ystematic **E**volution **of L**igands **by Ex**ponential enrichment**)**

Aptamers with an affinity for the desired target are selected from a large oligonucleotide library *in vitro* through a process known as the Sequential Evolution of Ligands by Exponential Enrichment, or SELEX. The process begins with the synthesis of a large single-stranded oligonucleotide library consisting of fully or partially randomised oligonucleotide sequences of fixed length flanked by defined regions which serve as primer binding sites, enabling the PCR amplification of the randomised sequences.[51,53,76] After the single-stranded oligonucleotide library is heated and cooled slowly to renature oligonucleotides with stable secondary and tertiary structures, the randomised library is then incubated with the immobilised target ligand to allow oligonucleotide-target binding.[76,77] Unbound oligonucleotides will be washed from the immobilised target and the remaining oligonucleotides with good binding affinities are retained by the target ligand. These selected oligonucleotide sequences are then exposed to denaturing conditions, re-amplified by PCR and converted into single-stranded oligonucleotides for the next round of selection processes.[61,78,79] In classical SELEX, multiple rounds of SELEX (from 5 to 16 cycles) are typically performed with increasing stringency of target binding to identify aptamers with the optimal binding affinity for the target.[53,77]

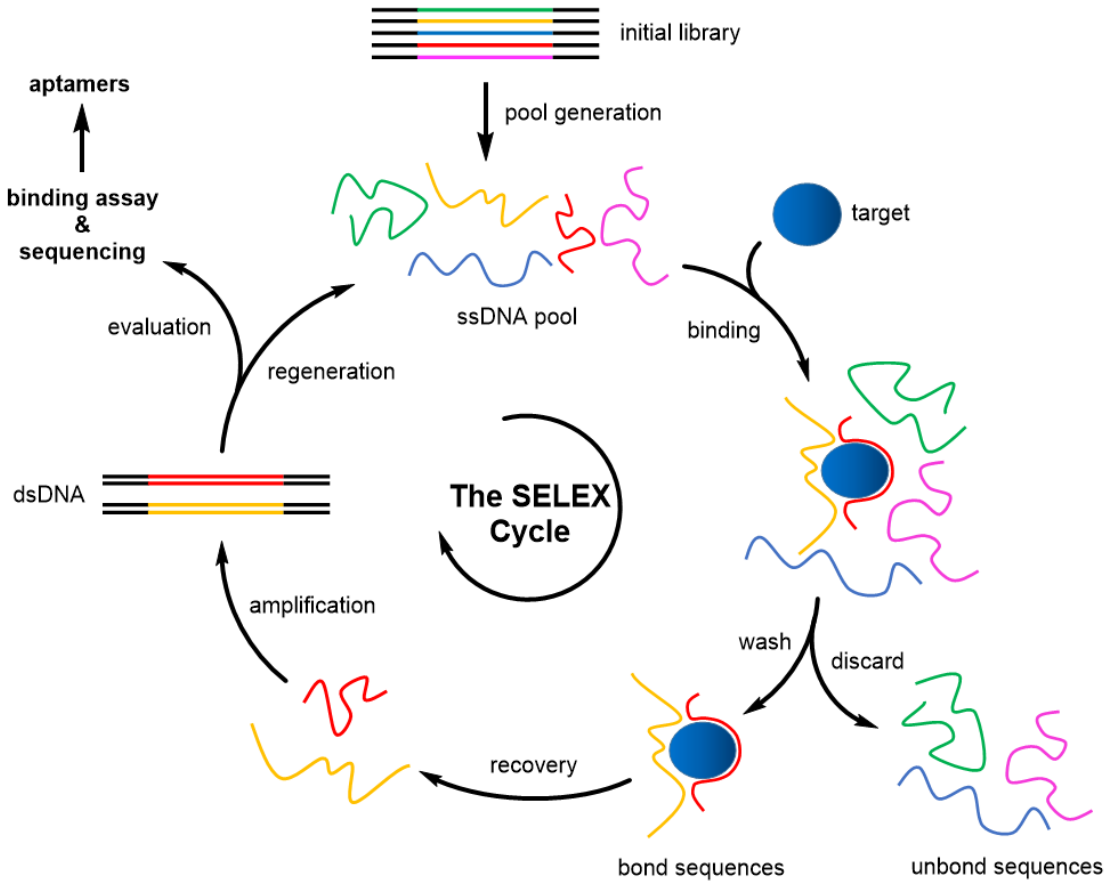



**Figure 1.17** Aptamers are self-encoding and can be identified using an *in vitro* selection method known as SELEX (**S**ystematic **E**volution of **L**igands by **Ex**ponential enrichment).

## 1.6 DNA sequencing

Oligonucleotides are selected and enriched through the SELEX process. Aptamer candidates, after several rounds of selection, are then identified by the sequencing of the DNA library obtained. DNA sequencing is a process to determine the order of oligonucleotide bases in a strand of DNA. This technique reveals the information encoded at the most fundamental level of a gene or genome. In 1977, Frederick Sanger and colleagues developed the dideoxy method of DNA sequencing which earned him his second Nobel prize.[80,81] Sanger's dideoxy DNA sequencing method was the first practical method of laboratory-based DNA sequencing. Nowadays, as highly efficient sequencing technologies are developing rapidly, DNA sequencing has since become indispensable in many areas of biological research.

### 1.6.1 Sanger sequencing

Sanger sequencing is the process of the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during *in vitro* DNA replication. The dideoxy method requires a ssDNA template to be sequenced, a DNA primer that is labelled at the 5′-end with $^{32}P$ phosphate, a DNA polymerase, normal deoxynucleoside triphosphates (dNTPs), and dideoxynucleoside triphosphates (ddNTPs). The DNA sample is divided into four individual sequencing reactions and each reaction contains all four of the normal dNTPs (dATP, dGTP, dCTP, and dTTP) and only one of the four ddNTPs (ddATP, ddGTP, ddCTP, or ddTTP). As DNA polymerase does not

discriminate fully between dNTPs and ddNTPs, either can be added at each step of the oligonucleotide chain termination process. If a dNTP is added then DNA chain extension continues. However, if a ddNTP is incorporated, the DNA synthesis reaction is terminated due to the lack of a 3′-hydroxyl group in the DNA chain. Each of the four separate extension reactions is a mixture of oligonucleotides of different lengths which are all terminated by a specific ddNTP. All four sequencing reactions are collected and separated (analysed) by gel electrophoresis based on their respective sizes. These $^{32}$P-labelled fragments produce an image of the reverse-complement sequence of the DNA template. This technique is limited to the sequencing of around 300 bases of DNA due to relatively low throughput and limited resolution. The modern equivalent of Sanger sequencing pioneered by Applied Biosystems, involves incorporating a different fluorescent dye into each of the four ddNTPs (on the nucleobases) and using a laser to excite the dyes attached to the DNA fragments as they elute through the gel. In this way only one sequencing reactions is needed and only one gel is required for the analysis.

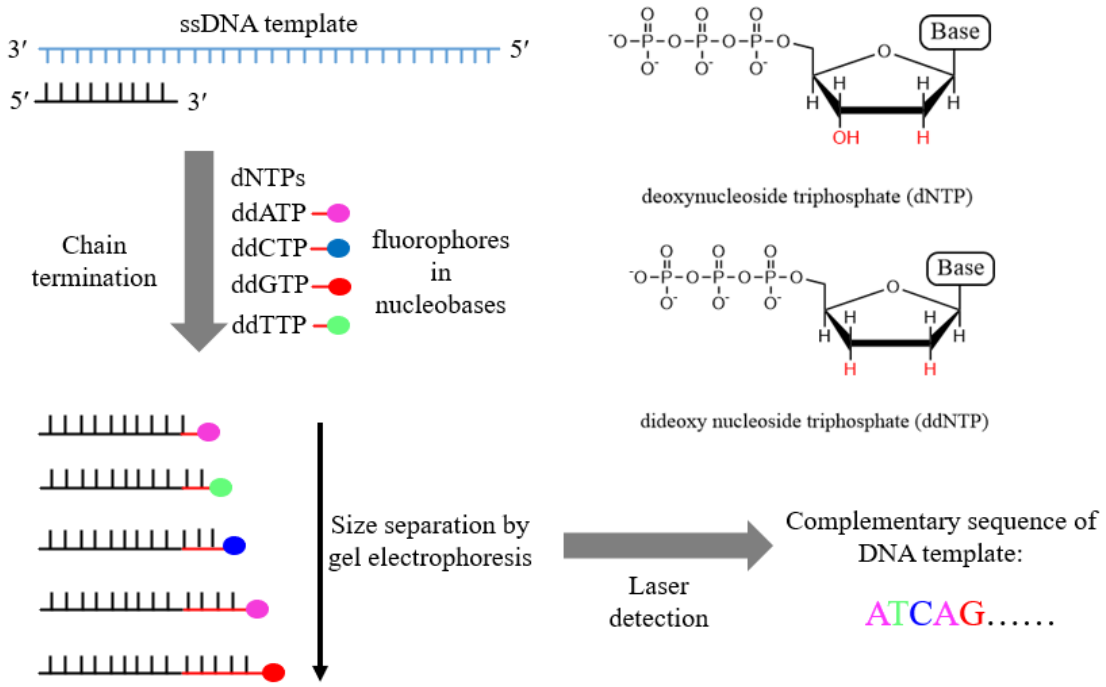



**Figure 1.18** Sanger's dideoxy sequencing method generally follows three steps: (1). Chain extension and termination reactions using dNTPs and ddNTPs; (2). Size separation of oligonucleotide fragments using gel electrophoresis; (3). Sequence identification by laser excitation and computer analysis.

## 1.6.2 Next-generation sequencing

Next-generation sequencing (NGS), also known as high throughput sequencing, is a large scale parallel sequencing technology that allows up to 100 million sequences to be determined in a single experiment. The Human Genome Project, which employed fluorescent Sanger sequencing to achieve the sequencing of 3.2 billion base pairs of the human genome took some 13 years (1990 to 2003) to complete and cost an estimated $3bn.[82] By using NGS, it is now possible to sequence the entire human genome in ten days at a cost of less than $10m.[83] NGS has been widely used in PCR product

sequencing, biomarker discovery, and notably, in aptamer selection.[84–86] The ultra-high throughput of NGS allows candidate sequences from all rounds of SELEX cycling to be analysed. A list of the most frequently found aptamers in each round of selection can thus be used to determine the best candidates. NGS allows accurate tracking of the distribution of nucleobases at all positions in the random region of the library. The critical positions of specific nucleobases that are consistent with high-binding aptamers are therefore identified.

The most popular NGS platform was developed by Illumina.[87] The concept behind NGS is the incorporation of fluorescently labelled dNTPs into a DNA template strand using DNA polymerase during sequential cycles of DNA synthesis. It is closely related to the fluorescent Sanger sequencing in Figure 1.18, but the growing DNA strands are immobilised on a surface. The Illumina NGS process allows sequencing of millions of DNA fragments in a large scale parallel fashion.[88,89] The method follows four basic steps:

1. Library preparation: The sequencing library is prepared by the random fragmentation of a genomic DNA sample followed by ligation with 5′ and 3′ adaptors.
2. Cluster generation: The library is then loaded into a flow cell and each sequencing fragment is hybridised to surface-bound oligonucleotides which are complementary to the ligation adaptors. Each fragment is amplified into distinct clusters through bridge amplification.
3. Sequencing: Four reversible terminator-bound dNTPs are present during each

sequencing cycle, each of which is labelled on the nucleobase with a different fluorescent indicator. These four nucleotides are incorporated into the DNA template as dictated by natural competition with unmodified dNTPs. The image of the flow cell after each nucleotide addition is recorded during these incorporations.

4. Alignment and data analysis: The newly identified sequence reads are aligned to a reference genome using bioinformatics software. Many variations of data analysis are available post alignment.

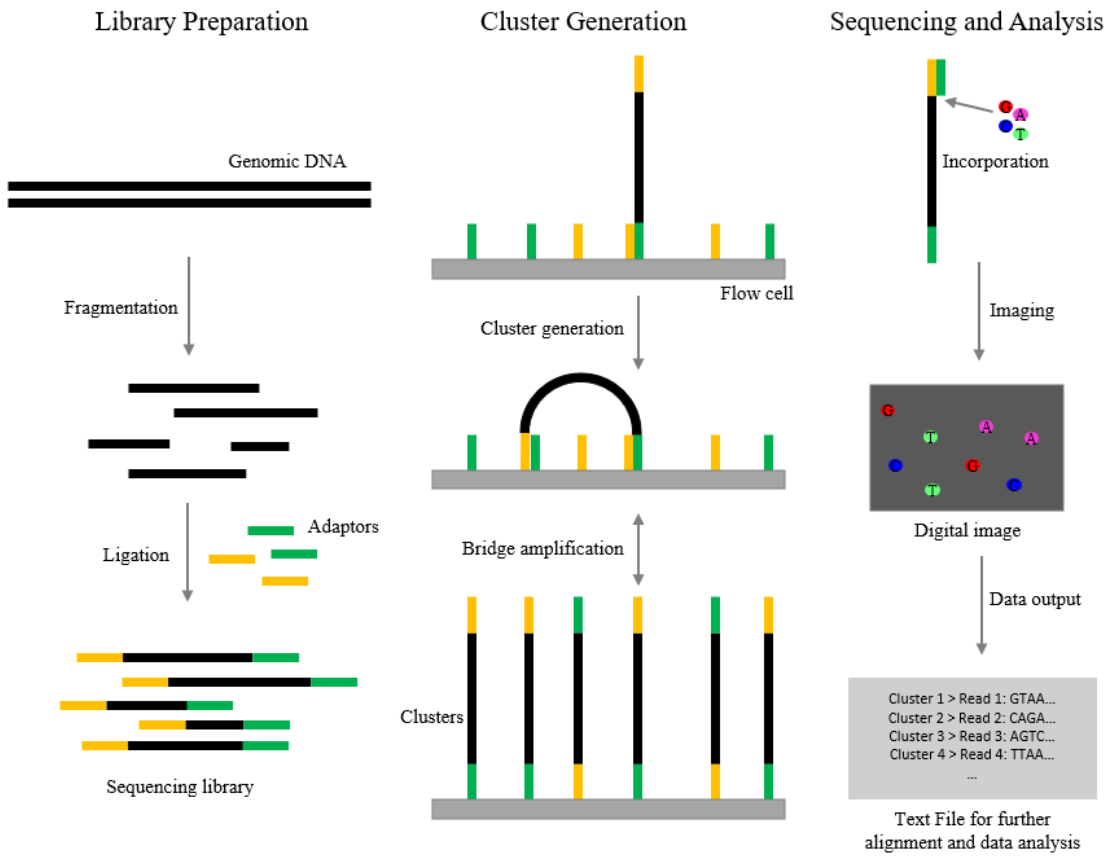



**Figure 1.19** Overview of Illumina next-generation sequencing (NGS). Illumina NGS generation follows four steps: 1. Library preparation; 2. Cluster amplification; 3. Sequencing; 4. Alignment and data analysis.

## 1.7 Research aims and thesis overview

In this project my objective was to increase the functional diversity of modified nucleic acids (e.g., aptamers) and thereby to improve the affinities of such modified nucleic acids for their targets. This was to be achieved by introducing a series of "non-natural" functional groups such as nitro groups, halogens, poly-substituted aromatic, and hydrophobic groups into DNA nucleobases using enzymatic (polymerase-based) approaches and DNA libraries. Through the incorporation of these functional groups, a wider range of supramolecular interactions should be attainable.

This thesis consists of seven chapters describing several interrelated projects. A series of derivatised 5-aminomethyl deoxyuridine (5-AM-dU) modified nucleic acids which contain "non-natural" functional groups attached to the amino group of the nucleobase were successfully designed and synthesised. The required modified monomers (dNTPs) were subsequently tested for their incorporation into DNA through linear primer-extension (PEX) and PCR amplification. Success was achieved in the incorporation of modified dNTPs that contain large hydrophobic functional groups. A fluorescent quadracyclic adenine analogue, qAN1, was also investigated and successfully used in PCR amplification in combination selected with 5-AM-dU analogues. The combination of qAN1 and modified 5-AM-dU was essential for efficient PCR. These modified nucleic acids were then used in SELEX approaches to select a new series of aptamers with good target binding affinities. In addition, the "non-natural" functional groups were extended to aminoethylglycine (AEG)-based ligands using an SABL library

approach. Both SELEX and SABL offer advantages and disadvantages in terms of target selection. Hence, by comparing the two approaches, a more comprehensive structural analysis of target binding through these "non-natural" functional groups can be established.

**Chapter 1** is an introduction to nucleic acid chemistry and associated techniques of relevance to this project.

**Chapter 2** describes the synthesis and characterisation of nucleic acids containing six dU derivatives (amU1 - amU6). These modified nucleic acids are compatible with DNA polymerases and have since been successfully incorporated into DNA, along with natural dNTPs, through the PEX and PCR.

**Chapter 3** reports an application of these modified nucleic acids. The PCR efficiency of incorporation of a strongly duplex-stabilising quadracyclic fluorescent adenine molecule, qAN1, was greatly improved by simultaneously introducing 5-AM-dU analogues with side chain functional groups that destabilise the DNA duplex. These results strongly support the hypothesis that the stabilising effect of qAN1 reduces PCR efficiency and that this problem can be solved by introducing a complementary nucleobase with a duplex destabilising effect to form DNA duplexes with moderate melting temperatures. Further, synthesis of a chemically modified DNA amplicon with full substitution of qAN1 for dA and amU2/amU3 for dT was achieved through PCR amplification.

**Chapter 4** introduces the chemically modified 5-AM-dU analogues into SELEX. Human Her2 protein was chosen as the target for aptamer selection. Four ssDNA libraries containing modified 5-AM-dU analogues were designed and generated. These libraries of oligonucleotide sequences were selected using Her2 protein targets by passing through a nitrocellulose membrane in 10 rounds of SELEX. The selections were monitored by qPCR. SELEX libraries were submitted for Illumina NGS sequencing and analysed to generate potential aptamers for target validation.

**Chapter 5** describes a new concept (SABL libraries) which are constructed by combining the concepts of DEL chemical diversity and aptamer-type biomolecular interactions and multivalency. Several functional groups, described in Chapters 2 to 4, were introduced into modified nucleotides and used in target binding and selection. In total, 12 different ligand building blocks for SABL libraries were designed and successfully synthesised with "non-natural" functional groups.

**Chapter 6** summarises the overall conclusions of the thesis and the achievements of the various projects. Suggestions are given for future avenues of research.

**Chapter 7** describes the experimental procedures and techniques used in this research.

# *Chapter 2.*

# *Synthesis and Analysis of 5-aminomethyl-dU Analogues*

# 2.1 Introduction and aims

## 2.1.1. Modification of nucleic acids in SELEX

When compared to antibodies (protein-based ligands), aptamers composed of naturally occurring nucleic acids are comprised of fewer building blocks (4 bases vs 20 amino acids), possess a polyanionic backbone with hydrophilic character, and have a relatively smaller repertoire of functional groups available for target recognition. There is thus a need to increase the functional diversity of nucleic acids used in the aptamer selection process. Several approaches towards increasing the chemical diversity of SELEX libraries have involved the introduction of additional base pairs. For instance, by expanding the four-letter genetic alphabet to six, the information density of such libraries is higher.[90] Additionally, aptamer avidity can be further increased via chemical modification of the nucleobases, introducing unnatural functional groups such as hydrophobic bases.[91]

Several modifications have been used on the phosphate backbone, sugar moiety, or nucleobases to generate nucleic acid-based ligands with improved binding affinity and metabolic stability (**Figure 2.1**).[92] Early attempts to modify aptamers focused on replacing the hydroxyl group of RNA with either 2′-fluoro or 2′-amino groups. The presence of a hydroxyl group at the 2′-position on the ribose sugar makes the adjacent phosphodiester bond vulnerable to enzymatic hydrolysis. This makes RNA extremely sensitive to nuclease degradation as compared to DNA. A replacement of the 2′-hydroxyl group with 2′-fluoro, 2′-amino or 2′-O-methyl groups led to analogues with

dramatically enhanced stability to RNase enzymes.[93–95] Protection from nuclease digestion can also be achieved by modifying the phosphate backbone e.g. to produce thioaptamers – wherein the non-binding oxygen in the phosphodiester linkage[96,97] is replaced by sulphur. In this context xenonucleic acids[98,99] have also been extensively studied. These are non-naturally occurring nucleic acids such as locked nucleic acids (LNAs)[100] and L-enantiomeric RNA (L-RNA)[101].

In parallel with phosphate backbone and sugar modification, chemical modification of the nucleobases has also expanded the available repertoire of nucleic acids. The introduction of non-native functional groups at positions orientated away from the hydrogen bonding face of the nucleobases, i.e. the 7- or 8-position of purines[102] or the 5-position of pyrimidines[103] is particularly useful. Modifications at these sites only minimally affect hydrogen bonding between nucleobases and thus maintain the secondary structures of oligonucleotides. Most current work in this area that is compatible with SELEX includes modified deoxyuridine (dU) libraries.[104–115] These modifications are introduced into oligonucleotides through modified phosphoramidites (for chemical solid-phase synthesis) and via nucleoside triphosphates (for enzymatic incorporation). Initial work with SELEX focused on using DNA libraries including a hydrophobic pentynyl group at the 5-position of deoxyuridine with thrombin as the target.[109] The resultant binding affinities were further improved by parallel studies in which the pentynyl group was substituted with positively charged primary amines with flexible linkers.[97,110,113] In particular, moieties with large hydrophobic and aromatic functional groups (*e.g.,* benzyl, naphthyl, tryptamino, isobutyl, *etc.*) that resemble the equivalent side chains in proteins have proven to be very efficient with a wide range of

protein targets.[114,116]

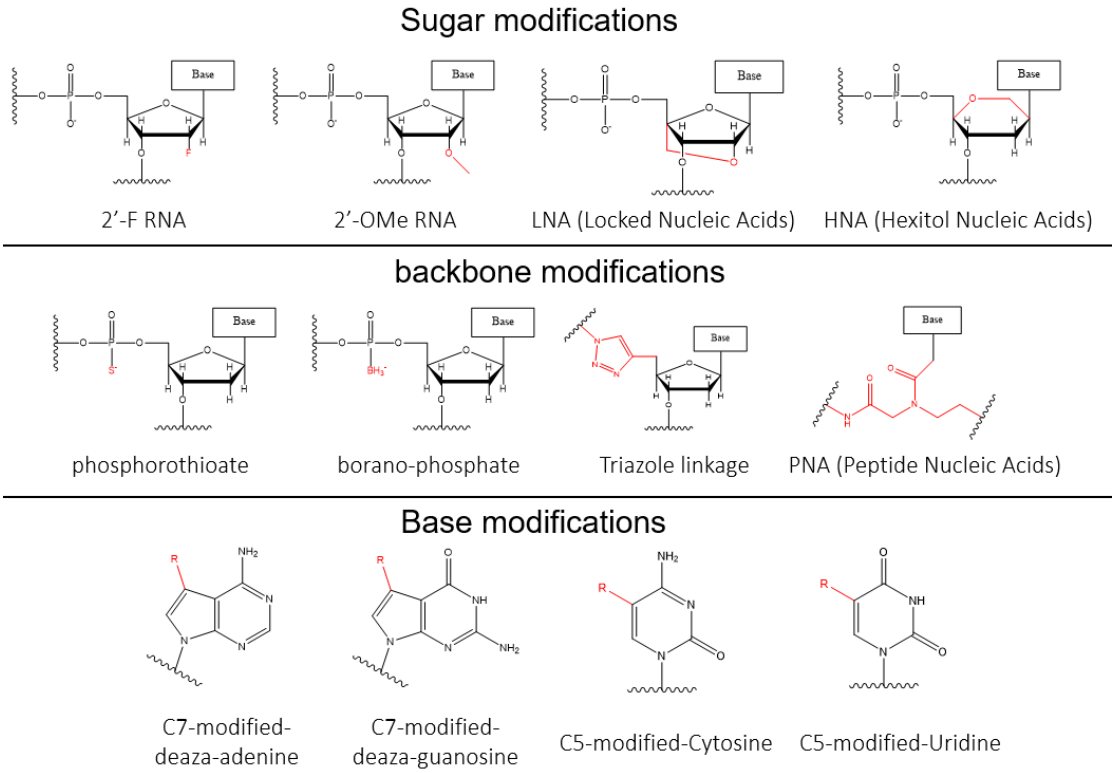



**Figure 2.1.** A schematic illustration of recent studies on attempting various of chemical modifications to nucleic acid structures.

The selection of these modified deoxyuridines through SELEX obviously requires these nucleotide analogues to be compatible with DNA polymerases due to the requirement for efficient and high-fidelity amplification by PCR. This can be precisely evaluated by primer extension as well as in specifically designed PCR assays. Post-SELEX chemistry can be used to alter selected aptamers with modifications that are not recognised during the *in vitro* selection process. Here a polymerase-compatible nucleotide triphosphate with a small reactive handle is incorporated during the SELEX process (e.g., click-SELEX). However, as the final modified forms of aptamers do not directly participate in the selection process, the identification of successful structures can be analytically challenging.[117]

### 2.1.2 Slow off-rate Modified Aptamers

**S**low **O**ff-rate **M**odified **A**ptamers, or SOMAmers, and the corresponding SomaScan assay were developed by *SomaLogic Inc* in 2010 as a 'next-generation' aptamer platform.[114] Like conventional aptamers, SOMAmers are short, single-stranded deoxyribooligonucleotides that are selected *in vitro* from large random libraries by their binding affinities for discrete molecular targets. SOMAmers are endowed with protein-like properties by adding functional groups with amino acid mimetic side chains.[117] These protein-like moieties increase the affinity and specificity of SOMAmers in binding to protein targets.[94,114] Another advantage of SOMAmers is that reagents can be selected with slow dissociation rates (*e.g.,* with $t_{1/2}$ > 30 min, hence their nomenclature) by introducing this second, kinetics-based element of specificity into the selection process. This allows a selective disruption of non-specific binding interactions by using a large excess of a polyanionic competitor.[118]

The structures of SOMAmers are engineered with dU (deoxyuridine) residues functionalised at the 5-position with different protein-like moieties (*e.g.,* benzyl, 2-napthyl, isobutyl, *etc.*) (**Figure 2.2**). SOMAmers with these modifications can interact with target molecules to form novel secondary and tertiary structures.[119] The affinities of SOMAmers for molecular targets are comparable to those of antibodies and hence they can be regarded as synthetic protein-binding reagents. The nuclease resistance of SOMAmers is also greatly improved relative to unmodified aptamers due to the hydrophobic modifications at the 5-position of dU residues.[120] The synthetic routes to

dU nucleotides modified with diverse functional groups at the 5-position were developed efficiently through a single step palladium ($Pd[PPh_3]_4$, 10 mol%)-catalysed carboxyamidation reaction on 5-iodo-2′-deoxyuridine under mild conditions.[121]

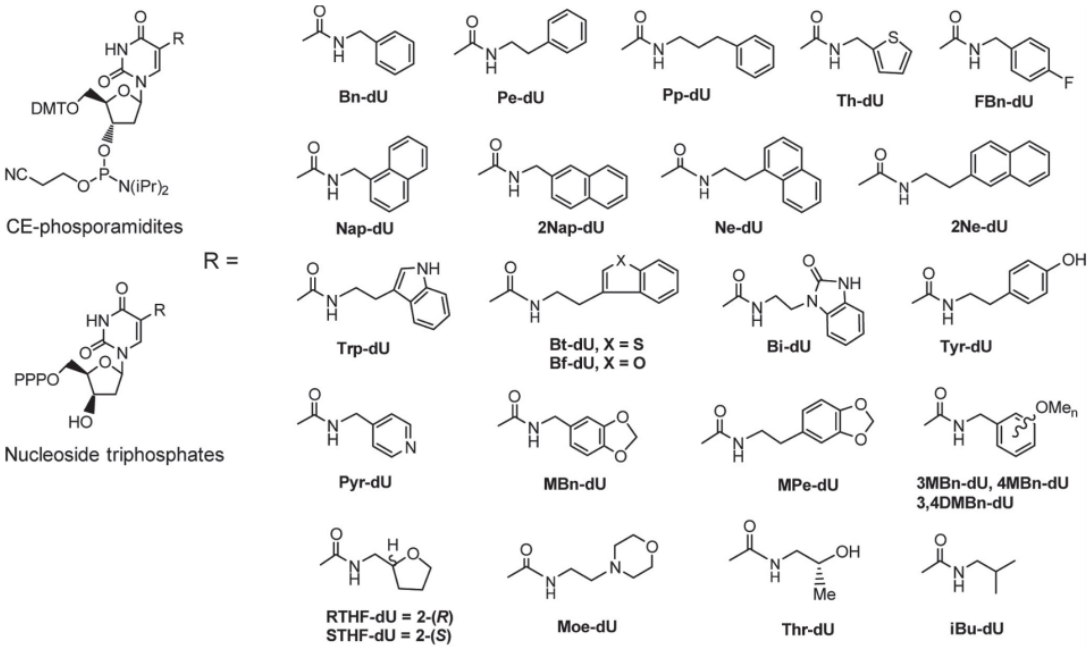



**Figure 2.2.** A partial listing of modifications at the 5-position of 2'-deoxyuridine that are available for SELEX and post-SELEX optimisation. The figure was adapted with permission from reference[122].

### 2.1.3 Structural designs of 5-AM-dU analogues

Based on the abundance of studies of modifications at the 5-position of deoxyuridine and the resulting success of SOMAmers, 5-aminomethyl-deoxyuridine (5-AM-dU) was selected as the backbone for a series of ligands using the SELEX approach. The modified 5-AM-dU structure, with a free amine group at the 5-position, will enable further chemical attachment of a series of moieties.[123] For this project, carboxylic acid

derivatives of various functional groups were attached to the 5-AM-dU structure through amide coupling reactions. When compared to SOMAmers, the carbonyl group conjugated to the pyrimidine ring (which commonly functions as a hydrogen bond acceptor) is separated from direct conjugation to the nucleobase in the 5-AM-dU analogues. Theoretically, this should provide the functionalities coupled to 5-AM-dU with greater conformational flexibility than in SOMAmers, and enable different three-dimensional oligonucleotide structures to form.

As a proof-of-principle, 5-AM-dU, together with five derivatives (amU1 – amU6), were designed and synthesised (**Figure 2.3**). The functional groups we elected to attach to 5-AM-dU via amide coupling are predominately hydrophobic or aromatic and are absent in unmodified aptamers.[120] Analogues with adamantanyl and nonanoyl groups are strongly hydrophobic, possessing either a bulky or a long linear alkyl chain, while analogues with substituted aromatic groups are expected to form extensive networks of π-π interactionsin oligonucleotide structures with associated stacking effects.[124,125] These 5-AM-dU analogues were designed to be introduced into oligonucleotides through modified phosphoramidites and via nucleoside triphosphates. The PCR-based method is essential for SELEX and the solid-phase phosphoramidite approach is required for the large-scale synthesis of the candidate aptamers identified by SELEX. The enzymatic incorporation of these analogues was investigated with several DNA polymerases using defined templates.

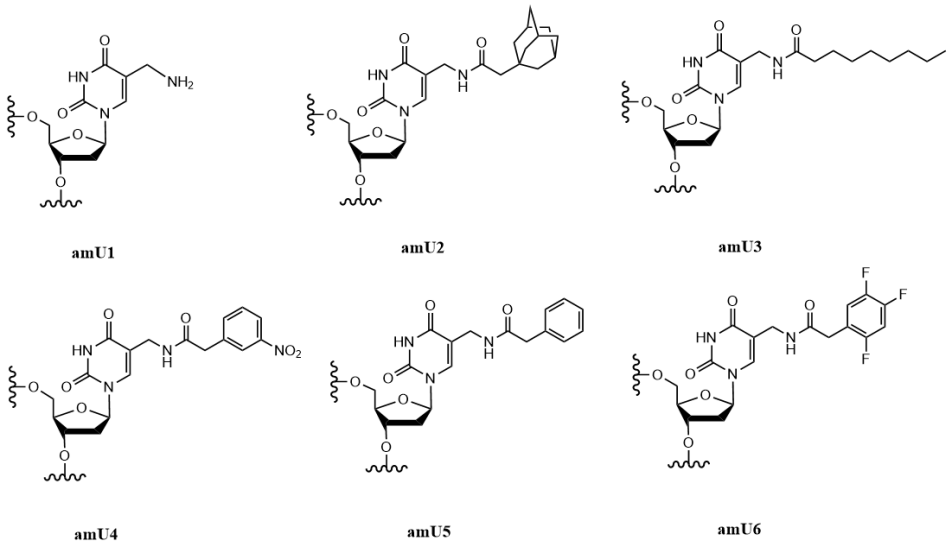



**Figure 2.3.** Structure of modified 5-AM-dU (amU1) with five analogues (amU2 – amU6) containing either hydrophobic or substituted aromatic functional groups at the 5-position of the pyrimidine ring.

# 2.2 Chemical synthesis of 5-aminomethyl-dU triphosphates and phosphoramidites

## 2.2.1 Synthesis of 5-aminomethyl-dU phosphoramidite monomers

The phosphoramidite method of solid-phase oligonucleotide synthesis requires each nucleoside monomer to have a protecting 5′-DMTr group. 5′-DMTr protected 5-aminomethyl-deoxyuridine (5-AM-dU) was synthesised by a 6-step process with an overall yield of 32% (**Figure 2.4**) starting with thymidine (T). The 3′ and 5′ hydroxyl groups were first protected using acetyl groups[126] prior to reacting the 5-position of the nucleobase. A radical bromination reaction at the 5-position was performed using Wohl-Ziegler reaction conditions in which azabisisobutyronitrile (AIBN) —the radical initiator—and N-bromosuccinimide (NBS) supply the bromine radicals.[127] The reaction is commonly performed using anhydrous $CCl_4$ as a solvent and often results in high yields.[128,129] However, the use of $CCl_4$ is restricted by its high toxicity, possible carcinogenicity, and ozone layer depleting effects.[130] Thus, dry acetonitrile was chosen as a suitable alternative to anhydrous $CCl_4$ after comparison with test reactions run in dry $CH_2Cl_2$ and $CHCl_3$.

The bromo group was then substituted by an azido group by reaction with sodium azide in DMF. The product was purified by silica column chromatography and then dissolved in 7N ammonia solution in methanol under pressure to deprotect the 3′ and 5′-OH groups. Selective protection of the 5′-OH was achieved using DMTr-Cl in the presence of base, which was followed by reduction of the azido group by Pd/C catalysed

hydrogenation.

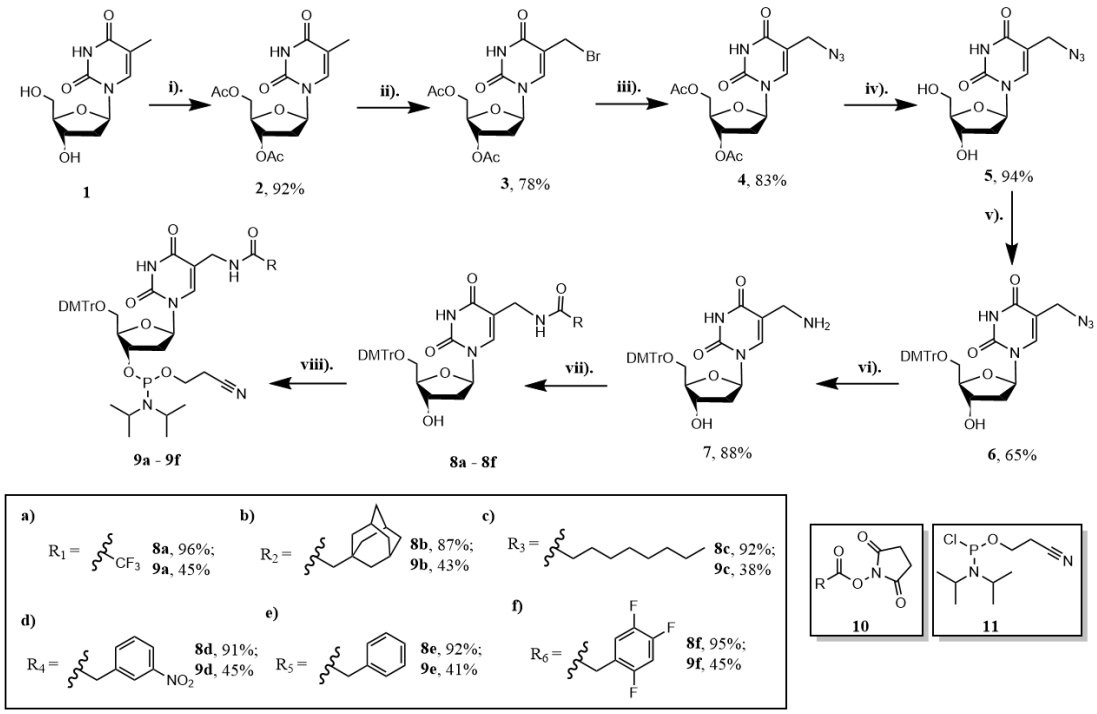



**Figure 2.4.** Synthesis of 5-AM-dU phosphoramidite monomers.

Reagents and conditions: i). $Ac_2O$, Pyridine, RT, 92%; ii). NBS, $CCl_4$, AIBN, Reflux, 78%; iii). $NaN_3$, DMF, 83%; iv). $NH_3$, methanol, 94%; v). DMTr-Cl, pyridine, 65%; vi). 10% Pd/C, MeOH, hydrogen atmosphere, RT, 88%; vii). NHS ester (compound 10), DIPEA, DMF; viii). anhydrous DCM, anhydrous DMF, DIPEA, 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (compound 11), RT, 2 h.

Five different protected 5-AM-dU analogues were synthesised through an amide coupling reaction between 5-AM-dU and a compound possessing either a hydrophobic or a poly-substituted aromatic side chain (carboxylic acid derivative).

Coupling reactions of the side chain can also be achieved after incorporation of the AM-dU phosphoramidite or triphosphate monomers into DNA. Protection of the amine with

a $COCF_3$ group in phosphoramidite monomer **2-9a** (**Figure 2.4**) is necessary for phosphoramidite solid-phase oligonucleotide synthesis. Removal of the protecting $COCF_3$ protecting group is quantitative by treatment of the oligonucleotide with concentrated aqueous ammonium hydroxide at 55 ºC for 5 hours during the final deprotection step. Phosphoramidites of five different 5-aminomethyl-dU analogues (**2-9 b - f**) were also synthesised via coupling of **2-7** with $R_2$ – $R_6$ side chains as NHS esters, respectively, under basic conditions in DMF. All of these analogues were converted to phosphoramidites and were stable during solid-phase oligonucleotide synthesis, requiring no further protection of their respective side chains.

It is important to note that synthesis of phosphoramidite monomers such as **2-9 a - f** requires an oxygen and water-free environment and relatively short reaction times to avoid formation of by-products.[131–133] To achieve this, the phosphitylating agent 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite was added dropwise to 5-AM-dU in anhydrous DCM and the resulting reaction was allowed to continue for no more than 2 hours. This reaction was then followed by TLC and the product was purified by silica column chromatography under basic, inert conditions. The purified phosphoramidite monomers were characterised by TLC, LCMS and NMR ($^1$H, $^{13}$C and $^{31}$P NMR) and were stored under argon at -20 °C.

## 2.2.2 Synthesis of 5-AM-dU triphosphates

Current methods for the chemical synthesis of 2′-deoxynucleoside triphosphates

generally follow the "one-pot, three steps" strategy developed by Ludwig as a modification of the Yoshikawa phosphorylation method.[134] This involves (**Figure 2.5**) the (*i*) generation of an activated nucleoside monophosphate with a phosphorylating agent (usually phosphorus oxychloride ($POCl_3$) in trimethylphosphate solvent); (*ii*) the reaction with bis-(tri-*n*-butylammonium) pyrophosphate to form a cyclic triphosphate intermediate; (*iii*) hydrolysis of the cyclic triphosphate by treatment with 1 M triethylammonium bicarbonate (TEAB) buffer (pH = 7.5), or else addition of a solution of iodine in pyridine and $H_2O$ to oxidise the Pα (III) centre followed by a 10% aqueous solution of $NaS_2O_3$ to quench the excess $I_2$. Deprotection is then performed in concentrated ammonium hydroxide ($NH_4OH$) at 55 ºC for 5 hours prior to precipitation of the crude triphosphate.[135]

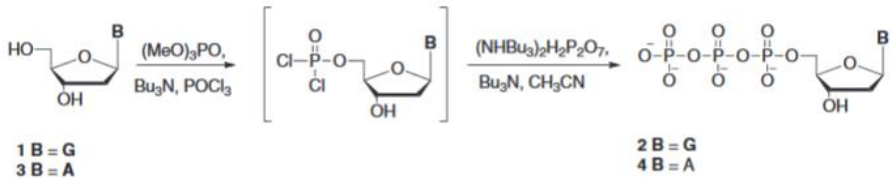



**Figure 2.5.** One-pot synthesis of deoxynucleoside triphosphates[134].

As an alternative, nucleobase-modified triphosphates can be synthesised using another phosphorylating reagent, 2-chloro-1,3,2-benzodioxaphosphorin-4-one (salicyl phosphorochloridite), which was developed as part of the Ludwig-Eckstein approach for the synthesis of various nucleobase-modified nucleoside triphosphates.[136] Reaction of the salicyl phosphorochloridite (**Figure 2.6**) with a nucleoside yields a more stable cyclic phosphite intermediate. This method for generating the monophosphate

derivative is a valid alternative to phosphorus oxychloride ($POCl_3$), as both purification (by re-distillation) and storage of fresh $POCl_3$ can be time-consuming and complex.

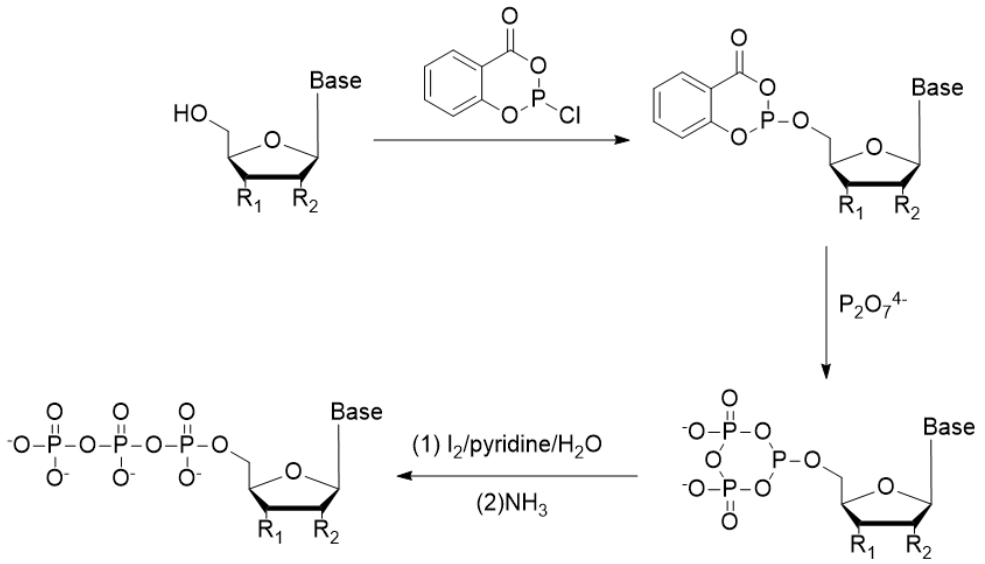



**Figure 2.6.** Another triphosphate synthesis method using 2-chloro-1,3,2-benzodioxaphosphorin-4-one.[136] Reagents and conditions: 1.1 equiv of 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one, DMF/pyridine/dioxane 1:1:4, rt, 10 min; 1.5 equiv of tributylammoniun pyrophosphate, tributylamine, DMF, rt, 10min; 0.02 M $I_2$, THF/pyridine/$H_2O$ 70:20:10, rt. 15 min; conc. $NH_4OH$, 1 h.

An alternative source of pyrophosphate is tris{bis(triphenylphosphoranylidene) ammonium} (PPN) chloride (**Figure 2.7**).[137] The PPN salt is prepared via the slow addition of an aqueous solution of PPN-Cl into a solution of tetrasodium pyrophosphate in distilled water at 50 ºC, and the pH is adjusted to 5.0 with 1 M hydrochloric acid. After cooling the resulting solution to room temperature, the PPN salt is collected and centrifuged at 4,000 rpm for 20 min.

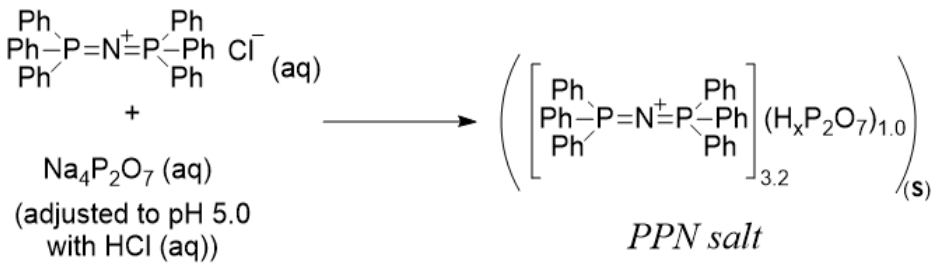



**Figure 2.7.** Crystalline PPN salt obtained by mixing tetrasodium pyrophosphate and PPN-Cl in aqueous solution.

I developed two synthetic routes for the large-scale synthesis of 5-AM-dUTP triphosphate analogues. Prior to triphosphate synthesis, the 5′-DMTr protected 5-AM-dU analogues first undergo an acid treatment to remove the DMTr group. The resulting product is then dried over potassium hydroxide (KOH) under vacuum overnight together with other phosphorylating agents which are then ready to use.

In one protocol (**Figure 2.8 (1)**), I dissolved the nucleoside in dry trimethyl phosphate at -15 ºC under argon (to minimise the production of 3' and 5' diphosphate by-products). Then fresh phosphorus oxychloride was added dropwise and the reaction was stirred at -15 ºC under argon for 1 hour. A solution of bis-(tri-*n*-butylammonium) pyrophosphate in dry DMF and *n*-tributylamine was subsequently added. The reaction was stirred for 3 hours allowing it to slowly reach room temperature.[138] 1.0 M triethylammonium bicarbonate (TEAB) buffer (pH = 7.5) was added to quench the reaction, followed by the addition of concentrated ammonium hydroxide.[139,140] The solution was stirred at 55 ºC for 5 hours and the solvent was removed *in vacuo*. The residue was diluted in 3 ml 0.3 M NaCl solution and precipitated by the addition of 9 ml ethanol. The product was

stored at -20 °C overnight to promote precipitation.[141] The resulting 5-AM-dUTP product was then collected by centrifugation and purified using anion-exchange HPLC and RP-HPLC with an overall yield of 16%.

In a different protocol (**Figure 2.8 (2)**), I used 2-chloro-1,3,2-benzodioxaphosphorin-4-one and PPN salt for phosphorylation. The nucleoside was then dissolved in a mixture of anhydrous pyridine and dioxane under argon. The ratio of pyridine and dioxane were set to around 1:3 for this specific experiment and may need to be varied depending on the solubility of the various nucleosides (pyridine is favoured for nucleosides whereas dioxane is preferred for the phosphorylating reagent). 2-Chloro-1,3,2-benzodioxaphosphorin-4-one in anhydrous dioxane was injected dropwise and the reaction was stirred at room temperature under argon for 1 hour.[142] The PPN salt in dry DMF and *n*-tributylamine were added and the reaction was stirred for a further 3 hours. A solution of 1% iodine in pyridine/water (10 mL, 98:2, v/v) was then added. After stirring for 20 min, the excess iodine was quenched by the continuous addition of a 5% aqueous $NaHSO_3$ solution (5 mL) until the yellow coloration disappeared. Concentrated ammonium hydroxide (10 mL) was added to the solvent and the solution was stirred at 55 ºC for 5 hours and the solvent was subsequently removed *in vacuo*. The residue was diluted in 3 ml of 0.3 M NaCl solution and precipitated by the addition of 9 ml of ethanol. The product was stored at -20 °C overnight to promote precipitation. The resulting 5-AM-dUTP was collected by centrifugation and purified using anion-exchange HPLC and RP-HPLC with an overall yield of 8%.

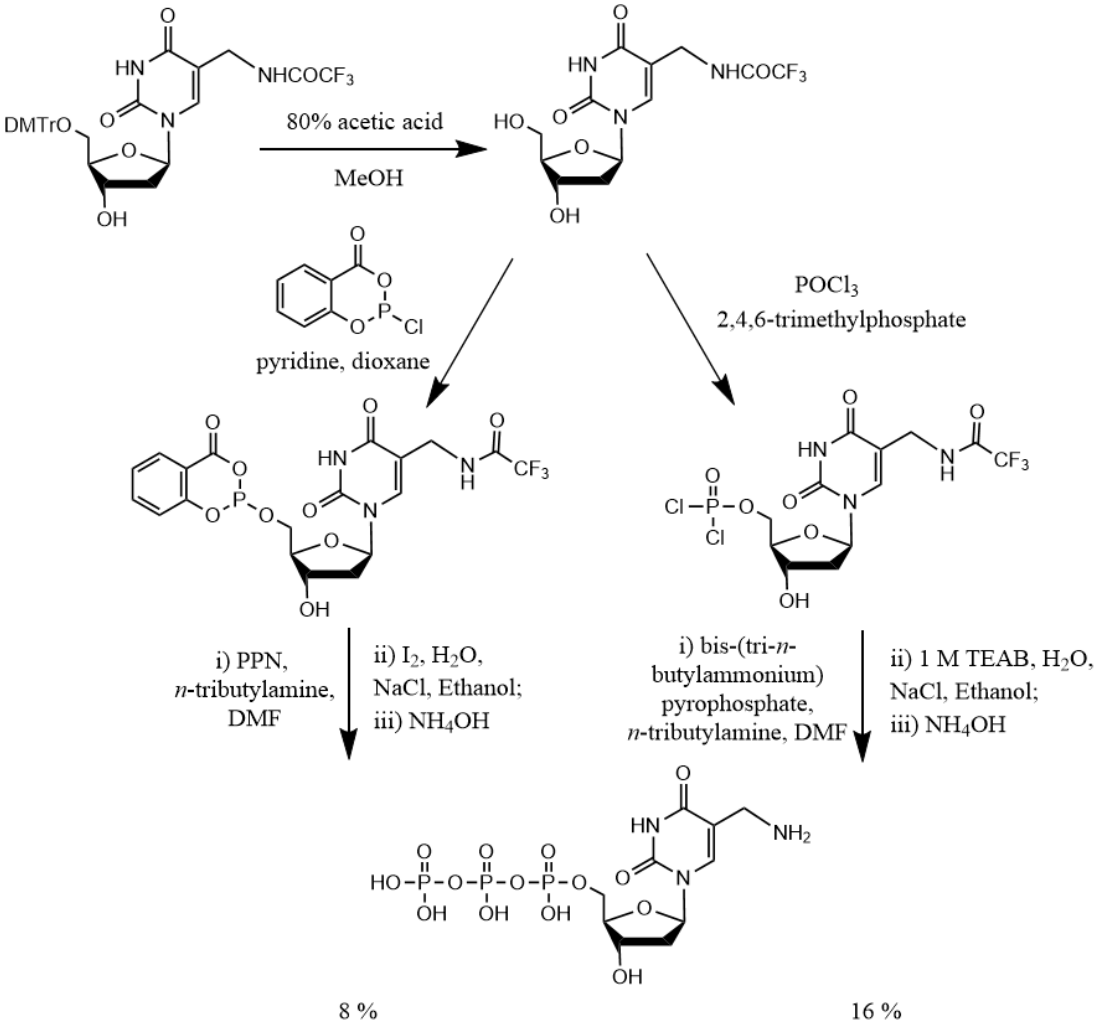



**Figure 2.8.** Two methods for triphosphate synthesis.

(1) 2-chloro-1,3,2-benzodioxaphosphorin-4-one, pyridine:dioxane = 1:3, DMF, 0 °C to RT; PPN salt, $Et_3N$, DMF; $I_2$; $NH_4OH$, 55 °C, 5 h.
(2) $(MeO)_3PO$, proton sponge, $POCl_3$, 2 h, °C; bis-(tri-*n*-butyl-ammonium) pyrophosphate, $Et_3N$, DMF; 1.0 M TEAB; $NH_4OH$, 55 °C, 5 h.

Both triphosphate synthesis protocols produced suitably modified 5-AM-dUTPs. However, using the PPN salt as the pyrophosphate reagent was found to create solubility issues, resulting in a lower yield. The excess of PPN salt required for the reaction also affected the isolation of the final product, as some salt co-precipitated with

the nucleotide upon ethanol addition. The strategy incorporating 2-chloro-1,3,2-benzodioxaphosphorin-4-one as the phosphorylating reagent for monophosphate synthesis was instead optimised. In relation to previously described methods, this offers a safer and more convenient alternative to phosphorus oxychloride for longer term, large-scale synthesis.

The final optimised triphosphate synthesis protocol for 5-AM-dU and its analogues was as follows: The 5-AM-dU nucleoside with its 5-position protected (100.0 mg, 280 μmol, 1.0 eq) was dried by coevaporation with pyridine overnight. The nucleoside was then suspended in dioxane (1.5 mL) and pyridine (0.5 mL) under argon. A solution of 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one (85.0 mg, 420 μmol, 1.5 eq) in dioxane (0.2 mL) was added to the mixture. The reaction mixture was stirred vigorously for 1 hr. A well-mixed solution of bis-(tri-*n*-butylammonium) pyrophosphate (900.0 mg, 1400 μmol, 5.0 eq) in *n*-tributylamine (400 μL, 1120 μmol, 4.0 eq) and 2 mL anhydrous DMF was added thereafter. After 3 hrs, a 2% $I_2$ solution in pyridine/$H_2O$ (98:2, v/v) (3 mL) was added. The solution was mixed periodically over 20 min, and excess iodine was quenched by the dropwise addition of 5 mL of 5% aqueous sodium bisulfite ($NaHSO_3$). The reaction was concentrated *in vacuo* to dryness. The resulting solid was then resuspended in $H_2O$ (5 mL) and ammonium hydroxide (10 mL, 30 %) was added. After stirring at room temperature for an additional 3 hrs at 55 °C with constant pressure, the solution was concentrated *in vacuo* to dryness. The product was purified first by anion-exchange HPLC then by RP-HPLC.

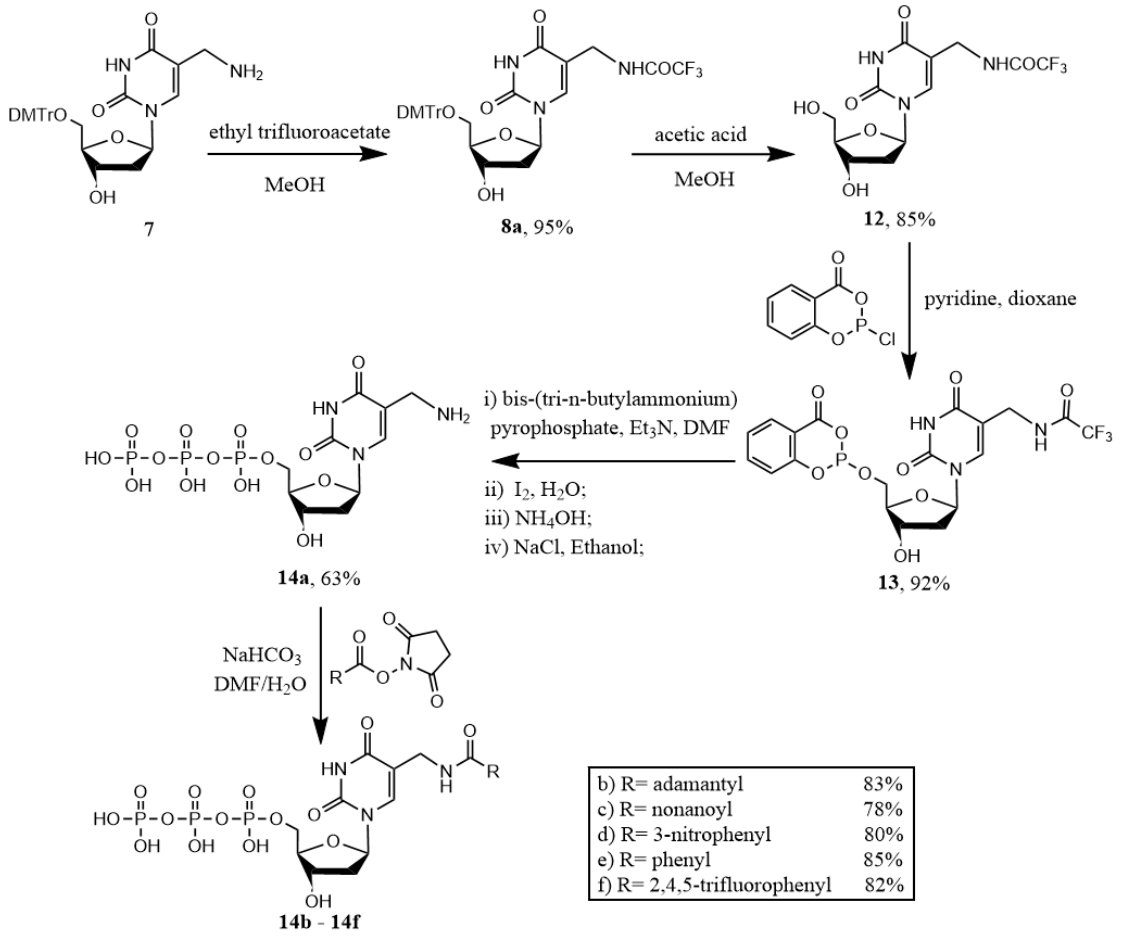



**Figure 2.9.** Final optimised triphosphate synthesis scheme for 5-AM-dUTP and its analogues.

The final optimised triphosphate synthesis scheme for 5-AM-dUTP and its analogues is shown in **Figure 2.9**. 5-aminomethyl-dUTP (**2-14a**) was obtained via a 4-step synthesis starting from 5′-DMTr protected 5-aminomethyl-dU (**2-7**). The free amine of compound **2-7** was first protected by reaction with ethyl trifluoroacetate in methanol. The 5′-DMTr protecting group was then removed under acidic conditions. Protocols using 2-chloro-1,3,2-benzodioxaphosphorin-4-one as the phosphorylating reagent to synthesise the monophosphate (**2-13**) were optimised as a safer and more convenient substitute for phosphorus oxychloride for longer term, large-scale synthesis.[134,136] This was followed by reaction with bis-(tri-*n*-butylammonium) pyrophosphate to produce

the cyclic triphosphate. Hydrolysis of the resulting cyclic triphosphate intermediate was achieved by the addition of a solution of iodine in pyridine and $H_2O$, followed by a 10% aqueous solution of $NaS_2O_3$ to quench any excess $I_2$.[135,137–142] A deprotection step was performed with concentrated ammonium hydroxide ($NH_4OH$) at 55 ºC for 5 hrs before precipitation of the crude trisphosphate.[135] The resulting 5-aminomethyl-dUTP (**2-14a**) was obtained as a white solid and collected by centrifugation. The product was purified by anion-exchange HPLC and RP-HPLC with an overall yield of 23% (**Figure 2.10**).

Compound **2-14a** could then be reacted with a suitable NHS ester (**Figure 2.9**) to provide five 5-aminomethyl-dUTP analogues (**2-14b – 2-14f**). The amide coupling reaction of 5-AM-dUTP with the NHS esters occurred in aqueous solution in the presence of sodium bicarbonate. The very low solubility of triphosphates in DMF solution makes coupling of triphosphates differ in detail from 5-AM-dU analogues made in the route to the phosphoramidites, which can utilise HATU or other coupling reagents such as DCC or EDC as well as NHS esters.[123,143–145]

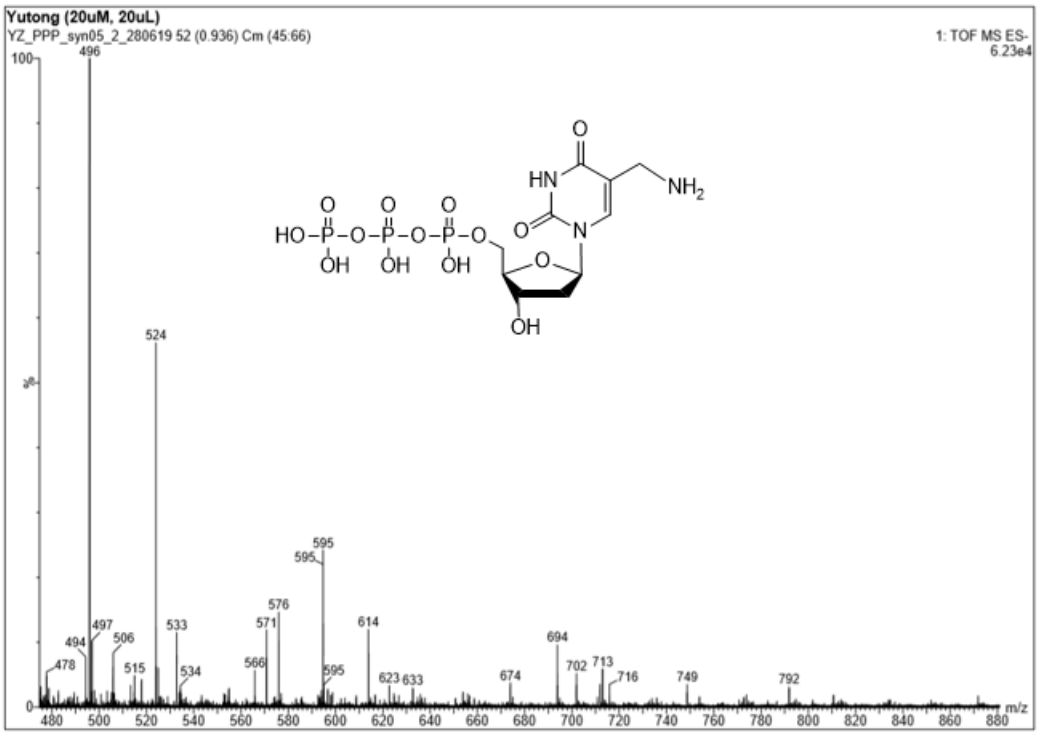



**Figure 2.10.** MS analysis of 5-AM-dUTP: mass calculated: 497, mass found: 496.

## 2.3 Primer extension (PEX) using modified 5-AM-dUTPs

Linear primer extension reactions were performed to test whether the synthesised 5-AM-dU analogues could be recognised by DNA polymerases and hence be incorporated into DNA by PCR. Primer extension is a method used to synthesise a specific DNA sequence by annealing an oligonucleotide primer to a suitable DNA template and adding a DNA polymerase and dNTPs to produce a cDNA copy. In our case we carried out the procedure in the presence of at least one modified dNTP.

The modifications at the C5 position of the pyrimidine base are not close to the H-bonding residues which are involved in Watson-Crick base pairing, and so the 5-AM-dU modified triphosphates are good analogues of dTTP. The groups at the 5-position of the nucleobases should protrude into the major groove of the DNA duplex without disrupting the normal B-conformation of the double helix.[146] In total, six modified 5-AM-dUTPs (amU1 – amU6 in triphosphates, **Figure 2.11**) with different functional groups present at the 5-position of uridine were synthesised. Unfunctionalized amU1 is a modified dTTP with an aminomethyl group on the 5-position of nucleobase, amU2 and amU3 are 5-AM-dUTP analogues bearing strongly hydrophobic adamantanyl and nonanoyl groups, respectively; while the other three triphosphates (amU4 to amU6) are 5-AM-dUTP analogues with substituted aromatic functional groups that are expected to form extensive networks of π-π interactions with associated stacking effects in structured oligonucleotides (aptamers).

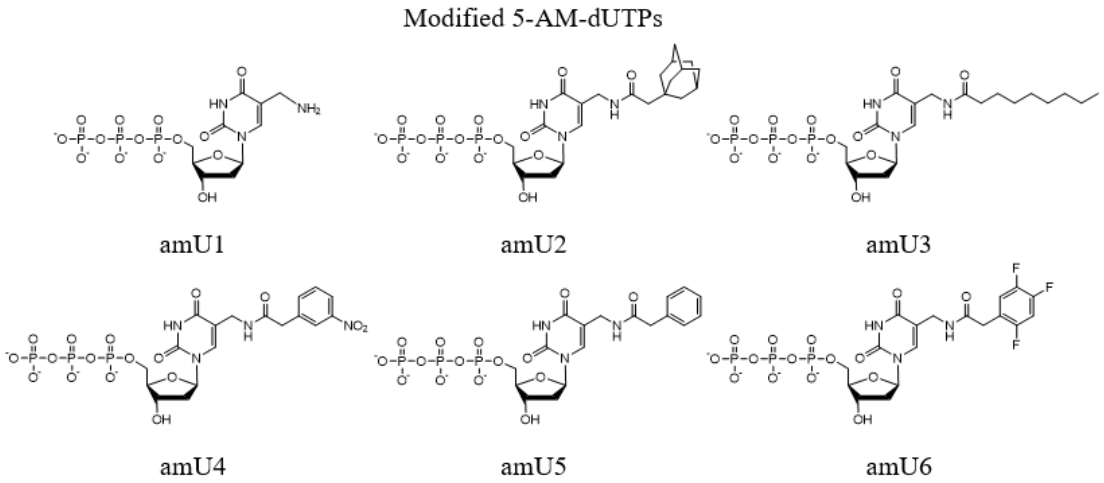



**Figure 2.11.** Six modified 5-AM-dUTP triphosphates synthesised for PEX and PCR tests.

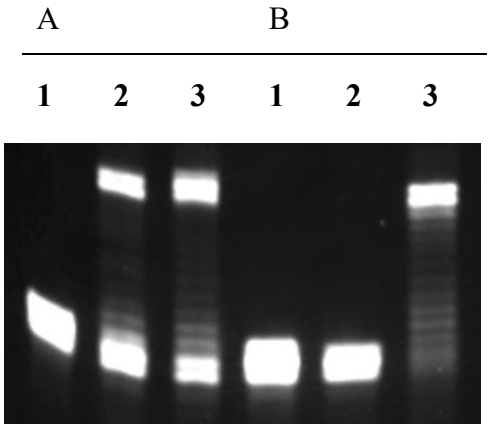
The sequences of primers and template used in the primer extension tests are shown in **Table 3.1**. A 6-carboxyfluorescein (FAM) fluorophore was attached at the 5'-end of the primer for visualisation during gel electrophoresis analysis. The 29-mer template (Res2465) for primer extension has three adenine sites (highlighted in **Table 2.1**) for incorporation of the modified 5-AM-dUTPs when an 18-mer FAM-labelled primer is attached. These three adenine sites are well separated, the extension region of the template starts with an adenine (A), and has all three adenine bases evenly distributed. The natural dNTPs (dATP, dCTP, dGTP and dTTP) and the modified 5-AM-dUTPs were used in conjunction in this primer extension test in the presence of various DNA polymerases (GoTaq and Vent exo-).

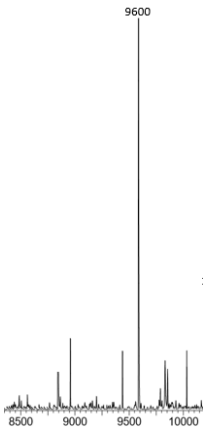
| Oligo name | Sequence (5'-3') | Length | Mass calc. | Mass found |
|---|---|---|---|---|
| Res2465 (Template 1) | C**A**GTC**A**CTGT**A**CTGCCGACACAC ATAACC | 29 | 8776 | 8775 |
| Res2843 (FAM-primer) | FAM-GGTTATGTGTGTCGGCAG | 18 | 6138 | 6138 |
| Res8470 (PCR template) | GCATTCGAGCAACGTAAGATCCTT ACCACACAATCTCACACTCTGGA ATTCACACTGACAATACTGCCGAC ACACATAACC | 81 | 24675 | 24675 |
| Res8471 (PCR primer) | GGTTATGTGTGTCGGCAG | 18 | 5602 | 5603 |
| Res8472 (PCR primer) | GCATTCGAGCAACGTAAG | 18 | 5533 | 5532 |

**Table 2.1.** Sequences of oligonucleotides used in primer extension and PCR reactions (FAM represents the 6-carboxyfluorescein, MW: 537.5).

The general protocols for primer extension using modified 5-AM-dUTPs are listed in the experimental section (page 222).

Two kinds of thermostable DNA polymerases (GoTaq and KOD) were used initially. PAGE and mass spectrometry confirmed that both GoTaq and KOD thermostable polymerases successfully incorporated amU1 into the oligonucleotide template. By comparing these two polymerases at different reaction times, it was found that GoTaq polymerase incorporates amU1 in a shorter period of time (5 min) but with lower efficiency (many primer molecules remain unreacted after 30 min). In contrast, KOD polymerase required more time to extend the primer in the presence of amU1 yet achieved excellent results after 30 min with all primers extended to their full length (**Figure 2.12**).




Template: 5'-***CAGTCACTGTA***CTGCCGACACACATAACC-3'
Primer: GACGGCTGTGTGTATTGG-FAM
Result: 3'-*GXCAGXGACAX*GACGGCTGTGTGTATTGG-FAM

**Figure 2.12.** PEX of triphosphate amU1 using GoTaq and KOD DNA polymerase. Top left: primer extension of triphosphate amU1 using GoTaq (A) and KOD (B) DNA polymerase. Lane 1: polymerase only with primer and template in the absence of any dNTPs; Lane 2: amU1 with template, primer and all other dNTPs (dATP + dCTP + dGTP) added with reaction times of 5 min; Lane 3: amU1 with template, primer and all other dNTPs (dATP + dCTP + dGTP) added with a reaction time of 30 min.Top right: Mass spectra of fully extended primers: amU1 full length product (calculated mass: 9601, actual: 9600). Bottom: Sequences of oligonucleotides used in primer extension test (X represents modified 5-AM-dUTP).

Further studies were carried out using GoTaq and KOD DNA polymerases to compare the incorporation of oligonucleotides using amU1 and dTTP. The reaction mixtures were heated at 60 °C for 90 min. The results for dTTP and amU1 on primer extension using GoTaq polymerase were promising, suggesting an efficiency comparable to those of natural dTTP (**Figure 2.13A**). However, results from the primer extension using KOD polymerase were unusual, in that even the band for the initial FAM-labelled primer is missing (**Figure 2.13B**). This phenomenon can be explained by the existence

of a 3′ → 5′exonuclease in KOD. The 3′ → 5′ exonucleases proofread by cleaving DNA at the 3′- end through the hydrolysis of phosphodiester bonds. In this case, the primer extension reaction was probably left too long, hence the 3′ → 5′ exonuclease activity hydrolysed all the oligonucleotides. Differs to KOD, which is a family B-DNA polymerase and possesses only 3′ → 5′ exonuclease activity, GoTaq is a family A-DNA polymerases with 5′ → 3′ exonuclease activity an only weak 3′ → 5′ exonuclease properties.[147] Therefore, primer extension reactions using GoTaq, despite being performed for 90 min, were relatively unaffected by this degradation phenomenon during these tests.

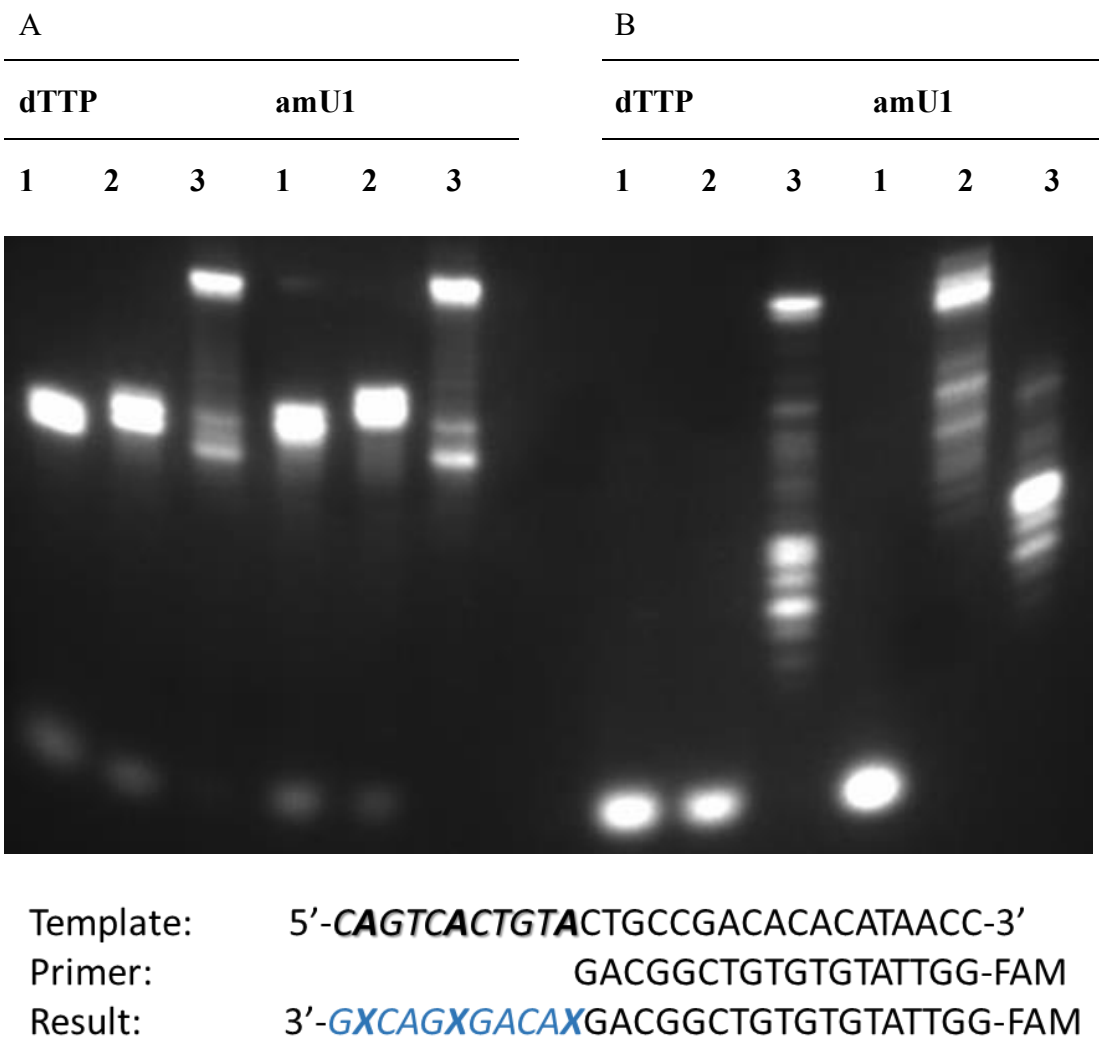



**Figure 2.13.** Primer extension test comparing amU1 and dTTP for GoTaq (A) and KOD (B) DNA polymerase. The reaction mixtures were heated at 60 °C for 90 min. Lane 1 contained only polymerase, primer and template in the absence of any dNTPs. Lane 2 contained polymerase, primer and template and only dTTP or amU1 without other dNTPs. Lane 3 contains polymerase, primer and template dTTP or amU1 with all other dNTPs (dATP + dCTP + dGTP) added. Note: the incorporation of amU1 without other dNTPs using KOD polymerase (column B, amU1, line 2) gives unexpected full-length results may due to contaminations of dNTPs solution.

Vent exo- (Vent DNA polymerase with 3′→ 5′exonuclease activity eliminated) is another family B-DNA polymerase that was used in place of KOD for primer extension

tests for oligonucleotides containing amU1 and dTTP over different reaction times (**Figure 2.14**). The PAGE and MS analysis concluded that, by replacing dTTPs into amU1 triphosphate, DNA polymerases such as GoTaq and Vent exo- could both recognise the modified triphosphates and successfully incorporate amU1 into the template during primer extension. The performance of Vent exo- with amU1 was found to be better than for GoTaq, since nearly all of the primers provided were successfully extended in a relatively short timeframe, whereas some primers remained unextended when using GoTaq.

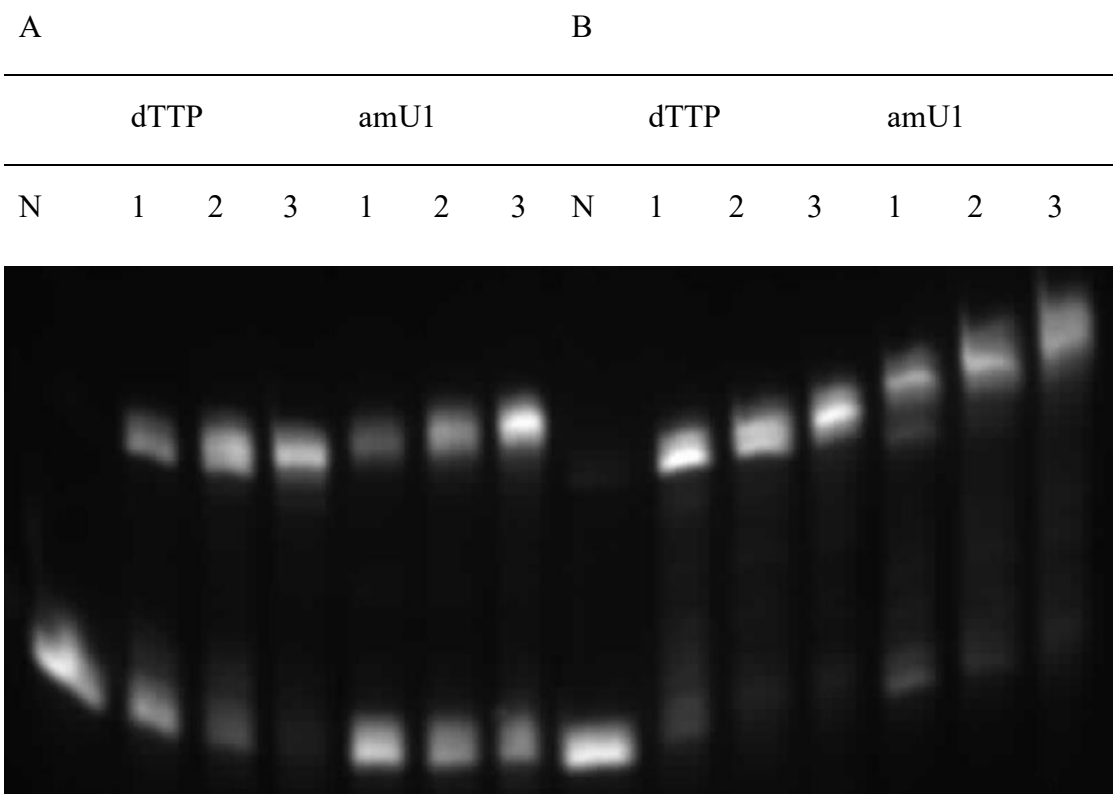



**Figure 2.14.** Primer extension reactions comparing amU1 and dTTP with GoTaq (A) and Vent exo- (B) DNA polymerases. The reaction mixtures were heated at 65 °C. Lane N: polymerase only, primer and template were added without any dNTPs; Lane 1: dTTP or amU1 added with all other dNTPs with a reaction time of 15 min. Lane 2: primer and template, dTTP or amU1 added with all other dNTPs with a reaction time of 45 min. Lane 3: primer and template, dTTP or amU1 added with all other dNTPs with a reaction time of 90 min.

Considering the promising results of amU1 incorporation during primer extension, five other modified 5-AM-dUTPs (amU2 – amU6) possessing hydrophobic or aromatic functional groups at the 5-position of uridine were tested for their compatibility with polymerases using the previously described protocol. HPLC-MS analysis (**Table 2.2**) confirmed that all modified 5-AM-dUTPs were successfully incorporated into oligonucleotides after a 45 min reaction, forming extended primers of full-length using Vent exo- DNA polymerase, although some produced either multiple or broad bands on the PAGE gels (**Figure 2.15**).

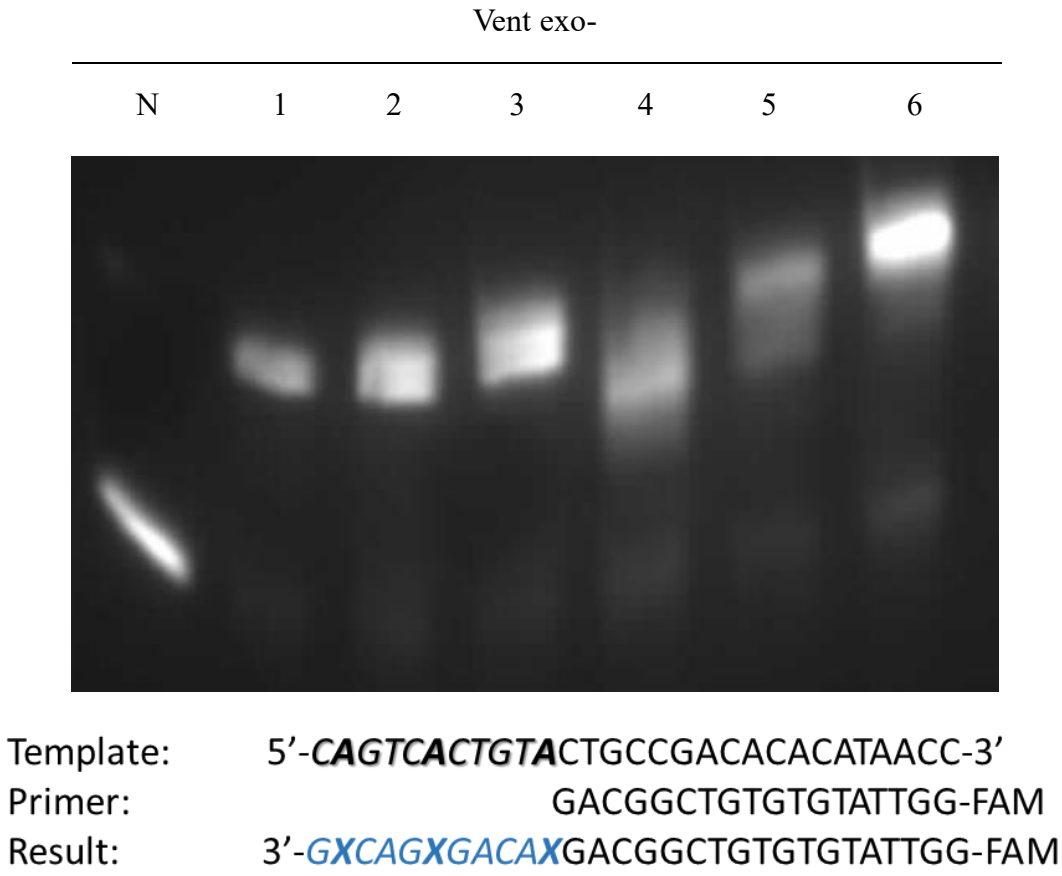



**Figure 2.15.** Primer extension tests using Vent exo- DNA polymerase with modified 5-AM-dUTPs (amU1 – amU6). The reaction mixtures were heated at 65 °C for 45 min. Lane N contained only polymerase, primer and template without any dNTPs. Lane 1 contained amU1 together with all other dNTPs (dATP + dCTP + dGTP). Lane 2 contained primer and template, amU2 with all other dNTPs. Lane 3 contained primer and template, amU3 with all other dNTPs. Lane 4 contained primer and template, amU4 with all other dNTPs. Lane 5 contained primer and template, amU5 with all other dNTPs. Lane 6 contained primer and template, amU6 with all other dNTPs.

| Modified dUTPs (X) | 29-mer with 5'-FAM, sequence (5'-3') | Mass calc. | Mass found |
|---|---|---|---|
| amU1 | GGTTATGTGTGTCGGCAGX$_1$ACAGX$_1$GACX$_1$G | 9601 | 9600 |
| amU2 | GGTTATGTGTGTCGGCAGX$_2$ACAGX$_2$GACX$_2$G | 10129 | 10129 |
| amU3 | GGTTATGTGTGTCGGCAGX$_3$ACAGX$_3$GACX$_3$G | 10021 | 10020 |
| amU4 | GGTTATGTGTGTCGGCAGX$_4$ACAGX$_4$GACX$_4$G | 10090 | 10091 |
| amU5 | GGTTATGTGTGTCGGCAGX$_5$ACAGX$_5$GACX$_5$G | 9955 | 9954 |
| amU6 | GGTTATGTGTGTCGGCAGX$_6$ACAGX$_6$GACX$_6$G | 10117 | 10116 |

**Table 2.2.** Sequences and mass spec results of fully extended primers containing modified 5-AM-dUTPs in primer extension tests (FAM = 6-carboxyfluorescein MW = 537.5).

## 2.4 PCR and qPCR analysis

### 2.4.1 PCR amplification of modified 5-AM-dUTPs

Studies of modified 5-AM-dUTPs (amU1 – amU6) using linear extension tests as delineated in the previous section demonstrate that all six 5-AM-dUTPs can be incorporated into DNA oligonucleotides by primer extension. Encouraged by these results, PCR amplification reactions using these modified 5-AM-dUTPs were evaluated. Compared to the previous primer extension studies, the PCR requirements for these modified 5-AM-dUTPs are more stringent, as chemical modifications of nucleobases must be tolerated by DNA polymerases when they are in the template strand as well as when they are being incorporated into the template as dNTPs. This is because in PCR

DNA polymerases must use a template containing modified 5-AM-dUTPs for subsequent cycles of PCR amplification after the first extension step, in contrast to the primer extension reactions where this is not a factor.

An 81-mer DNA template (Res8470, **Table 2.1**) was designed for PCR reactions using two 18-mer PCR primers (**Table 2.1**), one of which was complementary to the template at the 3′ end (Res8471) and the other identical to the template at the 5′ ends (Res8472). In total, three types of DNA polymerase were used to analyse PCR efficiency in the presence of 5-AM-dUTPs. The polymerases used were GoTaq G2 DNA polymerase, Phusion High-Fidelity DNA Polymerases, and Vent (exo-) DNA polymerase.

The general protocol using 5-AM-dUTPs in the PCR amplification was as follows:
To each 50 μL PCR reaction was added 0.2 μM of both forward and reverse primers, 40 nM (1 ng) of template (the amount of template required in a PCR reaction was determined through an optimisation test as shown in **Figure 2.16**), 200 μM of each dNTP (including natural dNTPs and modified 5-AM-dUTPs), followed by the addition of 1 unit of DNA polymerase and its corresponding 1x buffer with 2 mM $Mg^{2+}$ solution ($MgCl_2$ for GoTaq G2 and $MgSO_4$ for Vent exo- and Phusion).

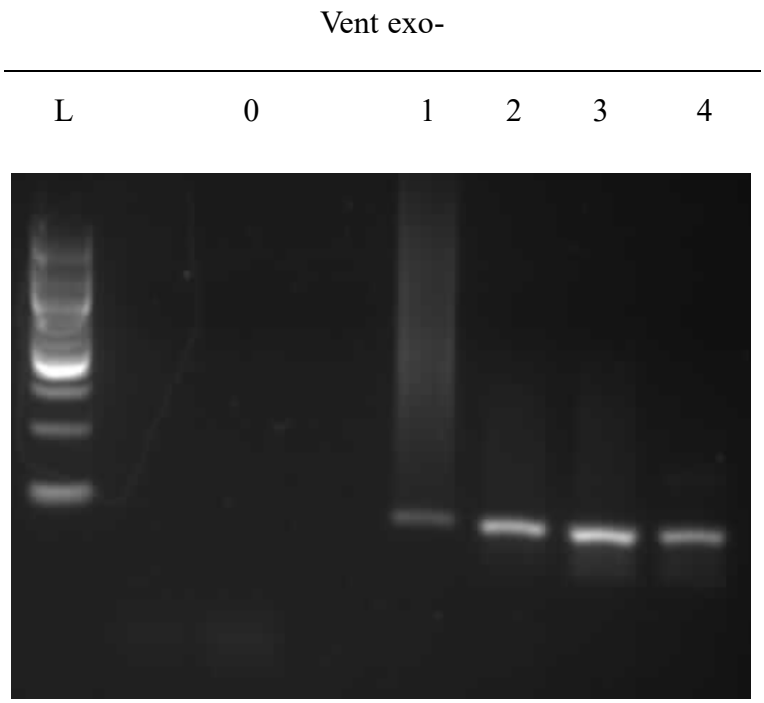



**Figure 2.16.** PCR optimisation test for template concentration using Vent (exo-) with natural dNTPs (reaction conditions showed on **Figure 2.17** below). Lane L: 100 bp DNA ladder. Lane 0: negative control with only template added so that no PCR reaction occurs. Lane 1: PCR using 10 pg of template (0.4 nM). Lane 2: PCR using 100 pg of template (4 nM). Lane 3: PCR using 1 ng of template (40 nM). Lane 4: PCR using 10 ng of template (0.4 μM).

The PCR amplification reaction was performed using a Bio-Rad T100 Thermal Cycler (PCR instrument). PCR thermal cycling conditions consisted of an initial denaturation step at 95 °C for 3 min, which was followed by 30 cycles of denaturation (95 °C for 30 s); a primer annealing step (55 °C for 30 s); and a template elongation/extension step (72 °C for 30 s). A further extension was performed at 72 °C for 5 min and the final reaction mixture was maintained at 4 °C (**Figure 2.17**). The samples were analysed by electrophoresis run at constant voltage (126 V) in 1× TBE buffer using a 2% agarose gel that had been pre-stained with 5 μL SYBR gold. The resulting bands containing the

fully extended oligonucleotides were excised, extracted, and desalted using NAP gel-filtration columns prior to analysis by HPLC-MS.

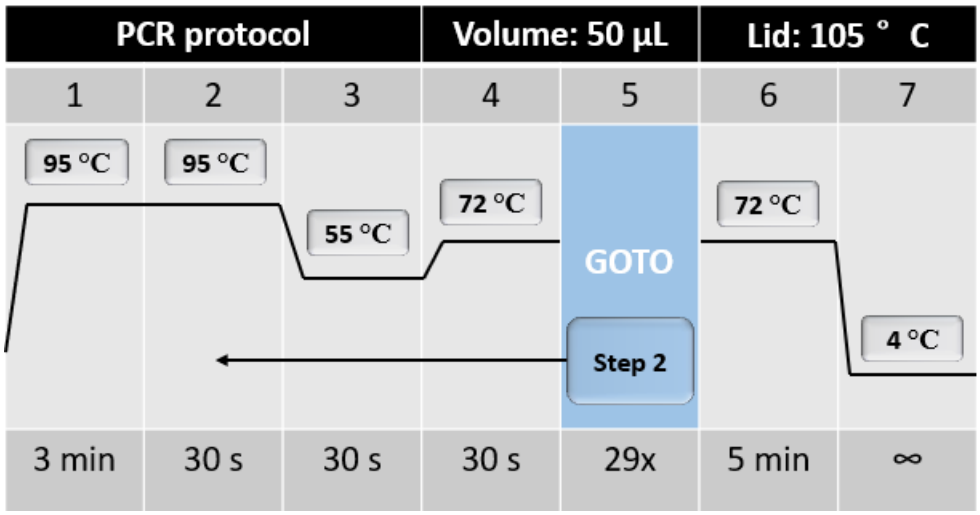



**Figure 2.17.** General PCR amplification conditions using a Bio-Rad T100 Thermal Cycler.

All six modified 5-AM-dUTPs (amU1 – amU6) were evaluated as substitutes for dTTP in PCR amplification experiments in the presence of dATP, dCTP and dGTP. In all cases, the amplification of an 81-mer DNA template was successful. GoTaq G2, Phusion, and Vent (exo-) DNA polymerases were all found to efficiently incorporate the modified 5-AM-dUTPs during PCR amplification. Vent (exo-) generally performed the best, with modified 5-AM-dUTPs giving similar product yields to dTTP under standard conditions.

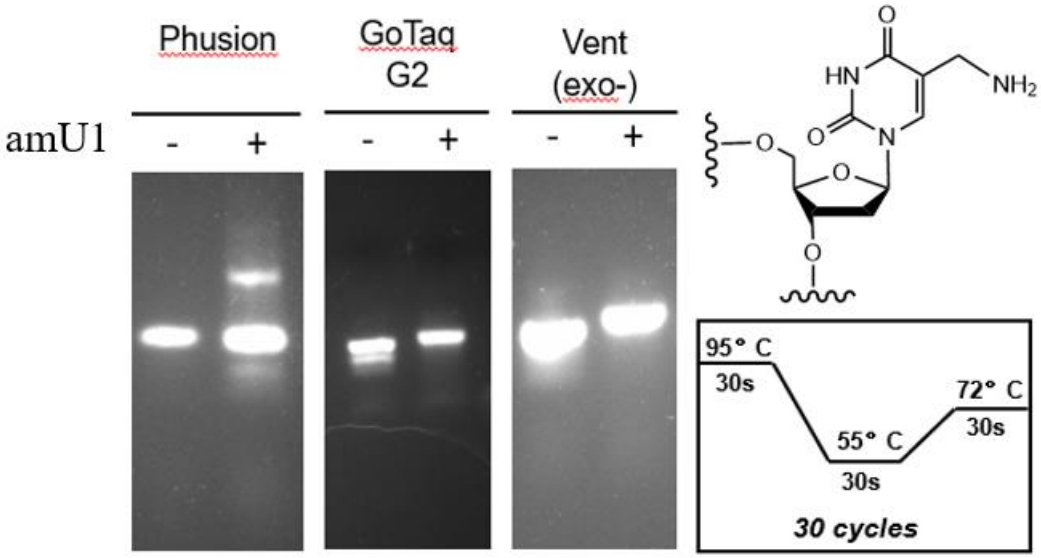



**Figure 2.18.** Left panel panel shows a comparison of the performance of three DNA polymerases, Phusion, GoTaq G2 and Vent (exo-) in PCR amplification reactions in the presence of dTTP (-) or amU1 (+). The right-hand panel shows the structure of the amU1nucleoside under which is the PCR amplification cycle used.

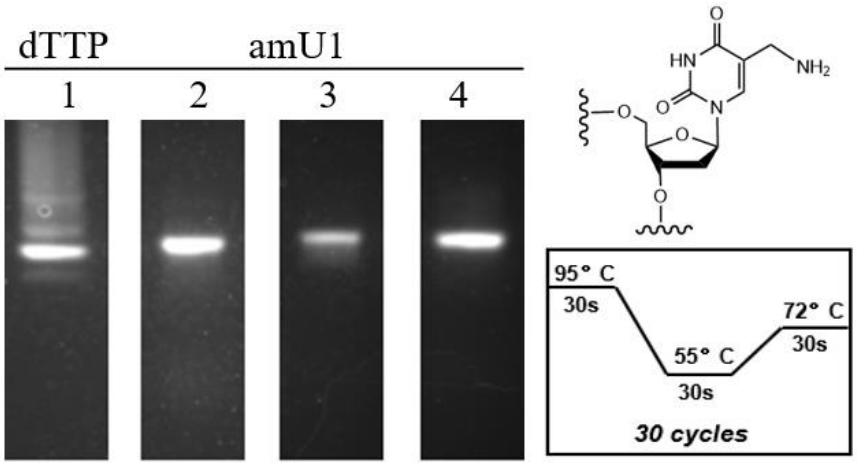



**Figure 2.19.** A successful PCR amplification incorporating amU1 to the template (full gel presented in experimental section). Left-hand panel: Lane 1: positive control with addition of dTTP together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 2. Addition of amU1 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 3: Addition of amU1 together with all other dNTPs (dATP + dCTP + dGTP) using Gotaq G2. Lane 4: Addition of amU1 together with all other dNTPs (dATP + dCTP + dGTP) using Phusion. The right-hand panel shows the structure of the amU1nucleoside under which is the PCR amplification cycle used. (The difference between **Figure 2.18** and **Figure 2.17** is that all four bands in **Figure 2.18** were performed in one complete gel.)

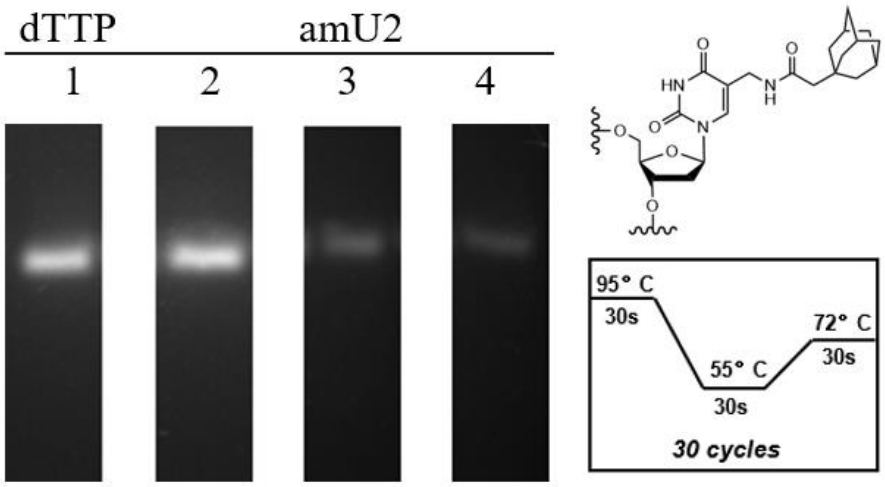



**Figure 2.20.** Successful PCR amplification incorporating amU2 into the template (full gel presented in experimental section). Left-hand panel: Lane 1: positive control with addition of dTTP together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 2: addition of amU2 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 3: addition of amU2 together with all other dNTPs (dATP + dCTP + dGTP) using Gotaq G2. Lane 4: addition of amU2 together with all other dNTPs (dATP + dCTP + dGTP) using Phusion. The right-hand panel shows the structure of the amU2 nucleoside under which is the PCR amplification cycle used.

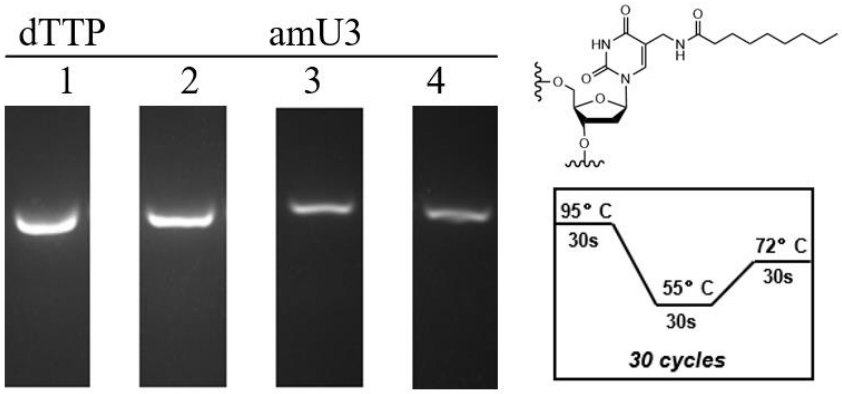



**Figure 2.21.** A successful PCR amplification reaction incorporating amU3 into the template (full gel presented in experimental section). Left-hand panel: Lane 1: positive control with addition of dTTP together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 2: addition of amU3 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 3: addition of amU3 together with all other dNTPs (dATP + dCTP + dGTP) using Gotaq G2. Lane 4: addition of amU3 together with all other dNTPs (dATP + dCTP + dGTP) using Phusion. The right-hand panel shows the structure of the amU3 nucleoside under which is the PCR cycle used.

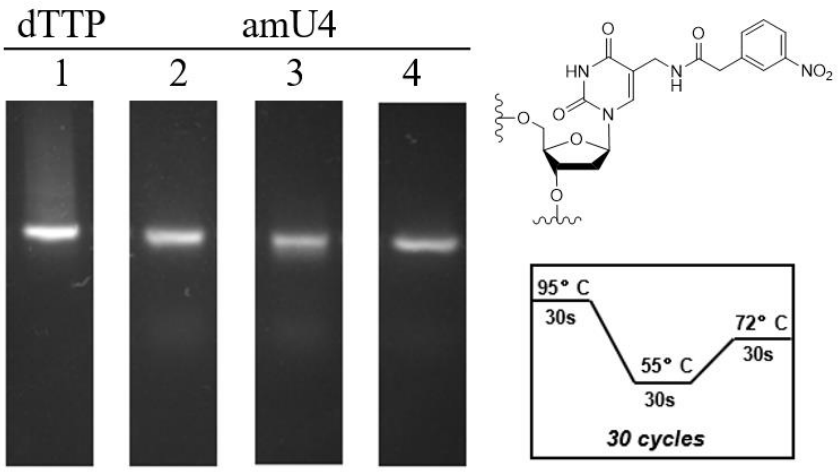



**Figure 2.22.** Successful PCR amplification incorporating amU4 into the template (full gel presented in experimental section). Left-hand panel: Lane 1: positive control by addition of dTTP together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 2: addition of amU4 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 3: addition of amU4 together with all other dNTPs (dATP + dCTP + dGTP) using Gotaq G2. Lane 4: addition of amU4 together with all other dNTPs (dATP + dCTP + dGTP) using Phusion. The right-hand panel shows the structure of the amU4 nucleoside under which is the PCR cycle used.

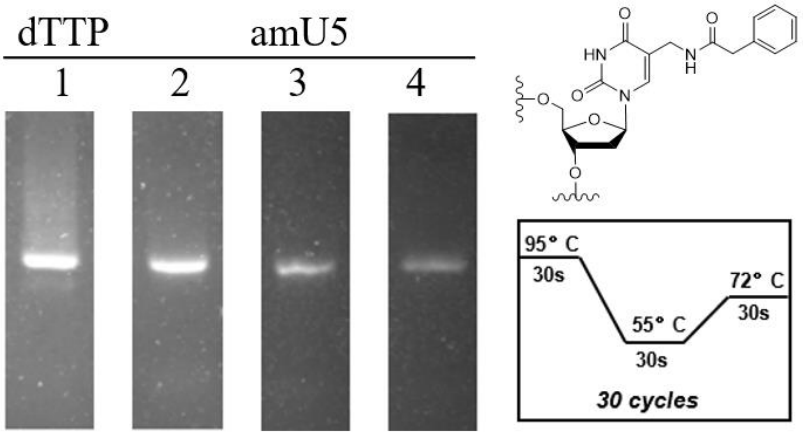



**Figure 2.23.** Successful PCR amplification incorporating amU5 into the template (full gel presented in experimental section). Left-hand panel: Lane 1: positive control with addition of dTTP together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 2: addition of amU5 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 3: addition of amU5 together with all other dNTPs (dATP + dCTP + dGTP) using Gotaq G2. Lane 4: addition of amU5 together with all other dNTPs (dATP + dCTP + dGTP) using Phusion. The right-hand panel shows the structure of the amU5 nucleoside under which is the PCR cycle used.

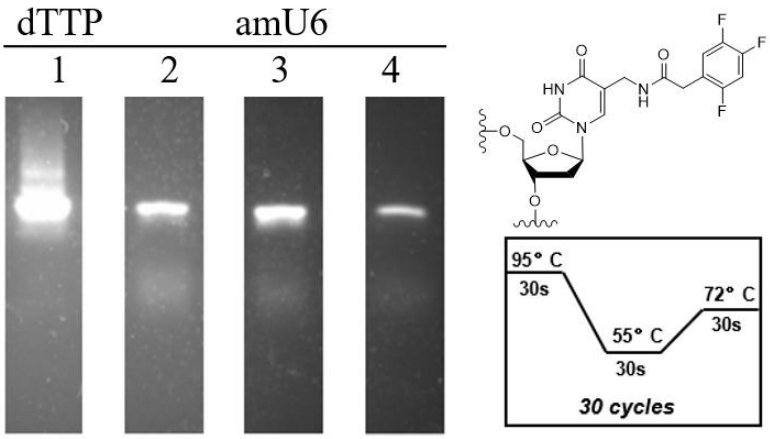



**Figure 2.24.** Successful PCR amplification incorporating amU6 into the template (full gel presented in experimental section). Left-hand panel: Lane 1: positive control with addition of dTTP together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 2: addition of amU6 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-). Lane 3: addition of amU6 together with all other dNTPs (dATP + dCTP + dGTP) using Gotaq G2. Lane 4: addition of amU6 together with all other dNTPs (dATP + dCTP + dGTP) using Phusion. The right-hand panel shows the structure of the amU6 nucleoside under which is the PCR cycle used.

**Figures 2.18** to **2.24** show that Vent (exo-) is the best polymerase for amplifying PCR templates when the modified amdUTPs are used in place of TTP.

| Modified 5-AM-dUTPs | Mass calc. of both strands | Found Mass of both strands |
|---|---|---|
| dTTP (positive control) | 24675, 25247 | 24674, 25247 |
| amU1 | 24870, 25577 | 24869, 25577 |
| amU2 | 27158, 29449 | 27157, 29448 |
| amU3 | 26690, 29657 | 26689, 29656 |
| amU4 | 26989, 29163 | 26989, 29162 |
| amU5 | 26404, 28173 | 26402, 28172 |
| amU6 | 27106, 29548 | 27106, 29547 |

**Table 2.3.** Sequences and mass spec analysis of fully extended primers containing modified 5-AM-dUTPs after PCR amplification. PCR product is an 81-mer.

## 2.4.2 Real-time PCR of modified 5-AM-dUTPs

Real-time PCR studies were conducted with all six 5-AM-dUTPs (amU1 – amU6) to analyse the $C_q$ values when these triphosphates are used (details about $C_q$ values and mechanisms of real-time PCR are presented in Chapter 1.5.2 under "real-time PCR/qPCR").

The amplification reactions in the presence of the modified 5-AM-dUTPs were performed using the following real-time PCR protocol:

With each PCR reaction of 20 µL was added 0.2 µM of both forward and reverse primers; 4 nM (100 pg) of template (the amount of template required in a PCR reaction was determined through an optimisation test shown in **Figure 2.25**), and 200 µM of each dNTP (natural dNTPs and modified 5-AM-dUTPs), followed by the addition of 1 unit of DNA polymerase and its corresponding 1x of 2 mM $Mg^{2+}$ solution buffer. 1x EvaGreen dye was added to each real-time PCR reaction. EvaGreen is a green fluorescent nucleic acid dye that is non-fluorescent as a free dye, but shows strong fluorescence upon binding to double-stranded DNA.

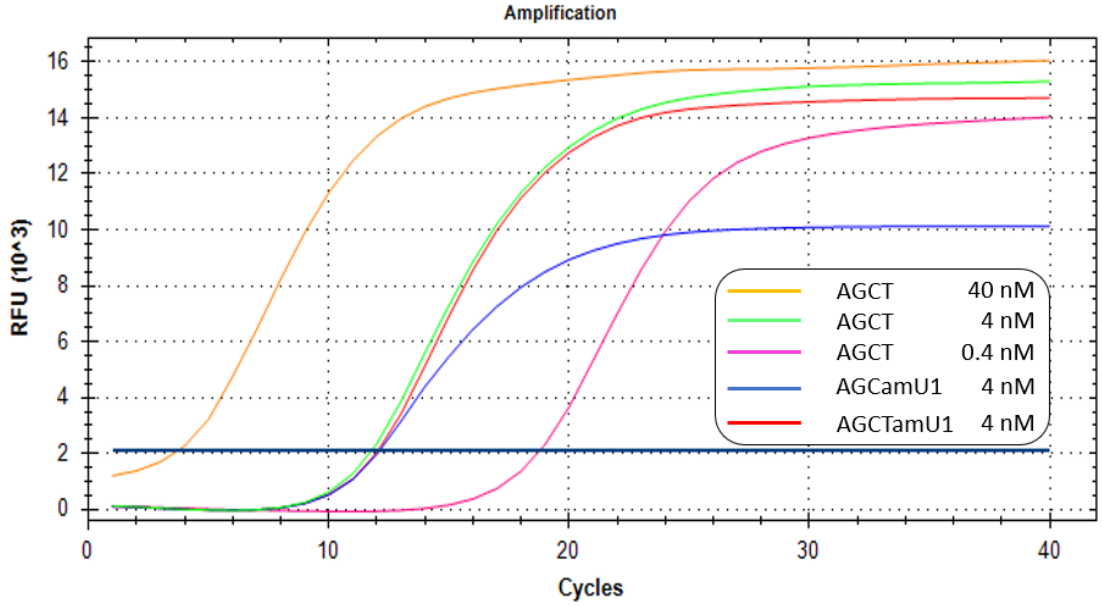



**Figure 2.25.** Real-time PCR optimisation test for template concentrations using Vent (exo-) in combination with naturally occurring dNTPs and amU1. The table inset shows the colour coding for each reaction based on the dNTPs used and the concentration of template in each reaction.

The PCR amplification was performed on a real-time PCR instrument (Bio-Rad CFX96). A generalised protocol for real-time PCR is presented in Figure 3.28 and is outlined as follows:

Samples containing modified 5-AM-dUTPs first underwent an initial denaturation at 95 °C for 3 min; followed by 40 cycles of denaturation at 95 °C for 30 s; primer annealing at 55 °C for 30 s; and subsequent extension at 72 °C for 30 s. A further extension was conducted at 72 °C for 5 min. During the reaction, a camera monitors the amplification of targeted DNA oligonucleotide by measuring the concentration of fluorescent EvaGreen dye after each annealing step. This was followed by melting experiments involving heating the amplification mixture to 95 °C for 30 s, cooling to 30 °C, and then increasing the temperature at 0.5 °C per second to 95 °C and holding the temperature at each step for 5 s. The samples were further analysed by

electrophoresis at a constant voltage (126 V) in 1× TBE buffer using a 2% agarose gel pre-stained with 5 µL of SYBR gold. The resulting bands with fully extended oligonucleotides were excised, the DNA was extracted with water, desalted on NAP columns, and then analysed by HPLC-MS.

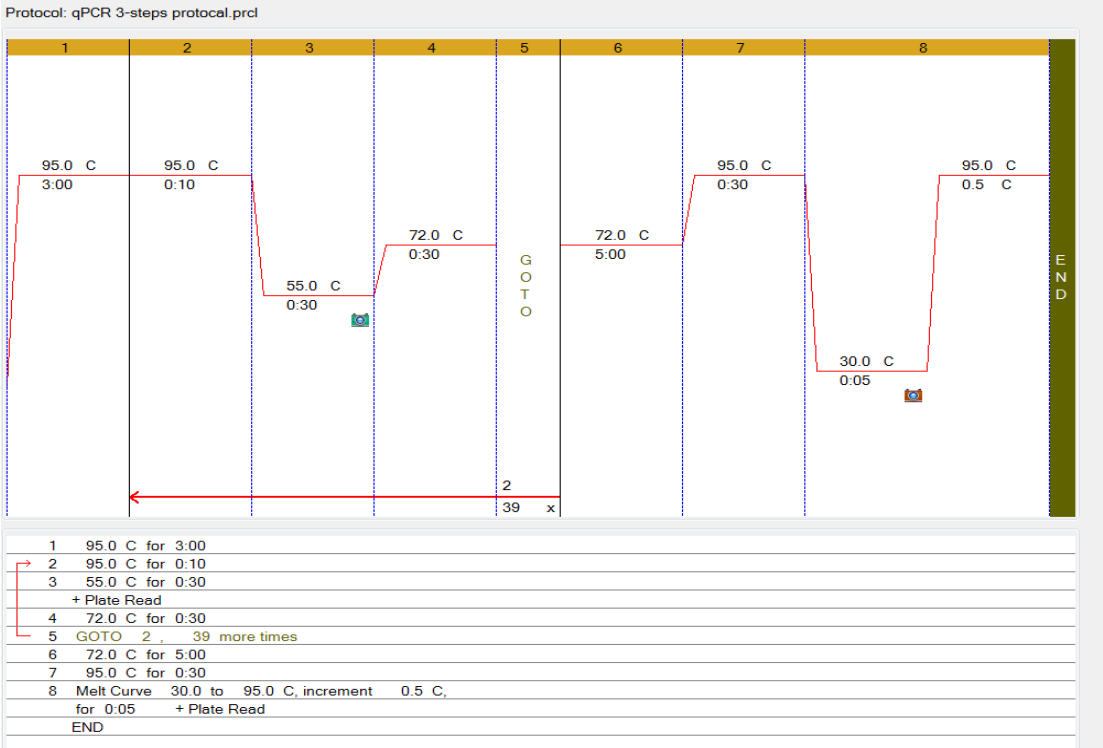
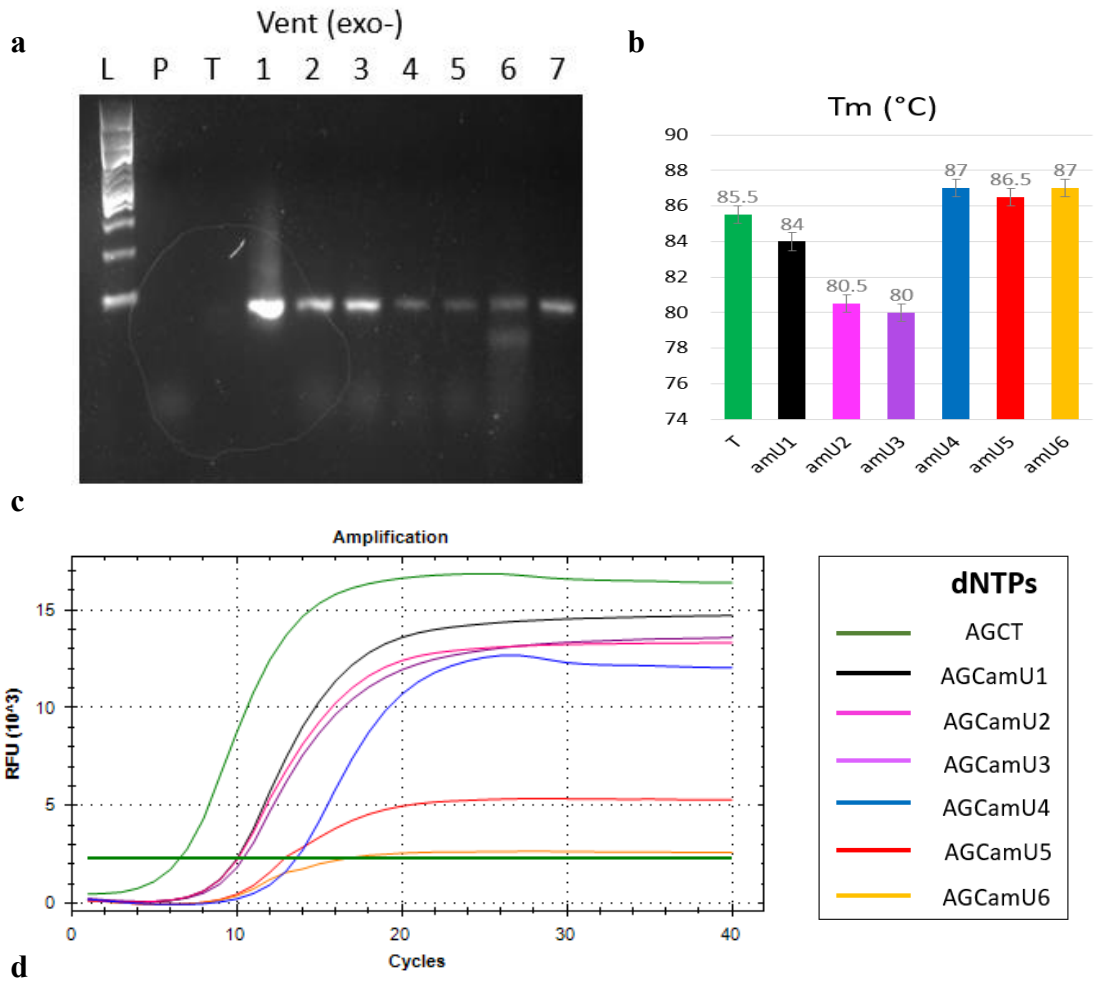



**Figure 2.26.** Protocol of real-time PCR (qPCR) amplification under standard conditions using a Bio-Rad CFX96.

A real-time PCR analysis of 5-AM-dUTPs and the obtained $C_q$ values are listed below in **Figure 2.27**. In summary, $C_q$ values for all six 5-AM-dUTPs were within acceptable ranges and were slightly higher than the $C_q$ values usually encountered for natural dTTP. All $C_q$ values for 5-AM-dUTPs presented were below 20 under standard conditions, indicating that the efficacy of incorporation into the template was relatively good and only marginally less efficient than for natural dTTP. $C_q$ values for all three 5-AM-

dUTPs (amU4 to amU6) containing polyaromatic side chains were found to be higher than for the other three 5-AM-dUTPs which contained either a free methylamine side chain (amU1) or two hydrophobic side chains (amU2 and amU3). This could be attributable to steric or electrostatic effects in the three 5-AM-dUTPs with aromatic side chains which could compromise interactions with the polymerase enzyme. The trifluorobenzyl analogue (amU6) led to a particularly odd amplification curve and high $C_q$ value. It is also possible that the binding of the fluorescent reporter dye is affected by the aromatic substituents which could also give rise to fluorescence quenching. Melting Temperature ($T_m$) studies of 5-AM-dUTPs and dTTPs reveal that the $T_m$ for amU1, amU2 and amU3 is decreased relative to dTTP, especially in the case of hydrophobic amU2 and amU3, wherein the $T_m$ of these two analogues is reduced by approximately 5 °C. On the other hand, the three analogues with aromatic side chains (amU4, amU5 and amU6) were found to have slightly increased $T_m$ values ($\Delta T_m$ ~ +1.5°C) (**Figure 2.27**).




| dNTPs | AGCT | AGCamU1 | AGCamU2 | AGCamU3 | AGCamU4 | AGCamU5 | AGCamU6 |
|---|---|---|---|---|---|---|---|
| $C_q$ values | 6.54 | 10.07 | 10.38 | 10.03 | 13.06 | 12.89 | 16.67 |
| Tm (°C) | 85.5 | 84.0 | 80.5 | 80.0 | 87.0 | 86.5 | 87.0 |

**Figure 2.27.** Real-time PCR reactions using Vent (exo-) in combination with naturally occurring dNTPs and modified 5-AM-dUTPs in place of dTTP. **(a)** PCR amplification using dTTP and modified 5-AM-dU analogues (amU1 – amU6) and Vent (exo-) DNA polymerase. Lane L: 100 bp DNA ladder; Lane P: primers only; Lane T: template only; Lane 1: positive control with addition of natural dNTPs; Lane 2: addition of amU1 together with all other dNTPs (dATP + dCTP + dGTP); Lane 3: addition of amU2 together with all other dNTPs; Lane 4: addition of amU3 together with all other dNTPs; Lane 5: addition of amU4 together with all other dNTPs; Lane 6: addition of amU5 together with all other dNTPs; Lane 7: addition of amU6 together with all other dNTPs; **(b)** bar chart showing melting temperatures ($T_m$) of PCR

products containing 5-AM-dUTP analogues relative to those of dTTP; **(c)** real-time PCR analysis of 5-AM-dUTP analogues. The table inset on the right illustrates the colour coding used to denote which dNTPs were used in each reaction. **(d)** Table showing $C_q$ and $T_m$ values for each 5-AM-dUTP analogue as compared to naturally occurring dTTP in real-time PCR amplification experiments.

## 2.5 Conclusions

A series of 5-aminomethyl-deoxyuridine (5-AM-dU) analogues were designed as functionalised dTTP analogues for use in SELEX. 5′-DMTr protected 5-aminomethyl-deoxyuridine (5-AM-dU) was synthesised in 6 steps with an overall yield of 32% from thymidine (T). Both phosphoramidite and nucleoside triphosphate forms of these 5-AM-dU analogues were synthesised to study their enzymatic incorporation into DNA oligonucleotides. The phosphoramidite of 5-AM-dU was obtained in an overall yield of 43%. In total, six modified 5-AM-dUTPs (amU1 – amU6) with either a free amine or a hydrophobic or aromatic group attached to the 5-position of the nucleobase were synthesised.

Primer extension studies were conducted to demonstrate that all six 5-AM-dUTP analogues with modifications at the 5-position of uridine are functionally accepted by DNA polymerases and can substitute for dTTP. PCR and real-time PCR amplification of an 81-mer DNA template using all six 5-AM-dUTP analogues alongside other natural dNTPs were all successful to varying degrees. All three DNA polymerases used, namely GoTaq G2, Phusion and Vent (exo-), efficiently incorporated these modified 5-AM-dUTPs onto the DNA template in PCR reactions, although Vent (exo-) polymerase generally performed best, resulting in similar yields to those obtained with dTTP under standard conditions. Real-time PCR studies of 5-AM-dUTPs provided similar results. The $C_q$ values obtained with all the 5-AM-dUTPs except amU6 were consistent with satisfactory amplification, with only slightly greater $C_q$ values than for natural dTTP.

Taken together, these results suggest that 5-AM-dU analogues possessing a range of unnatural side chains are functionally compatible with DNA polymerase enzymes and can be used as new analogues for *in vitro* selection by SELEX. We have not yet evaluated these analogues in RNA SELEX in which transcription, reverse transcription, and PCR are involved.

# *Chapter 3.*

# *DNA Amplification Studies on the Quadracyclic fluorescent adenine analogue qAN1*

# 3.1 Introduction and Aims

## 3.1.1 qAN1 as an adenine analogue in DNA

Enzymatic synthesis of nucleobase-functionalised nucleic acids using base-modified deoxyribonucleoside 5′-triphosphates (dNTPs) provides a convenient tool for introducing functional groups into a nucleic acid target of interest.[148–151] While Taq DNA polymerase is traditionally used in the enzymatic synthesis of base-modified oligonucleotides, some members of the B-family of polymerases (*e.g.,* KOD, Vent (exo-) and deep Vent (exo-)) give greater PCR efficiency.[148,152] Variables included the position of the modification, the type of functional group added, and the flexibility of the linker arm attachment, all of which strongly influence DNA polymerase efficiency.[149,153] Extensive investigations have revealed that modifications on nucleobases, notably of 5-substituted pyrimidine or 7-substituted 7-deazapurine dNTPs, serve as good substrates[102,154–158] for DNA polymerases and can be used for primer extension (PEX) and PCR amplification.[159] However, complete substitution of a modified dNTP for its natural counterpart in PCR amplification is commonly found to be problematic, as it tends to give very low or undetectable product formation.[103,149,160,161] This is due to a lower efficiency of nucleotide incorporation over multiple cycles, thereby inhibiting PCR.[148] Further, in most case studies, only one modified dNTP was used in the presence of the other three naturally occurring dNTPs. Several studies using two or more modified dNTPs simultaneously all reported greater difficulties in PEX and PCR amplification.[103,149,162–166]

qAN1 is a fluorescent quadracyclic adenine analogue (**Figure 3.1**) known to have a high fluorescence brightness value ($\varepsilon\Phi_F = 1700$).[167] This fluorescent base, as well as its other quadracyclic adenine analogues, have been successfully incorporated into DNA oligonucleotides via phosphoramidite chemistry and used as FRET-donors in base-base FRET pairings.[168–170] Circular dichroism (CD) studies show that duplexes modified with qAN1 adopt normal B-form DNA geometry. Melting temperature studies and mismatch data indicate that the quadracyclic adenine analogue qAN1 is selective for thymine (T) in terms of base pairing. This fluorescent qAN1 works as a FRET-donor with another quadracyclic adenine analogue $qA_{nitro}$, which is a non-fluorescent FRET-acceptor. Together they provide a brightness value of 510.[168] It is clear that qAN1 is a promising fluorescent DNA nucleobase analogue with potential use in SELEX.

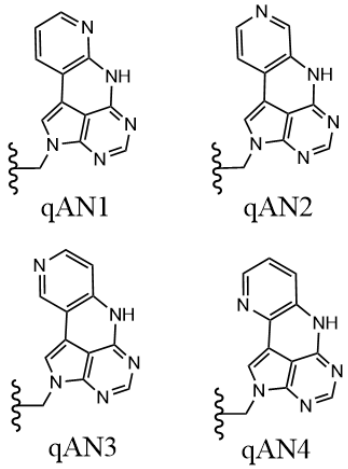



**Figure 3.1.** Structure of qAN1 and other quadracyclic adenine analogues.

We undertook such a study but initial results indicated that complete substitution of qAN1 for dATP in PCR amplification is problematic, with no detectable product formation (**Figure 3.6**). This is unfortunate, as incorporating qAN1 within a DNA

oligonucleotide template in its triphosphate form is essential for SELEX. Hence, solving the problem of low PCR efficiency with qAN1 triphosphate is a problem that must be solved.

Melting temperature ($T_m$) studies of qAN1-containing duplexes reveal that $T_m$ is increased, on average, by 2.9 °C for each qAN1-T base pair.[168] In some circumstances, such as when neighbouring base pairs are CC, CT and TT (underlined in **Figure 3.2**), the melting temperatures increases by more than 6 °C. Increased $T_m$ and lower PCR efficiency also occurs with unmodified DNA duplexes when using GC-rich DNA constructs (when 60% or more bases are either cytosine or guanine).[171] Several studies of high GC content genes using PCR indicate that these major problems are caused by the difficulty in achieving DNA denaturation, and the possibility of secondary structures forming within the DNA template.[172,173] The best solution currently available to overcome the problems associated with the PCR amplification of GC-rich sequences is the introduction of additives (*e.g.,* DMSO, formamide, glycerol, betaine or 7-deaza-dGTP) into the system which resolve the issue of the formation of complex secondary structures by destabilising DNA duplexes.[174–178] This well-developed concept of introducing organic molecules with destabilising (denaturing) effects to solve the problem of GC-rich DNA constructs was explored in the current study to improve the PCR efficiency of qAN1-containing DNA duplexes.

| sequence name[a] | DNA sequence[b] | $T_m^{qAN1}$ (°C)[c] | $T_m^{A}$ (°C)[c] | $\Delta T_m$ (°C) |
|---|---|---|---|---|
| AA | 5′-d(CGCAA(**qAN1**)ATCG)-3′ | 40.9 | 43.5 | −2.6 |
| AC | 5′-d(CGCAA(**qAN1**)CTCG)-3′ | 49.0 | 47.1 | 1.9 |
| AG | 5′-d(CGCAA(**qAN1**)GTCG)-3′ | 47.7 | 45.9 | 1.8 |
| AT | 5′-d(CGCAA(**qAN1**)TTCG)-3′ | 46.4 | 43.4 | 3.0 |
| CA | 5′-d(CGCAC(**qAN1**)ATCG)-3′ | 51.4 | 46.5 | 4.9 |
| CC | 5′-d(CGCAC(**qAN1**)CTCG)-3′ | 57.1 | 50.3 | 6.8 |
| CG | 5′-d(CGCAC(**qAN1**)GTCG)-3′ | 55.0 | 49.5 | 5.5 |
| CT | 5′-d(CGCAC(**qAN1**)TTCG)-3′ | 54.0 | 47.3 | 6.7 |
| GA | 5′-d(CGCAG(**qAN1**)ATCG)-3′ | 44.0 | 45.3 | −1.3 |
| GC | 5′-d(CGCAG(**qAN1**)CTCG)-3′ | 51.2 | 49.2 | 2.0 |
| GG | 5′-d(CGCAG(**qAN1**)GTCG)-3′ | 50.0 | 48.1 | 1.9 |
| GT | 5′-d(CGCAG(**qAN1**)TTCG)-3′ | 47.7 | 45.4 | 2.3 |
| TA | 5′-d(CGCAT(**qAN1**)ATCG)-3′ | 43.5 | 41.1 | 2.4 |
| TC | 5′-d(CGCAT(**qAN1**)CTCG)-3′ | 48.1 | 43.7 | 4.4 |
| TG | 5′-d(CGCAT(**qAN1**)GTCG)-3′ | 48.5 | 43.6 | 4.9 |
| TT | 5′-d(CGCAT(**qAN1**)TTCG)-3′ | 47.0 | 40.6 | 6.4 |

**Figure 3.2**. Differences between the melting temperatures ($T_m$) of qAN1 containing duplexes and unmodified duplexes. This study was conducted by Moa S. Wranne *et al.* from Prof. Marcus Wilhelmsson's laboratory in Chalmers University of Technology, Sweden. This figure was adapted with their permission from[168].

### 3.1.2 Combining duplex destabilising base pairs with qAN1 to improve PCR efficiency

The stability of DNA duplexes depends on a subtle balance between various interactions (described in detail in chapter 1.1.3 'Stability of DNA duplex'). Base-stacking interactions between adjacent bases are regarded as the primary source of DNA stability.[12] Interactions in terms of hydrogen bonding between nucleobases and between bases and surrounding water molecules are also important determinants of the stability of DNA duplexes. The structure of qAN1 possesses an extra pyrimidine ring when compared to adenine (A). The additional aromatic ring with a nitrogen in qAN1 inevitably increases base-stacking effects between qAN1 and adjacent bases. It also provides an extra hydrogen bonding position for qAN1. All of these additional DNA duplex stabilising effects of qAN1 relative to adenine are believed to be the cause of its positive effect on duplex melting temperature and is probably the major reason for its negative effect on PCR efficiency.

To improve PCR efficiency in the presence of qAN1 triphosphate, we decided to simultaneously incorporate a modified nucleobase with DNA destabilising effects to reduce (or neutralise) the strong stabilising effects of qAN1. The modified 5-AM-dUTPs, amU2 and amU3, which contain hydrophobic functional groups at the 5-position of pyrimidine are good candidates and should produce DNA duplexes with moderate melting temperatures, thus solving the problem of incomplete denaturation during PCR amplification. The hypothesis is that, in a single-stranded oligonucleotide, the hydrophobic groups of 5-AM-dUTPs will aggregate due to their strong

hydrophobicity. In contrast, in a double-stranded DNA structure, since the modifications are at the 5-position of thymine (T), the hydrophobic groups will lie in the major groove and point into an aqueous environment which will be entropically unfavourable. Therefore, DNA duplexes are destabilised by the introduction of these hydrophobic groups. The structures of amU2 and amU3, qAN1 and the qAN1-amU base pairing are below (**Figure 3.3**).

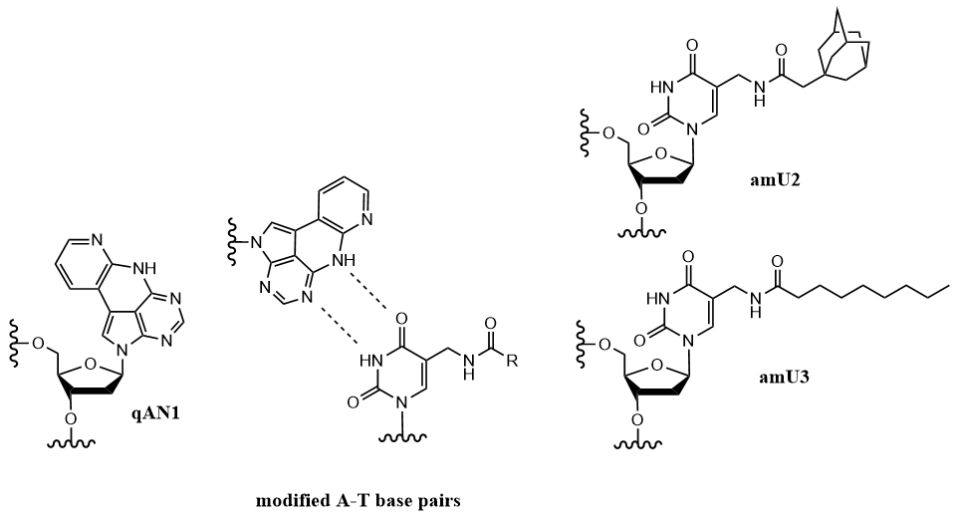



**Figure 3.3.** Structure of qAN1 and H-bonding between qAN1 and 5-AM-dU analogue base pairs.

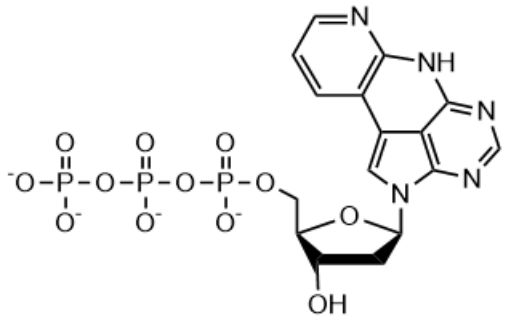
Since qAN1 triphosphate was found to be difficult to incorporate into the template in by PCR amplification under standard conditions (**Figure 3.5**), several changes in PCR conditions will be required during the optimisation tests. These may include increasing elongation temperature, increasing elongation time, lowering the concentration of $Mg^{2+}$ ($Mg^{2+}$ stabilises the duplex formed between the extended primer and the template[139]), adding DMSO as an duplex denaturant, and, for Phusion DNA polymerase, changing the buffer from HF to GC.

Two DNA polymerases were selected for this project, namely Phusion (with a standard 5X Phusion HF Buffer as well as 5X Phusion GC Buffer which can be used for complex or GC-rich templates) and Vent (exo-). Both of these polymerases are reported to exhibit high thermostability and favourable properties for resolving GC-rich sequences.[179]

## 3.2 PCR and qPCR analysis of qAN1 triphosphate

### 3.2.1 PCR using qAN1 triphosphate

The qAN1 nucleoside was supplied by Moa S. Wranne from Prof. Marcus Wilhelmsson's laboratory at Chalmers University of Technology, Sweden and synthesis of its triphosphate was carried out at ATDBio Ltd. The qAN1 triphosphate (**Figure 3.4**) was quantified by UV-vis spectroscopy and then dissolved in water to form a 2 mM solution for PCR amplification reactions. PCR amplification of qAN1 triphosphate was conducted on the same template (Res8470, 81-mer, in **Table 2.1**) and primers (Res8471 and Res 8472, 18-mer, in **Table 2.1**) which had been used for previous PCR amplifications on modified 5-AM-dUTPs. PCR amplification of qAN1 triphosphate was attempted in the presence of dCTP, dGTP and dTTP and in the absence of dATP.




qAN1
fluorescent adenine analogue

**Figure 3.4.** Structure of qAN1 triphosphate.

An initial attempt to incorporate qAN1 triphosphate into DNA by PCR amplification proved unsuccessful under normal PCR conditions (**Figure 3.5**). This failure of PCR

amplification with qAN1 triphosphate is not surprising given the duplex stabilising properties of qAN1.

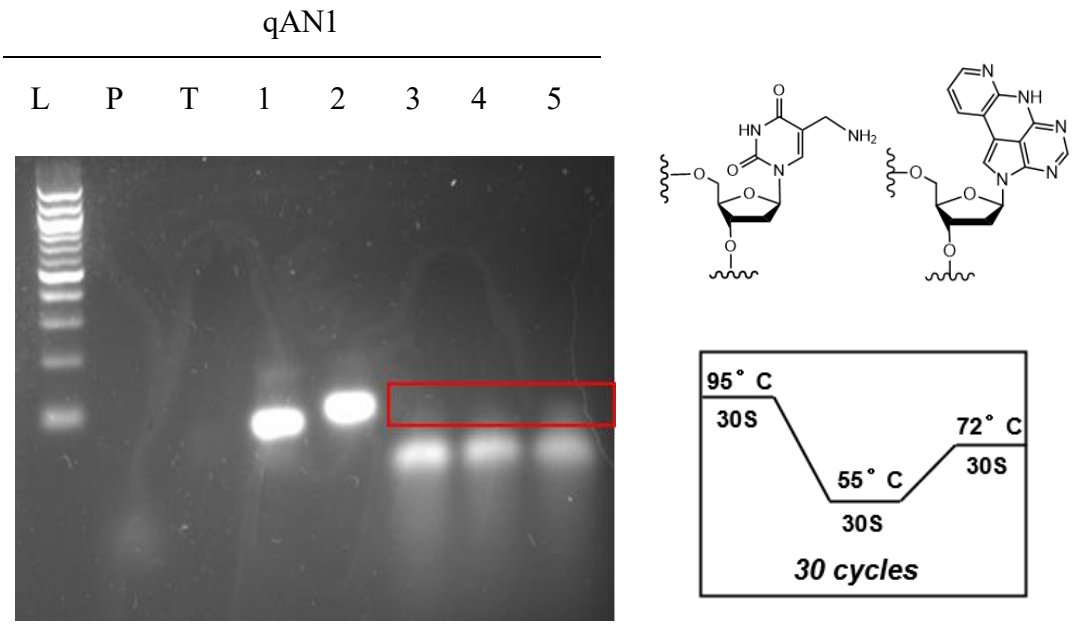



**Figure 3.5.** Failure of incorporation of qAN1 triphosphate into DNA via PCR amplification (see red box) under normal conditions. Left-hand panel: Lane L: 100 bp DNA ladder; Lane N: primers only; Lane T: template only; Lane 1: positive control by addition template and all natural dNTPs (dATP + dCTP + dGTP + dTTP) using Vent (exo-); Lane 2: addition of amU1 together with template and all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-); Lane 3: addition of qAN1 with template and all other dNTPs (dCTP + dGTP + dTTP) using Vent (exo-); Lane 4: addition of qAN1 with template and all other dNTPs using Phusion; Lane 5: addition of qAN1 with template and all other dNTPs using Phusion in the presence of GC buffer plus 2μL DMSO. Right-hand panel: Structure of nucleoside on amU1 and qAN1 (top) and conditions of PCR amplification reaction (bottom).

Several attempts at optimisation were made with respect to the reaction conditions for PCR amplification using qAN1. A range of PCR annealing temperatures were tested with Vent (exo-) DNA polymerase but these had little or no impact. Changing DNA polymerases, the addition of DMSO and decreasing the concentration of $Mg^{2+}$ were all

tried but were unsuccessful (**Figure 3.6**). Rather than the desired DNA PCR product containing qAN1, the resulting bands were characterised as being only primer dimers formed through non-specific amplification (highlighted in yellow box **Figure 3.6**).

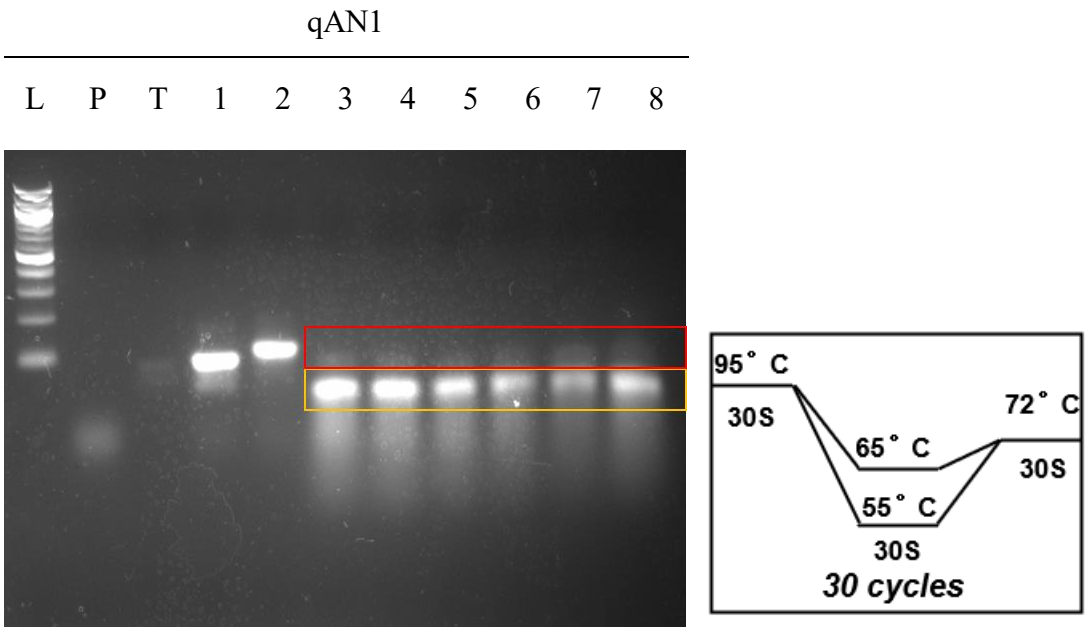



**Figure 3.6.** Various attempts to incorporate qAN1 into the template via PCR amplification under mild conditions with different DNA polymerases. No significant band appeared in expected area (red box) with only primer dimers formed (yellow box). Left-hand panel: Lane L: 100 bp DNA ladder; Lane N: primers only: Lane T, template only: Lane 1, positive control with addition of template and all natural dNTPs using Vent (exo-) with an annealing temperature of 60 °C; Lane 2: addition of $dX_1TP$ together with template and all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-) with an annealing temperature of 60 °C; Lane 3: addition of qAN1 with template and all other dNTPs using Vent (exo-) with an annealing temperature of 60 °C; Lane 4: addition of qAN1 with template and all other dNTPs using Vent (exo-) with an annealing temperature of 55 °C; Lane 5: addition of qAN1 with template and all other dNTPs using Vent (exo-) with an annealing temperature of 65 °C; Lane 6: addition of qAN1 with template and all other dNTPs (dCTP + dGTP + dTTP) using Phusion with an annealing temperature of 60 °C; Lane 7: addition of qAN1 with template and all other dNTPs using Phusion but with a reduced amount (500 µM) of $MgCl_2$ added with an annealing temperature of 60 °C; Lane 8: addition of qAN1 with template and all other dNTPs using Phusion but with GC buffer and the addition of 2µL DMSO with an annealing temperature of 60 °C. Right-hand panel: Conditions of PCR amplification reaction with different annealing temperatures used in these reactions.

The failure of these further attempts under moderate conditions with different DNA polymerases indicate that varying only one or two conditions from standard PCR protocols is not adequate to successfully incorporate qAN1 into DNA through PCR amplification. Therefore, a series of more extreme PCR conditions were tested.

The incorporation of qAN1 into the template through PCR amplification was finally achieved under more extreme conditions after a series of optimisation tests. For each PCR cycle the denaturing temperature was increased from 95 °C to 98 °C, the denaturing time also elongated from 30 s to 60 s; the annealing and extension temperatures varied from 55 °C to 65 °C. The relationship found between the increase in PCR extension temperature, and the increased efficacy of qAN1 triphosphate incorporation supports the hypothesis that the high stability of DNA duplexes containing qAN1 explains the low PCR efficiency. Another finding is that qAN1 triphosphate can be incorporated into the template using Phusion polymerase within its GC buffer but not using the HF buffer (**Figure 3.7**, lanes 6-8). This further supports the hypothesis that factors that improve the performance of templates containing GC-rich regions should also work for qAN1.

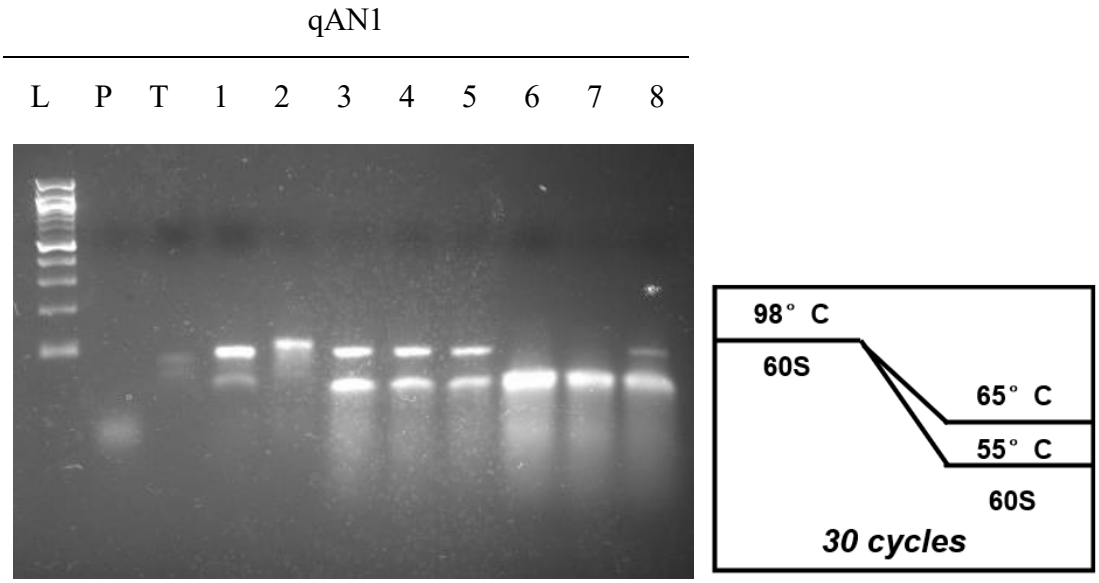



**Figure 3.7.** Successful incorporation of qAN1 into DNA through PCR amplification under extreme conditions with different DNA polymerases. Left-hand panel: Lane L: 100 bp DNA ladder; Lane N: primers only; Lane T: template only; Lane 1: positive control with addition of all natural dNTPs using Vent (exo-) with annealing temperatures of 60 °C; Lane 2: addition of template, amU1 together with all other dNTPs (dATP + dCTP + dGTP) using Vent (exo-) with an annealing temperature of 60 °C; Lane 3: addition of qAN1 with template, all other dNTPs (dCTP + dGTP + dTTP) using Vent (exo-) with an annealing temperature of 60 °C; Lane 4: addition of template, qAN1 with all other dNTPs using Vent (exo-) with an annealing temperature of 55 °C; Lane 5: addition of qAN1, template, with all other dNTPs using Vent (exo-) with an annealing temperature of 65 °C; Lane 6: addition of qAN1, template, with all other dNTPs using Phusion with an annealing temperature of 60 °C; Lane 7: addition of qAN1 with template, and all other dNTPs using Phusion + 2μL DMSO and an annealing temperature of 60 °C; Lane 8: addition of qAN1 with template and all other dNTPs using Phusion + GC buffer + 2μL DMSO with an annealing temperature of 60 °C. Right-hand panel: conditions of PCR amplification with very high denaturing and extension temperatures and longer annealing times.

Two different DNA polymerases, namely Phusion and Vent (exo-), were compared in terms of their efficiency in incorporating qAN1 triphosphate under these extreme reaction conditions (**Figure 3.8**). Although both polymerases to some degree successfully incorporated qAN1 triphosphate into the template through PCR (but still

gave rise to a major primer-dimer band), Vent (exo-) was much more effective than Phusion (with GC buffer and DMSO).

The final optimised protocol for qAN1 triphosphate incorporation via the PCR amplification reaction was achieved using vent (exo-) and annealing and extension temperature of 60 °C. The details of this protocol are listed in the experimental section (page 226).

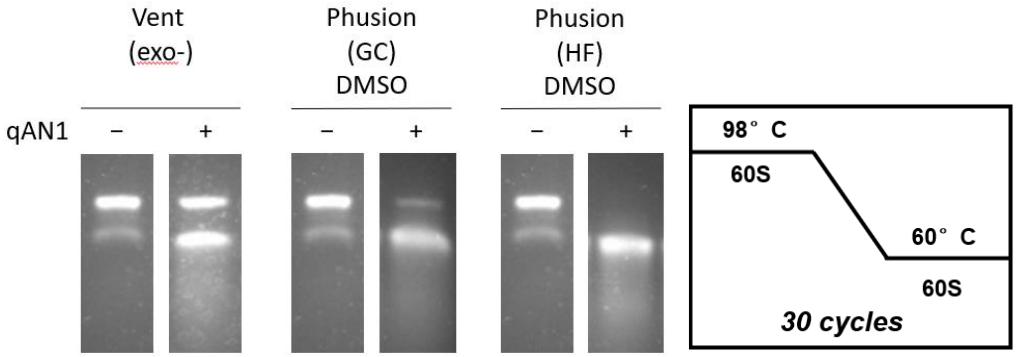



**Figure 3.8.** Comparison of two different DNA polymerases in terms of the successful incorporation of qAN1 into a DNA template (left-hand panel) under extreme PCR conditions (right-hand panel). Left-hand panel: – mean positive control with addition of all natural dNTPs; + mean addition of qAN1, template, with all other dNTPs.

### 3.2.2 Real-time PCR (qPCR) using qAN1 triphosphate

Real-time PCR amplification of qAN1 triphosphate using Vent (exo-) was conducted under extreme conditions relative to a standard PCR protocol (illustrated in **Figure 3.11**). For each PCR cycle the denaturing temperature was increased from 95 °C to 98 °C, the denaturing time also elongated from 30 s to 60 s; the annealing and extension temperatures varied from 55 °C to 65 °C.

Several conditions using different concentrations of template and primers for real-time PCR amplifications were analysed (**Figure 3.9**). A $C_q$ value for qAN1 triphosphate incorporated into a 4 nM template with 0.2 μM primers was 29.82. $C_q$ values obtained for two different optimisation condition reactions, each using 20 nM template and 0.4 μM primers, were 13.94 and 20.77, respectively. Although $C_q$ values were lowered by changing the concentration of template and primers, the fluorescence of the EvaGreen dye did not increase significantly. As the fluorescent EvaGreen dye is activated only upon binding to double-stranded DNA oligonucleotides, the low fluorescence emission indicates that, although DNA polymerases can recognise qAN1 triphosphate and incorporate it into the template during real-time PCR amplification, there is inhibition of the fluorescent signal, perpaps caused by a limited amount of PCR product or quenching of the EvaGreen dye.

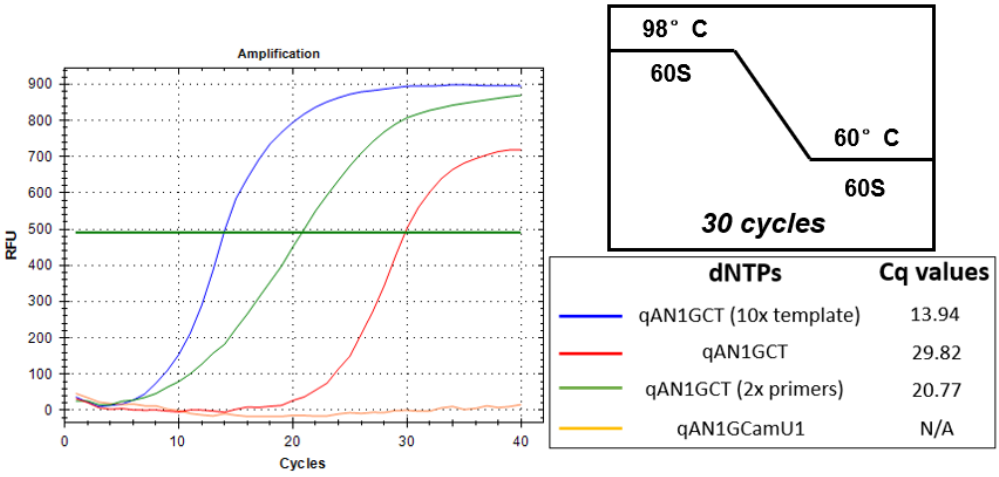



**Figure 3.9.** Real-time PCR reaction outcomes for qAN1 using Vent (exo-) DNA polymerase under extreme conditions in the presence of naturally occurring dNTPs and amU1. The table on the right-hand side illustrates, for each reaction labelled with different colours, which dNTPs were used along with their corresponding $C_q$ values. 10x template = 40nM, in the other reactions template concentration = 4 nM. The reaction of using Vent (exo-) DNA polymerase in the presence of both qAN1 and amU1 has failed (orange) probably because the reaction condition that works for qAN1 is too extreme for amU1.

# 3.3 PCR and qPCR analysis of pairing qAN1 and 5-AM-dUTPs

## 3.3.1 PCR for pairing qAN1 and 5-AM-dUTPs

After successfully incorporating qAN1 triphosphate into the PCR template using extreme conditions, modified 5-AM-dUTPs with strong hydrophobic groups (amU2 and amU3) were also introduced to the PCR reaction along with qAN1 triphosphate and other dNTPs. These PCR amplifications in which qAN1 triphosphate was paired with modified 5-AM-dUTP were attempted in the presence of dCTP and dGTP and the absence of dATP and dTTP.

The PCR amplification was first attempted (here we used the same sequences of primers and template that used for previous 5-am-dUTPs' PCR amplification experiments, the detailed sequences information can be found on table 2.1 on page 62) under the extreme conditions that allowed qAN1 to be incorporated into the template efficiently. However, the subsequent reaction failed, as none of 5-AM-dUTPs added with qAN1 triphosphate were present in a fully extended template after PCR amplification (**Figure 3.10**, lane 4-6) and only adding qAN1 triphosphate along with natural dTTP was successful (**Figure 3.10**, lane 3). This probably occurs because the PCR condition is too extreme for modified 5-AM-dUTPs to be incorporated together with the template, as a positive control for PCR amplification using only amU1 alongside with other natural dNTPs did not work effectively under these conditions (**Figure 3.10**, lane 2).

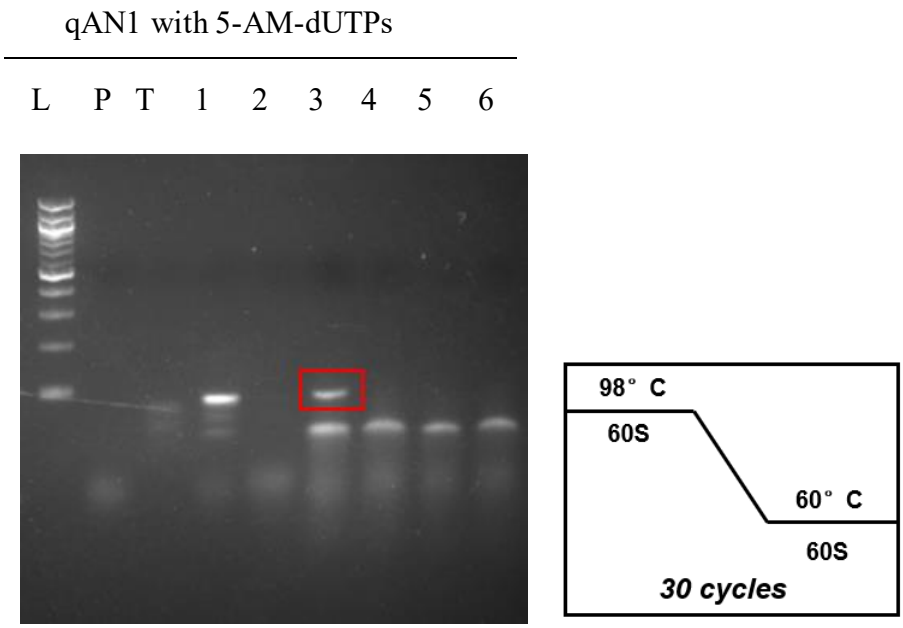



**Figure 3.10.** Failure to incorporate both qAN1 and 5-AM-dUTPs into the template through PCR amplification under extreme conditions using Vent (exo-) DNA polymerase. Left-hand panel: Lane L: 100 bp DNA ladder; Lane N: primers only; Lane T: template only; Lane 1: positive control with addition of template, primer and all naturally occurring dNTPs; Lane 2: addition of amU1 together with template, primer and all other dNTPs (dATP + dCTP + dGTP); Lane 3: addition of qAN1 with template, primer and all other dNTPs (dCTP + dGTP + dTTP); Lane 4: addition of template, primer and qAN1 + amU1 with all other dNTPs (dCTP + dGTP); Lane 5: addition of template, primer and qAN1 + amU2 with all other dNTPs (dCTP + dGTP); Lane 6: addition of template, primer and qAN1 + amU3 with all other dNTPs (dCTP + dGTP). Right-hand panel: Extreme conditions of PCR amplification used in this reaction.

This approach revealed a problem in that both 5-AM-dUTPs and qAN1 triphosphate have their own specific requirements in terms of PCR conditions for effective amplification. However, these two optimised conditions are mutually exclusive in that the PCR conditions that worked for qAN1 triphosphate did not do so for 5-AM-dUTPs (**Figure 3.10**) and *vice versa* (**Figure 3.6**). However, more moderate PCR conditions that could allow both forms of modified triphosphates to be functionally affected during

PCR amplification must be found if these two analogues are to be added concomitantly.

Based on previous studies using 5-AM-dUTPs and qAN1 triphosphate, the tolerance of PCR conditions is more restricted for qAN1 triphosphate than 5-AM-dUTPs, as most standard PCR conditions are inefficient in incorporating qAN1 triphosphate into the template. Therefore, in total, four divergent moderated conditions have been designed that are primarily based on qAN1's requirement for extreme conditions in PCR amplification (**Figure 3.11**). These four conditions gradually decrease the temperature and duration of extension for each cycle to a standard value and normalise the denaturing temperature from 98 °C back to a more conventional 95 °C.

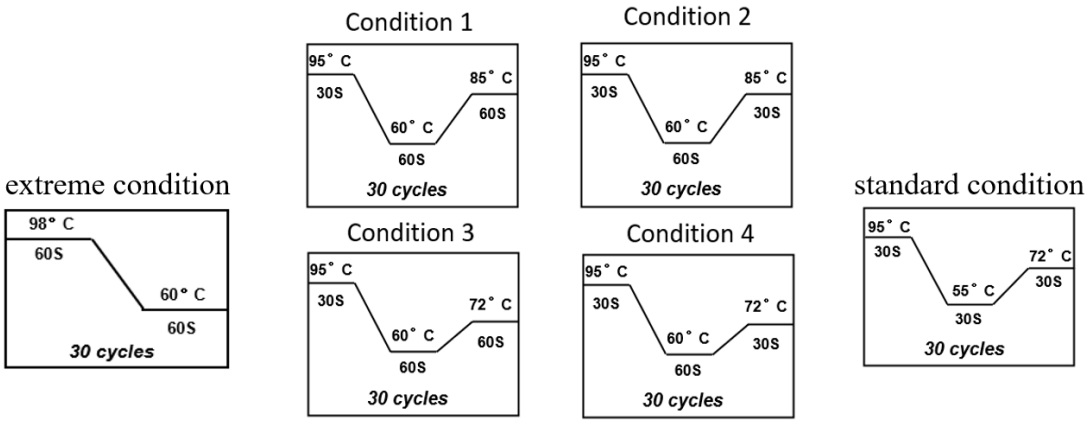



**Figure 3.11.** Four conditions used for incorporating both qAN1 and 5-AM-dUTPs into the DNA template during PCR amplifications. The extreme condition is the applied PCR condition only works for qAN1 triphosphate; the standard condition is a general PCR condition that tested to be applied for 5-AM-dU derivatives; and in the middle are four sets of PCR conditions tested for simultaneously incorporating qAN1 and 5-am-dUTPs derivatives. The temperature and time of the extension step were gradually lowered from extreme to standard along with these four PCR conditions.

By reducing both the extension and denaturing temperatures, all 4 moderated conditions were found to be successful in terms of incorporating both amU2 (**Figure 3.12**), amU3 (**Figure 3.13**) and qAN1 triphosphate into the template during PCR amplification. Furthermore, by incorporating together with amU2 or amU3 triphosphate in place of natural dTTP, the PCR efficiency of qAN1 triphosphate was found to be improved as it now could be functionalised under milder PCR conditions in the presence of amU2 or amU3 (**Figure 3.14**).

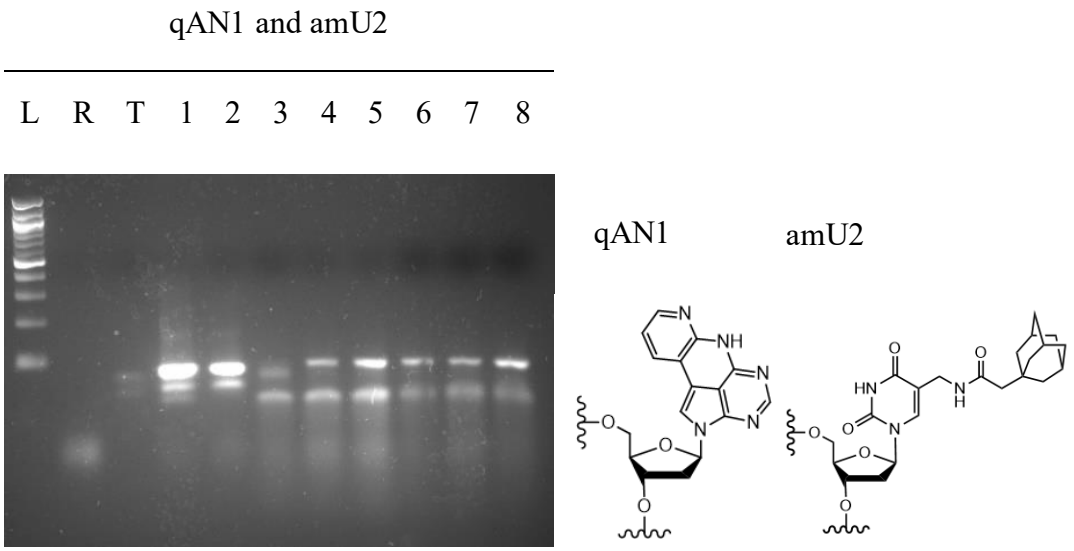



**Figure 3.12.** A successful incorporation of both qAN1 and amU2 into the DNA template via PCR amplification under 4 different reaction conditions using Vent (exo-). Left-hand panel: Lane L: 100 bp DNA ladder; Lane N: primers only; Lane T: template only; Lane 1: positive control with addition of template, primer and all natural dNTPs under condition 2; Lane 2: addition of amU2 together with template, primer and all other dNTPs (dATP + dCTP + dGTP) under condition 4; Lane 3: addition of qAN1 with template, primer and all other dNTPs (dCTP + dGTP + dTTP) under condition 1; Lane 4: addition of template, primer and qAN1 + amU2 with all other dNTPs (dCTP + dGTP) under condition 1; Lane 5: addition of template, primer and qAN1 + amU2 with all other dNTPs under condition 2; Lane 6: addition of template, primer and qAN1 + amU2 with all other dNTPs under condition 3; Lane 7: addition of template, primer and qAN1 + amU2 all other dNTPs under condition 4; Lane 8: comparison with addition of template, primer and qAN1 + amU2 with all other dNTPs under condition 2 but with Vent (exo+). Right-hand panel: structure of nucleosides qAN1 and amU2.

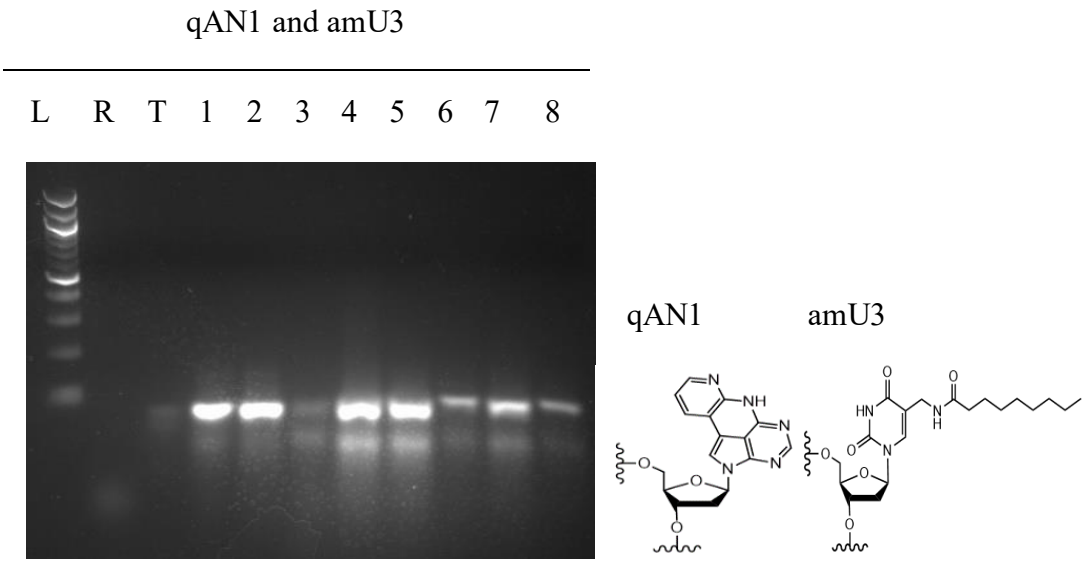



**Figure 3.13.** Successful incorporation of both qAN1 and amU3 to the DNA template through PCR amplification with 4 divergent reaction conditions using Vent (exo-). Left-hand panel: Lane L: 100 bp DNA ladder; Lane N: primers only; Lane T, template only; Lane 1: positive control with template, primer and addition of all natural dNTPs under condition 2; Lane 2: addition of amU3 together with template, primer and all other dNTPs (dATP + dCTP + dGTP) under condition 4; Lane 3: addition of qAN1 with template, primer and all other dNTPs (dCTP + dGTP + dTTP) under condition 1; Lane 4: addition of template, primer and qAN1 + amU3 with all other dNTPs (dCTP + dGTP) under condition 1; Lane 5: addition of template, primer and qAN1 + amU3 with all other dNTPs under condition 2; Lane 6: addition of template, primer and qAN1 + amU3 with all other dNTPs under condition 3; Lane 7: addition of template, primer and qAN1 + amU3 with all other dNTPs under condition 4; Lane 8: comparison with addition of template, primer and qAN1 + amU3 with all other dNTPs under condition 2 with Vent (exo+). Right-hand panel: structure of nucleosides qAN1 and amU3.

Further studies of the PCR efficiency of qAN1 triphosphate with amU2 and amU3 suggest that, although both amU2 and amU3 improve the PCR efficiency of qAN1 triphosphate, amU3 with a nonanoyl side chain added together with qAN1 triphosphate result in a slightly improved PCR efficiency than for amU2 which contains an adamantanyl side chain (**Figure 3.15**).

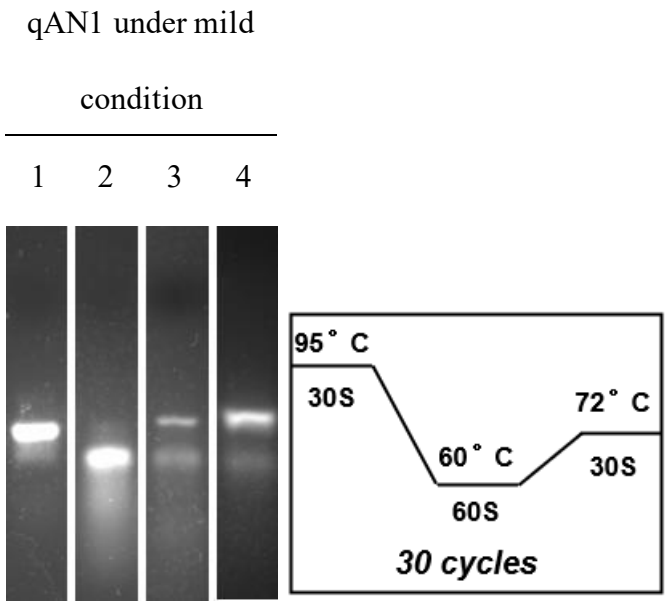



**Figure 3.14.** Comparison of incorporation of qAN1 into the template accompanied by natural dTTP, amU2 and amU3 during PCR amplification (lower bands are identified to be primer dimers). Left-hand panel: Lane 1: positive control with addition of template, primer and naturally occurring dNTPs; Lane 2: addition of qAN1 with template, primer and all other dNTPs (dCTP + dGTP + dTTP); Lane 3: addition of template, primer and qAN1 + amU2 with all other dNTPs (dCTP + dGTP); Lane 4, addition of template, primer and qAN1 + amU3 with all other dNTPs. Right-hand panel: Mild PCR conditions used.

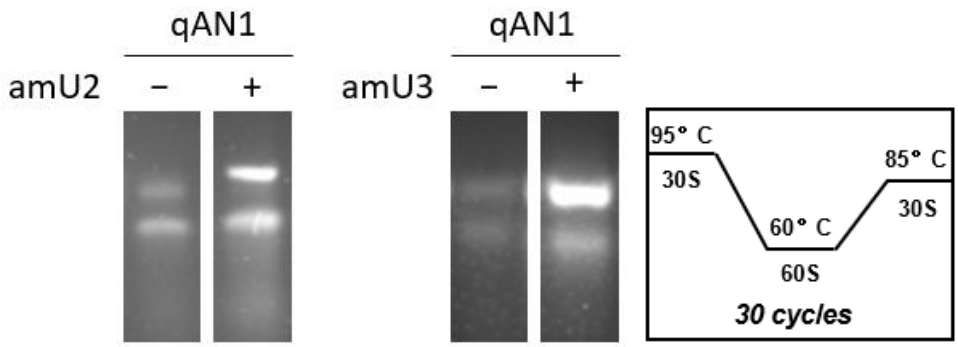



**Figure 3.15.** Incorporation of qAN1 into the DNA template accompanied with amU2 or amU3 during PCR amplification improved PCR efficacy (left-hand panel) under relatively mild PCR conditions (right-hand panel).

### 3.3.2 Real-time PCR (qPCR) for pairing qAN1 and 5-AM-dUTPs

A real-time PCR amplification was performed which incorporated both qAN1 triphosphate and 5-AM-dUTPs into the DNA template under moderate PCR conditions (based on PCR condition 1 in **Figure 3.11**). The general protocols for qPCR using modified 5-am-dUTPs and qAN1 are listed in the experimental section (page 228).

Only reactions of amU2 and amU3 that contain hydrophobic side chains added together with qAN1 triphosphate were successfully incorporated into a template with a good yield (**Figure 3.16**). The addition of other 5-AM-dUTPs did not help qAN1 triphosphate in its incorporation with the template. $C_q$ values for binding amU2 and amU3 with qAN1 triphosphate were 8.44 and 9.04, respectively, values that are comparable to the control group in which amU1 was added alone or with other natural dNTPs ($C_q$ value 8.58). Under this moderated condition the value even approached that of natural dNTPs ($C_q$ value 7.39).

Although this reaction was performed under moderate condition which, in theory, should be better for qAN1 triphosphate's incorporation, this result is still very close to optimal when considering the great improvement in $C_q$ values between adding qAN1 alone with other dNTPs (33.60) and both 5-AM-dUTPs (amU2 or amU3) concurrently with qAN1 triphosphate. These real-time PCR reactions further prove the hypothesis that, by incorporating with amU2 or amU3, which contain DNA duplexes that produce destabilising effects in the absence of natural dTTP, the PCR efficiency of qAN1 triphosphate will be strongly improved.

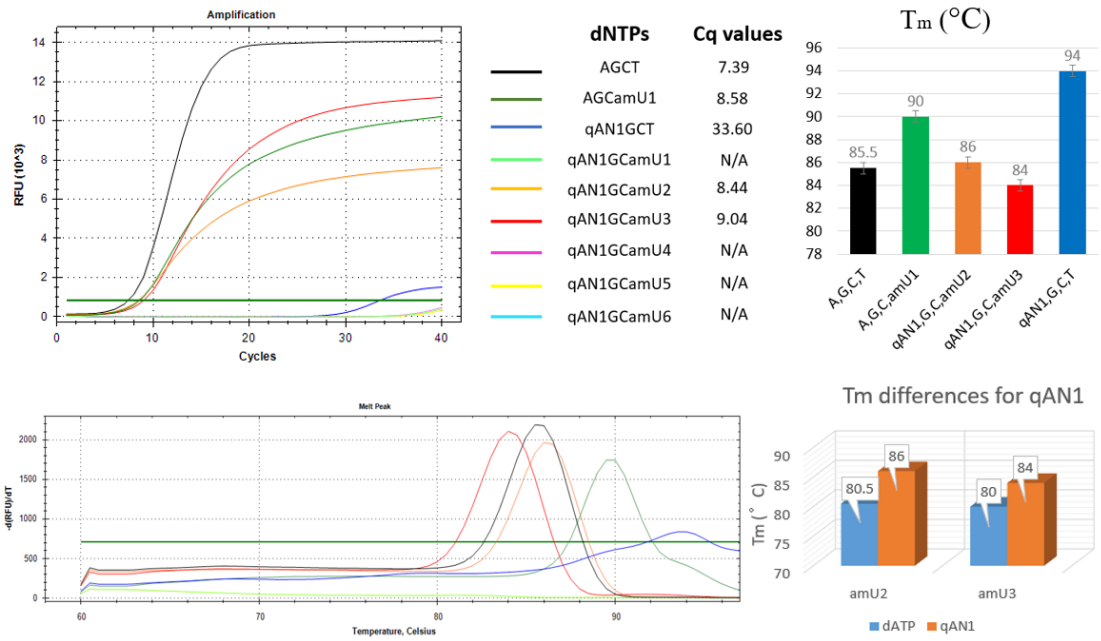



**Figure 3.16.** Real-time PCR reaction for qAN1 incorporated into a template together with natural dNTPs and 5-AM-dUTPs as substituents of dTTP using Vent (exo-). The table illustrates which dNTPs were used along with their respective $C_q$ values for each colour-coded reaction. The bar chart shows the Tm values for qAN1 with 5-AM-dUTP triphosphate pairs (top-right panel) and Tm differences between amU2 and amU3 with/without qAN1 triphosphate (bottom right).

## 3.4 Conclusions

In summary, the problem of low PCR efficiencies of incorporation of a strongly duplex-stabilising quadracyclic fluorescent adenine analogue, qAN1 triphosphate, was studied and it was shown that this modified triphosphate could only be incorporated with a DNA template under very extreme PCR conditions. Real-time PCR studies were also consistent with this observed difficulty relative to naturally occurring dNTPs. This low PCR efficiency is believed to be caused by its stabilising effect on DNA duplexes, as the problem faced when using qAN1 is found to be similar to those encountered when

using GC-rich DNA sequences in PCR amplification. As long as several changes in PCR conditions are met, including strongly increasing denaturing temperatures from 95 °C to 98 °C; increasing denaturing times from 30 s to 60 s; lowering the concentration of $Mg^{2+}$; adding DMSO as an environmental disruptor; and, for Phusion DNA polymerase, changing its buffer solution from HF to GC, the PCR efficiency of successfully incorporating qAN1 increased even though it was still not completely efficacious.

A new method has been generated based on the hypothesis that, by introducing a modified complementary nucleobase with duplex destabilising effects, it will effectively reduce/neutralise the stabilising effects of the qAN1 molecule upon DNA duplexes with moderate melting temperatures and hence improve its PCR efficiency. Two 5-AM-dUTPs with large hydrophobic groups (amU2 and amU3, whose synthesis was described in Chapter 2) were selected as modified dTTPs to test their availability in pairing with qAN1 triphosphate (as a modified dATP) in PCR amplification reactions.

PCR amplification studies under different conditions suggest that, by incorporating alongside amU2 or amU3 triphosphate instead of naturally occurring dTTP, the PCR efficiency of qAN1 triphosphate was found to be improved as it is now functionalised across a wider range of mild PCR conditions in the presence of amU2 or amU3. Real-time PCR reactions also reported significant changes in $C_q$ values from 30.60 to 8.44 or 9.04, respectively. These studies of the incorporation of qAN1 together with these four 5-AM-dUTPs serving as base pairs, reveal that two of these 5-AM-dU analogues (amU2 and amU3) that contain large hydrophobic functional groups dramatically

overcome the problem of low PCR efficiencies, whereas the other 5-AM-dUTPs did not offer any improvement in terms of PCR efficiency. This result, on the one hand, has proved our previous hypothesis of the stabilising effect in causing the low PCR efficiency of qAN1 while, on the other, has also illustrated the concept of introducing a pair of modified nucleotides each with opposing biophysical properties into DNA strands in order to neutralise their respective side effects upon biochemical processes.

The successful introduction of modified 5-AM-dU analogues with destabilising effects into the qAN1 molecule has not only solved the problem of low PCR efficiency for qAN1 moieties, but has also allowed us to explore a novel method for generating highly modified aptamers with a potentially greater avidity and tighter binding by simultaneously introducing a chemically modified DNA amplicon with full substitution of two nucleobases through PCR amplification. The successful incorporation of both a modified dTTP and a modified dATP concurrently into the template within a PCR amplification proves a wide range of applications in SELEX approaches for modified 5-AM-dU analogues as synthesised in Chapter 2 as well as further 5-AM-dU analogues containing side chains with different functional groups. These could, in turn, be designed based on the functional requirements and characteristics of other modified nucleobases introduced into the system.

# *Chapter 4.*

# *Incorporation of modified 5-aminomethyl-dU Analogues into SELEX approaches*

# 4.1 Introduction and Aims

## 4.1.1 Introducing modified nucleobases to target selection

Aptamers, compared to protein-based antibodies, contain fewer building blocks as they are composed of natural nucleic acids. They possess a polyanionic backbone with a relatively hydrophilic character and thus present a far more limited repertoire of functional groups available for target recognition. This limited chemical diversity in terms of nucleic acid libraries has constrained the utility of aptamers, not only in research, but also in diagnostic and therapeutic applications.[180] Given that, a lack of functional group diversity has long been a primary limitation in aptamer research in terms of increasing their chemical diversity and specificity.[181,182] The successful incorporation of 5-AM-dUTPs into DNA oligonucleotides through PEX and PCR amplification (as demonstrated in Chapter 2) positions 5-AM-dUTP derivatives as potential analogues for *in vitro* selection through SELEX. By introducing chemically modified 5-AM-dUTP analogues into SELEX approaches, nucleobases containing a unique hydrophobic/aromatic pattern for target recognition could be used as new components in what was previously only a four-letter genetic alphabet. By this approach we aim to expand the chemical diversity of aptamers and discover high-affinity ligands for protein targets through SELEX approaches.

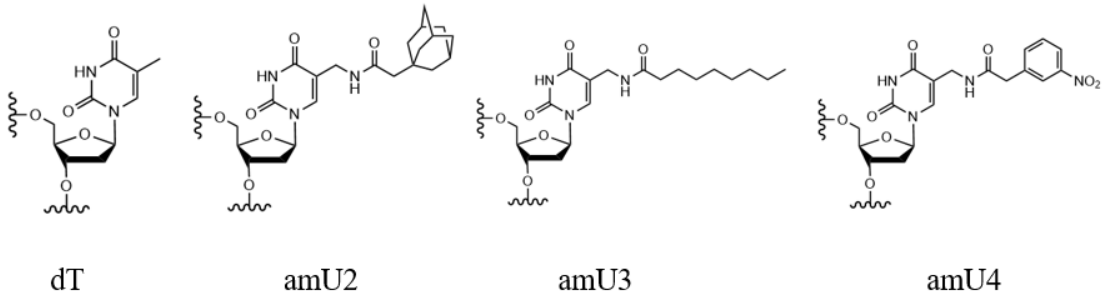



**Figure 4.1** Structure of Thymidine (dT) and three 5-AM-dU analogues (amU2 – amU4) used as nucleobase building blocks in aptamer selections.

Two of the 5-AM-dU analogues contain hydrophobic functional groups (amU2 and amU3) and can be fully substituted into oligonucleotide sequences together with qAN1. These two analogues were thus retained for this project for further SELEX-based selections. Another 5-AM-dU analogue selected was amU4 and this contains a nitrophenol functional group and can thus be used to study the effect of aromatic modifications in SELEX. Since the selection process is very time-consuming, the other three synthesised 5-AM-dU analogues (amU1 which is structurally similar to natural dTTP; and amU5 and amU6 which are structurally related to amU4) were not employed in this project.

### 4.1.2 Design of a 5-AM-dU ssDNA library

The design of this particular nucleic acid library is an important starting point for SELEX approaches. Many of the success criteria for aptamers, including binding affinities, nuclease resistance, and the relative convenience of chemical modifications, are largely affected by library design.[183] Therefore, a well-designed nucleic acid library has a great impact on the overall efficiency of the selection.

The ssDNA library for SELEX approaches will comprise a series of random regions (labelled N bases) in the middle and two fixed primer-binding sites at either end of an oligonucleotide. The sequence space and structural diversity are both related to the length of the random region. The theoretical diversity of a random region of N nucleotides is $4^N$ (e.g., a random 28-mer oligonucleotide library will theoretically have a diversity of $4^{28} \sim 7 \times 10^{16} \approx 0.1$ μmol of ssDNA).[184] Longer sequences are more likely to fold into more complex structures that may be needed to form a target binding domain. However, they are unable to cover all the possible sequences extensively within their respective library. Thus, a balance should be kept between the diversity of sequences and the complexity of tertiary structures formed by these sequences. For the *in vitro* selection of aptamers, nucleotides with randomised regions of 30−50-nt in length are commonly considered to be the most abundant.[185] For this project, the designed oligonucleotides contained a random region N which was set to be around 40-nt in length associated with two fixed 20-mer primer-binding sites at either end of the variable region (**Table 4.1**). The sequences of two primer-bind sites were designed

carefully to avoid self-association or secondary structure formation and to ensure efficient polymerase extension.[186]

| | Sequences (5′ → 3′) | |
|---|---|---|
| P1 (forward) | FAM-AGCAGCACAGAGGTCAGATG | 20mer |
| P2 (reverse) | Poly-A(20mer)-HEG-TTCACGGTAGCACGCATAGG | 40mer |
| Original Template | Poly-A(20mer)-HEG-AGCAGCACAGAGGTCAGATG-N40-CCTATGCGTGCTACCGTGAA | 100mer |
| Desired Template | FAM-AGCAGCACAGAGGTCAGATG-N40-CCTATGCGTGCTACCGTGAA (contains modified 5-AM-dUTPs instead of dTTP) | 80mer |

**Table 4.1** Schematic representation of ssDNA library generation using the size separation method followed by denaturing PAGE. The original library was composed of a 40-mer randomised region with a HEG and poly-A tail at the 5′-end. A FAM-labelled 20-mer primer and a 40-mer primer that also contains HEG and a poly-A tail were employed. The random region in the copied ssDNA library will contain an equal distribution of dA, dG, dC and dT/5-AM-dU analogues.

The capacity to generate ssDNA libraries from dsDNA products after each round of PCR amplification is another challenge for library design. As we demonstrated in chapter 1.6.2, it is ssDNAs with stable secondary and tertiary structures that binds to specific protein targets for selection in SELEX approaches. However, after each round

of the selection process, aptamers are amplified via PCR and thus create dsDNA forms. Therefore, the complementary DNA strands of each selected aptamer must be removed before they can be introduced into the next round of selection as a library pool. Various methods have been reported for the generation of a ssDNA library, including asymmetric PCR[187], lambda (λ) exonuclease digestion[92], streptavidin-biotin magnetic bead separation[188], and size separation followed by denaturing urea polyacrylamide gel electrophoresis (denaturing PAGE)[189]. In this project, size separation was selected as the most appropriate method for separating DNA strands after considering accessibility and time costs. Based on this method, one of the primers used in PCR amplification was attached to a spacer (an 18-carbon hexa-ethylene glycol, HEG) containing an extension of 20 nucleotides in length in the form of a string of adenosines (a 20-mer poly-A tail). Therefore, the two strands acquired from PCR amplification will possess different lengths (one strand possessing an extra HEG spacer with a poly-A tail and can thus be excised on a denaturing polyacrylamide gel (**Figure 4.2**)).

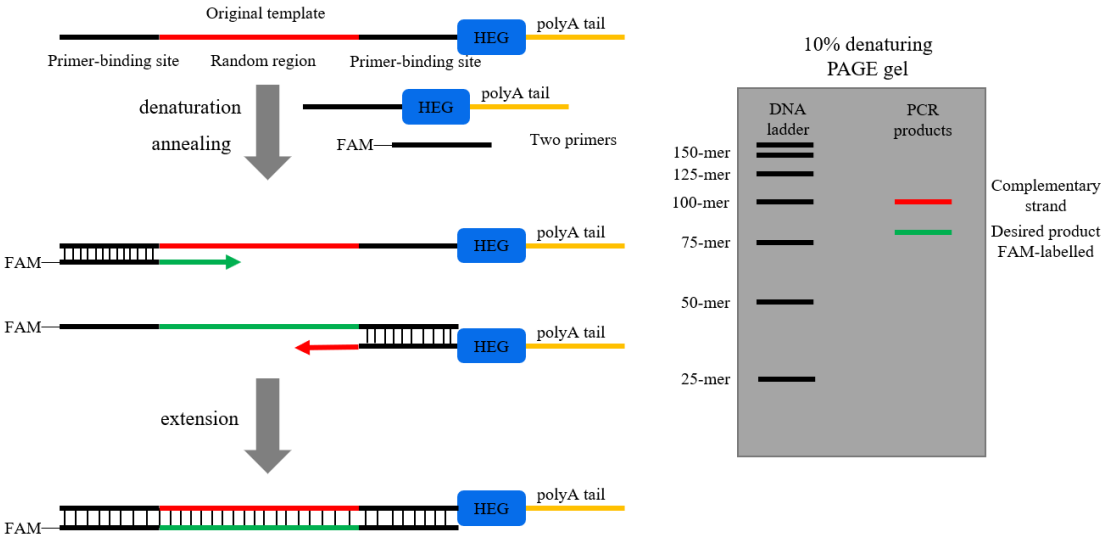



**Figure 4.2** ssDNA generation by PCR and size separation method using an 18-carbon spacer (HEG) and a 20-mer poly-A tail.

### 4.1.3 Target selection

Modified 5-AM-dUTP analogues, along with natural dNTPs, were used as nucleotide building blocks in order to examine their functions within SELEX processes. This could be achieved by comparing the differences between selected aptamers using natural dTTP and 5-AM-dUTPs. Therefore, a protein target that has been studied previously using aptamers comprising natural dNTPs was considered as a model for this project. After comparing several putative protein targets for aptamer selection, commercially available human epidermal growth factor 2 (also called HER2 or erbB-2) was selected as the target for aptamer selection incorporating 5-AM-dU analogues.[67,190–192] HER2 is a receptor tyrosine kinase and is found to be overexpressed in approximately 20−30% of all breast cancers.[193,194] Several aptamers targeting HER2 have been reported to date. Kim and Jeong published a 34-mer 2′-fluorine-modified RNA aptamer that was specific to the extracellular domain of the HER2 protein that possessed a binding affinity ($K_d$) of 3.5 nM.[195] They have since reported the generation of two further $^{18}$F-labelled HER2 specific aptamers for HER2 targeting.[193,196] In 2012, Liu *et al.* reported the generation of an 86-nucleotide DNA aptamer which bound to an epitope of HER2 with a $K_d$ of 18.9 nM,[194] while in 2013, Georg *et al.* discovered a trimeric version (42 nucleotides) of selected aptamers that also binds to the HER2 target.[197] Since HER2 has proven to be a good target for aptamer selection, it presents an ideal model target through which to study the properties of aptamers containing 5-AM-dU analogues.

During SELEX approaches, the protein targets are attached to a solid support during

each round of selection in order to separate target-bound oligonucleotide sequences from unbound sequences. The general method of using carboxylic acid magnetic beads as an immobilised solid support during the selection process was deemed to be too laborious. Thus a method employing a nitrocellulose membrane-based SELEX process was introduced for the purposes of this project to capture target-bound oligonucleotide sequences.[76,190] The selection was started by the incubation of recombinant HER2 protein within an ssDNA library. The recombinant was injected into an Eppendorf through a 0.45 µm nitrocellulose membrane filter (Millipore$^{TM}$) using a 1mL syringe (**Figure 4.3a**). The resulting ssDNA-HER2-bound complexes were captured within the nitrocellulose membrane and the selected ssDNA was eluted by heating the membrane in water at 95 °C for 15 min. An additional round of selection for negative controls is required when employing this method[198] as the ssDNA-HER2 complexes and some of ssDNA also bind to nitrocellulose membrane.[199] Thus, in a counter selection round (**Figure 4.3b**), recombinant bovine serum albumin (BSA) rather than HER2 was introduced into the ssDNA library and, after incubation, was first injected through the nitrocellulose membrane filter and only ssDNA eluent was collected for further incubation of recombinant HER2 protein with the ssDNA library. Through this counter selection process, oligonucleotide sequences that bound to BSA or the nitrocellulose membrane, but not specifically to the HER2 target protein, were eliminated from the ssDNA library.

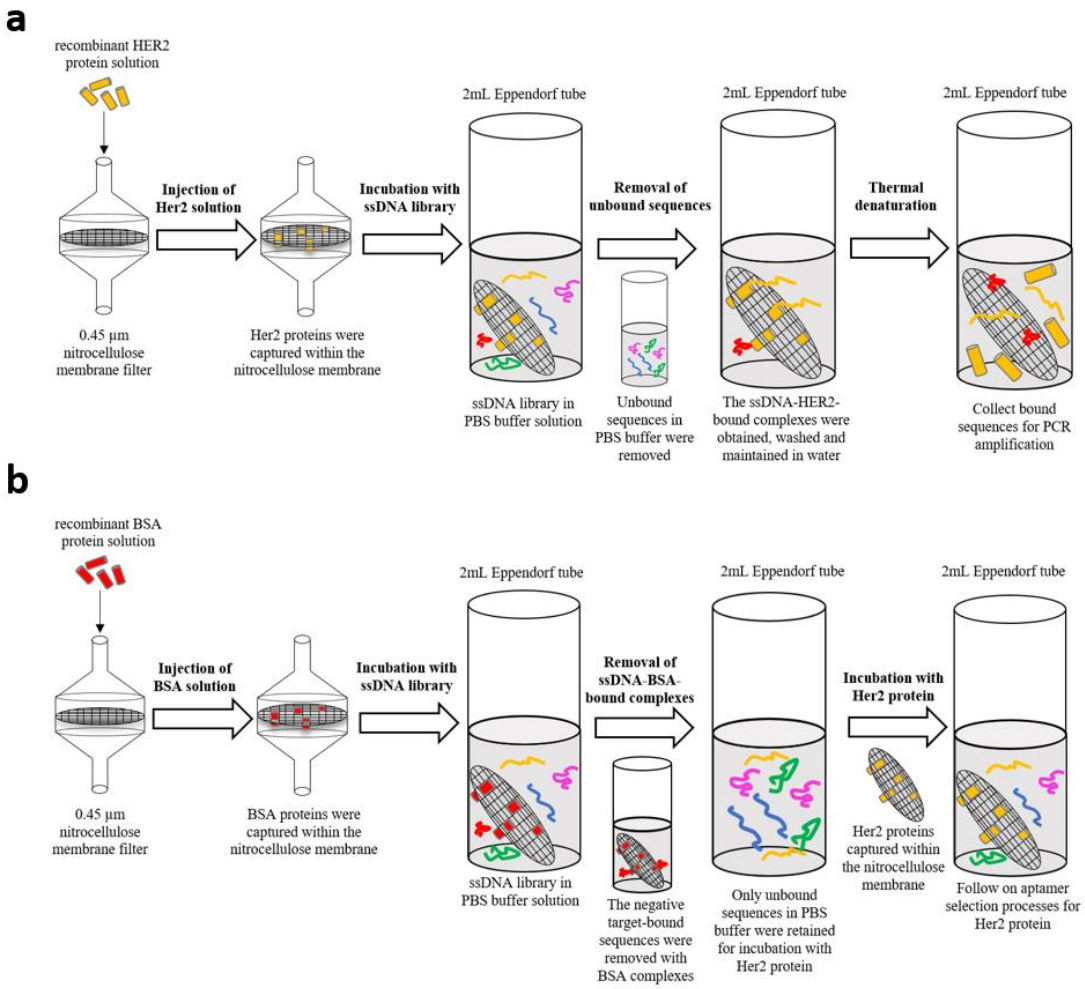



**Figure 4.3** Aptamer selection of Her2 protein target through a 0.45 µm nitrocellulose membrane filter.

**(a)** In each round of selection, aptamers that bound with Her2 proteins were collected and followed on a PCR amplification process; **(b)** In a counter selection round, recombinant bovine serum albumin (BSA) rather than HER2 was introduced into the ssDNA library as a negative control in order to remove all negative target-bound sequences attached to the membrane.

## 4.2 Selection of 5-AM-dU modified aptamers

### 4.2.1 ssDNA library generation

As outlined in section 4.1.2, an ssDNA original template library containing an extra HEG spacer and a poly-A tail was designed with two corresponding primer regions on each side. The desired oligonucleotide sequences that contained modified 5-AM-dU analogues within the randomised section are the complementary strands (80-mer labelled with FAM) to the original template strands (**Table 4.1**). Therefore, creation of the initial ssDNA library was achieved through a PCR amplification step followed by size separation of the resulting ssDNA amplicons using denaturing PAGE (**Figure 4.4**). The general protocols for size separation of resulting 80-mer ssDNA sequences through PCR are shown in the experimental section (page 227).

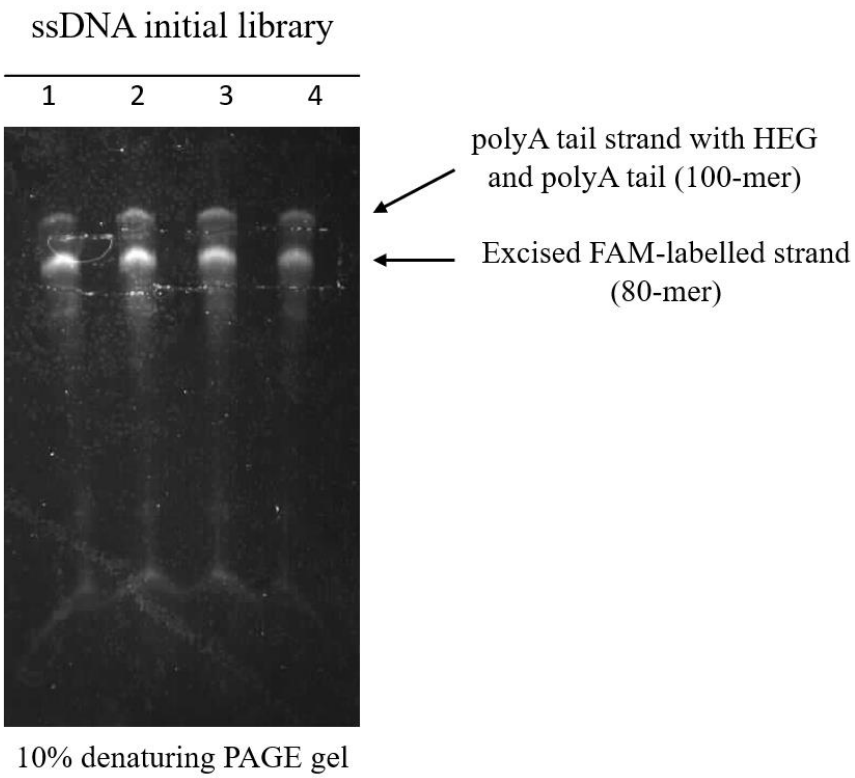



**Figure 4.4** Outcome of size separation protocol using 10% denaturing PAGE gel (stained with SYBR Gold) revealing separation of 80-mer FAM-labelled strands and 100-mer Poly-A tail strands after PCR amplification. Lane 1: strands of oligonucleotide sequences consisting of natural dNTPs; Lane 2: strands of oligonucleotide sequences consisting of amU2 in place of dTTP; Lane 3: strands of oligonucleotide sequences consisting of amU3 in place of dTTP; Lane 4: strands of oligonucleotide sequences consisting of amU4 in place of dTTP.

In total, four initial ssDNA libraries were constructed for this project. These enzymatically synthesised libraries of oligonucleotide sequences consisted of dA, dG, dC and dT (for library 1); amU2 (for library 2); amU3 (for library 3); amU4 (for library 4) as nucleobase 'building blocks' with equal distributions throughout the random region. An ssDNA library that contained only natural dNTPs (library 1) was initially used in the aptameric selection process to optimise the approach and to serve as a benchmark for the subsequent lateral selections of aptamers containing 5-AM-dU analogues (libraries 2 – 4).

### 4.2.2 Selection and optimisation using natural dNTPs

Human HER2/ErbB2 protein (Tag free, MW: 70.1 kDa, purchased from ACROBiosystems™) was used as a protein target in aptamer selection. HER2 solution (0.1 μM, 100 μL) was injected through a 0.45 μm nitrocellulose membrane filter (Millipore™) to capture Her2 protein targets in the membrane. To ensure aptamers with unique three dimensional shapes bind to Her2 target, ssDNA libraries stored in phosphate-buffered saline (PBS) buffer were heated at 95 °C for 3 min and immediately followed by a shock freezing in liquid nitrogen. This resulting solution of ssDNA sequences was subsequently incubated with the HER2-bound nitrocellulose membrane for 1 hr at room temperature. This method maintains folding structures of unimolecular oligonucleotide during the incubation process. The nitrocellulose membrane containing ssDNA-HER2 complexes was obtained, washed, and maintained in $H_2O$. The target-bound sequences were eluted from the HER2 and nitrocellulose membrane by thermal denaturation at 95 °C for 15 min.

In total, 10 rounds of selection were performed for each ssDNA library as part of the SELEX approach. Among these, two counter selection rounds (rounds 5 and 9) were introduced in order to remove non-specific binding aptamers. In a counter selection round, ssDNA sequences were first incubated with a nitrocellulose membrane containing a BSA protein and all negative target-bound sequences attached to the membrane were removed. Unbound sequences were then collected and incubated with HER2 target-bound nitrocellulose membranes for 1 hr. Unbound sequences were

washed away and only ssDNA-HER2 complexes were thus retained.

qPCR was used for the sequence enrichment processes for monitoring the quantification of target-bound oligonucleotides during SELEX. This 'in real-time' monitoring is necessary for library enrichment, as over-amplification can lead to non-template specific amplification and, subsequently, to a complete loss of the desired aptamers. Therefore, the amplification of oligonucleotides was ceased for further 5 cycles after it reaches its maximum amplification (*i.e.,* the cycle that presents highest value of fluorescence intensity in qPCR).

During every round of selection, the initial concentrations of ssDNA libraries that bind to the protein target are fixed to 0.5 μM in 500 μL PBS buffer solution. However, due to the differential binding efficiencies of ssDNA libraries to their target after several rounds of selection, the concentrations of target-bound oligonucleotide sequences that are finally added for qPCR amplification (as PCR templates) varies for each round of selection. The ssDNA library that contains more target-bound oligonucleotide sequences after selection will achieve a higher initial concentration during qPCR amplification and therefore reach its maximum amplification in a shorter time. Hence, the progress of aptameric selection after several rounds can be directly visualised by monitoring the shift in the maximum amplification (from right to left) for each round of selection. For example, in the selection of ssDNA library 1 which contains modified oligonucleotide sequences containing amU1, the cycles of maximum amplification for each round have shifted from 20 (in round 1, **Figure 4.5**) to 15 (in round 10, **Figure 4.6**). This result indicates that more and more target-bound oligonucleotide sequences

are preserved and enriched after several rounds of selection and, to some extent, this proves the efficacy of the selection process.

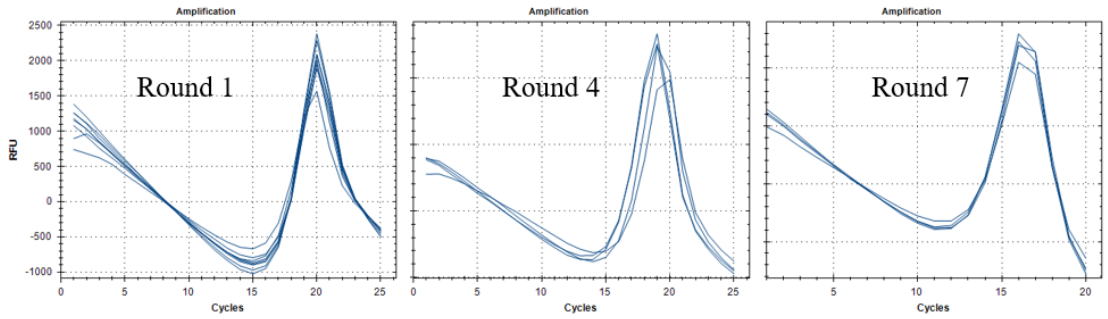



**Figure 4.5** The qPCR results illustrate a clear shift in terms of the cycle of maximum amplification through 10 rounds of selections for ssDNA library 1. The fluorescence intensity drop at the beginning is probably due to the dual effects of the two fluorophores used as FAM has a similar emission peak with EvaGreen although possessing a greater fluorescence intensity.[200,201] Therefore, the fluorescence intensity differences arising between dsDNA and ssDNA in the presence of EvaGreen are only be detectable after several cycles.

### 4.2.3 Selection of 5-AM-dU modified aptamers for Her2 protein

After 10 rounds of selection for an ssDNA library 1 directed against the Her2 target protein (containing oligonucleotide sequences all consisting of naturally occurring dNTPs as a control), the other three ssDNA libraries which incorporated modified nucleotides amU2 – amU4, respectively, were subsequently tested with the Her2 protein target (**Figure 4.6**). The target selection and qPCR amplification processes for each round of selection generally followed the previous protocol reported for ssDNA library 1. During the first round of selection, qPCR results reveal that both libraries 2 and 3 exhibited very late maximum amplifications, suggesting that most oligonucleotide sequences present in these libraries did not bind to the Her2 target successfully. However, after only 4 rounds of selection, the cycle of maximum amplification for libraries 2 and 3 were shifted to 18, with only slight progress in terms of the lateral selections. These results indicate that the aptamer selection process for libraries 2 and 3 were completed at a very early stage and that most of unbounded sequences were washed off before round 4, whereas the selection process for library 4 did not show any signification trend in terms of maximum amplification cycle shift during the 10 rounds of selection. Although library 4 yielded reasonable cycles for maximum amplification during the selection (which indicates a sufficient quantity of selected oligonucleotide sequences), this result led to a concern over a potential lack of highly specific aptameric candidates in library 4.

As only qPCR products from the final round of selection (round 10) for each ssDNA

library were sent for subsequent DNA sequencing, the primers used for qPCR amplification in round 10 were different. This is because dsDNA samples are required for DNA sequencing. Hence, two 20-meric DNA primers without any FAM or HEG-poly-A tail modifications (as any of these modifications will create difficulties for the lateral Illumina NGS process when attaching adapters into dsDNA) were synthesised specifically for qPCR amplification during round 10.

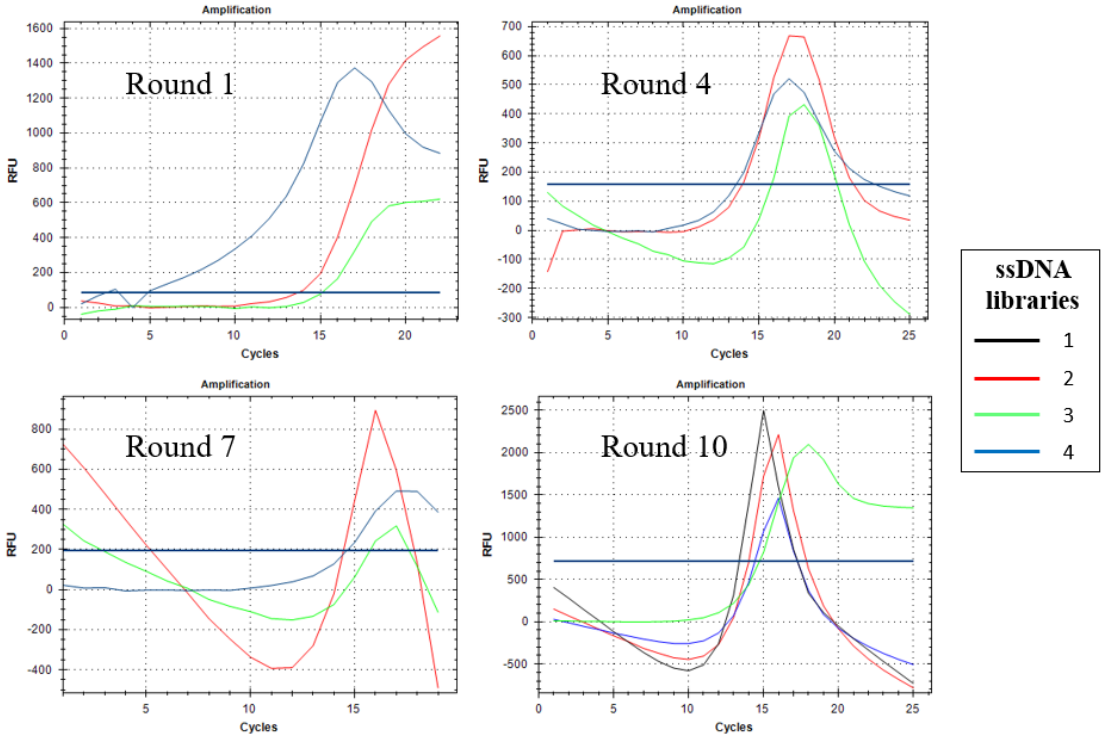



**Figure 4.6** qPCR results from four ssDNA libraries each containing dTTP and the modified nucleotides amU2 – amU4, respectively.

# 4.3 Identification of aptamer candidates

## 4.3.1 Base composition

In total, four samples of oligonucleotide library each containing dTTP or the modified nucleotides amU2 – amU4 (named library01 – library04 respectively) that recovered from 10 rounds of SELEX were analyzed by NGS. Sequencing experiments and were performed on an Illumina MiSeq instrument by Oxford Genomics Centre. Each sample of dsDNA amplicons that sent for NGS was further purified using Qiagen purification kits and normalized in a form of 20-30ng/µl in 33ul of 10mM Tris-Cl, pH 8.5. A quality control for these four submitted samples was first analyzed by the Oxford Genomics Centre. The quality (Q) Score is defined as the equation below:

$$Q\ Score = -10 \times \log_{10}\left(\frac{p}{1-p}\right) \qquad P: the\ probability\ error\ of\ a\ base$$

The average quality score of four submitted samples were tested to be 36.0. Then $P \approx 10^{-\frac{36}{10}} \approx 0.00025$. This means that there is one wrong base call in every 4000 base calls with a Q Score of 36. This proves a good quality of tested amplicons on NGS.

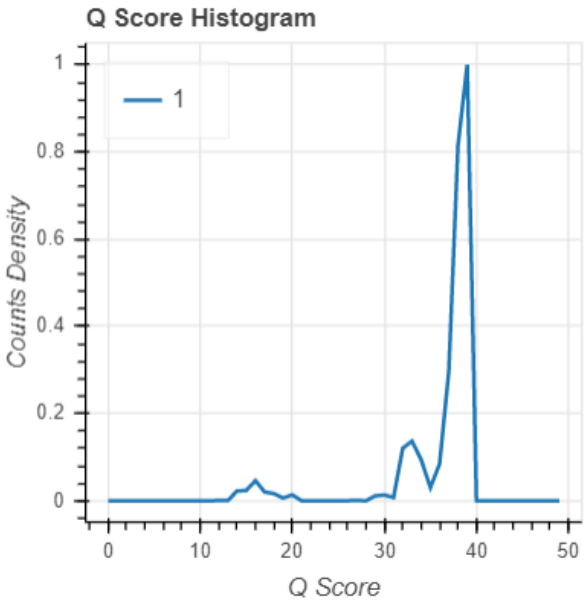



**Figure 4.7** The quality score histogram of tested amplicons.

The analyses of sequencing data were performed using MATLAB. The base composition of sequences read in each ssDNA library after SELEX was analysed. Since the sequences sent for NGS tested were in dsDNA form, the resulting rough sequencing data were composed of both selected 80-mer oligonucleotides and their complementary strands. Therefore, for NGS data of each library, only desired sequences that start with 5′-AGCAGCACAGAGGTCAGATG-3′ (forward primer) were selected and analysed. These sequences were ssDNA oligonucleotides that possess unique three-dimensional conformations which bind to Her2 target molecules through the previous SELEX process. The total number of each nucleobase in these resulting sequences were counted in four different tested libraries (**Figure 4.8**). In all of four ssDNA libraries, the distribution of nucleotide bases observed with more enrichment of dA and dG compare to dC and dT (or modified dT: amU2/amU3/amU4). The percentage of nucleobase dT in ssDNA library01 is 17.4%, whereas the percentage of three 5-AM-dU analogues

(amU2 in ssDNA library02; amU3 in ssDNA library03; amU4 in ssDNA library04) are 15.3%、14.8% and 15.3% respectively. The differences of base frequency between dT and three 5-AM-dU analogues are not significant, which suggest there is no strong selection pressure against 5-AM-dU analogues during SELEX.

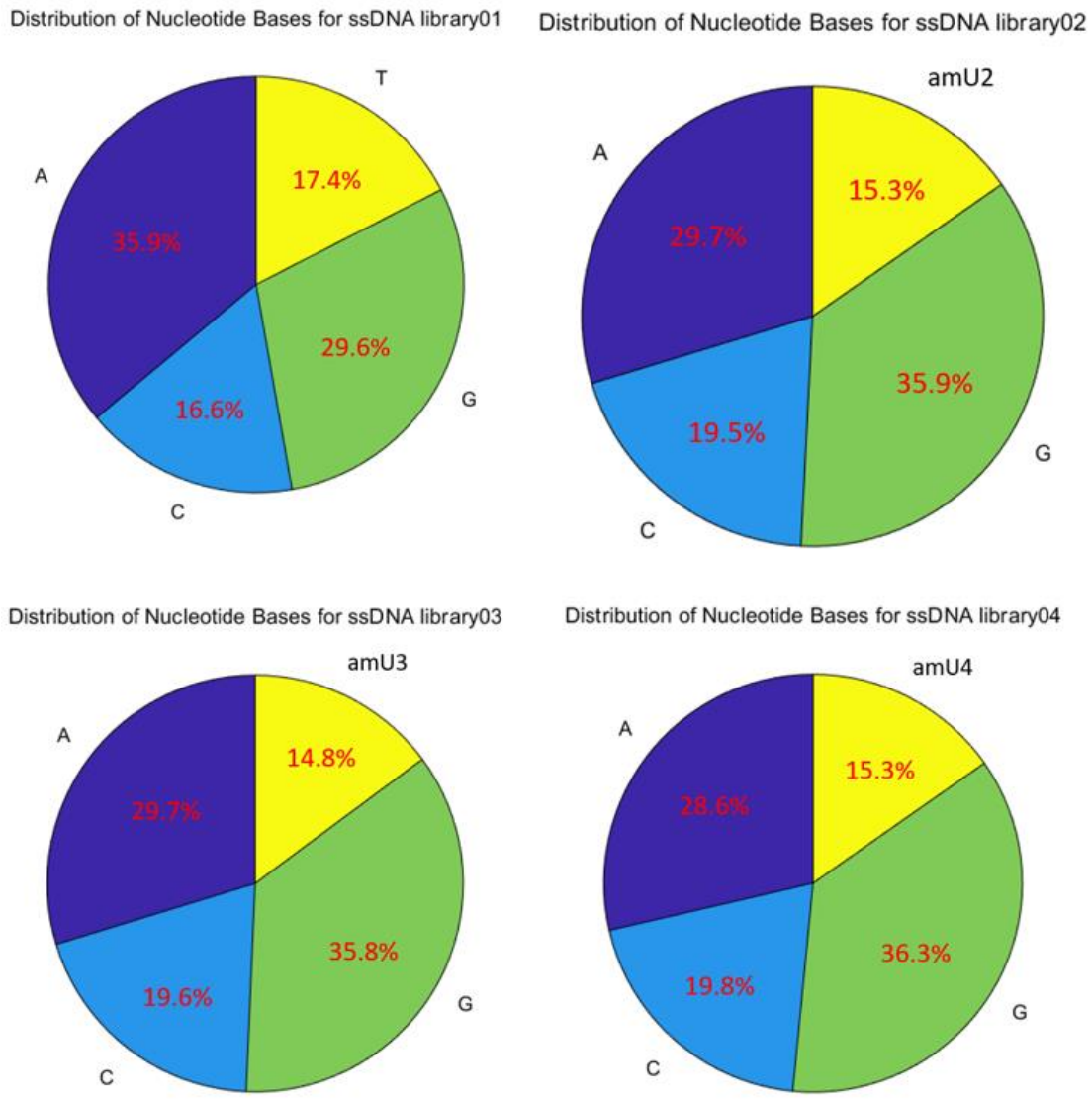



**Figure 4.8** The base frequencies of sequences in four different ssDNA libraries after 10-round of SELEX. The yellow columns in histogram represents Thymidine (dT) and three 5-AM-dU analogues (amU2 – amU4) respectively from library01 to library04.

The number of nucleobases in each selected sequence were also analysed (**Figure 4.9**).

In library01, most of 80-mer oligonucleotides composed of less than 20 of dC or dT, more than 20 of dG, and around 30 of dA. In other three libraries, the presence of dG is significantly higher than other nucleobases in most of sequences. This phenomenon of high dG frequency could be due to the presence of G-quadruplex on sequences selected through SELEX.

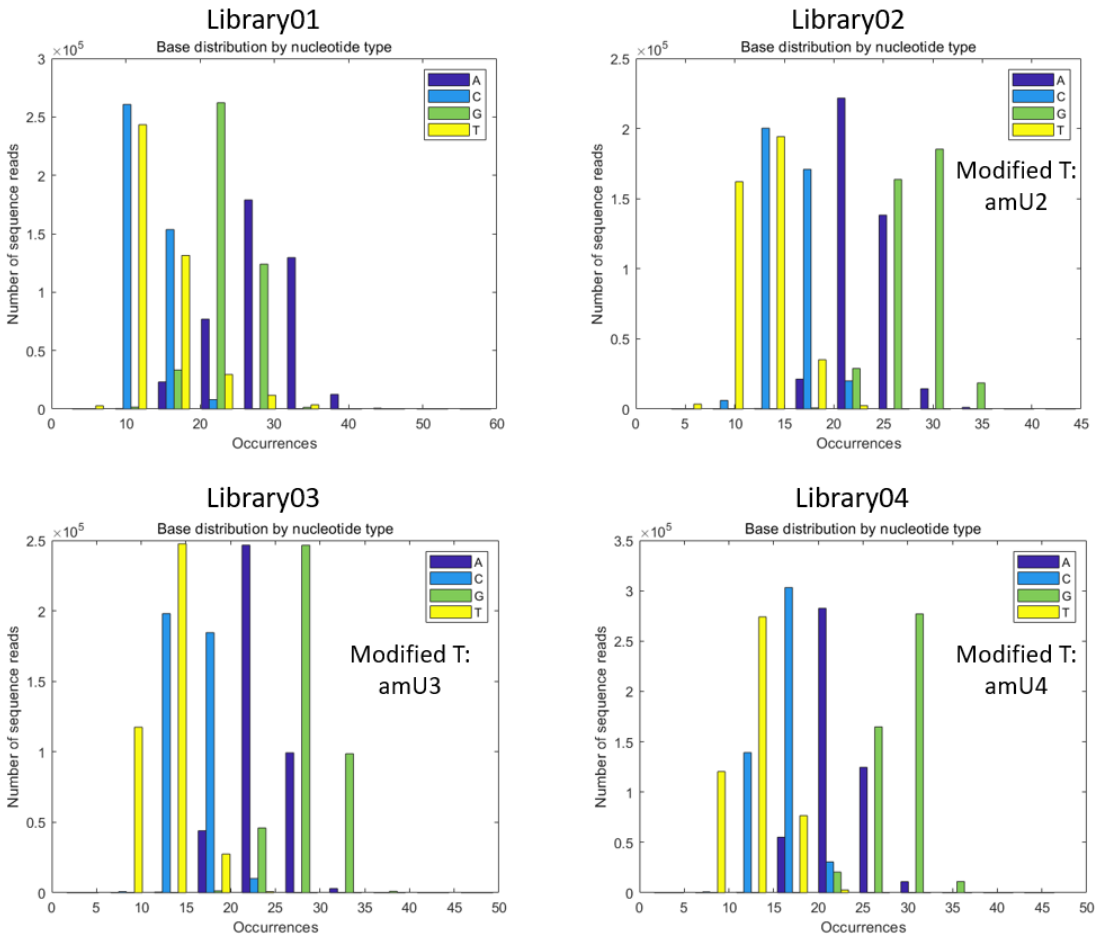



**Figure 4.9** The base composition of the sequences read. The x-axis shows number of sequences in a library; the y-axis shows the occurrences of each nucleobase in an 80-mer DNA sequence. The yellow columns in histogram represents Thymidine (dT) and three 5-AM-dU analogues (amU2 – amU4) respectively from library01 to library04.

### 4.3.2 Unique sequence enrichment

The analysis of unique sequence enrichment is the most common method used for aptamer sequencing. Sequences with good affinities to target proteins were enriched during SELEX. Unique sequence enrichment analyses identify these sequence reads that occur multiple times in NGS. The MATLAB programs for analyzing sequences occurrences in libraries are shown in the experimental section (page 230). Among four libraries, 10 most enriched sequences (＞50 reads per million) were plotted as a heat-map (**Figure 4.10**) to visualize their occurrences in different libraries. The random region (N40) of these 10 most enriched sequences were listed in **Table 4.2**.

| Number | Sequences (5′ → 3′) |
|---|---|
| N1 | AAAGAATAAAGAAACAAGAAAAAGGGGAAAGGGAAAAACT |
| N2 | AAGAAGAACACGAGAAGGAAGTAAGAAAAAAATAGAAGAA |
| N3 | AGAAGAGAAGAAAAGAGGAAAAAATGAGAGAAAAGTATGT |
| N4 | TCTTTAGTCTATTATTTCTCTCTTTTTGTATTCTTGTTTT |
| N5 | GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTTGG |
| N6 | AGGCGGACGAGCTGAGAGGAAATGGCGTGGAGGGCGGTAT |
| N7 | GTACAGGGGTGTAACGCGTCAGGAAGAGTGGACGACCTCA |
| N8 | CCTATGCGTGCTACCGTGAAAGATCGGAAGAGCACACGTC |
| N9 | TACGACGATGACGATAAGCTGGAGGCCTATGCGTGCTACCG |
| N10 | GGGTAGGGGAGCCACAAAGGGAAAGGACGAGGAGGTGTG |

**Table 4.2** Schematic representation of 10 most enriched sequences (in random region) selected from four libraries. (Nucleobase T in table represents T/amU2/amU3/amU4 according to which ssDNA library was analysed)

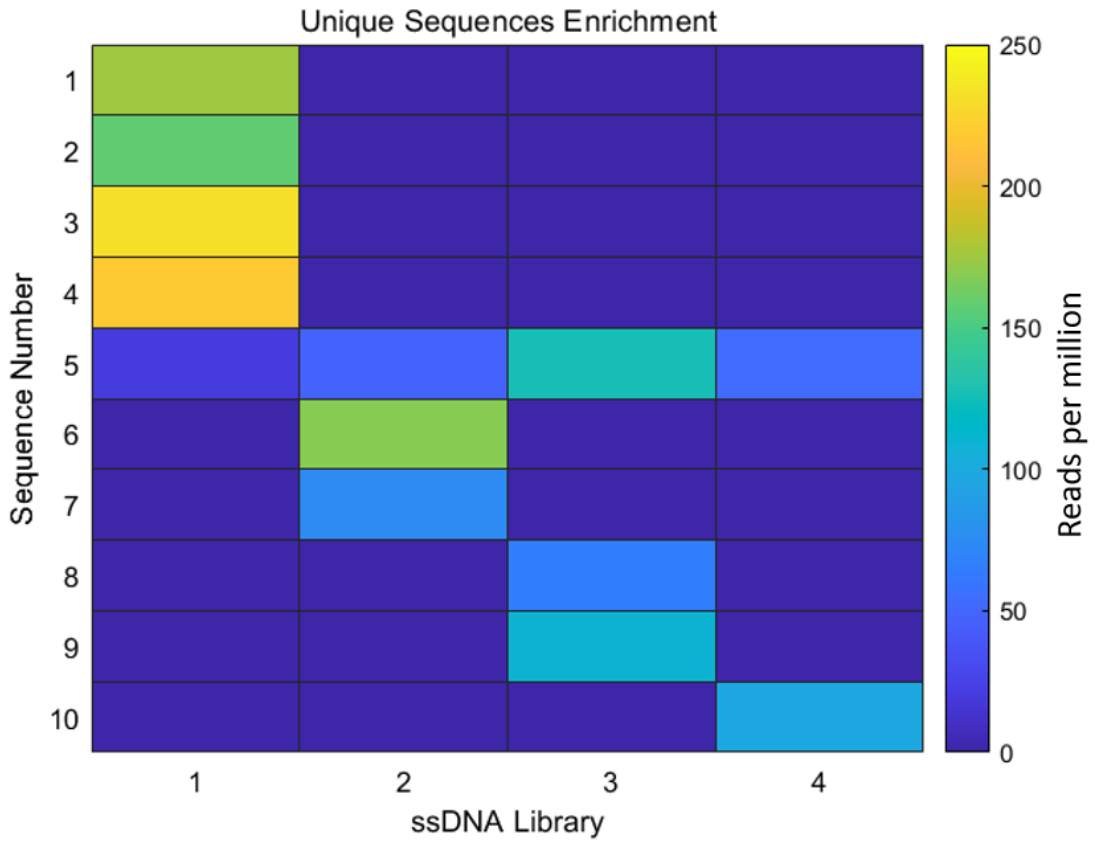



**Figure 4.10** The color-coded heat-map of unique sequences enrichment in different ssDNA library. (The base sequences of these oligonucleotides were listed in **Table 4.2**)

The first four sequences (N1 – N4) are selected as most enriched sequences in library01. N6 and N7 are sequences selected from library02 which is the library that using amU2 as modified dT during selection process. Similarly, N8 and N9 are sequences selected from library03 and N10 are selected from library04. N5 is a unique sequence was found to be enriched in all four libraries though each library possesses different nucleobase T. The universal affinity of N5 sequence to Her2 target in all four ssDNA libraries could be due to the forming of G-quadraplex structure. The unimolecular G-quadraplex motif was generally identified as: $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$.[202,203] Sequence of N5 contains this G-quadraplex motif and is very likely to form strong G-quadraplex secondary structure during SELEX. Whereas the structural differences of the sequence

caused by different nucleobases T/amU2/amU3/amU4 has only a minor effect compared to G-quadraplex secondary structure.

## 4.4 Conclusions and future work

In this chapter, three modified 5-AM-dU analogues were introduced into SELEX approaches for target selection. In total, four ssDNA libraries, each containing oligonucleotide sequences that consist of natural dNTPs (canonical) or 5-AM-dU analogues in place of dTTP (non-canonical), were designed and enzymatically synthesised via PCR amplification followed by a separation method based on size. Human Her2 protein was chosen as the target protein for aptameric selection. Target-bound oligonucleotide sequences were bound to nitrocellulose membrane and subsequently eluted. These sequence candidates were amplified by qPCR for sequence enrichment and used for the next round of selection. In total, 10 rounds of SELEX were achieved for each of the four ssDNA libraries. qPCR results indicate a generally robust selection process. The final samples of candidate oligonucleotide sequences were amplified and submitted to the Oxford Genomics Centre for sequencing.

The sequencing results of four submitted samples were analysed. Little difference of base frequency between dT and three 5-AM-dU analogues state no strong selection pressure against 5-AM-dU analogues during SELEX. For the sequencing results derived from each library, a method of unique sequence enrichment was applied to identify those unique sequences that have been enriched during the selection processes. A unique sequence (N5) is found to be enriched in all four samples of ssDNA library though each sample of library contains different nucleobase T. Sequence of N5 contains G-quadruplex motif and could form strong G-quadraplex secondary structure during

SELEX. The enrichment of N5 in all samples of ssDNA library illustrate the structural differences caused by modified dT did not affect the formation of strong G-quadraplex secondary structure in this sequence.

| Number | Sequences (5′ → 3′) |
|---|---|
| S1 | AGAAGAGAAGAAAAGAGGAAAAAATGAGAGAAAAGTATGT |
| S2 | TCTTTAGTCTATTATTTCTCTCTTTTTGTATTCTTGTTTT |
| S3 | GCGGGAAGGCAGGGAAGGGGGATGGGGCGGGACATTTGG |
| S4 | GCGGGAAGGCAGGGAAGGGGGA3GGGGCGGGACA333GG |
| S5 | AGGCGGACGAGC2GAGAGGAAA2GGCG2GGAGGGCGG2A2 |
| S6 | 3ACGACGA3GACGA3AAGC3GGAGGCC3A3GCG3GC3ACCG |
| S7 | GGG4AGGGGAGCCACAAAGGGAAAGGACGAGGAGG4G4G |

**Table 4.3** Potential aptamer sequences selected after NGS analysis. Nucleobase in red colour represent modified 5-AM-dU analogues (2 means amU2; 3 means amU3; 4 means amU4).

In future, a target validation process will be proceeded to several potential aptamer sequences (**Table 4.3**) designed based on NGS analysis results. A Biacore T2000 will be used for a SPR binding analysis of aptamers and its Her2 protein target in order to identify any aptamer(s) with a strong binding affinity for the target.

# *Chapter 5.*

# *Ligand Construction of Spontaneously Assembling Bivalent Ligand (SABL) Libraries*

# 5.1 Introduction and Aims

## 5.1.1 DNA-encoded library and SELEX

The concept of DNA encoding was first described by Brenner and Lerner in 1992 as a powerful drug screening method, one which linked combinatorial chemistry and molecular biology.[204] This DNA-encoded library (DEL) approach has since been validated as an alternative to traditional small molecule drug discovery.[205,206] In DELs, drug molecules are conjugated to DNA fragments which serve as identification barcodes that record the synthetic conditions used to produce the library members. The addition of a target to the library, separation of those molecules that bind to the target, and analysis of the barcodes by high throughput sequencing (HTS) enable entire libraries to be screened far more readily and rapidly when compared to conventional screening methods.[207]

Nucleic acid aptamers are single-stranded DNA or RNA oligonucleotides with unique three-dimensional conformations that bind to specific target molecules.[208–211] Although the applications of nucleic acid aptamers are not as advanced as those of antibodies, aptamers possess several desirable features which antibodies lack in that they (i) can be produced by chemical synthesis on a large scale; (ii) reform their secondary structure, and thus possess better storage properties; (iii) have a lower molecular weight, little immunogenicity and lower toxicity; and, most importantly, (iv) aptamers are self-encoding and can be selected by SELEX processes.[61,78,79] However, since aptamers are

essentially DNA or RNA strands, their libraries lack true diversity. Aptamers are polyanionic, largely confined to the A/T/G/C nucleotides (or their close analogues), and some motifs (e.g., G-quadruplex) are overrepresented. Currently, the affinities of aptamers to their substrates are, in most cases, lower than for those of antibodies. SELEX bears a resemblance to DEL in that both encode the ligand within an oligonucleotide sequence and use PCR and DNA sequencing to amplify the binding signal, thereby enabling analysis of the composition of successful target binding conformations. The key difference is that DELs can possess a greater chemical diversity but are of lower affinity as they are essentially small molecules.[212]

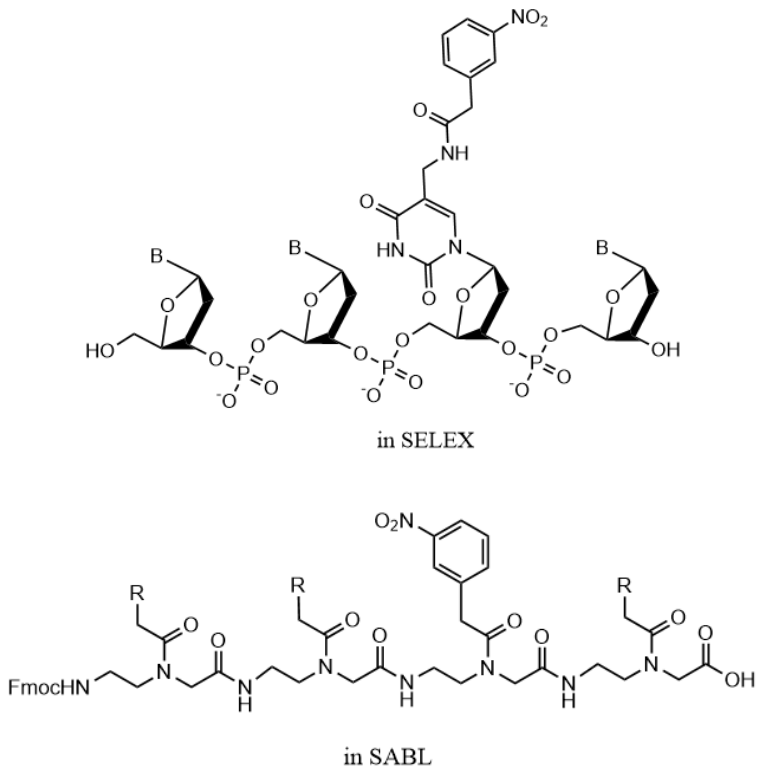



**Figure 5.1** An illustration of introducing the same "non-natural" functional group (nitrophenyl) into the AEG backbone within the SABL process and into the nucleic acid backbone in the SELEX approach.

### 5.1.2 Multivalency

Finally, we introduce the concept of multivalency, which is the linking of the identities of two (bivalency) or more ligands that bind cooperatively to the same target, thereby bringing greater affinities to the library. The improved binding efficiencies of multivalent structures as compared to monovalent species however depends on the mechanism of target binding. Several interactions are potentially involved in multivalency[213]:

(i) Sub-site binding: A multivalent structure with variously structured binding moieties could bind to different sub-sites on the same target.

(ii) Multiple-target binding: A multivalent structure with different binding moieties can also bind to entirely different targets (*e.g.,* different cell surface proteins) instead of different sites on the same target.

(iii) Chelation: A multivalent structure allows binding to the target via the chelation of multiple sites. The length and the flexibility of the linker is the key factor in chelation, permitting both sites to bind to the same target simultaneously.

(iv) Clustering: A multivalent structure can induce the clustering of targets at the surface, consequently increasing its effective concentration within that region.

(v) Statistical rebinding: By having multiple binding regions within the same structure,

the local concentration of 'binders' increases around the target, allowing the structure to rebind to the target quickly if one of its regions dissociates (in other words, presenting a lower effective $K_d$).

Rather than improving binding affinity, the potential benefits of incorporating multivalency into the structure of DEL also include simultaneously increasing the chemical diversity of the library. The best example of the greater avidity achieved with multiple ligand binding arises in nature in the form of the antibody proteins IgE, IgG and IgD.[214] These immunoglobulins contain two identical epitope-binding regions and hence are considered to be bivalent. This concept has also been explored within the aptamer field,[214–222] wherein two aptamers connected by a long flexible linker were used to target thrombin exosites to enhance binding affinity[223–226] (**Figure 5.2**).

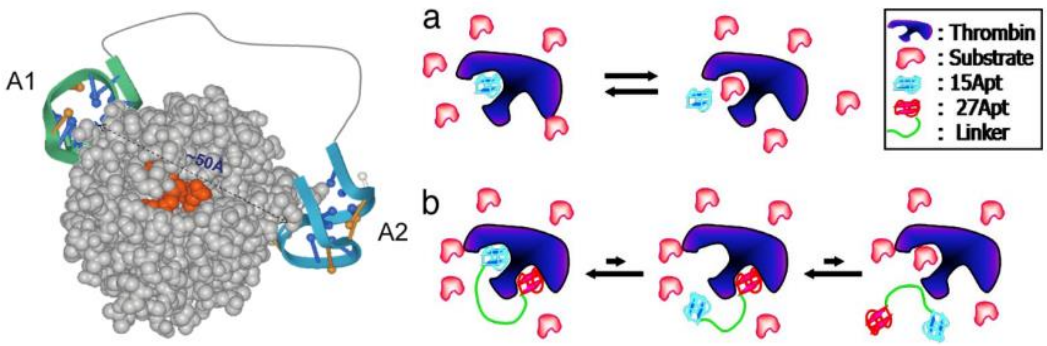



**Figure 5.2.** An illustration of the increased avidity caused by the binding of bivalent ligands to thrombin. A complex of bivalent ligands containing a 15-base long aptamer (15Apt) binding to exosite 1 (A1) and a 27-base long aptamer (27Apt) binding to exosite 2 (A2) connected by a long, flexible linker. As a monovalent ligand, 15Apt could easily diffuse into the bulk solution immediately after dissociation from thrombin, resulting in a low binding affinity. owever, after 15Apt is linked to 27Apt to form a bivalent ligand, 15Apt returns to the binding site quickly after dissociation since 27Apt is still bound to thrombin. Hence the equilibrium shifts to the left (pictures adapted with permission from[215,222]).

### 5.1.3 SABL concept

By combining the concepts of DEL chemical diversity, aptameric biomolecular interactions and multivalency into a single format, the affinity for any target is expected to improve significantly relative to previous systems. To achieve this goal, we will develop a novel concept in ligand library design, namely **S**pontaneously **A**ssembling **B**ivalent **L**igand (SABL) libraries. In SABL libraries two ligands bind cooperatively to the same target. The principle underlying the construction of SABL libraries combines four key concepts, namely the: (i) enormous power of DNA encoding; (ii) increased avidity resulting from multivalency (in this case bivalency); (iii) combinatorial expansion of library diversity caused by ligand pairing; (iv) exploration of a wider range of supramolecular interactions arising between ligands and targets, including some "non-biological" functional groups.

In order to achieve this SABL library concept, we will construct novel DNA-based sequences containing several regions for target binding (of modified ligands), ligand encoding and complementary regions for ligand pairing. In theory, a visualisation tag such as a fluorophore or an anchoring group like biotin can also readily be incorporated with the DNA construct. A solid-phase 'split-and-mix' strategy often used in combinatorial chemistry was employed to generate two sub-libraries consisting of ligands and their corresponding encoded DNA oligonucleotides (via the stepwise assembly of ligand building blocks and encoding DNA bases). One of the sub-library oligonucleotide syntheses will be in the 3'- to 5'- direction, while another sub-library

oligonucleotide synthesis will take place in the reverse direction (5'- to 3'-). The complementary DNA regions occurring between two sub-libraries will enable them to form bivalent ligand libraries spontaneously via simple mixing.

The library size and diversity will be dependent on the variety of building blocks used, the number of additions made and, most significantly, by the combinatorial pairing of ligand sub-libraries. For example, a SABL library with four different chemical building blocks and constructed sub-libraries of ligands with six building block additions will give rise to $4^6 = 4{,}096$ different ligands in each of two sub-libraries, corresponding to a SABL library of $4{,}096^2 = 16{,}777{,}216$ permutations in total. Although more than six building blocks could be added, the total yield will also decrease along with each coupling. For our initial studies, these sub-libraries with six building blocks were believed to constitute an adequate library size that could be used directly for target testing. At the same time, more AEG-based monomers will be made in order to increase the chemical diversity of SABL ligands.

### 5.1.4 Structures of PNA

Peptide Nucleic Acids (PNA) are nucleic acid analogues of DNA which were first created by Peter E. Nielsen in 1991. In these PNAs, the sugar phosphate backbone found in DNA and RNA is replaced by an uncharged, achiral *N*-(2-aminoethyl)-glycine (AEG) backbone.[227] PNA oligonucleotides also hybridise with complementary DNA (cDNA) and RNA oligonucleotides via Watson-Crick base pairing, although with a generally higher binding affinity and specificity than are found within normal DNA/RNA oligonucleotides. This is due to the reduced electrostatic repulsion arising between oligonucleotides, since the PNA backbone does not contain negatively charged phosphates (and hence, to some extent, PNA is not truly a nucleic *acid*).[228–230] The resulting duplexes formed between PNA and DNA exhibit a higher thermostability than for DNA/DNA duplexes and are virtually independent of salt concentration (**Figure 5.3**).[231] PNA oligonucleotides possess analogous amino- (*N*) termini and carboxyl- (*C*) termini to the 5′- and 3′- termini of DNA. PNA oligonucleotides hybridise with DNA oligonucleotides in a predominantly antiparallel orientation. Therefore, by convention, PNA oligonucleotides are orientated, unlike peptides, with the *N*-terminus at the first (left) position and the *C*- terminus at the final (right) position.[230,232]

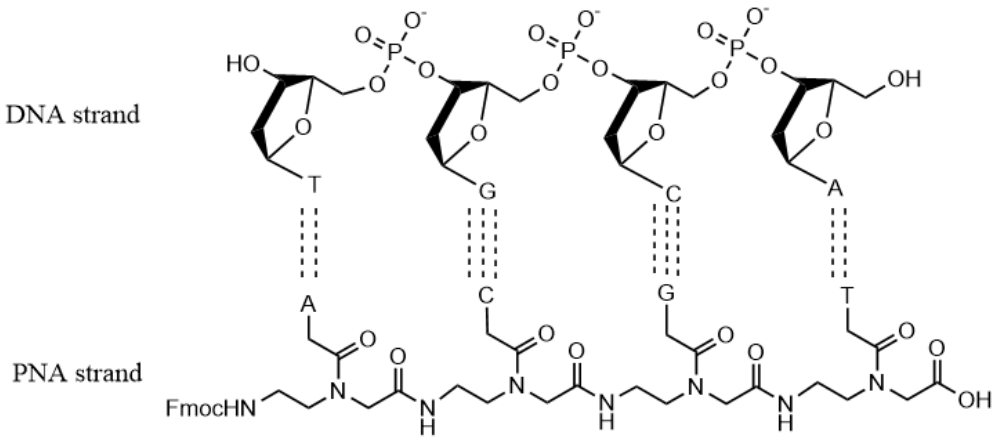



**Figure 5.3** Chemical structure of a duplex formed between PNA and DNA.

Due to the modification of the backbone, PNA oligonucleotides do not serve as substrates for enzymes such as polymerases, proteases or nucleases. For this reason, PNA oligonucleotides are incompatible with some molecular biological techniques such as PCR, since they cannot be used directly as PCR primers. However, it is also by virtue of this very property that PNA oligonucleotides are ideal candidates for antisense therapies given that they are resistant to biological degradation by nucleases.[232–235]

The chemical synthesis of PNA is usually performed via solid-phase peptide synthesis (SPPS).[236–243] When PNA monomers are assembled they form amide bonds between adjacent amino- and carboxyl groups via sequential peptide coupling reactions. The growing peptide chain is attached to the solid support (resin) via an ester or amide linkage, and the amine site of the new unit is protected by a 9-fluorenylmethyl-oxycarbonyl (Fmoc) group. These Fmoc protecting groups are base-labile and are removed by a solution of 20% piperidine in DMF during the deprotection stages (**Figure 5.4**). The primary amines of the heterocyclic bases occurring in PNA

monomers are protected by benzhydryloxycarbonyl (Bhoc) groups during synthesis. Coupling reagents like O-(7-azabenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HATU), O-(1*H*-benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HBTU), dicyclohexylcarbodiimide (DCC), and 1-hydroxybenzotriazole (HOBt) are generally used as activators for carboxylic acids in coupling reactions.[244]

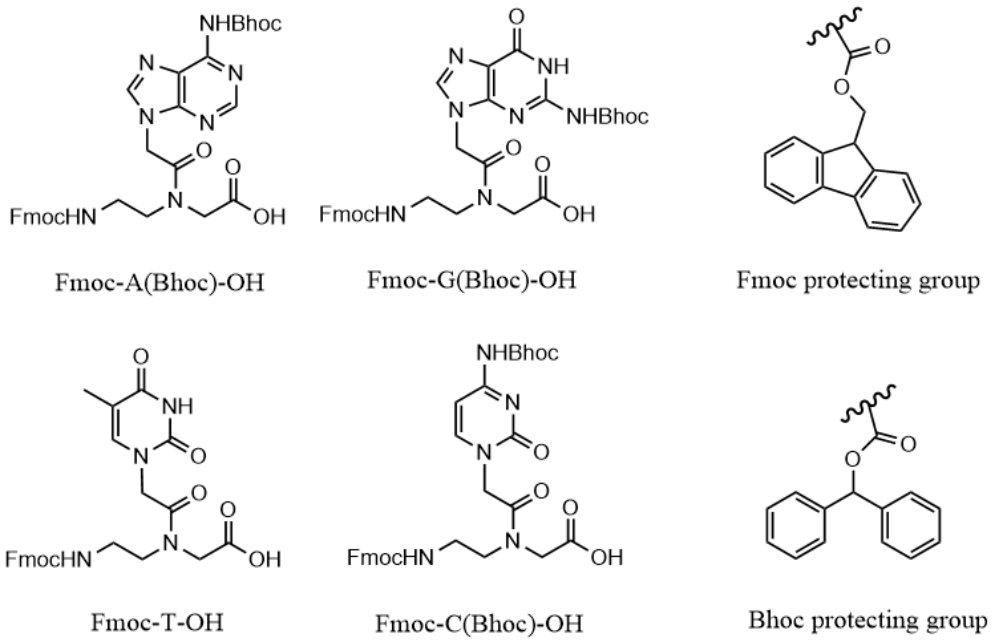



**Figure 5.4** Chemical structures of PNA monomers and associated protecting groups required for an Fmoc strategy.

### 5.1.5 Aim of project

The aim of the project is to construct a new class of ligands which are chemically and structurally more diverse than aptamers and also possess higher affinities that is comparable to antibodies. To achieve this goal, a new concept will be explored, that of creating SABL libraries. This SABL library approach involves linking two ligands which will bind to the same target. By introducing bivalency into the library design, its diversity will further be expanded due to combinatorial ligand pairing and the increased avidity when two ligands bind to the same target. Most of the chemical building blocks, or ligands, used in this project contain "non-natural" functional groups attached to the backbone of PNA structures, given that a wider range of supramolecular interactions can subsequently be investigated using SABL libraries.

## 5.2 Library designs

### 5.2.1 Structures of DNA encoding region

To construct this novel DNA-based library, two sub-libraries of oligomeric ligands will be generated by means of a solid-phase split-and-mix strategy. In each sub-library, ligand building blocks and their corresponding encoding DNA bases are assembled in a stepwise fashion (**Figure 5.5a**). For sub-library 1, oligonucleotide synthesis will be in the 3′- to 5′- direction and the reverse for sub-library 2. When the two sub-libraries are mixed, their cDNA regions will enable the DNA templates to assemble spontaneously to form the desired bivalent ligand libraries (the so-called "click-then-select" approach).[245] At the termini of two sub-libraries, an azide/alkyne group will be added for chemical ligation to generate a triazole linkage by the highly efficient CuAAC "click reaction" (**Figure 5.5c**).[246] The artificial triazole backbone is the first example of such artificial DNA linkages that can be read through by a polymerase and amplified via PCR.[247] This chemical ligation method provides a greater degree of flexibility in terms of reaction conditions than enzymatic ligation and also offers higher yields.[145,246–253] Alternatively, the two sub-libraries will be click-ligated to each other after they both bind to the same target. This "click-then-select" approach attaches sub-libraries prior to target selection so that it does not affect the target binding process itself and such ligand binding will be less conformationally restricted.

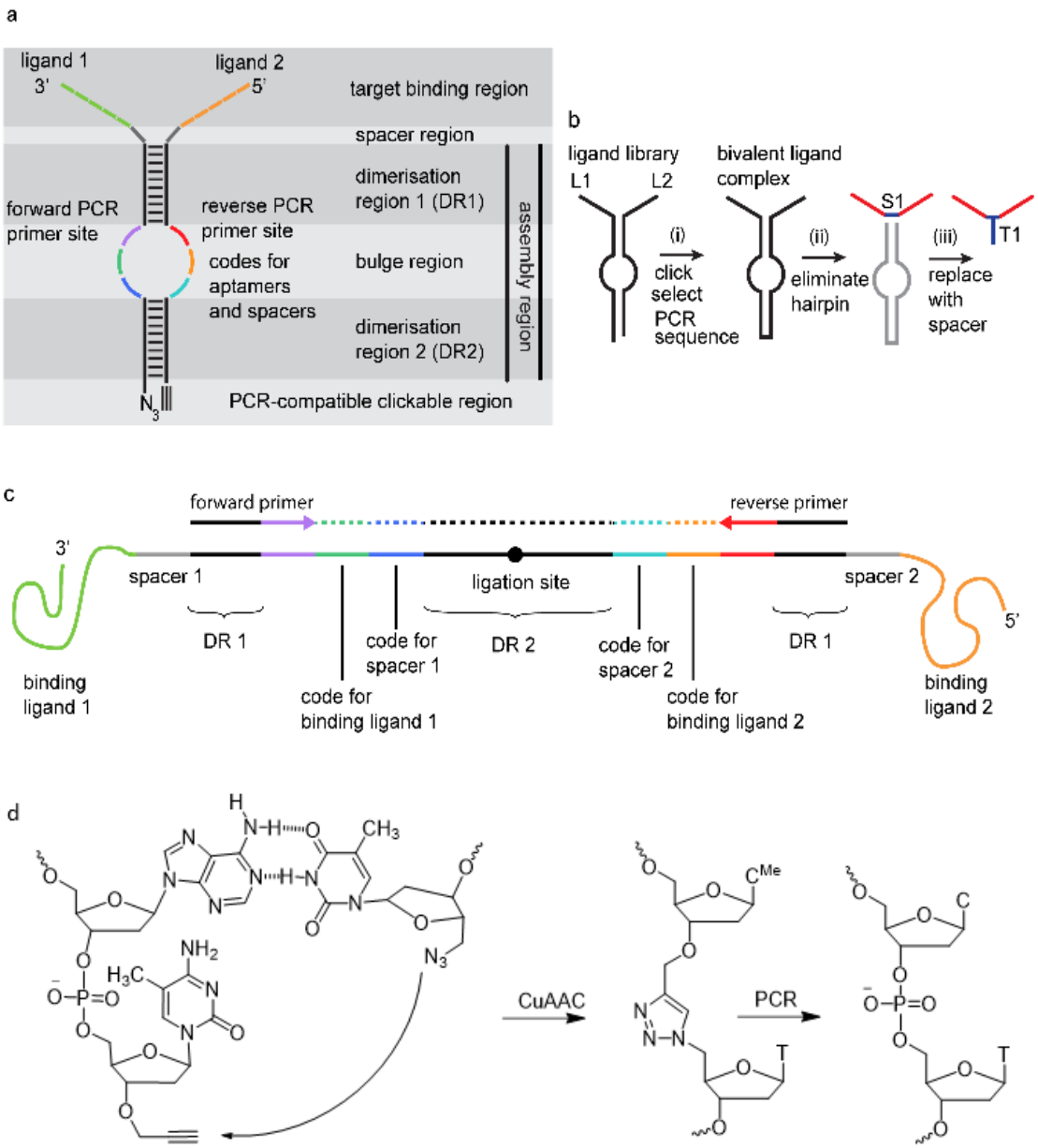



**Figure 5.5** SABL concept and its library design. **a.** Detailed construction of each SABL sub-library; **b.** two sub-libraries are 'clicked' together to generate bivalent libraries which can then undergo target binding. After separating good binders from bad, recovered ligands will be amplified by PCR and sequenced to identify their specific ligands; **c.** ligands and DNA codes are added stepwise to the oligonucleotide resin via SPPS and reverse phosphoramidites respectively; **d.** chemical ligation by the CuAAC click reaction (pictures adapted with permission from[254]).

### 5.2.2 Structures of ligand building blocks

Within SABL libraries, ligands are no longer restricted to four Watson-Crick nucleotides or 20 canonical amino acids, since the ligands do not have to be read through by DNA polymerases during the decoding process. This property allows SABL libraries to incorporate "non-natural" functionality, including nitrophenyl groups, halogens, and bulky lipophilic groups within the ligand system. The nitro group is a unique and versatile functional group in medicinal chemistry. Studies of nitrophenyl-group-containing compounds showed good activity as antioxidants and anticancer agents.[255,256] Halogen bonding is an attractive intermolecular interaction as it arises between a halogen atom that is covalently linked to an electron withdrawing substituent (which forms a localised sigma-hole on the opposite pole of the halogen atom) and another molecule (sigma-hole binding) with negative sites that act in much the same manner as a Lewis base (**Figure 5.6**).[255–257] Two ligands with non-covalent halogen bonding (XB) moieties serving as functional groups were designed and synthesised to study XB in protein recognition using a SABL approach.[260,261]

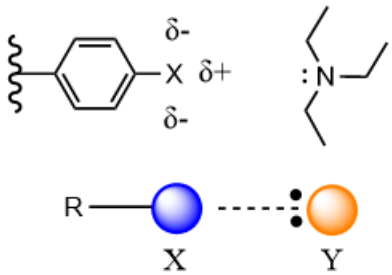



**Figure 5.6** Principle of halogen bonding. R refers to electron withdrawing substituent (e.g., an aromatic group); X is a halogen (e.g., Cl, Br, or I), and Y is a Lewis base (e.g., $Et_3N$) with free electron pairs.

The ligand scaffolds are also variable, as there are plenty of biologically stable scaffolds that could be used in SABL libraries, including D-peptide, β-peptides, peptoids and AEG-based (aminoethyl glycine-based). As a proof-of-principle, several Fmoc-protected AEG-based (aminoethyl glycine-based) monomers were prepared (**Figure 5.7**). AEG was chosen as it is also the backbone of PNA, wherein the sugar phosphate backbone of DNA is replaced with an uncharged, achiral polyamide chain.[238,241,242] Moreover, PNA contain amino- and carboxyl- ends which could be used in solid-phase peptide synthesis (SPPS) in a similar manner to methods which enable far simpler initial library construction. Each monomer contains a side chain of differing functionality and, for all these compounds, their functional groups do not require protection during SPPS. Fmoc was compulsory for the N of each individual ligand to enable stepwise coupling by the SPPS method in forming oligonucleotides.

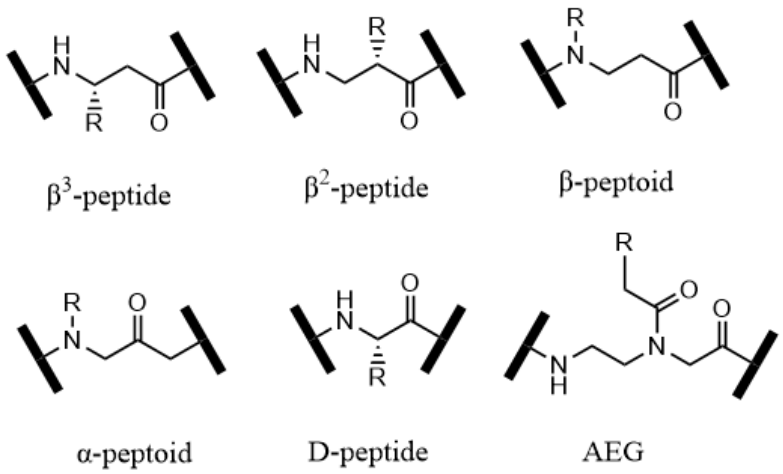



**Figure 5.7** Structures of potentially biologically stable scaffolds for SABL libraries.

For this project, a total of 12 different ligand building blocks (**Figure 5.8**) were designed and synthesised with the same Fmoc-protected AEG-based backbone bearing different types of side chain including halogen bond donors, basic groups, poly-substituted aromatic groups, hydrophobic group, etc.

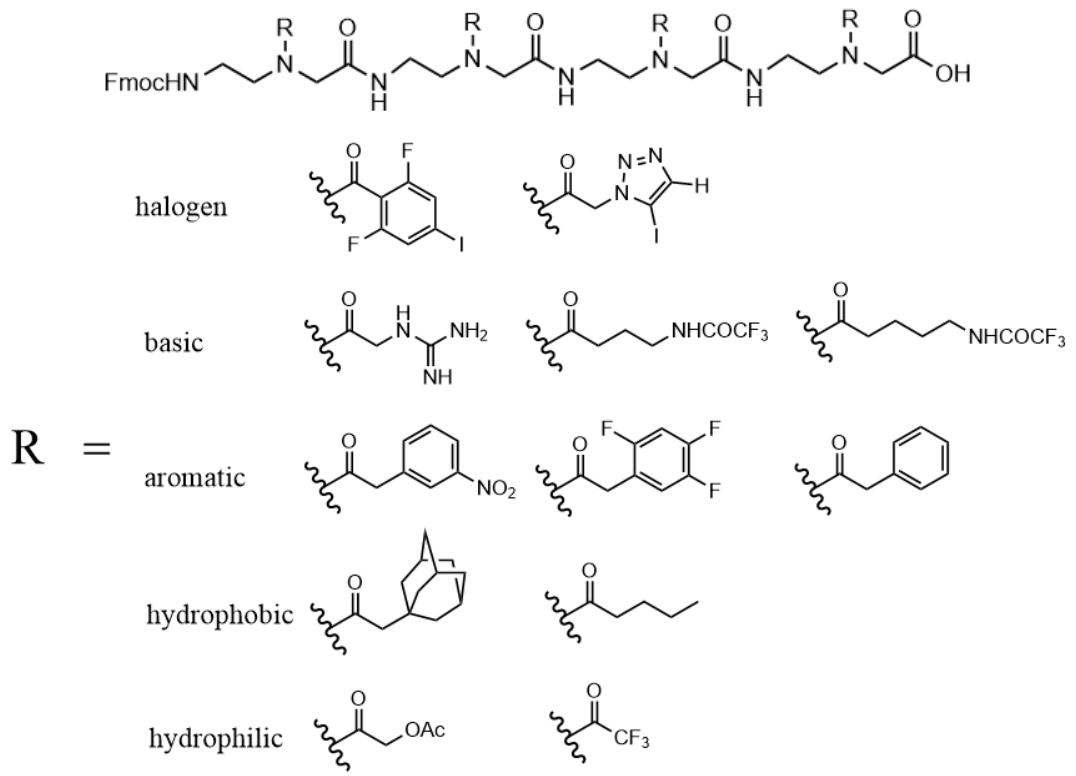



**Figure 5.8** Scaffolding of AEG-based backbone and associated functional groups attached to their side chains.

## 5.3 Synthesis of AEG-based chemical building blocks

Due to the large variety of monomers ultimately required, unfunctionalized Fmoc-AEG-OH blocks were synthesised. An extensive literature search revealed several published routes (**Figure 5.9**). The first requires three steps with an overall yield of around 72% (**Figure 5.9a**).[239] However, the starting material, N-(2-aminoethyl) glycine, is also quite costly. The second route requires four steps with an overall yield of 38% starting with ethylene diamine (**Figure 5.9b**).[240] Nevertheless, in our hands, using this method has proven very difficult to produce analytically pure material due to uncontrolled regioselectivity on Fmoc addition reaction and the number of by-products.[236,237]

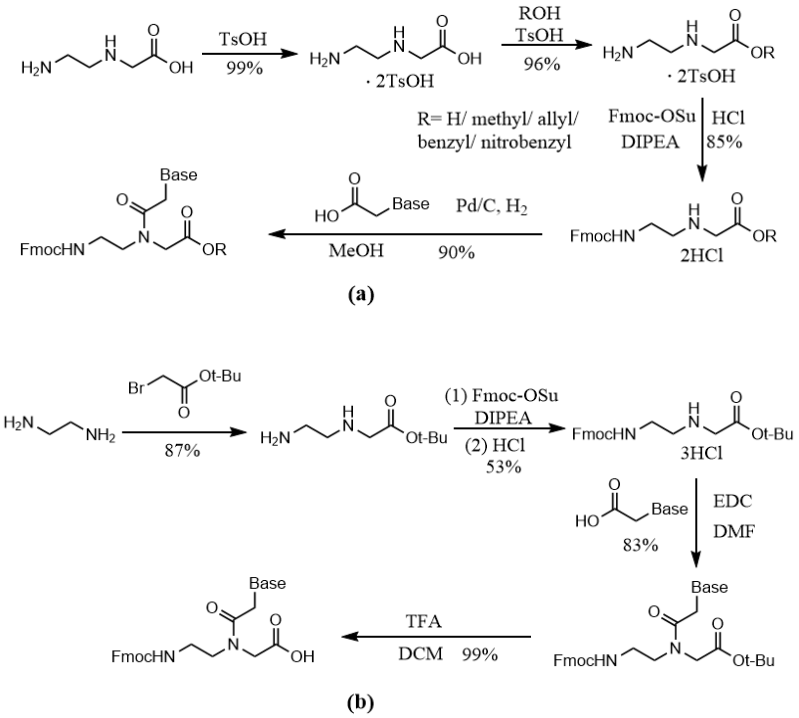



**Figure 5.9** Reported synthetic routes to a Fmoc-AEG-OR backbone. ***a.*** A synthetic route reported by Hudson *et al.*;[239] ***b.*** synthetic route reported by Thomson *et al.*[240]

Hence, an alternative route (**Figure 5.10**), similar to that of Thomson *et al.*[240] (**Figure 5.9b**), was considered. Instead of using Fmoc, a more base-stable Boc was used to protect the free amine initially before switching to Fmoc at the last step wherein both the formation of AEG backbone and the coupling of side chains are completed. By this route, the previous regioselectivity problems encountered by Thomson *et al.* were overcome. The bi-substitution of tert-butyl bromoacetate to ethylene diamine is no longer favourable, since Boc only undergoes single addition to ethylene diamine.[262] In order to achieve a higher yield, excess (7 eq.) ethylene diamine **5-10** was used and $Boc_2O$ added dropwise.[263] Similarly, for the second step, **5-11** should be used in excess (8 eq.) relative to tert-butyl bromoacetate.[264,265] All 12 AEG-based building blocks were successfully synthesised with overall yield varying from 35% to 45% (**Figure 5.11**). Given the success of the route, it is envisaged that many other AEG-based monomers can be made rapidly by this method in future.

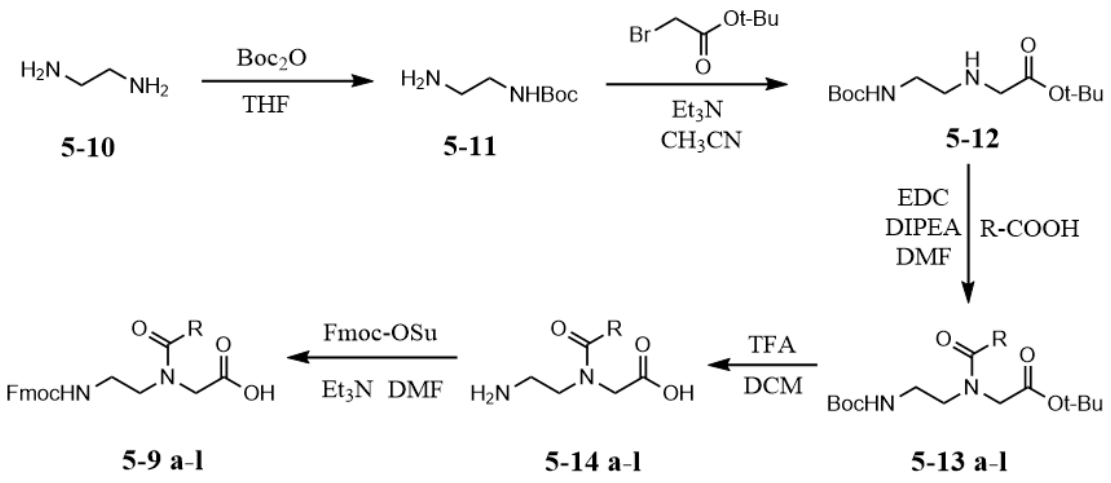



**Figure 5.10** Synthetic route for designed AEG-based building blocks with overall yields varying from 35% to 45%.

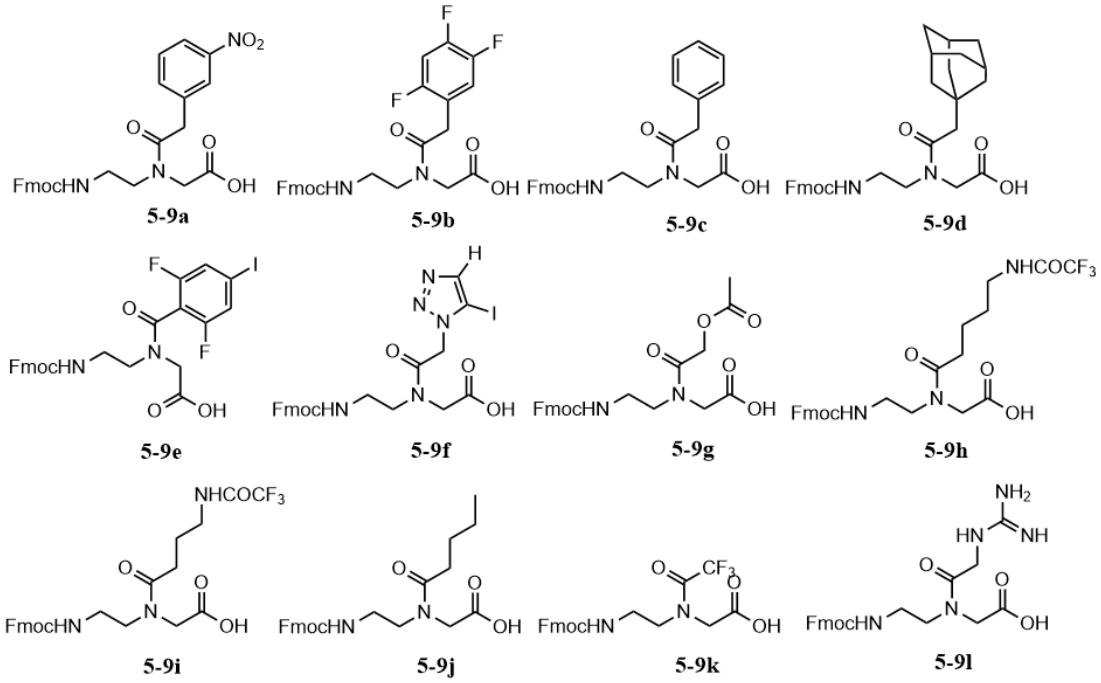



**Figure 5.11** Chemical structures of 12 synthesised AEG-based building blocks (structure of compounds **5-9 a-l**).

## 5.4 Encoding and construction of ligand sub-libraries

After the successful synthesis of 12 AEG-based ligand building blocks, a range of conditions for linking them to oligonucleotides were tested. Since every building block will have its own corresponding encoding DNA sequence recorded within the same oligonucleotides, it is first necessary to couple each individual AEG-based building block stepwise to the oligonucleotide chain. Initially, the AEG-based monomers were coupled to the free amine group of a modified amino-C6-dT base on a resin bound oligonucleotide, as shown in **Figure 5.12**. A short poly-(dT) sequence Res6249 (sequences of oligonucleotides are presented in **Table 5.1**) was chosen to avoid a benzoyl side-reaction (*e.g.,* benzoyl protection group on the other nucleotides could be shifted by an intramolecular reaction with free amino-C6-dT) before a more varied A/T/G/C sequence Res6595 was tested.

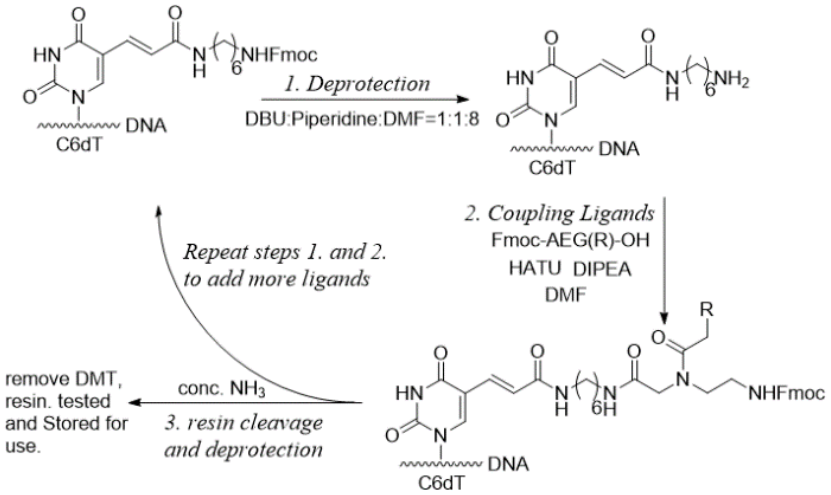



**Figure 5.12** General methods for coupling amino acids into oligonucleotides $C_6$dT.

The coupling reagents HATU, HBTU and PyBOP in the presence of HOBt were tested due to their frequent use in peptide coupling.[266–268] Compound 9a (with a nitrophenyl side chain) was used to test the four coupling conditions in order to standardise and optimise the conditions for the general coupling reactions of all ligands. **Figure 5.13a** shows that the coupling of amino acids to oligonucleotides by HATU was slightly better than for HBTU and PyBOP (conditions i, iii and iv, respectively). The addition of 1 eq. of HOBt did not have a significant effect (condition ii as compared to i) as expected, since HOBt is commonly added (with EDC/DCC) to minimise racemisation of amino acids, although the structure of HATU itself already contains the OBt moiety and so is unlikely to be affected by it.[266] Hence, condition i (using HATU as the coupling reagent) was used as the standard coupling condition for the binding of all ligands.

From the panel of 12 ligands, four were selected (**9a-9d**) to test whether, and to what extent, the structural differences in side chains affects the efficiency of coupling. The yield of compound **9a** (with a nitrophenyl side chain); **9b** (with a trifloro-substituted phenyl side chain); and **9c** (with a phenyl side chain) were all around 80% efficient, whereas the yield of **9d** (with a hydrophobic adamantanyl group) was less than 50% (**Figure 5.13b**). The yields were generally determined by the integration of both products and starting material peaks in HPLC traces (**Figure 5.14**). In the event that the product and starting material coincide on HPLC traces, the entire peak was deconvoluted and the height of product and starting material peaks were compared to measure relative yields (**Figure 5.13**). Although specific values in relation to yields may not be precise owing to the method of measurement, there was a very clear difference in terms of the coupling efficiency between ligands with poly-substituted

aromatic groups and those with hydrophobic adamantanyl groups. These results suggest that ligands with different side chains will possess different coupling efficiencies during the split-and-mix protocol, although ligands were added into the sub-libraries in a stepwise fashion. Although this conclusion will not theoretically affect the basics of sub-library design, it should be taken into consideration that, during the split-and-mix protocol, some ligands (*e.g.,* with hydrophobic adamantanyl groups on their respective side chains) may need to use their special coupling conditions rather than the standardized coupling conditions, as the latter would exhibit poor coupling efficiencies and lead to very low yields of oligo-peptide conjugates after capping.

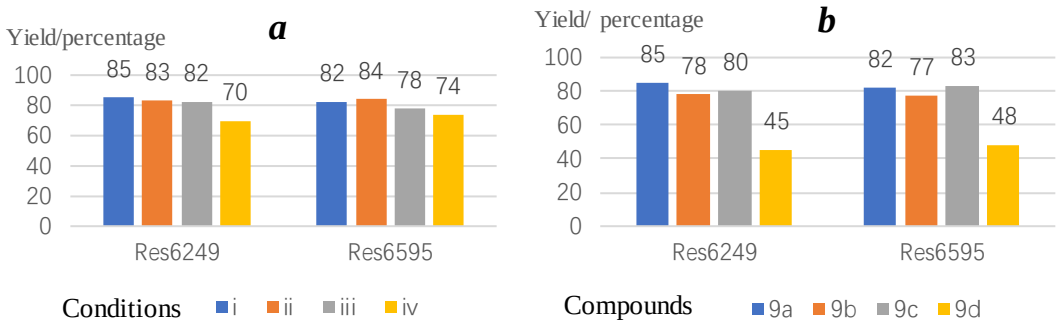



**Figure 5.13** Study of coupling conditions ***a.*** Yields of compound 9a under different coupling conditions: (i) eq. of oligo: amino acid: HATU: DIPEA: HOBt= 1:30:29:90:0 in DMF at 55 °C; (ii) eq. of oligo: amino acid: HATU: DIPEA: HOBt= 1:30:29:90:30 in DMF at 55 °C; (iii) eq. of oligo: amino acid: HBTU: wDIPEA: HOBt= 1:30:29:90:0 in DMF at 55 °C; (iv) eq. of oligo: amino acid: PyBOP: DIPEA: HOBt= 1:30:29:90:30 in DMF at 55 °C; ***b.*** yields of each individual compound (9a-d) incorporated into oligonucleotides under fixed conditions as stipulated in i.

In addition to AEG-based ligands, a Cy3 dye ester was also successfully coupled to the same oligonucleotide using Cy3-NHS-ester (**Table 5.1**), indicating that the synthesis of active esters of AEG monomers might also be useful, while the dye itself could be useful for monitoring reactions and labelling in future studies.

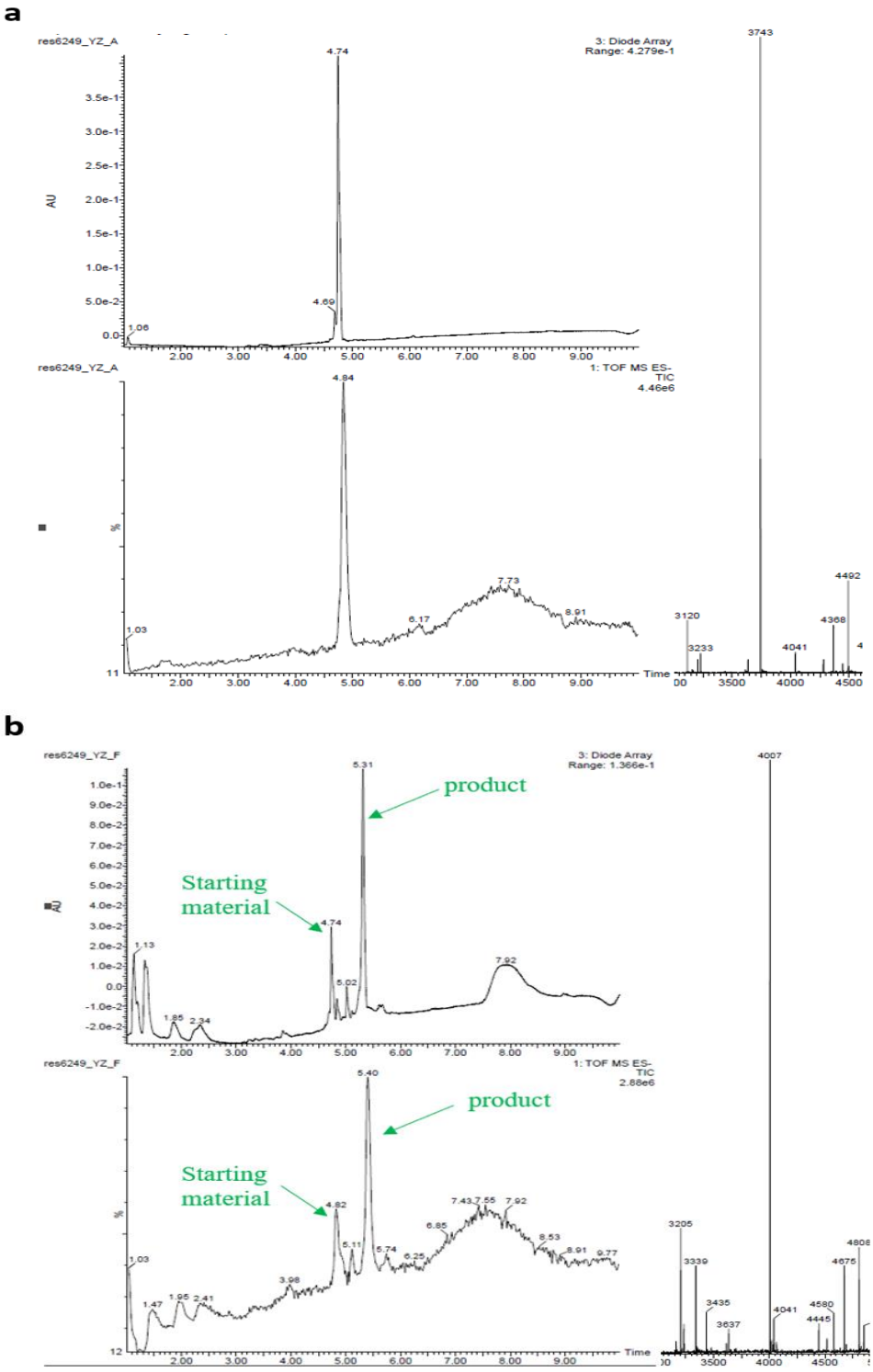



**Figure 5.14** An example (compound **9a**) of MS results obtained for the oligonucleotide (5'-TTTTTT**5(9a)**TTTTT-3') synthesised by general coupling conditions (eq. of oligo: amino acid: HATU: DIPEA: HOBt= 1:30:29:90:0 in DMF at 55 °C). **(a)** Starting material (5'-TTTTTT**5**TTTTT-3'), calculated mass = 3,743; found mass = 3,743; **(b)** product, calculated mass = 4,007; found mass = 4,007. Summary of details of all compounds can be found in **Table 5.1**.

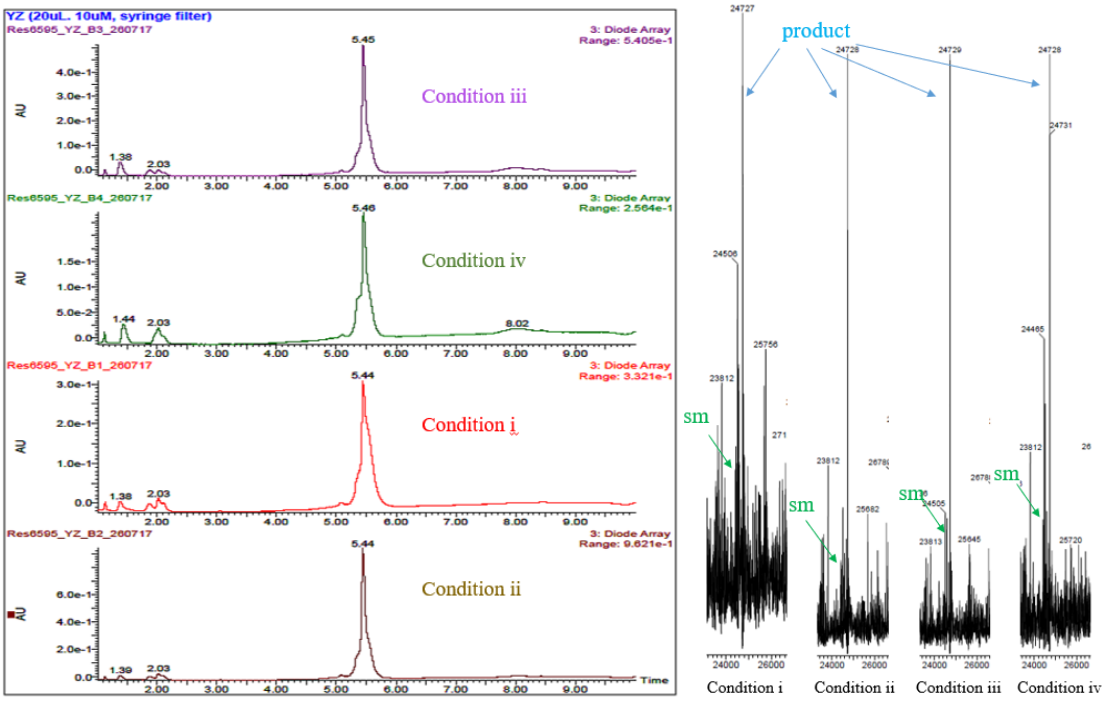



**Figure 5.15** Example of coupling conditions studies for oligonucleotide (Res6595) coupled to compound **5-9a** (calculated mass = 24,727; found mass = 24,727). Four different conditions were compared (labelled "sm" for starting material) and a summary of all coupling conditions used is shown in **Table 2.2** below.

| Oligo name | Sequence (5'-3') | Length | Mass calc. | Mass found |
|---|---|---|---|---|
| Res6249 | TTTTTT**5**TTTTT | 12 | 3742.5 | 3743 |
| Res6249 (9a) | TTTTTT**5(9a)**TTTTT | 12 | 4005.3 | 4007 |
| Res6249 (9b) | TTTTTT**5(9b)**TTTTT | 12 | 4014.3 | 4016 |
| Res6249 (9c) | TTTTTT**5(9c)**TTTTT | 12 | 3960.3 | 3961 |
| Res6249 (9d) | TTTTTT**5(9d)**TTTTT | 12 | 4018.3 | 4019 |
| Res6595 | CATGATATGCTACGCGATACTCAGCACA GGGTTGTCCGATTTTTT**5**TTTTTAAAGTC GTATAGCGCATCGATCATAGCA | 79 | 24463 | 24464 |
| Res6595 (9a) | CATGATATGCTACGCGATACTCAGCACA GGGTTGTCCGATTTTTT**5(9a)**TTTTTAAAG TCGTATAGCGCATCGATCATAGCA | 79 | 24726 | 24727 |
| Res6595 (9b) | CATGATATGCTACGCGATACTCAGCACA GGGTTGTCCGATTTTTT**5(9b)**TTTTTAAA GTCGTATAGCGCATCGATCATAGCA | 79 | 24735 | 24736 |
| Res6595 (9c) | CATGATATGCTACGCGATACTCAGCACA GGGTTGTCCGATTTTTT**5(9c)**TTTTTAAAG TCGTATAGCGCATCGATCATAGCA | 79 | 24681 | 24682 |
| Res6595 (9d) | CATGATATGCTACGCGATACTCAGCACA GGGTTGTCCGATTTTTT**5(9d)**TTTTTAAA | 79 | 24739 | 24741 |

| | | | | |
|---|---|---|---|---|
| | GTCGTATAGCGCATCGATCATAGCA | | | |
| Res6595($Cy_3$) | CATGATATGCTACGCGATACTCAGCACA<br>GGGTTGTCCGATTTTTT**5($Cy_3$)**TTTTTAAA<br>GTCGTATAGCGCATCGATCATAGCA | 79 | 24902 | 24901 |

**Table 5.1.** Sequences of oligonucleotides used in the synthesis of amino acid-linked oligonucleotides (**5** in poly-T regions represents the $C_6$dT amino acid).

| Experimental conditions | Oligonucleotide / eq. | AEG-based building blocks / eq. | Coupling reagents / eq. | DIPEA (eq.) | HOBt (eq.) | Temp. (°C) | **Yield (%)** |
|---|---|---|---|---|---|---|---|
| i | Res6595 / 1 | 9a / 30 | HATU / 29 | 90 | 0 | 55 | **85** |
| i | Res6595 / 1 | 9b / 30 | HATU / 29 | 90 | 0 | 55 | **78** |
| i | Res6595 / 1 | 9c / 30 | HATU / 29 | 90 | 0 | 55 | **80** |
| i | Res6595 / 1 | 9d / 30 | HATU / 29 | 90 | 0 | 55 | **45** |
| ii | Res6595 / 1 | 9a / 30 | HATU / 29 | 90 | 30 | 55 | **83** |
| iii | Res6595 / 1 | 9a / 30 | HBTU / 29 | 90 | 0 | 55 | **82** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| iv | Res6595 / 1 | 9a / 30 | PyBOP / 29 | 90 | 30 | 55 | **70** |
| i | Res6249 / 1 | 9a / 30 | HATU / 29 | 90 | 0 | 55 | **82** |
| i | Res6249 / 1 | 9b / 30 | HATU / 29 | 90 | 0 | 55 | **77** |
| i | Res6249 / 1 | 9c / 30 | HATU / 29 | 90 | 0 | 55 | **83** |
| i | Res6249 / 1 | 9d / 30 | HATU / 29 | 90 | 0 | 55 | **48** |
| ii | Res6249 / 1 | 9a / 30 | HATU / 29 | 90 | 30 | 55 | **84** |
| iii | Res6249 / 1 | 9a / 30 | HBTU / 29 | 90 | 0 | 55 | **78** |
| iv | Res6249 / 1 | 9a / 30 | PyBOP / 29 | 90 | 30 | 55 | **74** |

**Table 5.2.** Details of each set of coupling conditions.

## 5.5 Conclusions and Further works

SABL libraries were constructed by combining the concepts of DEL chemical diversity, aptameric biomolecular interactions, and multivalency. By this approach, oligo-peptide-conjugated sub-libraries were designed with several specialised regions including target binding (for modified ligands), ligand encoding, and complementary regions for ligand pairing. Special ligand building blocks for SABL libraries were designed with the same Fmoc-protected AEG-based backbone but incorporating different types of side chain including halogen bond donors, basic groups, poly-substituted aromatic groups, lipophilic groups, etc.

In total, 12 different Fmoc-protected AEG-based amino acids (ligand building blocks) were successfully synthesised from ethylene diamine with an overall yield of 35% to 45%. Among these 12 different building blocks with distinctive side chains, four were used to test their coupling efficiencies to oligonucleotides. Of the various conditions considered, HATU gave the best results and hence was chosen as the general coupling method for further use in making SABL sub-libraries. These methods are easily transferable to other building blocks and should allow the rapid construction of future libraries.

In future, all 12 building blocks rather than 4 will be tested for their corresponding coupling conditions that could achieve highest coupling efficiency. Afterwards, a true sub-library will be prepared by coupling AEGs and concurrently encoding the corresponding DNA through split-and-mix method (**Figure 5.16**). The oligo structures

of sub-libraries constructed as twin prerequisites for a universal DNA scaffold exhibited (i) sufficient structural stability to maintain a fixed distance between the two attached ligands; and (ii) the ability to denature during PCR to allow amplification and subsequent sequencing. Therefore, a DNA hairpin stretch of about 20 base pairs fits these criteria perfectly.

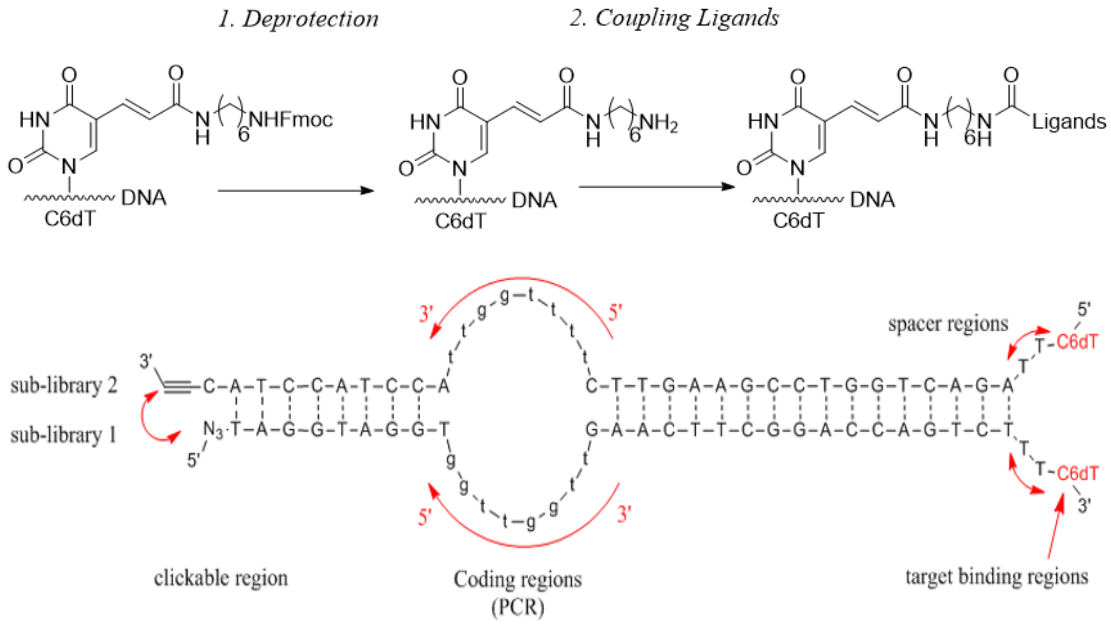



**Figure 5.16** For sub-library 1, oligonucleotide synthesis will be in the 3'- to 5'- direction, with sub-library 2 synthesis in the reverse direction. When the two sub-libraries are mixed, their cDNA regions will enable the DNA templates to be assembled spontaneously to form the desired bivalent ligand libraries. Ligands were attached to the C6dT region.

After forming suitable SABL sub-libraries, SABLs will be tested against biologically relevant targets (*e.g.,* thrombin and Her2), with the optimisation of the separation of good binders from bad being a crucial step. This will then be followed by biophysical studies to establish relative affinities and the role of the various "non-natural" functional groups in the ligand selected.

# *Chapter 6.*

# *Conclusions and Further work*

# *Chapter 6. Conclusions and Further work*

The aim of the project is to construct a new class of ligands to satisfy the increasing requirement for functional diversity and higher affinities when using nucleic acids for the aptamer selection process. This will be achieved by introducing "non-natural" functionality including nitro groups, halogens, poly-substituted aromatic groups, and bulky lipophilic groups into the ligand system. Ligands with these "non-natural" functionality will thus provide a wider range of supramolecular interactions that can be further explored in target binding studies.

In one project, modified 5-aminomethyl-deoxyuridines (5-AM-dU) comprising five analogues were designed and synthesised. Functional groups attached to 5-AM-dU through coupling reactions are predominately hydrophobic or aromatic. Both the phosphoramidite and nucleoside triphosphate forms of these 5-AM-dU were synthesized. All of six 5-AM-dU analogues were tested and proved to be compatible with DNA polymerases and have since been successfully incorporated into DNA, along with natural dNTPs, through the PEX and PCR.

One application for these various 5-AM-dU analogues was their introduction into a system of qAN1 molecule to improve its PCR efficiency. This application for 5-AM-dU moieties lend strong support to the hypothesis that the low PCR efficiency observed with qAN1 molecule arises mainly due to its strongly duplex-stabilising effects. The successful incorporation of modified 5-AM-dU analogues and qAN1 molecules

simultaneously in PCR amplification enabled the full substitution of qAN1. Through this method, a chemically modified DNA amplicon with full substitution of qAN1 for dA and amU2/amU3 for dT was achieved through PCR amplification.

Following this, these chemically modified 5-AM-dU analogues were introduced into SELEX approaches to form a series of new aptamers through *in vitro* selection. ssDNA libraries containing modified 5-AM-dU analogues were synthesised. These libraries of oligonucleotide sequences underwent 10 rounds of SELEX with Her2 proteins as binding targets, and the final sequences were sent for DNA sequencing. Base composition analysis of NGS results proves there are no significant selection pressure against 5-AM-dU analogues during SELEX. Unique sequence enrichment analysis gives 10 unique sequences that enriched the most during SELEX and could be designed as potential aptamer candidates for further target validation test. Due to time constraints, this project temporary stopped at this step. In future, binding affinity tests with Her2 target proteins using SPR will be assigned for these selected DNA aptamers. If any aptamer that contained modified 5-AM-dU analogues is found to possess a relatively high $K_d$ value, further studies of the interactions between the functional groups on 5-AM-dU analogues in that aptamer with the target protein will be conducted.

In a parallel project, the concept of an SABL library construction was achieved by combining the advantages of DEL chemical diversity, aptameric biomolecular interactions, and multivalency. In total, 12 different ligand building blocks for SABL libraries were newly designed and successfully synthesised with the same Fmoc-protected AEG-based backbone but bearing different types of side chain, including:

halogen bond donors, basic groups, poly-substituted aromatic rings, lipophilic groups, *etc*. In SABL libraries, all of these ligands are encoded with DNA barcodes and hence do not need to be recognised by DNA polymerases. Such sub-libraries will be prepared by coupling these building blocks and encoding them using the corresponding DNA at the same time via the 'split-and-mix' method. After forming the SABL sub-libraries, SABLs will be tested against biologically relevant targets to select good binders. This will then be followed by biophysical studies to establish the affinity and role of the various "non-biological" functional groups within the ligand selected.

In these projects, a series of "non-natural" functional groups were attached into DNA nucleobases using enzymatic (polymerase-based) approaches and DNA libraries (AEG-based ligands). Through the incorporation of these functional groups, a more comprehensive structural and chemical understanding of target binding of these "non-natural" functional groups was established. Through the incorporation of these functional groups, the increase of functional diversity of modified nucleic acids (e.g., aptamers) was achieved and the affinities of such modified nucleic acids for their targets could also be improved using this method.

# *Chapter 7.*

# *Experimental*

## 7.1 General Information

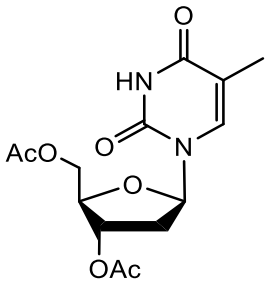
All starting materials were obtained either from Sigma-Aldrich or Fluorochem and used without any additional purification. The following solvents, DCM, THF and pyridine were taken directly from the communal MBraun SPS-800 solvent drying systems. The TLC plates used were 60G F254 (Merck) which are pre-coated silica gels containing fluorescence indicator F254. Most of compounds formed were thereby revealed under short-wave UV illumination, whereas some containing free amines were detected by ninhydrin. Purifications were performed via silica gel (40 – 60 μm, SDS) column chromatography.

$^{1}H$ NMR spectra were measured either by Bruker Avance-400 MHz spectrometers or Bruker Avance-500 MHz spectrometers. $^{13}C$ NMR spectra were measured at 101 MHz using Bruker Avance-400 MHz spectrometers. Chemical shifts (δ) in NMR were reported in parts per million (ppm) units downfield from TMS (δ = 0.00) and the deuterated solvents used were chloroform-d ($CDCl_3$, δ = 7.26 ppm for $^{1}H$ and δ = 77.0 ppm for $^{13}C$) and DMSO-d6 (δ = 2.50 ppm for $^{1}H$ and δ = 39.5 ppm for $^{13}C$). Coupling constants (J) were given in Hertz (Hz). The FTIR/ATR spectra were obtained by means of a PerkinElmer Spectrum 100 machine. Melting points were tested using a Buchi-B510 melting point apparatus. Low-resolution mass spectra (LRMS) were tested in HPLC grade methanol and recorded using electrospray ionisation ($ESI^+$ or $ESI^-$) on a Waters ZMD quadrupole mass spectrometer. High-resolution mass spectra (HRMS) were tested using HPLC grade methanol and electrospray ionisations (EI) on a Bruker APEX III FT-ICR mass spectrometer. Data were processed using MaxEnt. Gel pictures

were taken using a Syngene G: BOX with gene-snap imagine software.

# 7.2 Organic Synthesis section

## 7.2.1 Experiments conducted for Chapters 2, 3 and 4



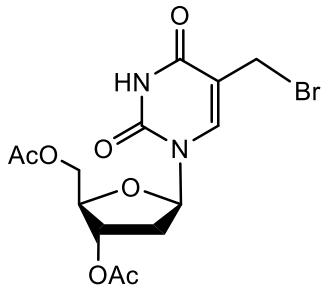

Chemical Formula: $C_{14}H_{18}N_2O_7$
Exact Mass: 326.11
Molecular Weight: 326.31

***Synthesis of 3',5'-O-diacetylthymidine (1)***

Thymidine (10.00 g, 41.3 mmol, 1.0 eq) was dissolved in 50.0 mL anhydrous pyridine. Acetic anhydride (12.60 g, 124 mmol, 3.0 eq) was then added drop wise into this solution and the resulting mixture was stirred at room temperature for 6 hrs. The mixture was concentrated on a rotary evaporator under vacuum to approximately 1/3 of its original volume to remove most of the pyridine. The mixture was then diluted with 150 mL of EtOAc and washed with diluted HCl three times for extraction. The organic layer extracted was then washed with water and brine and dried over $Na_2SO_4$. This resulting solution was concentrated on a rotary evaporator under reduced pressure to yield a colourless crystalline material which formed a white solid (12.33g, 92%) after trituration with hexanes (Rf = 0.60, DCM).

1H NMR (400 MHz, CDCl3, 25 °C): δ = 7.19 (s, 1H), 5.85 (t, *J* = 7.0 Hz, 1H), 5.58 (t, *J* = 7.0 Hz, 1H), 4.92 (m, 1H), 4.40-4.05 (m, 2H), 2.44 (m, 1H), 2.34 (s, 3H), 2.19 (m,

1H), 2.04 (s, 3H), 2.02 (s, 3H). 13C NMR (100 MHz, CDCl3, 25 °C): δ = 170.2, 163.7, 150.8, 136.0, 93.9, 87.2, 74.4, 62.1, 37.9, 21.0, 12.4. HRMS (ESI+) was calculated for C14H18N2O7Na: 349.1011 and found to be 349.1012.

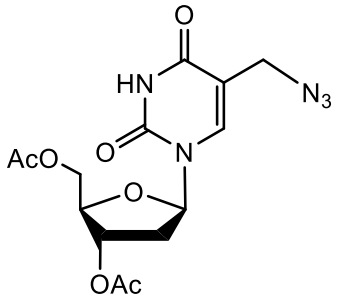



Chemical Formula: $C_{14}H_{17}BrN_2O_7$
Exact Mass: 404.02
Molecular Weight: 405.20

***Synthesis of 3',5'-O-diacetyl-5-(bromomethyl)-2'-deoxyuridine (2)***

Solid N-bromosuccinamide (NBS) (8.20 g, 46.0 mmol, 1.5 eq) was added to a solution of **1** (10.00 g, 30.6 mmol, 1 eq) in dry CDCl3 (Aldrich, 80 mL), followed by addition of azabisisobutyronitrile (AIBN) (750 mg, 4.6 mmol, 0.15 eq). The mixture was heated to reflux at 75 ºC under nitrogen for 4 hr. The reaction was concentrated under reduced pressure on a rotary evaporator while preventing the introduction of moisture by releasing the vacuum with dry nitrogen. The residue that was left after concentration was diluted with EtOAc (50 mL × 2) and repeatedly concentrated using the rotary evaporator. This mixture was purified by column chromatography over silica gel (50% EtOAc/hexane) to produce an orange solid 2 (9.68 g) with a 78% yield (Rf = 0.40 in 1:1 EtOAc/hexane).

1H NMR (400 MHz, CDCl3, 25 °C): δ = 7.19 (s, 1H), 5.85 (t, *J* = 7.0 Hz, 1H), 5.58 (t, *J* = 7.0 Hz, 1H), 4.92 (m, 1H), 4.40-4.05 (m, 2H), 3.97 (s, 2H), 2.44 (m, 1H), 2.19 (m, 1H), 2.04 (s, 3H), 2.02 (s, 3H). 13C NMR (100 MHz, CDCl3, 25 °C): δ = 170.2, 163.7, 150.8, 136.0, 93.9, 87.2, 74.4, 62.1, 37.9, 33.1, 21.0. HRMS (ESI+) calculated for C14H17BrN2O7Na: 427.0116; and found to be 427.0119.

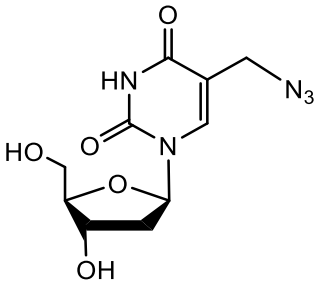



Chemical Formula: $C_{14}H_{17}N_5O_7$
Exact Mass: 367.11
Molecular Weight: 367.32

***Synthesis of 3',5'-O-diacetyl-5-(azidomethyl)-2'-deoxyuridine (3)***

Orange solid **2** (6.00 g, 15 mmol, 1.0 eq) was dissolved in 70 mL anhydrous acetonitrile and sodium azide (1.5 g, 45 mmol, 1.5 eq) was then added into the solution. The mixture was heated to 70 ºC, stirring for 3 hrs. The mixture was allowed to cool to room temperature and then diluted with EtOAc (50 mL). The mixture was washed with water (25 mL x 2) and brine. The organic layer was separated, dried over Na2SO4 and evaporated under *vacuo*. The new non-polar spot was isolated by column chromatography over silica gel using an initial gradient of 50% EtOAc/hexane followed by 80% of the same solvent mixture to produce a yellow solid **3** (4.65 g) with an 83% yield (Rf = 0.85, 8:2 EtOAc/hexane).

1H NMR (400 MHz, CDCl3, 25 °C): δ = 7.19 (s, 1H), 5.85 (t, **J** = 7.0 Hz, 1H), 4.40 (t, **J** = 7.0 Hz, 1H), 3.94 (m, 1H), 3.85 (m, 1H), 3.55 (m, 2H), 2.35 (m, 1H), 2.12 (m, 1H), 2.04 (s, 3H), 2.02 (s, 3H), 2.00 (s, 2H). 13C NMR (100 MHz, CDCl3, 25 °C): δ = 170.2, 163.7, 150.8, 136.0, 93.9, 87.2, 74.4, 62.1, 50.0, 37.9, 21.0. MS (ESI+) m/z calculated for C14H18N5O7: 368.1206 and found to be 368.1210.

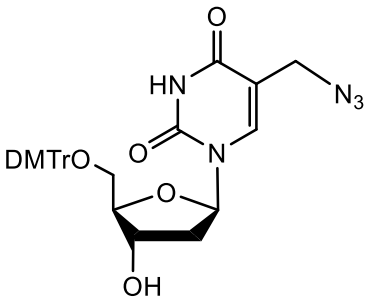



Chemical Formula: $C_{10}H_{13}N_5O_5$
Exact Mass: 283.09
Molecular Weight: 283.24

***Synthesis of 5-(azidomethyl)-2'-deoxyuridine (4)***

Yellow solid **3** (3.50 g, 10 mmol, 1.0 eq) was dissolved in 30 mL ammonia dissolved in 7N methanol. The mixture was stirred for 3 hrs under constant pressure. The resulting mixture was concentrated under reduced pressure on a rotary evaporator. This mixture was subsequently purified by column chromatography over silica gel (50% EtOAc/hexane) to produce an orange solid **4** (2.50 g) with a 94% yield (Rf = 0.45, 1:1 EtOAc/hexane).

1H NMR (400 MHz, CDCl3, 25 °C): δ = 7.19 (s, 1H), 5.85 (t, **J** = 7.0 Hz, 1H), 4.92 (m, 1H), 4.40-4.05 (m, 2H), 2.44 (m, 1H), 2.19 (m, 1H), 2.04 (s, 3H), 2.02 (s, 3H), 2.00 (s, 2H). 13C NMR (100 MHz, CDCl3, 25 °C): δ = 163.7, 150.8, 137.5, 109.7, 94.0, 89.1, 70.5, 61.3, 52.0. MS (ESI+) m/z calculated for C10H13N5O5Na: 306.0814 and found to be 306.0812.

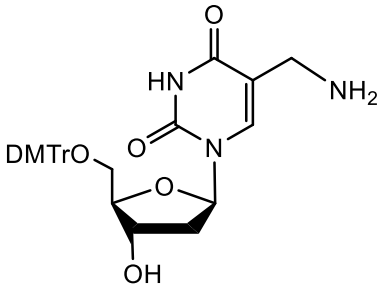



Chemical Formula:
$C_{31}H_{31}N_5O_7$
Exact Mass: 585.22
Molecular Weight: 585.62

***Synthesis of 5'-O-(4,4'-Dimethoxytrityl)-5-(azidomethyl)-2'-deoxyuridine (5)***

Orange solid **4** (2.00 g, 7.20 mmol, 1.0 eq) was dissolved in 20 mL anhydrous pyridine. A mixture of DMTr-Cl (2.90 g, 8.60 mmol, 1.2 eq) in 20 mL anhydrous pyridine was added drop wise to the solution and the mixture was stirred at room temperature for 4 hrs and the reaction was then quenched by adding 20mL MeOH. The mixture was concentrated on a rotary evaporator under vacuum to approximately 1/3 volume to remove most of pyridine before dilution with DCM. The organic layer was washed with aqueous $NaHCO_3$ (saturated) and brine, dried over $Na_2SO_4$, and concentrated under reduced pressure on a rotary evaporator. This mixture was purified by column chromatography over silica gel (10% MeOH/DCM) to give a light yellow solid **5** (2.70 g) with a 65% yield (Rf = 0.75, 1:1 EtOAc/hexane).

$^1$H-NMR (400 MHz, CDCl3, 25 °C): δ = 8.31 (1H, s), 7.48–6.87 (13H, m), 5.85 (1H, t, J = 7.0 Hz), 4.61 (1H, m,), 4.11 (1H, m,), 3.80 (6H, s), 3.51 (2H, m), 3.40 (2H, m), 2.44 (1H, m), 2.33 (1H, m); $^{13}$C-NMR (400 MHz, CDCl3, 25 °C): δ = 162.3, 158.8, 149.9, 144.2, 138.9, 135.2, 130.1, 128.1, 127.3, 113.3, 109.8, 87.1, 86.2, 85.0, 72.3, 63.3, 55.3, 46.6, 41.2. MS (ESI+) m/z calculated for $C_{31}H_{31}N_5O_7Na$: 608.2121 and found to be 608.2118.

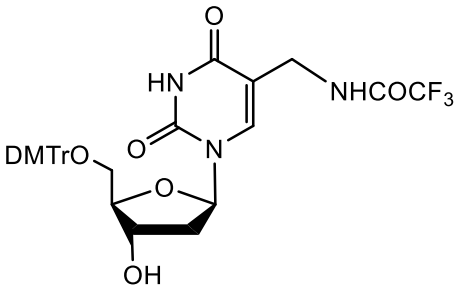



Chemical Formula: $C_{31}H_{33}N_3O_7$
Exact Mass: 559.23
Molecular Weight: 559.62

***Synthesis of 5'-O-(4,4'-Dimethoxytrityl)-5-(aminomethyl)-2'-deoxyuridine (6)***

A mixture of **5** (2.00 g, 3.45 mmol, 1.0 eq) and Pd/C (10%, 73 mg, 0.17 mmol, 0.05 eq) in 10 mL MeOH was stirred under $H_2$ at atmospheric pressure at room temperature. After 2 hrs, the catalyst was filtered off with Celite. The filtrate was evaporated under reduced pressure. This mixture was then purified by column chromatography over silica gel (5% to 10% MeOH/DCM) to give a white solid 6 (1.70 g) with an 88% yield (Rf = 0.20, 1:9 MeOH/CHCl3).

$^1$H-NMR (400 MHz, DMSO-$d_6$) δ 8.75 (2H, br), 7.48–6.82 (14H, m), 6.02 (1H, d, $J$ = 6.8), 5.31 (1H, d, $J$ = 7.0), 4.23 (1H, m), 3.87 (1H, m), 3.72 (6H, s), 3.17 (2H, m), 3.07 (2H, m), 2.17 (2H, m); $^{13}$C-NMR (100 MHz, DMSO-$d_6$) δ = 163.2, 158.1, 130.2, 144.7, 135.5, 129.7, 127.8, 127.7, 126.7, 116.0, 113.2, 100.8, 85.7, 85.4, 83.9, 79.1, 70.5, 55.0. MS (ESI+) m/z calculated for $C_{31}H_{33}N_3O_7Na$: 582.2216 and found to be 582.2224.

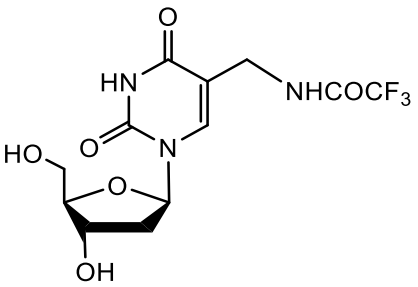



Chemical Formula: $C_{33}H_{32}F_3N_3O_8$
Exact Mass: 655.21
Molecular Weight: 655.63

***Synthesis of 5'-O-(4,4'-Dimethoxytrityl)-5-[N-(trifluoroacetyl)-aminomethyl]-2'-deoxyuridine (7)***

Ethyl trifluoroacetate (0.35 mL, 3.0 mmol, 1.2 eq) was added to a solution of **6** (1.40 g, 2.50 mmol, 1.0 eq) in triethylamine (0.42 mL, 3.0 mmol, 1.2 eq) and 25 mL anhydrous DCM at room temperature. The mixture was stirred at room temperature for 2 hrs. The mixture was concentrated on a rotary evaporator under vacuum and purified by column chromatography over silica gel (5% to 10% MeOH/DCM) to give a white solid **7** (1.51 g) with a 95% yield (Rf = 0.5, 1:9 MeOH/CHCl3).

$^1$H-NMR (400 MHz, DMSO-$d_6$, 25 °C): δ = 11.51 ( 1H, s), 9.42 (1H, s), 7.67 (1H, s), 7.39–6.85 (13H, m), 6.17 (1H, d, $J$ = 6.8 Hz), 5.35 (1H, d, $J$ = 4.4 Hz), 4.23 (1H, t, $J$ = 5.0 Hz), 3.87 (1H, m), 3.71 (6H, s), 3.64 (2H, m), 3.18 (2H, m), 2.18 (2H, m); $^{13}$C-NMR (100 MHz, DMSO-$d_6$, 25 °C): δ = 162.5, 158.1, 156.1, 150.1, 144.7, 139.5, 135.3, 129.7, 127.9, 127.7, 126.7, 113.2, 108.6, 85.8, 85.4, 84.2, 79.2, 70.3, 63.8, 55.0. MS (ESI+) m/z calculated for $C_{33}H_{32}N_3O_8F_3Na$: 678.2039 and found to be 678.2043.

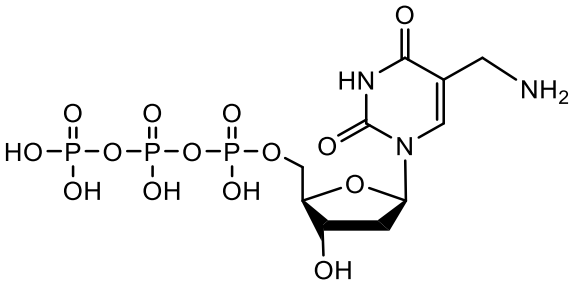



Chemical Formula: $C_{12}H_{14}F_3N_3O_6$
Exact Mass: 353.08
Molecular Weight: 353.25

***Synthesis of 5-[N-(trifluoroacetyl)-aminomethyl]-2'-deoxyuridine (8)***

White solid **7** (1.00 g, 1.50 mmol) was dissolved in 4:1 parts of acetic acid/water (5 mL) and stirred at room temperature for 30 min. The mixture was concentrated on a rotary evaporator under vacuum and purified by column chromatography over silica gel (80% EtOAc/Hexane) to give a white solid **8** (0.42 g) with an 80% yield (Rf = 0.31; 1:9 MeOH/DCM).

$^{1}$H NMR (400 MHz, CDCl3, 25 °C): δ = 9.24 (s, 1H), 8.23 (s, 1H), 7.76–7.67 (m, 1H), 6.26 (dd, *J* = 7.8, 6.1 Hz, 1H), 5.39–5.34 (m, 1H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.8 Hz, 1H), 3.89 (dd, *J* = 6.1 Hz, 1H), 2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100 MHz, CDCl3, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 141.2, 115.7, 109.3, 86.4, 86.0, 75.2, 62.4, 38.3, 36.5, 20.9. MS (ESI+) m/z calculated for C12H14F3N3O6Na: 376.0732; actual 376.0736.



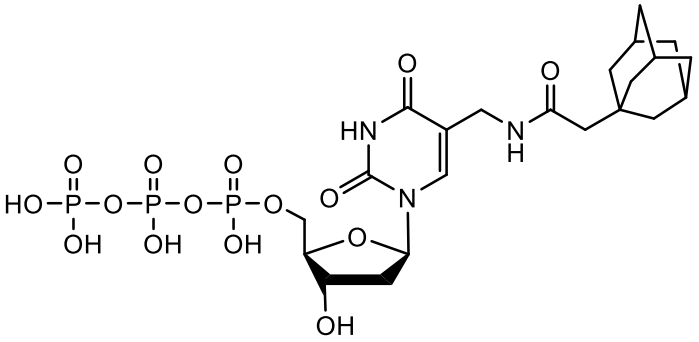
Chemical Formula: $C_{10}H_{18}N_3O_{14}P_3$
Exact Mass: 497.00
Molecular Weight: 497.18

***Synthesis of 5-(aminomethyl)-2'-deoxyuridine triphosphate (9)***

White solid nucleoside **8** (100.0 mg, 280 μmol, 1.0 eq) was dried by coevaporation with pyridine overnight. Then this nucleoside 8 was suspended in dioxane (1.5 mL) and pyridine (0.5 mL) under $N_2$. A solution of 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one (85.0 mg, 420 μmol, 1.5 eq) in dioxane (0.2 mL) was added to the mixture. The reaction mixture was stirred vigorously for 1 hr. A well-mixed solution of tri-n-butylammonium pyrophosphate (900.0 mg, 1400 μmol, 5.0 eq) in tri-n-butylamine (400 μL, 1120 μmol, 4.0 eq) and 2 mL anhydrous DMF was then added, followed by an additional 3 hrs of stirring before 2% $I_2$ in pyridine/$H_2O$ (98:2) solution (3 mL) was added. The solution was mixed periodically over 20 min, and excess $I_2$ was quenched by the dropwise addition of 5 mL aqueous sodium bisulfite solution (5 %). The reaction was concentrated on a rotary evaporator under vacuum to dryness. The resulting solid was then resuspended in $H_2O$ (5 mL) and ammonium hydroxide (10 mL, 30 %) was added to the mixture. After stirring at room temperature for an additional 3 hrs in 55 °C under constant pressure, the solution was concentrated to dryness. The product was first purified via anion-exchange HPLC (eluent A: 10 mM TEAB buffer, B: 1 M TEAB buffer, 0-80 % buffer B in 40 min) and then through RP-HPLC (eluent A: 0.1 M TEAB buffer, B: 40 % acetonitrile in 0.1 M TEAB buffer, 3.5-35 % buffer B in 20 min).

Compound **9** was obtained as a white solid with a 23 % yield (32 mg, 64 μmol). $R_f$: 0.25 [n-propanol: ammonium hydroxide: $H_2O$=6:1:4 v/v/v].

$^{1}$H NMR (400MHz, $D_2O$, 25 °C): δ = 8.23 (s, 1H), 6.26 (t, *J* = 6.1 Hz, 1H), 4.65–4.50 (m, 2H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.4, 3.3 Hz, 1H), 3.89 (dd, *J* = 7.4, 3.3 Hz, 1H), 2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100MHz, $D_2O$, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 145.1, 98.6, 97.6, 85.5, 85.3, 69.6, 64.7, 39.5, 30.0 ppm. $^{31}$P NMR (100MHz, $D_2O$, 25 °C) δ -5.36 (d, $^{2}J_{PP}$ = 21.0 Hz), -10.42 (d, $^{2}J_{PP}$ =19.2 Hz), -21.40 (t, $^{2}J_{PP}$ = 20.0 Hz) ppm. MS (HPLC) m/z calculated for $C_{10}H_{18}N_3O_{14}P_3$: 497.2 and found to be 495.6 (M-1)-. λmax = 265 nm.



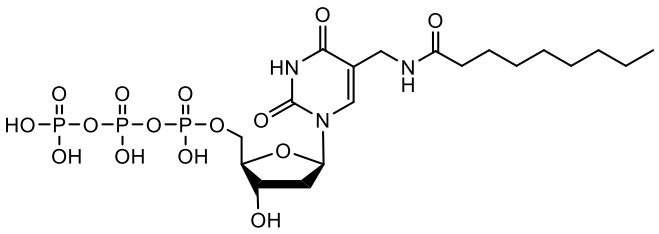
Chemical Formula: $C_{22}H_{34}N_3O_{15}P_3$
Exact Mass: 673.12
Molecular Weight: 673.44

***Synthesis of 5-[(2-(adamantan-1-yl)-acetamido) methyl]-2'-deoxyuridine triphosphate (10)***

A mixture of adamantic acid (1 mg, 5 μmol, 1.2 eq), HATU (2 mg, 5 μmol, 1.2 equiv) in 40 μL anhydrous DMF was added to a solution of compound **9** (2 mg, 4 μmol, 1.0 eq) in aqueous sodium bicarbonate (80 μL, 0.4 M). The resulting mixture was agitated in 37 °C for 3 hrs and was then freeze dried. The crude solid was washed with anhydrous DMF and resuspended in water. The product was purified first through anion-exchange HPLC (eluent A: 10 mM TEAB buffer, B: 1 M TEAB buffer, 0-80 % buffer B for 40 min) and then through RP-HPLC (eluent A: 0.1 M TEAB buffer, B: 40 % acetonitrile in 0.1 M TEAB buffer, 3.5-35 % buffer B for 20 mins). Compound **10** was obtained in white solid form with a 75 % yield (1.98 mg, 3 μmol). $R_f$: 0.35 [n-propanol: ammonium hydroxide: $H_2O$=6:1:4 v/v/v].

$^1$H NMR (400MHz, $D_2O$, 25 °C): δ = 8.23 (s, 1H), 6.26 (t, *J* = 6.1 Hz, 1H), 4.65–4.50 (m, 2H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.4, 3.3 Hz, 1H), 3.89 (dd, *J* = 7.4, 3.3 Hz,1H),

2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100MHz, $D_2O$, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 145.1, 98.6, 97.6, 85.5, 85.3, 69.6, 64.7, 39.5, 30.0 ppm. $^{31}$P NMR (100MHz, $D_2O$, 25 °C) δ -5.36 (d, $^2J_{PP}$ = 21.0 Hz), -10.42 (d, $^2J_{PP}$ =19.2 Hz), -21.40 (t, $^2J_{PP}$ = 20.0 Hz) ppm. MS (HPLC): m/z calculated for $C_{22}H_{34}N_3O_{15}P_3$: 673.4 and found to be 672.5 (M-1)-. λmax = 268 nm.



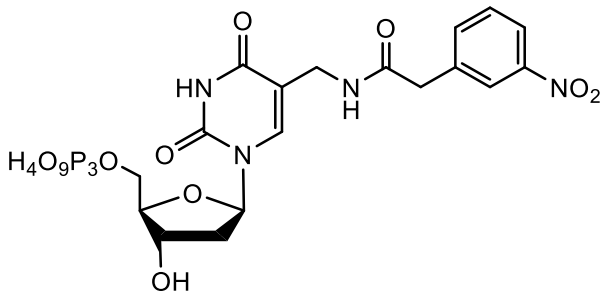
Chemical Formula: $C_{19}H_{34}N_3O_{15}P_3$
Exact Mass: 637.12
Molecular Weight: 637.41

***Synthesis of 5-(nonanamidomethyl)-2'-deoxyuridine triphosphate (11)***

A mixture of nonacetic acid (1 mg, 5 μmol, 1.2 eq), HATU (2 mg, 5 μmol, 1.2 equiv) in 40 μL anhydrous DMF was added to a solution of compound **9** (2 mg, 4 μmol, 1.0 eq) in aqueous sodium bicarbonate (80 μL, 0.4 M). The mixture was agitated at 37 °C for 3 hrs and then freeze dried. The crude solid was washed with DMF anhydrous and resuspended with water. The product was purified first via anion-exchange HPLC (eluent A: 10 mM TEAB buffer, B: 1 M TEAB buffer, 0-80 % buffer B for 40 min) and then through RP-HPLC (eluent A: 0.1 M TEAB buffer, B: 40 % acetonitrile in 0.1 M TEAB buffer, 3.5-35 % buffer B for 20 min). Compound **11** was obtained in white solid form with a 70 % yield (1.76 mg, 3 μmol). $R_f$: 0.35 [n-propanol: ammonium hydroxide: $H_2O$=6:1:4 v/v/v].

$^1$H NMR (400MHz, $D_2O$, 25 °C): δ = 8.23 (s, 1H), 6.26 (t, *J* = 6.1 Hz, 1H), 4.65–4.50 (m, 2H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.3, 4.1 Hz, 1H), 3.89 (dd, *J* = 7.3, 4.1 Hz, 1H), 2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100MHz, $D_2O$, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 145.1, 98.6, 97.6, 85.5, 85.3, 69.6, 64.7, 39.5, 30.0 ppm. $^{31}$P NMR (100MHz, $D_2O$, 25 °C) δ -5.36 (d, $^2J_{PP}$ = 21.0 Hz), -10.42 (d, $^2J_{PP}$

=19.2 Hz), -21.40 (t, $^{2}J_{PP}$ = 20.0 Hz) ppm. MS (HPLC): m/z calculated for $C_{19}H_{24}N_3O_{15}P_3$: 637.4 and found to be 636.2 (M-1)-. λmax = 268 nm.



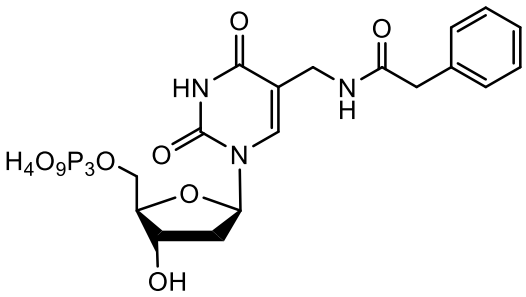

Chemical Formula: $C_{18}H_{23}N_4O_{17}P_3$
Exact Mass: 660.03
Molecular Weight: 660.31

***Synthesis of 5-[(3-(nitrophenyl)-acetamido) methyl]-2'-deoxyuridine triphosphate (12)***

A mixture of 4-nitrophenylacetic acid (1 mg, 5 μmol, 1.2 eq), HATU (2 mg, 5 μmol, 1.2 equiv) in 40 μL anhydrous DMF was added to a solution of compound **9** (2 mg, 4 μmol, 1.0 eq) in aqueous sodium bicarbonate (80 μL, 0.4 M). The mixture was agitated in 37 °C for 3 hrs and then freeze dried. The crude solid was washed with anhydrous DMF and resuspended in water. The product was purified first via anion-exchange HPLC (eluent A: 10 mM TEAB buffer, B: 1 M TEAB buffer, 0-80 % buffer B for 40 mins) and then through RP-HPLC (eluent A: 0.1 M TEAB buffer, B: 40 % acetonitrile in 0.1 M TEAB buffer, 3.5-35 % buffer B for 20 mins). Compound **12** was obtained in white solid form with a 75 % yield (1.83 mg, 3 μmol). $R_f$: 0.20 [n-propanol: ammonium hydroxide: $H_2O$=6:1:4 v/v/v].

$^1$H NMR (400MHz, $D_2O$, 25 °C): δ = 8.23 (s, 1H), 6.26 (t, *J* = 6.1 Hz, 1H), 4.65–4.50 (m, 2H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.6, 3.4 Hz, 1H), 3.89 (dd, *J* = 7.6, 3.4 Hz, 1H), 2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100MHz, $D_2O$, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 145.1, 98.6, 97.6, 85.5, 85.3, 69.6, 64.7, 39.5, 30.0 ppm. $^{31}$P NMR (100MHz, $D_2O$, 25 °C) δ -5.36 (d, $^2J_{PP}$ = 21.0 Hz), -10.42 (d, $^2J_{PP}$

=19.2 Hz), -21.40 (t, $^{2}J_{PP}$ = 20.0 Hz) ppm. MS (HPLC): m/z calculated for $C_{18}H_{23}N_3O_{17}P_3$: 660.3 and found to be 659.3 (M-1)-. λmax = 270 nm.

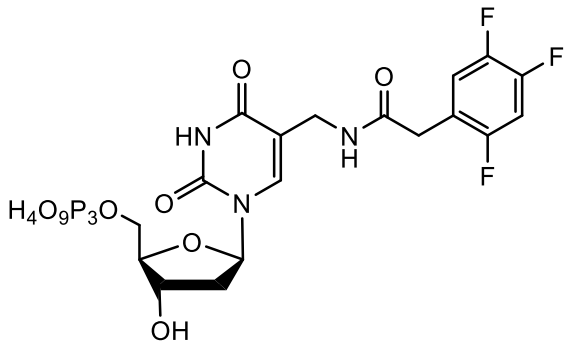



Chemical Formula: $C_{18}H_{24}N_3O_{15}P_3$
Exact Mass: 615.04
Molecular Weight: 615.32

***Synthesis of 5-[(2-phenyl-acetamido) methyl]-2'-deoxyuridine triphosphate (13)***

A mixture of phenylacetic acid (1 mg, 5 μmol, 1.2 eq) and HATU (2 mg, 5 μmol, 1.2 equiv) in 40 μL anhydrous DMF was added to a solution of compound **9** (2 mg, 4 μmol, 1.0 eq) in aqueous sodium bicarbonate (80 μL, 0.4 M). The mixture was then agitated in 37 °C for 3 hrs and then freeze dried. The crude solid was washed with anhydrous DMF and resuspended in water. The product was purified first via anion-exchange HPLC (eluent A: 10 mM TEAB buffer, B: 1 M TEAB buffer, 0-80 % buffer B for 40 min) and then through RP-HPLC (eluent A: 0.1 M TEAB buffer, B: 40 % acetonitrile in 0.1 M TEAB buffer, 3.5-35 % buffer B for 20 min). Compound **13** was obtained in white solid form with a 75 % yield (1.85 mg, 3 μmol). $R_f$: 0.20 [n-propanol: ammonium hydroxide: $H_2O$=6:1:4 v/v/v].

$^1$H NMR (400MHz, $D_2O$, 25 °C): δ = 8.23 (s, 1H), 6.26 (t, *J* = 6.1 Hz, 1H), 4.65–4.50 (m, 2H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.4, 3.3 Hz, 1H), 3.89 (dd, *J* = 7.4, 3.3 Hz, 1H), 2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100MHz, $D_2O$, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 145.1, 98.6, 97.6, 85.5, 85.3, 69.6, 64.7, 39.5, 30.0 ppm. $^{31}$P NMR (100MHz, $D_2O$, 25 °C) δ -5.36 (d, $^2J_{PP}$ = 21.0 Hz), -10.42 (d, $^2J_{PP}$ =19.2 Hz), -21.40 (t, $^2J_{PP}$ = 20.0 Hz) ppm. MS (HPLC): m/z calculated for $C_{18}H_{24}N_3O_{15}P_3$: 615.3 and found to be 614.3 (M-1)-. λmax = 270 nm.



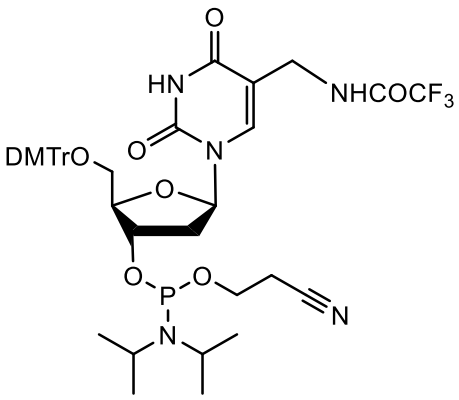

Chemical Formula: $C_{18}H_{21}F_3N_3O_{15}P_3$
Exact Mass: 669.01
Molecular Weight: 669.29

***Synthesis of 5- [(2-(2,4,5- trifluorophenyl)-acetamido) methyl]-2'-deoxyuridine triphosphate (14)***

A mixture of 2,4,5-trifluorophenylacetic acid (1 mg, 5 μmol, 1.2 eq) and HATU (2 mg, 5 μmol, 1.2 equiv) in 40 μL anhydrous DMF was added to a solution of compound **9** (2 mg, 4 μmol, 1.0 eq) in aqueous sodium bicarbonate (80 μL, 0.4 M). The mixture was agitated at 37 °C for 3 hrs and then freeze dried. The crude solid was washed with anhydrous DMF and resuspended in water. The product was purified first via anion-exchange HPLC (eluent A: 10 mM TEAB buffer, B: 1 M TEAB buffer, 0-80 % buffer B for 40 min) and then through RP-HPLC (eluent A: 0.1 M TEAB buffer, B: 40 % acetonitrile in 0.1 M TEAB buffer, 3.5-35 % buffer B for 20 min). Compound **14** was obtained in white solid form with a 65 % yield (1.45 mg, 3 μmol). $R_f$: 0.22 [n-propanol: ammonium hydroxide: $H_2O$=6:1:4 v/v/v].

$^1$H NMR (400MHz, $D_2O$, 25 °C): δ = 8.23 (s, 1H), 6.26 (t, *J* = 6.1 Hz, 1H), 4.65–4.50 (m, 2H), 4.23–4.13 (m, 3H), 3.96 (dd, *J* = 7.4, 3.3 Hz, 1H), 3.89 (dd, *J* = 7.4, 3.3 Hz, 1H), 2.54–2.47 (m, 1H), 2.33–2.42 (m, 1H), 2.09 (s, 3H). $^{13}$C NMR (100MHz, $D_2O$, 25 °C): δ = 170.6, 164.0, 158.1, 150.3, 145.1, 98.6, 97.6, 85.5, 85.3, 69.6, 64.7, 39.5,

30.0 ppm. $^{31}$P NMR (100MHz, $D_2O$, 25 °C) δ -5.36 (d, $^2J_{PP}$ = 21.0 Hz), -10.42 (d, $^2J_{PP}$ =19.2 Hz), -21.40 (t, $^2J_{PP}$ = 20.0 Hz) ppm. MS (HPLC): m/z calculated for $C_{18}H_{21}N_3O_{15}P_3$: 669.3 and found to be 668.2 (M-1)-. λmax = 270 nm.



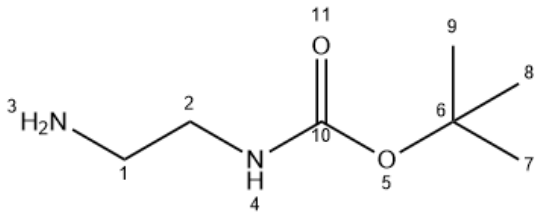

Chemical Formula: $C_{42}H_{49}F_3N_5O_9P$
Exact Mass: 855.32
Molecular Weight: 855.85

***Synthesis of 5-(aminomethyl)-2'-deoxyuridine phosphoramidite (15)***

White solid compound **7** (200 mg, 0.21 mmol, 1.0 equiv) was dried overnight and then dissolved in anhydrous 1,2-dichloroethane (15 mL) and DIPEA (300 μL, 1.76 mmol, 4.9 equiv) before 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (112 μL, 0.50 mmol, 1.4 equiv) was added under $N_2$. The mixture was stirred at room temperature for 1 hr. TLC was used to monitor reaction progress to avoid further oxidation. The reaction was quenched with EtOAc (50 mL) and concentrated on a rotary evaporator under vacuum to dryness. The organic layer was dried over anhydrous $Na_2SO_4$ and concentrated to dryness under reduced pressure. First the residue was purified by silica gel column chromatography (80%, EtOAc/Hexane) under a $N_2$ atmosphere to produce compound 15 in white solid form (142 mg, 45% yield). Rf = 0.45 (8:2, EtOAc/Hexane); 31P NMR (162 MHz, CDCl3) δ 148.9, 148.7. MS (ESI+) m/z calculated for $C_{42}H_{50}N_5O_9F_3P$: 856.3298 and found to be 856.3311.

## 7.2.2 Experiments for Chapter 5

***N-tert-Butoxycarbonyl-1,2-ethanediamine (11)***[263]

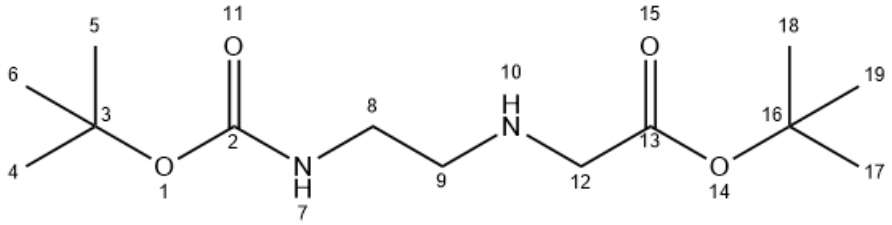



Chemical Formula: $C_7H_{16}N_2O_2$
Molecular Weight: 160.22

A solution of di-tert-butyl dicarbonate (11 g, 50 mmol) in THF (80 mL) was added slowly to a solution of 1,2-diaminoethane (25 mL, 400 mmol) in THF (100 mL) over a period of 3 hrs. The solution was stirred vigorously overnight before $H_2O$ (150 mL) was added to quench the reaction. Solid $K_2CO_3$ (10 g) was added and the resulting compound was extracted into an organic layer using DCM. The organic layer was then dried by $Na_2SO_4$, filtered and then concentrated *in vacuo* to yield (**11**) (6.86 g, 86 %) as a light-yellow oil. The oil gradually solidified to yield a white solid.

m.p. 109 – 112 °C; $^1$H NMR (400 MHz; $CDCl_3$) δ: 4.79 (br, 1H, **H$^4$**), 3.11 (d, *J* = 5.8 Hz, 2H, **H$^2$**), 2.73 (d, *J* = 5.8 Hz, 2H, **H$^1$**), 1.38 (s, 9H, **H$^7$**, **H$^8$**, **H$^9$**); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 156.30 (**C$^{10}$**), 78.98 (**C$^6$**), 53.44 (**C$^1$**), 41.33 (**C$^2$**), 29.43 (**C$^7$**, **C$^8$**, **C$^9$**); LRMS (ESI$^+$) calculated for $C_7H_{16}N_2O_2$: 160.2 and found to be 161.2 $[M+H]^+$, 100%.

***Tert-Butyl-N-(2-Boc-aminoethyl) glycinate (12)***[262]

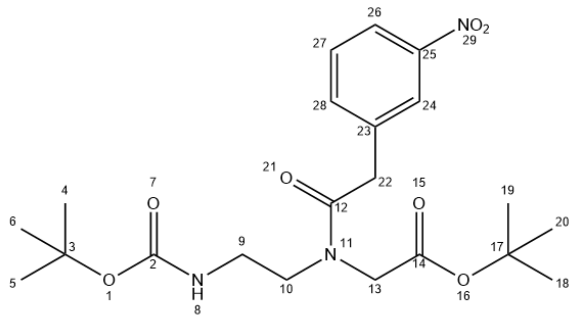



Chemical Formula: $C_{13}H_{26}N_2O_4$
Molecular Weight: 274.36

Compound **11** (8.00 g, 48 mmol) and $Et_3N$ (7 mL, 48mL) were dissolved together in ACN (100 mL). A solution of tert-butyl bromoacetate (2.5 mL, 16 mmol) in ACN (50 mL) was added slowly to the solution over a period of 1 hr. The resulting reaction mixture was heated to reflux overnight. Then EtOAc (150 mL) was added to quench the reaction after cooling. The reaction mixture was extracted using aqueous $K_2CO_3$ (10 g in 100 mL $H_2O$) and brine respectively. The organic layer was taken, dried over $Na_2SO_4$, filtered and concentrated *in vacuo*. The residues were purified by silica gel chromatography (EtOAc: Hexane, 70: 30) to yield **12** (6.30 g, 92 %) as a light-yellow oil.

TLC Rf 0.35 (ethyl acetate: hexane, 7:3); $^1$H NMR (400 MHz; $CDCl_3$) δ: 5.28 (br, 1H, **H**$^{10}$), 3.99 (br, 1H, **H**$^{8}$), 3.15 (s, 2H, **H**$^{12}$), 3.08 (t, $J$ = 7.1 Hz, 2H, **H**$^{8}$), 2.59 (t, $J$ = 7.1 Hz, 2H, **H**$^{9}$), 1.33 (s, 9H, **H**$^{4}$, **H**$^{5}$, **H**$^{6}$), 1.31 (s, 9H, **H**$^{17}$, **H**$^{18}$, **H**$^{19}$); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 171.64 (**C**$^{13}$), 156.01 (**C**$^{2}$), 81.09 (**C**$^{16}$), 78.80 (**C**$^{3}$), 51.13 (**C**$^{12}$), 48.63 (**C**$^{8}$), 40.66 (**C**$^{9}$), 28.31 (**C**$^{4}$, **C**$^{5}$, **C**$^{6}$), 27.97 (**C**$^{17}$, C$^{18}$, **C**$^{19}$); LRMS (ESI$^+$) calculated for $C_{13}H_{26}N_2O_4$: 274.3 and found to be 275.2 $[M+H]^+$, 100%.

***Tert-butyl N-(2-((tert-butoxycarbonyl) amino) ethyl)-N-(2-(3-nitrophenyl) acetyl) glycinate*** (13a)



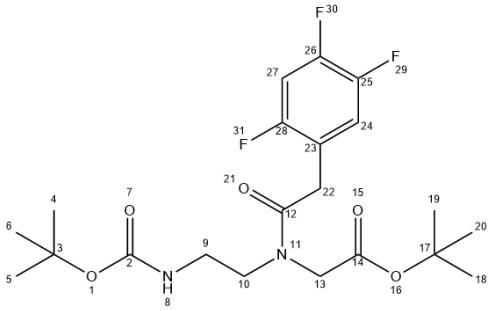

Chemical Formula: $C_{21}H_{31}N_3O_7$
Molecular Weight: 437.49

A solution of 2-(3-nitrophenyl) acetic acid (1.2 g, 6.5 mmol) in DMF (10 ml) was cooled to 0°C for 5 min. Then **12** (1.5 g, 5.5 mmol), DIPEA (1 mL, 5.5 mmol) and EDC (1.5 g, 8.0 mmol) were added to the solution. The reaction mixture was warmed to RT and stirred for 3 hrs. The resulting solution was diluted with DCM (100 mL) and extracted with a KCl solution (half-saturated, 100 mL). The organic layer was taken, dried over $Na_2SO4$, filtered and concentrated *in vacuo*. The residues were purified by dry-load silica gel chromatography (EtOAc: Hexane, 70: 30) to yield **13a** (1.6 g, 67 %) as a light orange oil.

TLC Rf 0.52 (ethyl acetate: hexane, 7:3); $^1$H NMR (400 MHz; $CDCl_3$) δ: 8.09 (s, 2H, **H$^{24}$**, **H$^{26}$**), 7.58 (d, *J* = 7.6 Hz, 1H, **H$^{27}$**), 7.45 (s, 1H, **H$^{28}$**), 5.54 (br, 1H, **H$^{8}$**), 3.91 (s, 2H, **H$^{13}$**), 3.83 (s, 2H, **H$^{22}$**), 3.50 (t, *J* = 5.7 Hz, 2H, **H$^{10}$**), 3.25 (t, *J* = 5.7 Hz, 2H, **H$^{9}$**), 1.43 (s, 18H, **H$^{4}$**, **H$^{5}$**, **H$^{6}$**, **H$^{18}$**, **H$^{19}$**, **H$^{20}$**); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 169.54(**C$^{14}$**, **C$^{12}$**), 160.68 (**C$^{2}$**), 148.42 (**C$^{25}$**), 136.32 (**C$^{26}$**), 135.72 (**C$^{24}$**), 130.71 (**C$^{27}$**), 124.20 (**C$^{28}$**), 122.84 (**C$^{23}$**), 81.88 (**C$^{17}$**), 79.50 (**C$^{3}$**), 51.31 (**C$^{9}$**), 50.03 (**C$^{10}$**), 40.04 (**C$^{22}$**), 38.10 (**C$^{13}$**), 28.76 (**C$^{4}$**, **C$^{5}$**, **C$^{6}$**), 28.42 (**C$^{18}$**, **C$^{19}$**, **C$^{20}$**); MS (ESI$^-$) calculated for $C_{21}H_{31}N_3O_7$: 437.5

and found to be 436 $[M-H]^-$, 100%.

***Tert-butyl N-(2-((tert-butoxycarbonyl) amino) ethyl)-N-(2-(2,4,5-trifluorophenyl) acetyl) glycinate*** (13b)



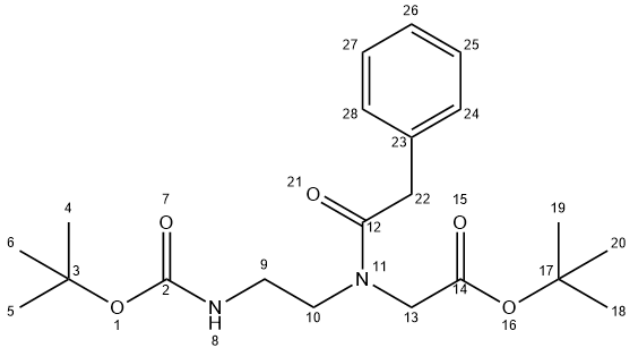

Chemical Formula: $C_{21}H_{29}F_3N_2O_5$
Molecular Weight: 446.47

A solution of 2-(2,4,5-trifluorophenyl) acetic acid (0.37 g, 2.00 mmol) in DMF (10 ml) was cooled to 0°C for 5 min. Then **12** (0.55 g, 2.00 mmol), DIPEA (0.17 mL, 2.00 mmol) and EDC (0.58 g, 3.00 mmol) were added to the solution. The reaction mixture was warmed to RT and stirred for 3 hrs. The solution was diluted with DCM (100 mL) and extracted with KCl solution (half-saturated, 100 mL). The organic layer was taken, dried over $Na_2SO4$, filtered and concentrated *in vacuo*. The residues were purified by dry-load silica gel chromatography (EtOAc: Hexane, 70: 30) to yield **13b** (0.63 g, 70 %) as a light orange oil.

TLC Rf 0.58 (ethyl acetate: hexane, 7:3); $^1$H NMR (400 MHz; $CDCl_3$) δ: 8.01(s, 1H, **H$^{24}$**), 6.90 (s,1H, **H$^{27}$**), 5.51 (br, 1H, **H$^{8}$**), 3.92 (s, 2H, **H$^{22}$**), 3.69 (t, *J* = 5.6 Hz, 2H, **H$^{10}$**), 2.95 (t, *J* = 5.6 Hz, 2H, **H$^{9}$**), 2.88 (s, 2H, **H$^{13}$**), 1.62 (s, 9H, **H$^{4}$**, **H$^{5}$**, **H$^{6}$**), 1.46 (s, 9H, **H$^{18}$**, **H$^{19}$**, **H$^{20}$**); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 171.09 (**C$^{14}$**, **C$^{12}$**), 169.93 (**C$^{2}$**), 168.35(**C$^{25}$**, **C$^{26}$**), 155.96 (**C$^{28}$**), 119.25 (**C$^{23}$**, **C$^{24}$**), 105.43 (**C$^{14}$**), 83.10 (**C$^{3}$**) 82.21 (**C$^{17}$**), 51.56 (**C$^{9}$**), 49.80 (**C$^{10}$**), 48.27 (**C$^{13}$**),38.81 (**C$^{22}$**), 28.31 (**C$^{4}$**, **C$^{5}$**, **C$^{6}$**), 27.92 (**C$^{18}$**, **C$^{19}$**, **C$^{20}$**); MS (ESI$^-$) calculated for $C_{21}H_{29}F_3N_2O_5$: 446.5 and found to be 445 [M-H]$^-$, 100%.

***Tert-butyl N-(2-((tert-butoxycarbonyl) amino) ethyl)-N-(2-phenylacetyl) glycinate***

(13c)

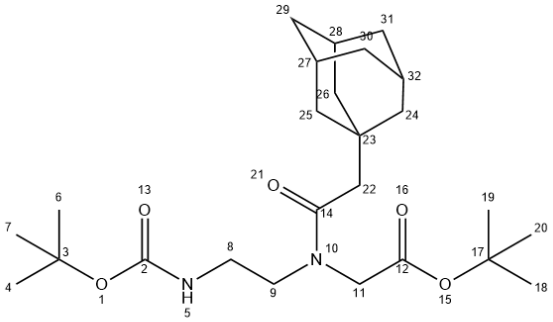



A solution of 2-phenylacetic acid (0.60 g, 3.00 mmol) in DMF (10 ml) was cooled to 0°C for 5 min. Then **12** (0.82 g, 3.00 mmol), DIPEA (0.50 mL, 3.00 mmol) and EDC (0.86 g, 4.50 mmol) were added to the solution. The reaction mixture was warmed to r.t. and stirred for 3 hr. The solution was diluted with DCM (100 mL) and extracted with KCl solution (half-saturated, 100 mL). The organic layer was taken, dried over $Na_2SO4$, filtered and concentrated *in vacuo*. The residues were purified by dry-load silica gel chromatography (EtOAc: Hexane, 70:30) to yield **13c** (1.02 g, 85 %) as colourless liquid.

TLC Rf .051 (ethyl acetate: hexane, 7:3); $^1$H NMR (400 MHz; $CDCl_3$) δ: 7.15-7.27 (m, 5H), 3.38 (s,2H, $\mathbf{H^{22}}$), 3.68 (s,2H, $\mathbf{H^{13}}$), 3.39 (t, $J$ = 6.8 Hz, 2H, $\mathbf{H^{10}}$), 3.14 (t, $J$ = 5.8 Hz, 2H, $\mathbf{H^{9}}$), 1.40 (s, 9H, $\mathbf{H^{4}}$, $\mathbf{H^{5}}$, $\mathbf{H^{6}}$), 1.39 (s, 9H, $\mathbf{H^{18}}$, $\mathbf{H^{19}}$, $\mathbf{H^{20}}$); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 169.42 ($\mathbf{C^{14}}$, $\mathbf{C^{12}}$), 168.23 ($\mathbf{C^{2}}$), 160.35 ($\mathbf{C^{26}}$), 148.42($\mathbf{C^{25}}$, $\mathbf{C^{28}}$), 130.02($\mathbf{C^{24}}$), 129.63($\mathbf{C^{27}}$), 129.20 ($\mathbf{C^{23}}$), 81.76 ($\mathbf{C^{3}}$), 76.60 ($\mathbf{C^{17}}$), 51.31 ($\mathbf{C^{9}}$), 50.03 ($\mathbf{C^{10}}$), 40.54 ($\mathbf{C^{13}}$), 38.12 ($\mathbf{C^{22}}$), 28.75 ($\mathbf{C^{4}}$, $\mathbf{C^{5}}$, $\mathbf{C^{6}}$), 28.38 ($\mathbf{C^{18}}$, $\mathbf{C^{19}}$, $\mathbf{C^{20}}$); MS (ESI$^-$) calculated for

$C_{21}H_{32}N_2O_5$: 392.5 and found to be 391 $[M-H]^-$, 100%.

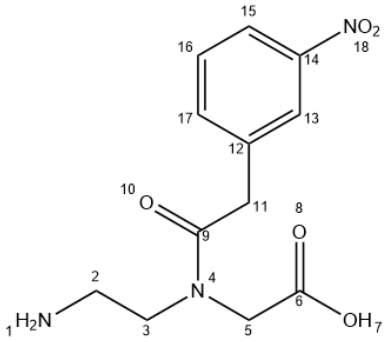
***tert-butyl N-(2-((3r,5r,7r)-adamantan-1-yl) acetyl)-N-(2-((tert-butoxycarbonyl) amino) ethyl) glycinate*** (13d)




A solution of 2-((3r,5r,7r)-adamantan-1-yl) acetic acid (1.20 g, 6.00 mmol) in DMF (10 ml) was cooled to 0°C for 5 min, and then **12** (1.40 g, 5.00 mmol), DIPEA (0.9 mL, 5.00 mmol) and EDC (1.45 g, 7.50 mmol) were added to the solution. The reaction mixture was warmed to RT and stirred for 3 hrs before the solution was diluted with DCM (100 mL) and extracted with KCl solution (half-saturated, 100 mL). The organic layer was taken, dried over $Na_2SO4$, filtered and concentrated by rotary evaporator. The residues were purified by dry-load silica gel chromatography (EtOAc: Hexane, 70: 30) to yield **13d** (0.6 g, 27 %) as a light-yellow oil.

TLC Rf 0.56 (ethyl acetate: hexane, 7:3); $^1$H NMR (400 MHz; $CDCl_3$) δ:5.55 (br, 1H, **H$^{8}$**), 4.11 (q, 2H, **H$^{22}$**), 3.85 (s,2H, **H$^{13}$**), 3.51 (t, $J$ = 6.4 Hz, 2H, **H$^{9}$**), 3.21 (t, $J$ = 6.4 Hz, 2H, **H$^{8}$**), 2.03-2.16 (m, 12H, **H$^{23\text{-}31}$**), 1.45 (s, 9H, **H$^{4}$**, **H$^{5}$**, **H$^{6}$**), 1.43 (s, 9H, **H$^{18}$**, **H$^{19}$**, **H$^{20}$**); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 171.34 (**C$^{14}$**, **C$^{12}$**), 168.50 (**C$^{2}$**), 81.76 (**C$^{3}$**), 76.60 (**C$^{17}$**), 51.31 (**C$^{9}$**), 50.03 (**C$^{10}$**), 48.46 (**C$^{13}$**), 40.54 (**C$^{22}$**), 38.12 (**C$^{24}$**, **C$^{25}$**, **C$^{26}$**), 36.72 (**C$^{27}$**, **C$^{28}$**, **C$^{32}$**), 33.98 (**C$^{29}$**, **C$^{30}$**, **C$^{31}$**), 28.76 (**C$^{4}$**, **C$^{5}$**, **C$^{6}$**), 28.33 (**C$^{18}$**, **C$^{19}$**, **C$^{20}$**); MS ($ESI^-$) calculated for $C_{25}H4_2N_2O_5$: 450.6 and found to be 450 $[M]^-$.

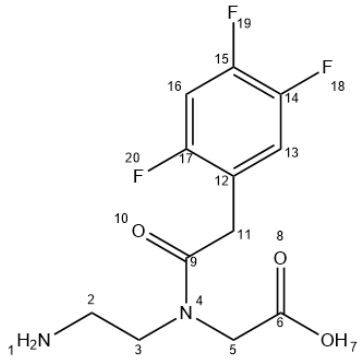
***N-(2-aminoethyl)-N-(2-(3-nitrophenyl) acetyl) glycine*** (14a)




Chemical Formula: $C_{12}H_{15}N_3O_5$
Molecular Weight: 281.27

TFA (15 mL) is added slowly to a solution of compound **13a** (1.5 g, 3.45 mmol) in DCM (15 mL) and stirred for 30 mins. The resulting solution was then added together with toluene (15 mL) to remove remaining TFA *in vacuo*. Then diethyl ether (30 mL) was added to the residues and cooled to -20 °C. A precipitate was formed and separated by suction filtration to yield **14a** as a white solid that was used directly in the next reaction to form **9a**.

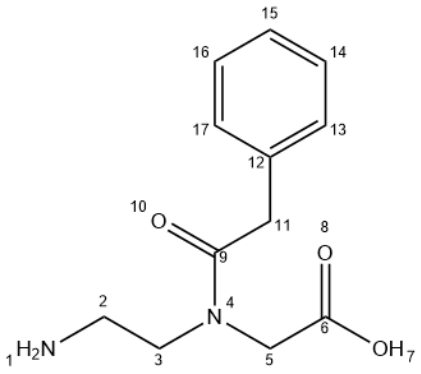
***N-(2-aminoethyl)-N-(2-(2,4,5-trifluorophenyl) acetyl) glycine*** (14b)




Chemical Formula: $C_{12}H_{13}F_3N_2O_3$
Molecular Weight: 290.24

TFA (5 mL) is added slowly to the solution of compound **13b** (0.6 g, 1.40 mmol) in DCM (10 mL) and stirred for 30 mins. The solution was then added together with toluene (10 mL) to remove any remaining TFA *in vacuo*. Then diethyl ether (30 mL) was added to the residues and cooled to -20 °C. A precipitate was formed and separated by suction filtration to yield **14b** as a white solid which was used directly into the next reaction to form **9b**.

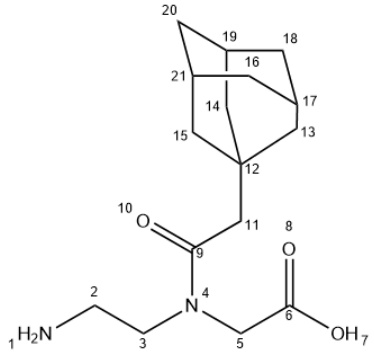
***N-(2-aminoethyl)-N-(2-phenylacetyl) glycine (14c)***




Chemical Formula: $C_{12}H_{16}N_2O_3$
Molecular Weight: 236.27

TFA (8 mL) is added slowly into a solution of Compound **13c** (1.0 g, 2.55 mmol) in DCM (8 mL) and stirred for 30 mins. The solution was then added together with toluene (10 mL) to remove remaining TFA *in vacuo*. Then diethyl ether (30 mL) was added to the residues and cooled to -20 °C. A precipitate was formed and separated by suction filtration to yield **14c** as a white solid which was used directly into the next reaction to form **9c**.

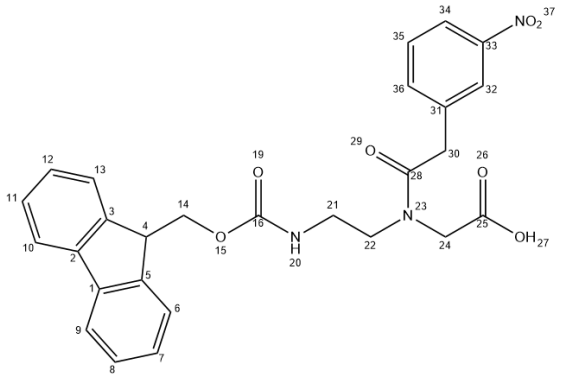
***N-(2-((3r,5r,7r)-adamantan-1-yl) acetyl)-N-(2-aminoethyl) glycine*** (14d)




Chemical Formula: $C_{16}H_{26}N_2O_3$
Molecular Weight: 294.40

TFA (5 mL) is added slowly into the solution of Compound **13d** (0.6 g, 1.40 mmol) in DCM (10 mL) and stirred for 0.5 hr. The resulting solution was then added together with toluene (10 mL) to remove the remaining TFA *in vacuo*. Then diethyl ether (30 mL) was added to the residue and cooled to -20 °C. A precipitate formed which was separated by suction filtration to yield **14d** as a white solid that was used directly in the next reaction to form **9d**.

***N-(2-((((9H-fluoren-9-yl) methoxy) carbonyl) amino) ethyl)-N-(2-(3-nitrophenyl) acetyl) glycine*** (9a)



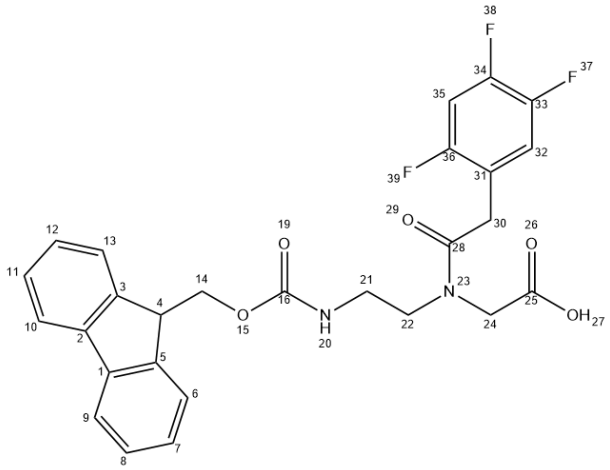

Chemical Formula: $C_{27}H_{25}N_3O_7$
Molecular Weight: 503.51

Fmoc-OSu (1.20 g, 3.50 mmol) and DIPEA (1.00 mL, 10.50 mmol) were added together into the solution of **14a** (1.00 g, 3.50 mmol) in DMF (10 mL). The resulting reaction mixture was stirred for 2 hrs before DCM (50 mL) was added to quench the reaction. The resulting solution was extracted with $H_2O$ (100 mL) and the organic layer was taken, dried over $Na_2SO_4$, filtered and concentrated *in vacuo* to remove any DMF solvent. Diethyl ether (20 mL) was added and then removed by filtration to yield **9a** (1.45 g, 80 %, overall yield 45 %) as a light-yellow oil.

$^1$H NMR (400 MHz; $CDCl_3$) δ: 8.10 (d, $J$ = 6.8 Hz, 1H, $\mathbf{H^{34}}$), 8.07 (s, 1H, $\mathbf{H^{32}}$), 8.02 (d, $J$ = 6.5 Hz, 2H, $\mathbf{H^{9}}$, $\mathbf{H^{10}}$), 7.75 (d, $J$ = 6.8 Hz, 2H, $\mathbf{H^{35}}$, $\mathbf{H^{36}}$), 7.58 (d, $J$ = 7.2 Hz, 2H, $\mathbf{H^{6}}$, $\mathbf{H^{13}}$), 7.51 (d, $J$ = 7.2 Hz, 2H, $\mathbf{H^{7}}$, $\mathbf{H^{12}}$), 7.38 (t, $J$ = 6.5, 7.4 Hz, 2H, $\mathbf{H^{8}}$, $\mathbf{H^{11}}$), 6.08 (br, 1H, $\mathbf{H^{20}}$), 4.42 (d, $J$ = 5.0 Hz, 2H, $\mathbf{H^{14}}$), 4.30 (d, $J$ = 5.0 Hz, 1H, $\mathbf{H^{4}}$), 4.21 (t, $J$ = 4.7 Hz, 2H, $\mathbf{H^{22}}$), 4.04 (t, $J$ = 4.7 Hz, 2H, $\mathbf{H^{21}}$), 3.79 (s, 2H, $\mathbf{H^{30}}$), 3.72 (s, 2H, $\mathbf{H^{24}}$); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 173.25 ($\mathbf{C^{25}}$, $\mathbf{C^{28}}$), 169.70 ($\mathbf{C^{16}}$), 157.60 ($\mathbf{C^{33}}$), 148.35 ($\mathbf{C^{3}}$,

**C**$^{5}$), 143.80 (**C**$^{1}$, **C**$^{2}$), 142.63 (**C**$^{31}$), 136.62 (**C**$^{35}$, **C**$^{36}$), 135.67 (**C**$^{32}$, **C**$^{34}$), 129.30 (**C**$^{9}$, **C**$^{10}$), 128.65 (**C**$^{8}$, **C**$^{11}$), 126.70 (**C**$^{7}$, **C**$^{12}$), 126.25 (**C**$^{6}$, **C**$^{13}$), 67.32 (**C**$^{14}$), 52.20 (**C**$^{24}$), 50.03 (**C**$^{22}$), 49.68 (**C**$^{21}$), 40.54 (**C**$^{4}$), 38.43 (**C**$^{30}$); LRMS (ESI$^{+}$) 504 $[M+H]^{+}$, 100%; HRMS (ESI$^{+}$), for $C_{27}H_{25}N_3O_7^{+}$, calculated 504.1692 and found to be 504.1770.

***N-(2-((((9H-fluoren-9-yl) methoxy) carbonyl) amino) ethyl)-N-(2-(2,4,5-trifluorophenyl) acetyl) glycine*** (9b)



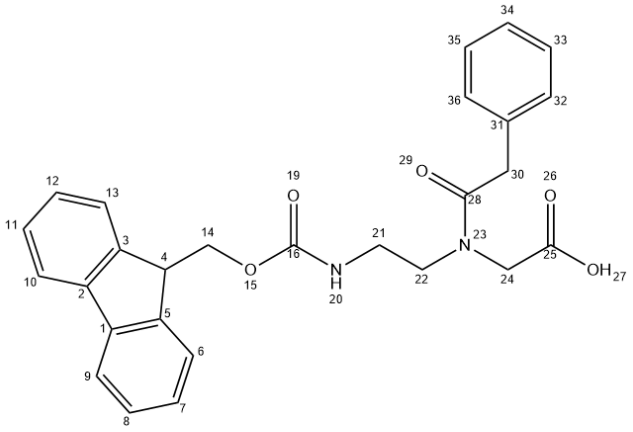
Chemical Formula: $C_{27}H_{23}F_3N_2O_5$
Molecular Weight: 512.49

Fmoc-OSu (0.70 g, 2.00 mmol) and DIPEA (0.50 mL, 6.00 mmol) were added to a solution of **14b** (0.50 g, 2.00 mmol) in DMF (5 mL). The reaction mixture was stirred for 2 hrs before DCM (20 mL) was added to quench the reaction. The solution was extracted with $H_2O$ (50 mL), the organic layer was taken, dried over $Na_2SO_4$, filtered and concentrated *in vacuo* to remove any DMF solvent. Diethyl ether (10 mL) was added and then removed by filtration to yield **9b** (0.76 g, 78 %, overall yield 41 %) as a light yellow solid.

$^1$H NMR (400 MHz; $CDCl_3$) δ: 8.02 (s, 1H, $\mathbf{H^{32}}$), 7.75 (s, 1H, $\mathbf{H^{35}}$), 8.69 (d, $J$ = 6.8 Hz, 2H, $\mathbf{H^{9}}$, $\mathbf{H^{10}}$), 7.38 (d, $J$ = 7.0 Hz, 2H, $\mathbf{H^{6}}$, $\mathbf{H^{13}}$), 7.31 (d, $J$ = 7.0 Hz, 2H, $\mathbf{H^{7}}$, $\mathbf{H^{12}}$), 6.08 (t, $J$ = 6.8, 7.4 Hz, 2H, $\mathbf{H^{8}}$, $\mathbf{H^{11}}$), 6.08 (br, 1H, $\mathbf{H^{20}}$), 4.36 (d, $J$ = 5.3 Hz, 2H, $\mathbf{H^{14}}$), 4.28 (d, $J$ = 5.3 Hz, 1H, $\mathbf{H^{4}}$), 4.24 (t, $J$ = 5.2 Hz, 2H, $\mathbf{H^{22}}$), 4.02 (t, $J$ = 5.2 Hz, 2H, $\mathbf{H^{21}}$), 3.75 (s, 2H, $\mathbf{H^{30}}$), 3.64 (s, 2H, $\mathbf{H^{24}}$); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 173.14 ($\mathbf{C^{25}}$, $\mathbf{C^{28}}$), 160.54 ($\mathbf{C^{16}}$), 160.21 ($\mathbf{C^{33}}$), 159.32 ($\mathbf{C^{34}}$), 157.63 ($\mathbf{C^{36}}$), 149.02 ($\mathbf{C^{3}}$, $\mathbf{C^{5}}$), 145.68 ($\mathbf{C^{1}}$,

**C**$^{2}$), 142.24 (**C**$^{31}$), 141.16 (**C**$^{35}$), 138.25 (**C**$^{32}$), 126.70 (**C**$^{9}$, **C**$^{10}$), 126.25 (**C**$^{8}$, **C**$^{11}$), 124.79 (**C**$^{7}$, **C**$^{12}$), 122.06 (**C**$^{6}$, **C**$^{13}$), 67.59 (**C**$^{14}$), 52.20 (**C**$^{24}$), 46.24 (**C**$^{22}$), 40.54 (**C**$^{21}$), 39.10 (**C**$^{4}$), 31.15 (**C**$^{30}$); LRMS (ESI$^{+}$) 535 $[M+Na]^{+}$, 100%; HRMS (ESI$^{+}$), for $C_{27}H_{23}F_3N_2O_5Na^{+}$, calculated 536.1457 and found to be 536.1535.

***N-(2-((((9H-fluoren-9-yl) methoxy) carbonyl) amino) ethyl)-N-(2-phenylacetyl) glycine*** (9c)



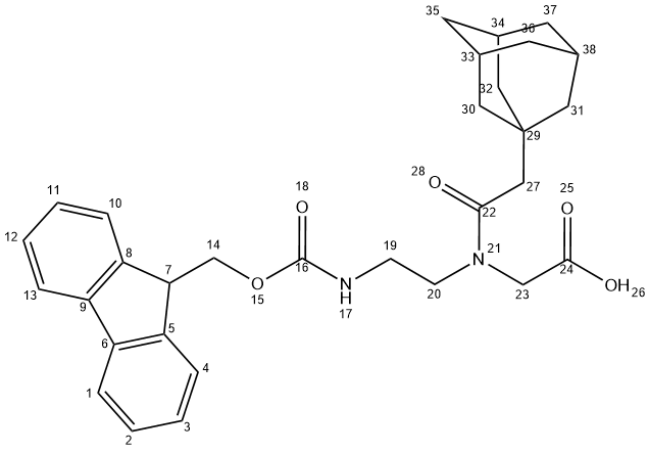
Chemical Formula: $C_{27}H_{26}N_2O_5$
Molecular Weight: 458.51

Fmoc-OSu (0.30 g, 1.00 mmol) and DIPEA (0.42 mL, 3.00 mmol) were added to the solution of **14c** (0.20 g, 1.00 mmol) in DMF (2 mL). The reaction mixture was stirred for 2 hrs before DCM (20 mL) was added to quench the reaction. The resulting solution was extracted with $H_2O$ (50 mL), the organic layer was taken, dried over $Na_2SO_4$, filtered and concentrated *in vacuo* to remove the DMF solvent. Diethyl ether (10 mL) was added and then removed by filtration to yield **9c** (0.34 g, 76 %, overall yield 42 %) as a colourless oil.

$^1$H NMR (400 MHz; $CDCl_3$) δ: 7.77 (d, *J* = 6.7 Hz, 2H, **H$^{9}$**, **H$^{10}$**),7.62 t, *J* = 6.3 Hz, 1H, **H$^{34}$**), 7.59 (d, *J* = 6.3 Hz, 2H, **H$^{33}$**, **H$^{35}$**), 7.52 (d, *J* = 7.1 Hz, 2H, **H$^{6}$**, **H$^{13}$**), 7.39 (d, *J* = 6.3 Hz, 2H, **H$^{32}$**, **H$^{36}$**), 7.31 (d, *J* = 7.1 Hz, 2H, **H$^{7}$**, **H$^{12}$**), 7.32 (t, *J* = 6.7 Hz, 2H, **H$^{8}$**, **H$^{11}$**), 5.30 (br, 1H, **H$^{20}$**), 4.39 (d, *J* = 5.5 Hz, 2H, **H$^{14}$**), 4.03 (d, *J* = 5.5 Hz, 1H, **H$^{4}$**), 3.71 (t, *J* = 5.1 Hz, 2H, **H$^{22}$**), 3.65 (t, *J* = 5.1 Hz, 2H, **H$^{21}$**), 3.32 (s, 2H, **H$^{30}$**), 2.72 (s, 2H, **H$^{24}$**); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 174.26 (**C$^{25}$**, **C$^{28}$**), 160.02 (**C$^{16}$**), 143.45 (**C$^{3}$**, **C$^{5}$**), 142.06 (**C$^{1}$**, **C$^{2}$**), 135.60 (**C$^{31}$**), 130.04 (**C$^{33}$**, **C$^{35}$**), 129.75 (**C$^{32}$**, **C$^{36}$**), 129.64 (**C$^{9}$**, **C$^{10}$**),

128.12 (**C**$^{8}$, **C**$^{11}$), 127.57 (**C**$^{7}$, **C**$^{12}$), 126.70 (**C**$^{6}$, **C**$^{13}$), 120.40 (**C**$^{34}$), 67.23 (**C**$^{14}$), 54.78 (**C**$^{24}$), 46.24 (**C**$^{22}$), 45.37 (**C**$^{21}$), 40.55 (**C**$^{4}$), 38.10 (**C**$^{30}$); LRMS (ESI$^{+}$) 459 $[M+H]^{+}$, 100%; HRMS (ESI$^{+}$), for $C_{27}H_{26}N_2O_5^{+}$, calculated 459.1842 and found to be 509.1920.

***N-(2-((((9H-fluoren-9-yl) methoxy) carbonyl) amino) ethyl)-N-(2-((3r,5r,7r)-adamantan-1-yl) acetyl) glycine*** (9d)



Chemical Formula: $C_{31}H_{36}N_2O_5$
Molecular Weight: 516.64

Fmoc-OSu (0.60 g, 2.00 mmol) and DIPEA (0.50 mL, 6.00 mmol) were added to the solution of **14d** (0.50 g, 2.00 mmol) in DMF (5 mL). The reaction mixture was stirred for 2 hrs before DCM (20 mL) was added to quench the reaction. The solution was extracted with $H_2O$ (50 mL), the organic layer was taken, dried over $Na_2SO_4$, filtered and concentrated *in vacuo* to remove DMF solvent. Diethyl ether (20 mL) was added and then removed by filtration to yield **9d** (0.80 g, 80 %, overall yield 44 %) as a white solid.

$^1$H NMR (400 MHz; $CDCl_3$) δ: 7.76 (d, *J* = 7.5 Hz, 2H, **H$^{1}$**, **H$^{13}$**), 7.61 (d, *J* = 6.6 Hz, 2H, **H$^{4}$**, **H$^{10}$**), 7.42 (d, *J* = 6.6 Hz, 2H, **H$^{3}$**, **H$^{11}$**), 7.34 (t, *J* = 7.5 Hz, 2H, **H$^{2}$**, **H$^{12}$**), 4.55 (d, *J* = 3.8 Hz, 2H, **H$^{14}$**), 4.34 (d, *J* = 3.8 Hz, 1H, **H$^{7}$**), 3.67 (t, *J* = 4.3 Hz, 2H, **H$^{20}$**), 3.38 (t, *J* = 4.3 Hz, 2H, **H$^{19}$**), 2.72 (s, 2H, **H$^{23}$**), 1.80-1.68 (m, 15H); $^{13}$C NMR (101 MHz; $CDCl_3$) δ: 173.31 (**C$^{22}$**, **C$^{24}$**), 171.42 (**C$^{16}$**), 157.86 (**C$^{5}$**, **C$^{8}$**), 143.31 (**C$^{6}$**, **C$^{9}$**), 142.64 (**C$^{1}$**, **C$^{13}$**), 126.74 (**C$^{2}$**, **C$^{12}$**), 126.35 (**C$^{3}$**, **C$^{11}$**), 124.30 (**C$^{4}$**, **C$^{10}$**), 67.15 (**C$^{14}$**), 52.85 (**C$^{23}$**),

48.44 (**C**$^{27}$), 46.24 (**C**$^{20}$), 43. 38 (**C**$^{30}$, **C**$^{31}$, **C**$^{32}$), 40.67 (**C**$^{19}$), 36.70 (**C**$^{7}$), 32.65 (**C**$^{30}$, **C**$^{35}$, **C**$^{37}$), 28.02 (**C**$^{33}$, **C**$^{34}$, **C**$^{38}$); LRMS (ESI$^{+}$) 517 $[M+H]^{+}$, 100%; HRMS (ESI$^{+}$), for $C_{31}H_{36}N_2O_5^{+}$, calculated 517.2624 and found to be 517.2703.

# 7.3 Biochemistry section

## 7.3.1 Experiments for Chapters 2, 3 and 4

Vent polymerase, Vent (exo-), Deep Vent (exo-), Phusion and GoTaq G2 polymerases, deoxynucleotide (dNTP) solution mix and the Quick-Load® Purple 1 kb DNA ladder were all purchased from New England Biolabs (UK). NAP-10 and NAP-25 columns were purchased from G.E. Healthcare Life Sciences. The 96-well black, polystyrene, microplates were purchased from Greiner Bio-one. SYBR™ Gold Nucleic Acid Gel Stain, SYBR™, Green I Nucleic Acid Gel Stain, and SYBR™ Green II RNA Gel Stain were purchased from ThermoFisher. The 5X Nucleic acid sample loading buffer, used for non-denaturing polyacrylamide gel electrophoresis experiments, was purchased from Bio-Rad. All PCR reactions were performed using a Bio-rad T100™ thermocycler and all qPCR reactions were conducted using a Bio-Rad CFX96 thermocycler.

**Linear Extension Experiments**

In a 20 μL reaction, 66 pmol of FAM-labelled DNA primer (Res2465), 132 pmol of template (Res2843), and 3.2 nmol of each modified 5-AM-dUTP or natural dNTPs were mixed with the DNA polymerase (1 unit for GoTaq and Vent exo-, or 0.5 units for KOD) and 1× buffer (as supplied with the corresponding enzymes). In the case of KOD polymerase, 1 mM $MgCl_2$ was added separately to the reaction mixture. The reaction mixtures were heated on a BIO-RAD T100$^{TM}$ Thermal Cycler at 60 °C for a period varying from 5 to 90 min, and 20 μL formamide was then added to quench the reaction. The reactions were stored at -20 °C before analysis via 20% denaturing polyacrylamide gel electrophoresis (PAGE) at a constant 20 W of power in 1× TBE buffer. Gel images were taken using a Syngene G: BOX with associated gene-snap imaging software on a transilluminator (at 302 nm) wherein oligonucleotide bends were cut off and analyzed by HPLC-MS.

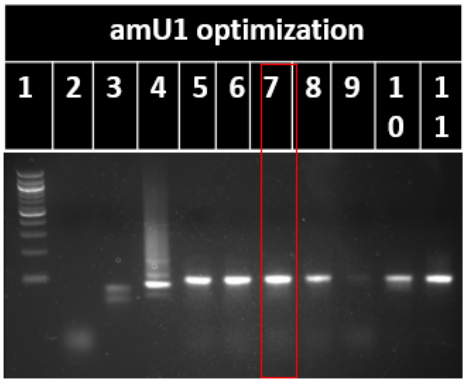
**PCR amplifications**

For total 50 μL PCR protocol, two primers Px (1 μL each, 0.2 M), template Tx (1 μL, 1ng), modified-dUTP or unmodified dNTPs (5 μL each, 2 mM), MgSO4 (1 μL, 2 mM), polymerase (0.5 μL, 1 Unit) and its corresponding buffers (5 μL, 1x) were added to 20.5 μL of water. Amplification was performed on a Bio-rad T100$^{TM}$ thermocycler via an initial denaturation step at 95 ºC for 3 min; followed by 30 cycles of denaturation at 95 ºC for 30 s; annealing at 55 ºC for 30 s, and elongation at 72 ºC for 30 s. Further extension was carried out at 72 ºC for 5 min. Samples were analysed via 2% agarose gel electrophoresis with SYBR$^{TM}$ Gold (200 ng/mL) at constant voltage (126 V) in 1× TBE buffer.

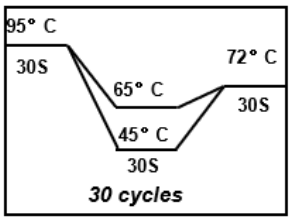
Modified PCR amplification protocol for qAN1 contained (per 50 μL PCR reaction): 0.2 μM of forward and reverse primers; 40 nM (1 ng) of template; and 200 μM of each dNTPs (including natural dNTPs and qAN1 triphosphate as modified dATP) followed by the addition of 1 unit of Vent (exo-) DNA polymerase and its corresponding 1x buffer in a solution containing 2 mM $Mg^{2+}$. The PCR amplification was performed on a Bio-Rad T100 Thermal Cycler with an initial denaturation at 98 °C for 3 min; followed by 30 cycles of denaturation at 98 °C for 60 s; followed by primer annealing and template elongation/extension at 60 °C for 60 s. A further extension was performed at 72 °C for 5 min and the reaction mixture was held under 4 °C for an indefinite period prior to analysis. The samples were analysed by electrophoresis at a constant voltage (126 V) in 1× TBE buffer using a 2% agarose gel containing 5 μL of SYBR gold nucleic acid stain. The resulting bands were excised, extracted to water, filtered by NAP columns,

and then characterised by HPLC-MS.

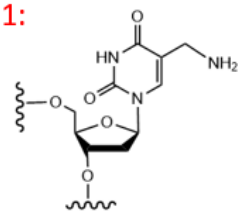
## *PCR test for amU1*



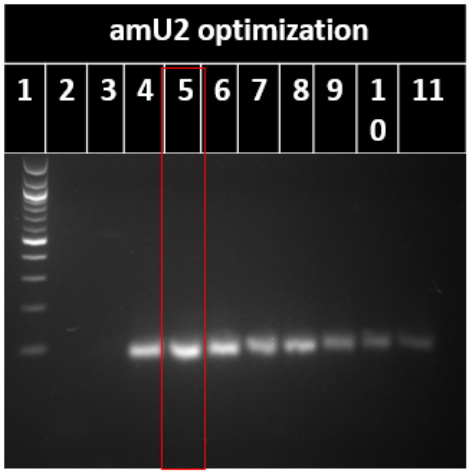

1. DNA Ladder (100 bp);
2. Only primers;
3. Only template;
4. AGCT, Vent exo-, 55° C;
5. AGCamU1 , Vent exo-, 45° C;
6. AGCamU1 , Vent exo-, 50° C;
7. AGCamU1 , Vent exo-, 55° C;
8. AGCamU1 , Vent exo-, 60° C;
9. AGCamU1 , Vent exo-, 65° C;
10. AGCamU1 , Gotaq G2, 55° C;
11. AGCamU1 , phusion HF, 55° C.



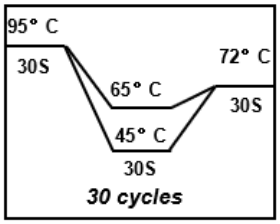

amU1:



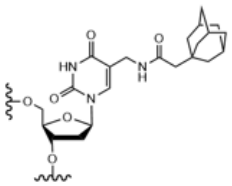
## *PCR test for amU2*



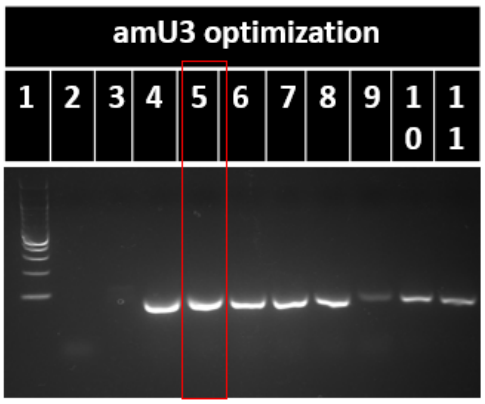

1. DNA Ladder (100 bp);
2. Only primers;
3. Only template;
4. AGCT, Vent exo-, 55° C;
5. AGCamU2 , Vent exo-, 45° C;
6. AGCamU2 , Vent exo-, 50° C;
7. AGCamU2 , Vent exo-, 55° C;
8. AGCamU2 , Vent exo-, 60° C;
9. AGCamU2 , Vent exo-, 65° C;
10. AGCamU2 , Gotaq G2, 55° C;
11. AGCamU2 , phusion HF, 55° C.



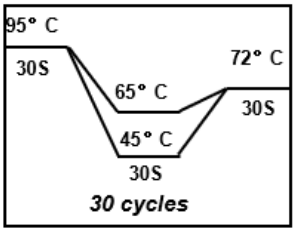

amU2:



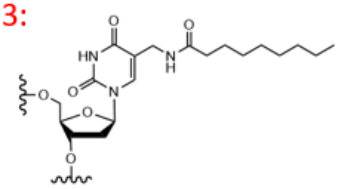
## *PCR test for amU3*



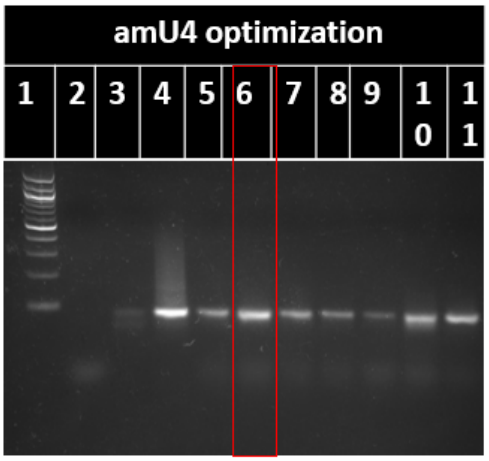

1. DNA Ladder (100 bp);
2. Only primers;
3. Only template;
4. AGCT, Vent exo-, 55° C;
5. AGCamU3 , Vent exo-, 45° C;
6. AGCamU3 , Vent exo-, 50° C;
7. AGCamU3 , Vent exo-, 55° C;
8. AGCamU3 , Vent exo-, 60° C;
9. AGCamU3 , Vent exo-, 65° C;
10. AGCamU3 , Gotaq G2, 55° C;
11. AGCamU3 , phusion HF, 55° C.



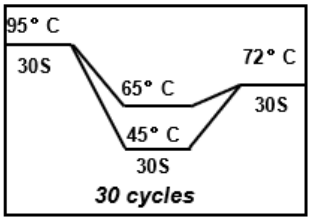

amU3:



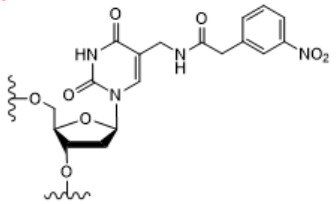
## *PCR test for amU4*



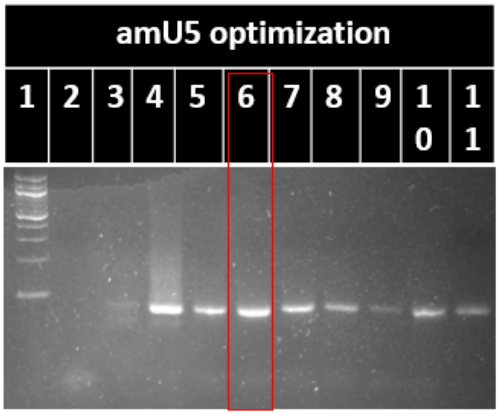

1. DNA Ladder (100 bp);
2. Only primers;
3. Only template;
4. AGCT, Vent exo-, 55° C;
5. AGCamU4 , Vent exo-, 45° C;
6. AGCamU4 , Vent exo-, 50° C;
7. AGCamU4 , Vent exo-, 55° C;
8. AGCamU4 , Vent exo-, 60° C;
9. AGCamU4 , Vent exo-, 65° C;
10. AGCamU4 , Gotaq G2, 55° C;
11. AGCamU4 , phusion HF, 55° C.



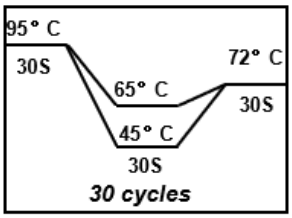

amU4:



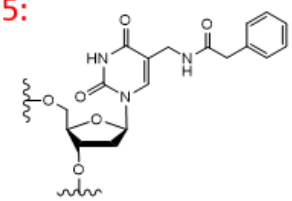
## *PCR test for amU5*



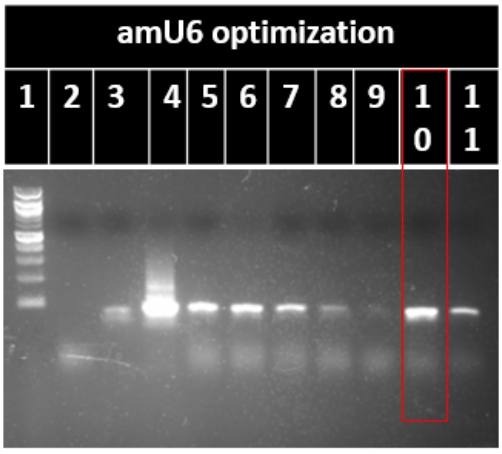

1. DNA Ladder (100 bp);
2. Only primers;
3. Only template;
4. AGCT, Vent exo-, 55° C;
5. AGCamU5 , Vent exo-, 45° C;
6. AGCamU5 , Vent exo-, 50° C;
7. AGCamU5 , Vent exo-, 55° C;
8. AGCamU5 , Vent exo-, 60° C;
9. AGCamU5 , Vent exo-, 65° C;
10. AGCamU5 , Gotaq G2, 55° C;
11. AGCamU5 , phusion HF, 55° C.



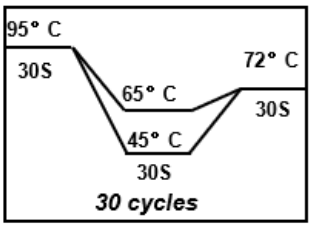

amU5:



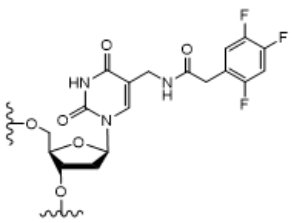
## *PCR test for amU6*




1. Only primers;
2. Only template;
3. DNA Ladder (100 bp);
4. AGCT, Vent exo-, 55° C;
5. AGCamU6 , Vent exo-, 45° C;
6. AGCamU6 , Vent exo-, 50° C;
7. AGCamU6 , Vent exo-, 55° C;
8. AGCamU6 , Vent exo-, 60° C;
9. AGCamU6 , Vent exo-, 65° C;
10. AGCamU6 , Gotaq G2, 55° C;
11. AGCamU6 , phusion HF, 55° C.




amU6:



**qPCR General Protocol**

For a total volume of 20 μL PCR protocol, two primers Px (0.4 μL each, 0.2 M), template Tx (1 μL, 1 ng/40 nM), modified-dUTP or unmodified dNTPs (2 μL each, 2 mM), MgSO4 (0.4 μL, 2 mM), EvaGreen dye (1 μL, 1 x), polymerase (0.5 μL, 1 Unit) and its corresponding buffers (2 μL, 1 x) were added to 5 μL of water. Amplification was performed on a Bio-Rad CFX96 real-time PCR instrument by an initial denaturation step at 95 ºC for 3 min; followed by 30 cycles of denaturation at 95 ºC for 30 s; annealing at 55 ºC for 30 s, and elongation at 72 ºC for 30 s. Further extension was carried out at 72 ºC for 5 min. This was followed by melting temperature measurements involving heating the reaction mixture to 95 °C for 30 s, cooling to 30 °C, then increasing the temperature to 95 °C at a rate of 1 or 0.5 °C/s and holding (monitoring) at each temperature for 5 s. Samples were analysed by 2% agarose gel electrophoresis with SYBR$^{TM}$ Gold (200 ng/mL) at constant voltage (126 V) in 1× TBE buffer.

The real-time PCR amplification incorporating both modified 5-am-dUTPs and qAN1 were performed using a real-time PCR instrument (Bio-Rad CFX96) with an initial denaturation step at 95 °C for 3 min; followed by 40 cycles of denaturation at 95 °C for 30 s; primer annealing at 60 °C for 60 s; and subsequent extension at 85 °C for 60 s. A further extension was carried out at 72 °C for 5 min. The concentration of template and both forward and reverse primers (within a 20 μL PCR reaction) were set at 4 nM (100 pg) for template and 0.2 μM for the primers, respectively. Addition of 200 μM of each dNTP (including the naturally occurring dNTPs, qAN1 triphosphate and modified

5-AM-dUTPs) were followed by 0.5 μL (1 unit) of Vent (exo-) DNA polymerase and its corresponding 1× buffer in a 2 mM $Mg^{2+}$ solution. A further 1× EvaGreen dye was added to each reaction for a real-time PCR.

**Size separation of ssDNA libraries**

Each 50 μL PCR reaction contained 0.2 μM of both forward and reverse primers, 1μL of template and 200 μM of each dNTPs, followed by the addition of 1 unit of DNA polymerase (Vent exo-) and its corresponding 1× buffer with a 2 mM $Mg^{2+}$ solution. The PCR amplification reaction was performed on a Bio-Rad T100 Thermal Cycler PCR instrument with an initial denaturation step at 95 °C for 3 min; followed by 25 cycles (to avoid non-specific amplification) of denaturation at 95 °C for 30 s; primer annealing at 60 °C for 30 s; and, finally, template elongation/extension at 72 °C for 30 s. A further extension was performed at 72 °C for 2 min and the resulting reaction mixture was maintained at 4 °C for an indefinite period before carrying out an analysis. A 10% denaturing PAGE step was conducted and the desired strands with FAM-labelling at the 5′ site could readily be visualised and quantified by fluorescence intensity, if necessary (the amount of DNA is low and undetectable by Nanodrop). The band containing these 80-meric ssDNA sequences was excised; purified through NAP column; and freeze dried for further use as initial library for aptamer selection.

**Aptamer selection of Her2 protein target using nitrocellulose membrane filter**

Human HER2/ErbB2 protein (Tag free, MW: 70.1 kDa, purchased from ACROBiosystems™) was used as a protein target in aptamer selection. An HER2 solution (0.1 μM, 100 μL) was injected to an Eppendorf through a 0.45 μm nitrocellulose membrane filter (Millipore™) using a 1mL syringe, and the membrane (with HER2 target proteins attached) was kept for further incubation with an ssDNA library dissolved in 1× phosphate-buffered saline (PBS) buffer (0.5 μM, 500 μL) at 95 ℃ for 3 min followed by shock freezing in liquid nitrogen. This resulting solution of ssDNA sequences was subsequently incubated with the HER2-bound nitrocellulose membrane for 1 hr at room temperature. The nitrocellulose membrane containing ssDNA-HER2 complexes was obtained, washed, and maintained in $H_2O$ (500 μL). The target-bound sequences were eluted from the HER2 and nitrocellulose membrane by thermal denaturation at 95 ℃ for 15 min.

**qPCR amplification for each round of aptamer selection process**

The qPCR amplification stage was performed on a Bio-Rad CFX96 instrument using the following procedure: after an initial denaturation at 95 °C for 3 min; followed by ~25 cycles (the number of cycles of enrichment depend on the actual state of the library when it reaches its maximum amplification) of denaturation at 95 °C for 30 s; primer annealing at 60 °C for 30 s; and extension at 72 °C for 30 s. A further extension was carried out at 72 °C for 5 min. A camera monitors the amplification of targeted DNA oligonucleotides through the measurement of quantitative fluorescent EvaGreen dye after each annealing step. The samples were further analysed using a 10% denaturing PAGE gel containing a SYBR gold (0.5× in TBE buffer) nucleic acid stain. The band of 80-mer FAM-labelled oligonucleotides were excised, purified through a NAP column, and freeze dried for use in the next round of selection.

## MATLAB Program for Analyzing Library Diversity from a High-Throughput Sequencing Result

Open fastq. Files:

```
filename = 'D:\sequencing_data\WTCHG_903385_73065378_1.fastq';

info = fastqinfo(filename)

reads = fastqread(filename)
```

Surveying base composition of sequences reads:

```
N = info.NumberOfEntries;
seqs = {reads.Sequence};
readsLen = cellfun(@length, seqs);

figure(); hist(readsLen);
xlabel('Number of bases'); ylabel('Number of sequence reads');
title('Length distribution of sequence reads')


nt = {'A', 'C', 'G', 'T'};
pos = cell(4,N);

for i = 1:4
    pos(i,:) = strfind(seqs, nt{i});
end

count = zeros(4,N);
for i = 1:4
    count(i,:) = cellfun(@length, pos(i,:));
end

%=== plot nucleotide distribution
```

```
figure();
subplot(2,2,1); hist(count(1,:)); title('A'); ylabel('Number of
sequence reads');
subplot(2,2,2); hist(count(2,:)); title('C');
subplot(2,2,3); hist(count(3,:)); title('G'); xlabel('Occurrences');
ylabel('Number of sequence reads');
subplot(2,2,4); hist(count(4,:)); title('T'); xlabel('Occurrences');

figure(); hist(count');
xlabel('Occurrences');
ylabel('Number of sequence reads');
legend('A', 'C', 'G', 'T');
title('Base distribution by nucleotide type');
```

Unique sequences occurrences:

```
%=== determine read frequency
[uReads,~,n] = unique({reads.Sequence});
numUnique = numel(uReads)
readFreq = accumarray(n(:),1);
figure(); hist(readFreq, unique(readFreq));
xlabel('Occurrences'); ylabel('Number of sequence reads');
title('Read occurrences');

%=== identify multiply-occurring sequence reads
d = readFreq > 1;
dupReads = uReads(d)'
dupFreq  = readFreq(d)'
```

Selecting oligonucleotides start with specific sequences:

```
fr = fopen("D:\sequencing_data\WTCHG_903385_73065378_1.fastq");
fw = fopen("D:\sequencing_data\ssDNA01.fastq","w");
while ~ feof(fr)
    line1 = fgetl(fr);
    line2 = fgetl(fr);
```

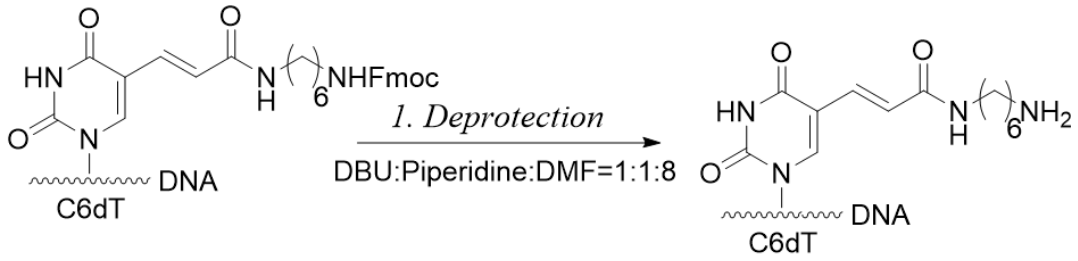
```
        line3 = fgetl(fr);
        line4 = fgetl(fr);
        if startsWith(line2,"AGCAGCACA")
            fprintf(fw,'%s\n',line1);
            fprintf(fw,'%s\n',line2);
            fprintf(fw,'%s\n',line3);
            fprintf(fw,'%s\n',line4);
        end
    end
```

## 7.3.2 Experiments for Chapter 5

### *Experimental Oligonucleotide Synthesis*

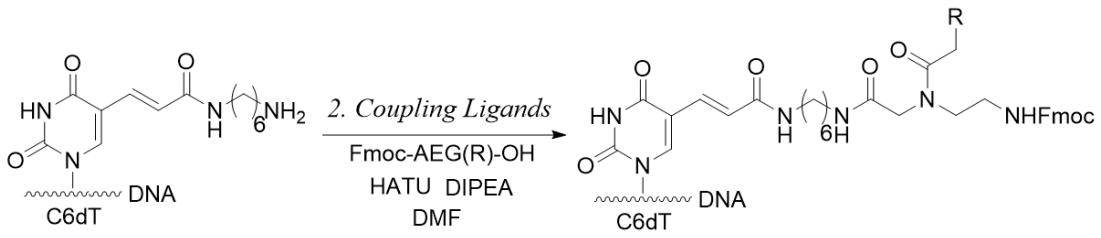
All DNA reagents, including standard DNA phosphoramidites and CPG resin for solid-phase oligonucleotide synthesis, were obtained from Link Technologies Ltd. An oligonucleotide synthesiser was used which is an automated DNA/RNA synthesiser (model 394) as supplied by *Applied Biosystems Inc*. Phosphoramidite cycles, including detritylation, activation and coupling, capping and iodine oxidation steps were undertaken at the 0.2 or 1.0 μmol scale. Phosphoramidite monomers used in oligonucleotide sequencing were all standard β-cyanoethyl phosphoramidite. Coupling efficiencies and overall yields of synthesised oligonucleotides were ≥ 98.0% in all cases. These were determined by the automated trityl cation conductivity monitoring facility of the synthesiser. During the solid-phase oligonucleotide synthesis, standard phosphoramidite monomers were dissolved in ACN (anhydrous) to a concentration of 0.1 M and used immediately. The coupling time for A, T, G and C monomers was set to 60 s, and for modified monomers was set to 600 s.

***Fmoc-deprotection general method***



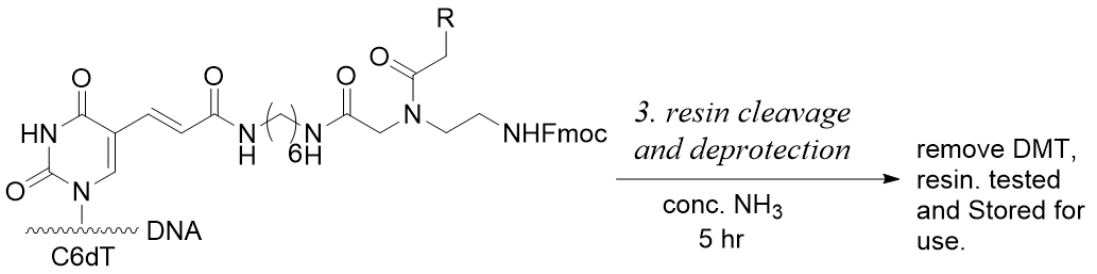

Solid resin containing Fmoc-protected oligonucleotides was placed in an Eppendorf tube and flashed with argon. The tube was washed with DMF (0.1 mL) via syringe and excess DMF was then removed by syringe needle. The deprotection solution (0.1 mL, a mixture of DBU:piperidine:DMF in a ratio of 1:1:8) was added into the tube via syringe in order to remove Fmoc groups. After 5 min, the deprotection solution was removed by syringe and DMF (0.1 mL) was added and consequentially removed to wash the resin. The deprotection solution (0.1 mL) was added once again to the tube for 5 min and then removed. DMF (0.1 mL) was added and removed followed by ACN (0.1 mL) to wash the resin. Then resin was flashed and dried with argon.

***Coupling reaction general method***




For each molar equivalent of oligonucleotide, the Fmoc-protected AEG-based amino acid (30 eq.), HATU (29 eq.) and DIPEA (90 eq.) were mixed in DMF (0.1 mL). The reaction mixture was added to the resin in an Eppendorf tube and left for 1 hr at 55 °C. Afterwards, the reaction mixture was removed by syringe and the resin washed with DMF (0.1 mL) and ACN (0.1 mL) 3 times and then dried by argon.

***Resin cleavage general method***




The resin was heated at 55 °C in aliquoted ammonia for 5 hr. Then the cleaved resin was concentrated by rotary evaporator, filtered and then stored at -20 °C.

# *References*

## *References*


1. Clayden, J., Greeves, N., Warren, S. & Wothers, P. McMurry's Special Method in Elementary Science Special Method in Elementary Science for the Common School. Charles A. McMurry . *The American Naturalist* **40**, 455–456 (1906).

2. Vella, F. Nucleic Acids in Chemistry and Biology. *Biochemical Education* **19**, 97–98 (1991).

3. Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6**, 201–209 (1950).

4. Wilkins, M. H. F., Seeds, W. E., Stokes, A. R. & Wilson, H. R. Helical structure of crystalline deoxypentose nucleic acid. *Nature* **172**, 759–762 (1953).

5. Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741 (1953).

6. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

7. Dickerson, R. E. *et al.* The anatomy of A-, B-, and Z-DNA. *Science* **216**, 475–485 (1982).

8. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core.

*Nature* **423**, 145–150 (2003).

9. Seeman, N. C. Nucleic acid junctions and lattices. *Journal of Theoretical Biology* **99**, 237–247 (1982).

10. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Journal of the American Chemical Society* **120**, 9401–9409 (1998).

11. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* **34**, 564–574 (2006).

12. Šponer, J. *et al.* Nature of base stacking: Reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chemistry - A European Journal* **12**, 2854–2865 (2006).

13. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* **34**, 564–574 (2006).

14. Hagen, J. Thermodynamics of DNA Duplex Formation. *California Polytechnic State University Chemistry Department Lab Workbook* 4–6 (2014).

15. Zipper, H., Brunner, H., Bernhagen, J. & Vitzthum, F. Investigations on DNA

intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research* **32**, (2004).

16. Giglio, S., Monis, P. T. & Saint, C. P. Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR. *Nucleic acids research* **31**, (2003).

17. Pattyn, F., Speleman, F., De Paepe, A. & Vandesompele, J. RTPrimerDB: The real-time PCR primer and probe database. *Nucleic Acids Research* vol. 31 122–123 Preprint at https://doi.org/10.1093/nar/gkg011 (2003).

18. Merrifield, B. SOLID-PHASE SYNTHESIS. *Chemica scripta* (1985).

19. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149–2154 (1963).

20. Pon, R. T. Solid-phase supports for oligonucleotide synthesis. *Methods in molecular biology (Clifton, N.J.)* **20**, 465–496 (1993).

21. Michelson, A. M. & Todd, A. R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′: 5′-internucleotidic linkage. *Journal of the Chemical Society (Resumed)* 2632–2638 (1955) doi:10.1039/JR9550002632.

22. Hall, R. H., Todd, A. & Webb, R. F. Nucleotides. Part XLI. Mixed anhydrides as intermediates in the synthesis of dinucleoside phosphates. *Journal of the Chemical*

*Society (Resumed)* 3291–3296 (1957) doi:10.1039/jr9570003291.

23. Gilham, P. T. & Khorana, H. G. Studies on Polynucleotides. I. A New and General Method for the Chemical Synthesis of the C5′-C3′ Internucleotidic Linkage. Syntheses of Deoxyribo-dinucleotides. *Journal of the American Chemical Society* **80**, 6212–6222 (1958).

24. Reese, C. B. The chemical synthesis of oligo- and poly-nucleotides by the phosphotriester approach. *Tetrahedron* **34**, 3143–3179 (1978).

25. Pless, R. C. & Letsinger, R. L. Solid support synthesis of ollgothymidylates using phosphorochloridates and 1-alkylimidazoles. *Nucleic Acids Research* **2**, 773–786 (1975).

26. LETSINGER, R. L., FINNAN, J. L., HEAVNER, G. A. & LUNSFORD, W. B. ChemInform Abstract: NUCLEOTIDE CHEMISTRY PART 20, PHOSPHITE COUPLING PROCEDURE FOR GENERATING INTERNUCLEOTIDE LINKS. *Chemischer Informationsdienst* **6**, no-no (1975).

27. Matteucci, M. D. & Caruthers, M. H. Synthesis of Deoxyoligonucleotides on a Polymer Support. *Journal of the American Chemical Society* **103**, 3185–3191 (1981).

28. Sonveaux, E. The organic chemistry underlying DNA synthesis. *Bioorganic*

*Chemistry* vol. 14 274–325 Preprint at https://doi.org/10.1016/0045-2068(86)90038-6 (1986).

29. Vongsutilers, V., Daft, J. R., Shaughnessy, K. H. & Gannett, P. M. A general synthesis of C8-arylpurine phosphoramidites vorasit vongsutilers. *Molecules* **14**, 3339–3352 (2009).

30. Caruthers, M. H. A brief review of DNA and RNA chemical synthesis. in *Biochemical Society Transactions* vol. 39 575–580 (2011).

31. Paul-Dauphin, S. *et al.* Probing size exclusion mechanisms of complex hydrocarbon mixtures: The effect of altering eluent compositions. *Energy and Fuels* **21**, 3484–3489 (2007).

32. CHROMacademy. The Theory of HPLC: Chromatographic Parameters. *Crawford Scientific* 23 (2014).

33. Majors, R. E. The cleaning and regeneration of reversed-phase HPLC columns. *LC-GC Europe* vol. 16 404–409 Preprint at (2003).

34. Lee, Y. C. High-performance anion-exchange chromatography for carbohydrate analysis. *Analytical Biochemistry* vol. 189 151–162 Preprint at https://doi.org/10.1016/0003-2697(90)90099-U (1990).

35. Petrov, A., Tsa, A. & Puglisi, J. D. Analysis of RNA by analytical polyacrylamide

gel electrophoresis. in *Methods in Enzymology* vol. 530 301–313 (2013).

36. Albright, L. M. & Slatko, B. E. Denaturing polyacrylamide gel electrophoresis. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **Appendix 3**, (2001).

37. Bartlett, J. M. S. & Stirling, D. A Short History of the Polymerase Chain Reaction. in *PCR Protocols* 3–6 (2003). doi:10.1385/1-59259-384-4:3.

38. Saiki, R. K. *et al.* Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).

39. Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491 (1988).

40. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of Bacteriology* **127**, 1550–1557 (1976).

41. Brock, T. D. & Freeze, H. Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of bacteriology* **98**, 289–297 (1969).

42. U.G. PCR protocols — A guide to methods and applications. *Trends in Biochemical Sciences* **15**, 405–406 (1990).

43. Bullock, M. R. *et al.* Neurosurgery: Introduction. *Neurosurgery* **58**, (2006).

44. Dorak, M. T. *Real-time PCR*. *Real-time PCR* (2007). doi:10.4324/9780203967317.

45. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Research* **6**, 986–994 (1996).

46. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**, 611–622 (2009).

47. Shangguan, D. *et al.* Aptamers evolved from live cells as effective molecular probes for cancer study. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11838–11843 (2006).

48. Prakash, J. S. & Rajamanickam, K. Aptamers and their significant role in cancer therapy and diagnosis. *Biomedicines* vol. 3 248–269 Preprint at https://doi.org/10.3390/biomedicines3030248 (2015).

49. Ku, T. H. *et al.* Nucleic acid aptamers: An emerging tool for biotechnology and biomedical sensing. *Sensors (Switzerland)* vol. 15 16281–16313 Preprint at https://doi.org/10.3390/s150716281 (2015).

50. Cerchia, L. *et al.* Neutralizing aptamers from whole-cell SELEX inhibit the RET receptor tyrosine kinase. *PLoS Biology* **3**, 0697–0704 (2005).

51. Wolter, O. & Mayer, G. Aptamers as valuable molecular tools in neurosciences. *Journal of Neuroscience* **37**, 2517–2523 (2017).

52. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

53. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

54. Ku, T. H. *et al.* Nucleic acid aptamers: An emerging tool for biotechnology and biomedical sensing. *Sensors (Switzerland)* vol. 15 16281–16313 Preprint at https://doi.org/10.3390/s150716281 (2015).

55. Piganeau, N. & Schroeder, R. Aptamer structures: A preview into regulatory pathways? *Chemistry and Biology* vol. 10 103–104 Preprint at https://doi.org/10.1016/S1074-5521(03)00028-0 (2003).

56. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. *Science* vol. 287 820–825 Preprint at https://doi.org/10.1126/science.287.5454.820 (2000).

57. Patel, D. J. *et al.* Structure, recognition and adaptive binding in RNA aptamer complexes. *Journal of Molecular Biology* vol. 272 645–664 Preprint at https://doi.org/10.1006/jmbi.1997.1281 (1997).

58. Feigon, J., Dieckmann, T. & Smith, F. W. Aptamer structures from A to ζ.

*Chemistry and Biology* vol. 3 611–617 Preprint at https://doi.org/10.1016/S1074-5521(96)90127-1 (1996).

59. Mallikaratchy, P. Evolution of complex target SELEX to identify aptamers against mammalian cell-surface antigens. *Molecules* vol. 22 Preprint at https://doi.org/10.3390/molecules22020215 (2017).

60. Ireson, C. R. & Kelland, L. R. Discovery and development of anticancer aptamers. *Molecular Cancer Therapeutics* vol. 5 2957–2962 Preprint at https://doi.org/10.1158/1535-7163.MCT-06-0172 (2006).

61. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* vol. 24 381–403 Preprint at https://doi.org/10.1016/j.bioeng.2007.06.001 (2007).

62. Birch, J. R. & Racher, A. J. Antibody production. *Advanced Drug Delivery Reviews* vol. 58 671–685 Preprint at https://doi.org/10.1016/j.addr.2005.12.006 (2006).

63. Jellineky, D. *et al.* Potent 2'-Amino-2'-deoxypyrimidine RNA Inhibitors of Basic Fibroblast Growth Factor. *Biochemistry* **34**, 11363–11372 (1995).

64. Padilla, R. & Sousa, R. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs.

*Nucleic acids research* **30**, (2002).

65. Yan, X., Gao, X. & Zhang, Z. Isolation and characterization of 2'-amino-modified RNA aptamers for human TNFalpha. *Genomics, proteomics & bioinformatics / Beijing Genomics Institute* **2**, 32–42 (2004).

66. Dellafiore, M. A., Montserrat, J. M. & Iribarren, A. M. Modified nucleoside triphosphates for in-vitro selection techniques. *Frontiers in Chemistry* vol. 4 Preprint at https://doi.org/10.3389/fchem.2016.00018 (2016).

67. Morita, Y., Leslie, M., Kameyama, H., Volk, D. E. & Tanaka, T. Aptamer therapeutics in cancer: Current and future. *Cancers* vol. 10 Preprint at https://doi.org/10.3390/cancers10030080 (2018).

68. Volk, D. E. & Lokesh, G. L. R. Development of phosphorothioate DNA and DNA thioaptamers. *Biomedicines* vol. 5 Preprint at https://doi.org/10.3390/biomedicines5030041 (2017).

69. Argos, P. *et al.* Thermal Stability and Protein Structure. *Biochemistry* **18**, 5698–5703 (1979).

70. SantaLucia, J. & Hicks, D. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure* vol. 33 415–440 Preprint at https://doi.org/10.1146/annurev.biophys.32.110601.141800 (2004).

71. Pontius, B. W. & Berg, P. Rapid renaturation of complementary DNA strands mediated by cationic detergents: A role for high-probability binding domains in enhancing the kinetics of molecular assembly processes. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 8237–8241 (1991).

72. Song, K. M., Lee, S. & Ban, C. Aptamers and their biological applications. *Sensors* vol. 12 612–631 Preprint at https://doi.org/10.3390/s120100612 (2012).

73. Ciesiolka, J., Gorski, J. & Yarus, M. Selection of an RNA domain that binds Zn2+. *Rna* **1**, 538–550 (1995).

74. Liu, Z. *et al.* Novel HER2 Aptamer Selectively Delivers Cytotoxic Drug to HER2-positive Breast Cancer Cells in Vitro. *Journal of Translational Medicine* **10**, (2012).

75. Shangguan, D. *et al.* Aptamers evolved from live cells as effective molecular probes for cancer study. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11838–11843 (2006).

76. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

77. Stoltenburg, R., Schubert, T. & Strehlitz, B. In vitro selection and interaction

studies of a DNA aptamer targeting Protein A. *PLoS ONE* **10**, (2015).

78. Ipe, B. I., Neurock, M. & Filhol, J. S. Preview: Angew. Chem. Int. Ed. 2/2006. *Angewandte Chemie International Edition* (2006).

79. Gbelcová, H. *et al.* RAS proteins key regulators of cell cycle. *Chemicke Listy* **109**, 364–370 (2015).

80. Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**, 441–448 (1975).

81. F. SANGER, S. NICKLEN, A. A. R. C. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* **74**, 5463–5467 (1977).

82. Leroy Hood & David Galas. The digital code of DNA. *Nature* **421**, 444–448 (2003).

83. Abbott, A. Human genome at ten: The human race. *Nature* **464**, 668–669 (2010).

84. Paolillo, C., Londin, E. & Fortina, P. Next generation sequencing in cancer: opportunities and challenges for precision cancer medicine. *Scandinavian journal of clinical and laboratory investigation. Supplementum* **245**, S84-91 (2016).

85. Schütze, T. *et al.* Probing the SELEX process with next-generation sequencing.

*PLoS ONE* **6**, e29604 (2011).

86. Quang, N. N., Perret, G. & Ducongé, F. Applications of high-throughput sequencing for in vitro selection and characterization of aptamers. *Pharmaceuticals* vol. 9 Preprint at https://doi.org/10.3390/ph9040076 (2016).

87. Illumina Inc. Illumina sequencing introduction. *Illumina sequencing introduction* 1–8 (2017).

88. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome biology* **14**, R51 (2013).

89. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

90. Sefah, K. *et al.* In vitro selection with artificial expanded genetic information systems. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1449–1454 (2014).

91. Kimoto, M., Yamashige, R., Matsunaga, K. I., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nature Biotechnology* **31**, 453–457 (2013).

92. Gawande, B. N. *et al.* Selection of DNA aptamers with two modified bases. *Proceedings of the National Academy of Sciences of the United States of America*

**114**, 2898–2903 (2017).

93. Hobbs, J., Sternbach, H., Sprinzl, M. & Eckstein, F. Polynucleotides Containing 2′-Amino-2′-deoxyribose and 2′-Azido-2′-deoxyribose. *Biochemistry* **12**, 5138–5145 (1973).

94. Rohloff, J. C. *et al.* Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. *Molecular Therapy - Nucleic Acids* **3**, e201 (2014).

95. Chen, T. *et al.* Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nature Chemistry* **8**, 556–562 (2016).

96. Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54–57 (1997).

97. 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).

98. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344 (2012).

99. Taylor, A. I. *et al.* Catalysts from synthetic genetic polymers. *Nature* **518**, 427–430 (2015).

100. Crouzier, L. *et al.* Efficient reverse transcription using locked nucleic acid nucleotides towards the evolution of nuclease resistant RNA aptamers. *PloS one* **7**, (2012).

101. Olea, C., Weidmann, J., Dawson, P. E. & Joyce, G. F. An L-RNA Aptamer that Binds and Inhibits RNase. *Chemistry and Biology* **22**, 1437–1441 (2015).

102. Sidorov, A. V., Grasby, J. A. & Williams, D. M. Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and amino functionalities. *Nucleic Acids Research* **32**, 1591–1601 (2004).

103. Jäger, S. & Famulok, M. Generation and enzymatic amplification of high-density functionalized DNA double strands. *Angewandte Chemie - International Edition* **43**, 3337–3340 (2004).

104. Eaton, B. E. & Pieken, W. A. Ribonucleosides and RNA. *Annual Review of Biochemistry* vol. 64 837–863 Preprint at https://doi.org/10.1146/annurev.bi.64.070195.004201 (1995).

105. Eaton, B. E. *et al.* Post-SELEX combinatorial optimization of aptamers. in *Bioorganic and Medicinal Chemistry* vol. 5 1087–1096 (1997).

106. 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).

107. Thomas, J. M., Yoon, J. K. & Perrin, D. M. Investigation of the catalytic mechanism of a synthetic DNAzyme with protein-like functionality: An RNaseA mimic? *Journal of the American Chemical Society* **131**, 5648–5658 (2009).

108. Battersby, T. R. *et al.* Quantitative analysis of receptors for adenosine nucleotides obtained via in vitro selection from a library incorporating a cationic nucleotide analog. *Journal of the American Chemical Society* **121**, 9781–9789 (1999).

109. Latham, J. A., Johnson, R. & Toole, J. J. The application of a modified nucleotide in aptamer selection: Novel thrombin aptamers containing -(1 -pentynyl)-2′-deoxyuridine. *Nucleic Acids Research* **22**, 2817–2822 (1994).

110. Masud, M. M., Kuwahara, M., Ozaki, H. & Sawai, H. Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorganic and Medicinal Chemistry* **12**, 1111–1120 (2004).

111. Ohsawa, K. *et al.* Arginine-modified DNA aptamers that show enantioselective recognition of the dicarboxylic acid moiety of glutamic acid. *Analytical Sciences* **24**, 167–172 (2008).

112. Santoro, S. W., Joyce, G. F., Sakthivel, K., Gramatikova, S. & Barbas, C. F. RNA cleavage by a DNA enzyme with extended chemical functionality. *Journal of the American Chemical Society* **122**, 2433–2439 (2000).

113. Shoji, A., Kuwahara, M., Ozaki, H. & Sawai, H. Modified DNA aptamer that binds the (R)-isomer of a thalidomide derivative with high enantioselectivity. *Journal of the American Chemical Society* **129**, 1456–1464 (2007).

114. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, (2010).

115. Mehan, M. R. *et al.* Highly multiplexed proteomic platform for biomarker discovery, diagnostics, and therapeutics. in *Advances in Experimental Medicine and Biology* vol. 734 283–300 (2013).

116. Vaught, J. D. *et al.* Expanding the chemistry of DNA for in vitro selection. *Journal of the American Chemical Society* **132**, 4141–4151 (2010).

117. Eaton, B. E. The joys of in vitro selection: Chemically dressing oligonucleotides to satiate protein targets. *Current Opinion in Chemical Biology* **1**, 10–16 (1997).

118. Brody, E. N., Gold, L., Lawn, R. M., Walker, J. J. & Zichi, D. High-content affinity-based proteomics: Unlocking protein biomarker discovery. *Expert Review of Molecular Diagnostics* vol. 10 1013–1022 Preprint at

https://doi.org/10.1586/erm.10.89 (2010).

119. Kraemer, S. *et al.* From SOMAmer-based biomarker discovery to diagnostic and clinical applications: A SOMAmer-based, streamlined multiplex proteomic assay. *PLoS ONE* **6**, (2011).

120. Rohloff, J. C. *et al.* Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. *Molecular Therapy - Nucleic Acids* vol. 3 e201 Preprint at https://doi.org/10.1038/mtna.2014.49 (2014).

121. Dewey, T. M., Zyzniewski, M. C., Mundt, A. A., Crouch, G. J. & Eaton, B. E. New Uridine Derivatives for Systematic Evolution of RNA Ligands by Exponential Enrichment. *Journal of the American Chemical Society* **117**, 8474–8475 (1995).

122. Rohloff, J. C. *et al.* Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. *Molecular Therapy - Nucleic Acids* vol. 3 e201 Preprint at https://doi.org/10.1038/mtna.2014.49 (2014).

123. Liu, E., Lam, C. H. & Perrin, D. M. Synthesis and enzymatic incorporation of modified deoxyuridine triphosphates. *Molecules* **20**, 13591–13602 (2015).

124. Davies, D. R. *et al.* Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19971–19976 (2012).

125. Gelinas, A. D., Davies, D. R. & Janjic, N. Embracing proteins: Structural themes in aptamer-protein complexes. *Current Opinion in Structural Biology* vol. 36 122–132 Preprint at https://doi.org/10.1016/j.sbi.2016.01.009 (2016).

126. Iyidogan, P., Sullivan, T. J., Chordia, M. D., Frey, K. M. & Anderson, K. S. Design, synthesis, and antiviral evaluation of chimeric inhibitors of HIV reverse transcriptase. *ACS Medicinal Chemistry Letters* **4**, 1183–1188 (2013).

127. Djerassi, C. Brominations with N-Bromosuccinimide and Related Compounds. The Wohl-Ziegler Reaction. *Chemical Reviews* **43**, 271–317 (1948).

128. Walling, C., Rieger, A. L. & Tanner, D. D. Positive Halogen Compounds. VIII. Structure and Reactivity in N-Bromosuccinimide Brominations. *Journal of the American Chemical Society* **85**, 3129–3134 (1963).

129. Amijs, C. H. M., Van Klink, G. P. M. & Van Koten, G. Carbon tetrachloride free benzylic brominations of methyl aryl halides. *Green Chemistry* **5**, 470–474 (2003).

130. Thapa, R., Brown, J., Balestri, T. & Taylor, R. T. Regioselectivity in free radical bromination of unsymmetrical dimethylated pyridines. *Tetrahedron Letters* **55**, 6743–6746 (2014).

131. Kistemaker, H. A. V., Meeuwenoord, N. J., Overkleeft, H. S., Van Der Marel, G. A. & Filippov, D. V. On the Synthesis of Oligonucleotides Interconnected through

Pyrophosphate Linkages. *European Journal of Organic Chemistry* **2015**, 6084–6091 (2015).

132. Gukathasan, R. *et al.* Large-scale synthesis of high purity 'phos reagent' useful for oligonucleotide therapeutics. in *Journal of Organometallic Chemistry* vol. 690 2603–2607 (2005).

133. Meher, G., Efthymiou, T., Stoopa, M. & Krishnamurthy, R. Microwave-assisted preparation of nucleoside-phosphoramidites. *Chemical Communications* **50**, 7463–7465 (2014).

134. Ludwig, J. A new route to nucleoside 5′-triphosphates. *Acta Biochimica et Biophysica Academiae Scientiarum Hungaricae* **16**, 131–133 (1981).

135. Kore, A. R., Shanmugasundaram, M., Senthilvelan, A. & Srinivasan, B. An improved protection-free one-pot chemical synthesis of 2′-deoxynucleoside-5′-triphosphates. *Nucleosides, Nucleotides and Nucleic Acids* **31**, 423–431 (2012).

136. 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. *Journal of Organic Chemistry* **54**, 631–635 (1989).

137. 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 and Silicon and the Related Elements* **190**, 758–762 (2015).

138. Gillerman, I. & Fischer, B. An improved one-pot synthesis of nucleoside 5′-triphosphate analogues. *Nucleosides, Nucleotides and Nucleic Acids* **29**, 245–256 (2010).

139. Ren, X., El-Sagheer, A. H. & Brown, T. Azide and trans-cyclooctene dUTPs: Incorporation into DNA probes and fluorescent click-labelling. *Analyst* **140**, 2671–2678 (2015).

140. Wu, W. *et al.* Termination of DNA synthesis by N6-alkylated, not 3′-O-alkylated, photocleavable 2′-deoxyadenosine triphosphates. *Nucleic Acids Research* **35**, 6339–6349 (2007).

141. Ren, X., Gerowska, M., El-Sagheer, A. H. & Brown, T. Enzymatic incorporation and fluorescent labelling of cyclooctyne-modified deoxyuridine triphosphates in DNA. *Bioorganic and Medicinal Chemistry* **22**, 4384–4390 (2014).

142. Hollenstein, M., Smith, C. C. & Räz, M. Nucleoside triphosphates - From synthesis to biochemical characterization. *Journal of Visualized Experiments* (2014) doi:10.3791/51385.

143. Lam, C. H., Hipolito, C. J., Hollenstein, M. & Perrin, D. M. A divalent metal-dependent self-cleaving DNAzyme with a tyrosine side chain. *Organic and Biomolecular Chemistry* **9**, 6949–6954 (2011).

144. Ren, X., El-Sagheer, A. H. & Brown, T. Efficient enzymatic synthesis and dual-colour fluorescent labelling of DNA probes using long chain azido-dUTP and BCN dyes. *Nucleic Acids Research* **44**, e79 (2016).

145. Ren, X., El-Sagheer, A. H. & Brown, T. Azide and trans-cyclooctene dUTPs: Incorporation into DNA probes and fluorescent click-labelling. *Analyst* **140**, 2671–2678 (2015).

146. Ren, X., El-Sagheer, A. H. & Brown, T. Efficient enzymatic synthesis and dual-colour fluorescent labelling of DNA probes using long chain azido-dUTP and BCN dyes. *Nucleic Acids Research* **44**, e79 (2016).

147. Filée, J., Forterre, P., Sen-Lin, T. & Laurent, J. Evolution of DNA polymerase families: Evidences for multiple gene exchange between cellular and viral proteins. *Journal of Molecular Evolution* **54**, 763–773 (2002).

148. Paul, N. & Yee, J. PCR incorporation of modified dNTPs: The substrate properties of biotinylated dNTPs. *BioTechniques* **48**, 333–334 (2010).

149. Ondruš, M., Sýkorová, V., Bednárová, L., Pohl, R. & Hocek, M. Enzymatic

synthesis of hypermodified DNA polymers for sequence-specific display of four different hydrophobic groups. *Nucleic Acids Research* **48**, 11982–11993 (2020).

150. Hottin, A. & Marx, A. Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Accounts of Chemical Research* vol. 49 418–427 Preprint at https://doi.org/10.1021/acs.accounts.5b00544 (2016).

151. Tasara, T. *et al.* Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. *Nucleic Acids Research* **31**, 2636–2646 (2003).

152. 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. *Bioorganic and Medicinal Chemistry Letters* **26**, 288–291 (2016).

153. Paul, N. & Yee, J. PCR incorporation of modified dNTPs: The substrate properties of biotinylated dNTPs. *BioTechniques* vol. 48 333–334 Preprint at https://doi.org/10.2144/000113405 (2010).

154. 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. *Journal of Organic Chemistry* **79**, 9914–9921 (2014).

155. Kielkowski, P., Fanfrlík, J. & Hocek, M. 7-aryl-7-deazaadenine 2′-deoxyribonucleoside triphosphates (dNTPs): Better substrates for DNA polymerases than dATP in competitive incorporations. *Angewandte Chemie - International Edition* **53**, 7552–7555 (2014).

156. Cahová, H., Panattoni, A., Kielkowski, P., Fanfrlík, J. & Hocek, M. 5-substituted pyrimidine and 7-substituted 7-deazapurine dNTPs as substrates for DNA polymerases in competitive primer extension in the presence of natural dNTPs. *ACS Chemical Biology* **11**, 3165–3171 (2016).

157. Bergen, K. *et al.* Structures of KlenTaq DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates. *Journal of the American Chemical Society* **134**, 11840–11843 (2012).

158. Kropp, H. M., Diederichs, K. & Marx, A. The Structure of an Archaeal B-Family DNA Polymerase in Complex with a Chemically Modified Nucleotide. *Angewandte Chemie - International Edition* **58**, 5457–5461 (2019).

159. Hocek, M. Enzymatic Synthesis of Base-Functionalized Nucleic Acids for Sensing, Cross-linking, and Modulation of Protein-DNA Binding and Transcription. *Accounts of Chemical Research* **52**, 1730–1737 (2019).

160. Thum, O., Jger, S. & Famulok, M. Functionalized DNA: A new replicable biopolymer. *Angewandte Chemie - International Edition* **40**, 3990–3993 (2001).

161. Jäger, S. *et al.* A versatile toolbox for variable DNA functionalization at high density. *Journal of the American Chemical Society* **127**, 15071–15082 (2005).

162. Hollenstein, M., Hipolito, C. J., Lam, C. H. & Perrin, D. M. A DNAzyme with three protein-like functional groups: Enhancing catalytic efficiency of M2+-independent RNA cleavage. *ChemBioChem* **10**, 1988–1992 (2009).

163. Hollenstein, M. Deoxynucleoside triphosphates bearing histamine, carboxylic acid, and hydroxyl residues-synthesis and biochemical characterization. *Organic and Biomolecular Chemistry* **11**, 5162–5172 (2013).

164. Balintová, J. *et al.* Benzofurazane as a new redox label for electrochemical detection of DNA: Towards multipotential redox coding of DNA bases. *Chemistry - A European Journal* **19**, 12720–12731 (2013).

165. Simonova, A. *et al.* Tuning of Oxidation Potential of Ferrocene for Ratiometric Redox Labeling and Coding of Nucleotides and DNA. *Chemistry - A European Journal* **26**, 1286–1291 (2020).

166. Wang, Y., Liu, E., Lam, C. H. & Perrin, D. M. A densely modified M2+-independent DNAzyme that cleaves RNA efficiently with multiple catalytic

turnover. *Chemical Science* **9**, 1813–1821 (2018).

167. Dumat, B. *et al.* Second-generation fluorescent quadracyclic adenine analogues: Environment-responsive probes with enhanced brightness. *Chemistry - A European Journal* **21**, 4039–4048 (2015).

168. Wranne, M. S. *et al.* Toward Complete Sequence Flexibility of Nucleic Acid Base Analogue FRET. *Journal of the American Chemical Society* **139**, 9271–9280 (2017).

169. Ban, Ž. *et al.* Flexibility and preorganization of fluorescent nucleobase-pyrene conjugates control DNA and RNA recognition. *Molecules* **25**, (2020).

170. Bood, M., Sarangamath, S., Wranne, M. S., Grøtli, M. & Wilhelmsson, L. M. Fluorescent nucleobase analogues for base-base FRET in nucleic acids: Synthesis, photophysics and applications. *Beilstein Journal of Organic Chemistry* vol. 14 114–129 Preprint at https://doi.org/10.3762/bjoc.14.7 (2017).

171. Kool, E. T. Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annual Review of Biophysics and Biomolecular Structure* vol. 30 1–22 Preprint at https://doi.org/10.1146/annurev.biophys.30.1.1 (2001).

172. Mamedov, T. G. *et al.* A fundamental study of the PCR amplification of GC-rich DNA templates. *Computational Biology and Chemistry* **32**, 452–457 (2008).

173. Assal, N. & Lin, M. PCR procedures to amplify GC-rich DNA sequences of mycobacterium bovis. *bioRxiv* Preprint at https://doi.org/10.1101/2020.02.18.953695 (2020).

174. Mytelka, D. S. & Chamberlin, M. J. Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. *Nucleic Acids Research* **24**, 2774–2781 (1996).

175. Ralser, M. *et al.* An efficient and economic enhancer mix for PCR. *Biochemical and Biophysical Research Communications* **347**, 747–751 (2006).

176. Musso, M., Bocciardi, R., Parodi, S., Ravazzolo, R. & Ceccherini, I. Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. *Journal of Molecular Diagnostics* **8**, 544–550 (2006).

177. Baskaran, N. *et al.* Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. *Genome Research* **6**, 633–638 (1996).

178. Sidhu, M. K., Liao, M. J. & Rashidbaigi, A. Dimethyl sulfoxide improves RNA amplification. *BioTechniques* **21**, 44–47 (1996).

179. Vigneault, F. & Drouin, R. Optimal conditions and specific characteristics of Vent exo- DNA polymerase in ligation-mediated polymerase chain reaction protocols. *Biochemistry and Cell Biology* **83**, 147–165 (2005).

180. Gawande, B. N. *et al.* Selection of DNA aptamers with two modified bases. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 2898–2903 (2017).

181. Odeh, F. *et al.* Aptamers chemistry: Chemical modifications and conjugation strategies. *Molecules* **25**, (2020).

182. Kratschmer, C. & Levy, M. Effect of Chemical Modifications on Aptamer Stability in Serum. *Nucleic Acid Therapeutics* **27**, 335–344 (2017).

183. Vorobyeva, M. A., Davydova, A. S., Vorobjev, P. E., Pyshnyi, D. V. & Venyaminova, A. G. Key aspects of nucleic acid library design for in vitro selection. *International Journal of Molecular Sciences* **19**, (2018).

184. Pobanz, K. & Lupták, A. Improving the odds: Influence of starting pools on in vitro selection outcomes. *Methods* **106**, 14–20 (2016).

185. Cowperthwaite, M. C. & Ellington, A. D. Bioinformatic analysis of the contribution of primer sequences to aptamer structures. *Journal of Molecular Evolution* **67**, 95–102 (2008).

186. Hall, B. *et al.* Design, synthesis, and amplification of DNA pools for in vitro selection. *Current Protocols in Molecular Biology* 1–27 (2009) doi:10.1002/0471142727.mb2402s88.

187. Davies, D. R. *et al.* Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19971–19976 (2012).

188. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, (2010).

189. Ohbayashi, T. *et al.* Expansion of repertoire of modified DNAs prepared by PCR using KOD Dash DNA polymerase. *Organic and Biomolecular Chemistry* **3**, 2463–2468 (2005).

190. Liang, S. *et al.* Measuring luteinising hormone pulsatility with a robotic aptamer-enabled electrochemical reader. *Nature Communications* **10**, 1–10 (2019).

191. Kim, M., Kim, D. M., Kim, K. S., Jung, W. & Kim, D. E. Applications of cancer cell-specific aptamers in targeted delivery of anticancer therapeutic agents. *Molecules* **23**, 1–20 (2018).

192. Su, J. L. *et al.* ssDNA aptamer-based surface plasmon resonance biosensor for the detection of retinol binding protein 4 for the early diagnosis of type 2 diabetes. *Analytical Chemistry* vol. 80 2867–2873 Preprint at https://doi.org/10.1021/ac800050a (2008).

193. Kim, H. J. *et al.* PET imaging of HER2 expression with an 18 F-fluoride labeled

aptamer. *PLoS ONE* **14**, 1–14 (2019).

194. Liu, Z. *et al.* Novel HER2 Aptamer Selectively Delivers Cytotoxic Drug to HER2-positive Breast Cancer Cells in Vitro. *Journal of Translational Medicine* **10**, (2012).

195. Kim, M. Y. & Jeong, S. In vitro selection of RNA aptamer and specific targeting of ErbB2 in breast cancer cells. *Nucleic Acid Therapeutics* **21**, 173–178 (2011).

196. Jeong, H. Y. *et al.* Development of HER2-specific aptamer-drug conjugate for breast cancer therapy. *International Journal of Molecular Sciences* **21**, 1–13 (2020).

197. Mahlknecht, G. *et al.* Aptamer to ErbB-2/HER2 enhances degradation of the target and inhibits tumorigenic growth. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8170–8175 (2013).

198. Andrew D. Ellington & Jack W.Szostak. Selection in vitro of single. stranded DNA molecules that fold into specific ligand.binding structures. *Nature* **335**, 850–852 (1992).

199. Irvine, D., Tuerk, C. & Gold, L. Selexion. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *Journal of Molecular Biology* **222**, 739–761 (1991).

200. Gracie, K. *et al.* Interaction of fluorescent dyes with DNA and spermine using fluorescence spectroscopy. *Analyst* **139**, 3735–3743 (2014).

201. Mao, F., Leung, W. Y. & Xin, X. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnology* **7**, 1–16 (2007).

202. Chu, B., Zhang, D. & Paukstelis, P. J. A DNA G-quadruplex/i-motif hybrid. *Nucleic Acids Research* gkz1008 (2019) doi:10.1093/nar/gkz1008.

203. Todd, A. K. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Research* **33**, 2901–2907 (2005).

204. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5381–5383 (1992).

205. Needels, M. C. *et al.* Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10700–10704 (1993).

206. Nielsen, J., Brenner, S. & Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *Journal of the American Chemical Society* **115**, 9812–9813 (1993).

207. Chan, A. I., McGregor, L. M. & Liu, D. R. Novel selection methods for DNA-encoded chemical libraries. *Current Opinion in Chemical Biology* **26**, 55–61 (2015).

208. Shangguan, D. *et al.* Aptamers evolved from live cells as effective molecular probes for cancer study. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11838–11843 (2006).

209. Prakash, J. S. & Rajamanickam, K. Aptamers and their significant role in cancer therapy and diagnosis. *Biomedicines* **3**, 248–269 (2015).

210. Ku, T. H. *et al.* Nucleic acid aptamers: An emerging tool for biotechnology and biomedical sensing. *Sensors (Switzerland)* vol. 15 16281–16313 Preprint at https://doi.org/10.3390/s150716281 (2015).

211. Cerchia, L. *et al.* Neutralizing aptamers from whole-cell SELEX inhibit the RET receptor tyrosine kinase. *PLoS Biology* **3**, 0697–0704 (2005).

212. Eidam, O. & Satz, A. L. Analysis of the productivity of DNA encoded libraries. *MedChemComm* **7**, 1323–1331 (2016).

213. Kiessling, L. L., Gestwicki, J. E. & Strong, L. E. Synthetic multivalent ligands in the exploration of cell-surface interactions. *Current Opinion in Chemical Biology* vol. 4 696–703 Preprint at https://doi.org/10.1016/S1367-5931(00)00153-8

(2000).

214. Vauquelin, G. & Charlton, S. J. Exploring avidity: Understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *British Journal of Pharmacology* **168**, 1771–1785 (2013).

215. Kim, Y., Cao, Z. & Tan, W. Molecular assembly for high-performance bivalent nucleic acid inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5664–5669 (2008).

216. Wichert, M. *et al.* Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nature Chemistry* **7**, 241–249 (2015).

217. Krauss, I. R. *et al.* Thrombin-aptamer recognition: A revealed ambiguity. *Nucleic Acids Research* **39**, 7858–7867 (2011).

218. Zhao, X., Lis, J. T. & Shi, H. A systematic study of the features critical for designing a high avidity multivalent aptamer. *Nucleic Acid Therapeutics* **23**, 238–242 (2013).

219. Mack, E. T., Perez-Castillejos, R., Suo, Z. & Whitesides, G. M. Exact analysis of ligand-induced dimerization of monomeric receptors. *Analytical Chemistry* **80**, 5550–5555 (2008).

220. Nathan, A. J. & Scobell, A. How China sees America. *Foreign Affairs* **91**, 13070–

13075 (2012).

221. Nimjee, S. M. & Sullenger, B. A. Aptamers in the clinic. *Drugs of the Future* vol. 34 897–901 Preprint at https://doi.org/10.1358/dof.2009.034.11.1445907 (2009).

222. Tian, L. & Heyduk, T. Bivalent ligands with long nanometer-scale flexible linkers. *Biochemistry* **48**, 264–275 (2009).

223. Varizhuk, A. M. *et al.* Synthesis, characterization and in vitro activity of thrombin-binding DNA aptamers with triazole internucleotide linkages. *European Journal of Medicinal Chemistry* **67**, 90–97 (2013).

224. Huntington, J. A. Natural inhibitors of thrombin. *Thrombosis and Haemostasis* **111**, 583–589 (2014).

225. Russo Krauss, I. *et al.* High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Research* **40**, 8119–8128 (2012).

226. Long, S. B., Long, M. B., White, R. R. & Sullenger, B. A. Crystal structure of an RNA aptamer bound to thrombin. *Rna* **14**, 2504–2512 (2008).

227. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254**, 1497–1500 (1991).

228. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. Peptide Nucleic Acids (PNA). Oligonucleotide Analogues with an Achiral Peptide Backbone. *Journal of the American Chemical Society* vol. 114 1895–1897 Preprint at https://doi.org/10.1021/ja00031a062 (1992).

229. Nielsen, P. E. & Haaima, G. Peptide nucleic acid (PNA). A DNA mimic with a pseudopeptide backbone. *Chemical Society Reviews* **26**, 73 (1997).

230. Egholm, M. *et al.* PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993).

231. N. Ganesh, K. & E. Nielsen, P. Peptide Nucleic Acids: Analogs and Derivatives. *Current Organic Chemistry* vol. 4 931–943 Preprint at https://doi.org/10.2174/1385272003375969 (2005).

232. Nielsen, P. E. & Egholm, M. An introduction to peptide nucleic acid. *Current issues in molecular biology* vol. 1 89–104 Preprint at https://doi.org/10.21775/cimb.001.089 (1999).

233. Good, L., Awasthi, S. K., Dryselius, R., Larsson, O. & Nielsen, P. E. Bactericidal antisense effects of peptide - PNA conjugates. *Nature Biotechnology* **19**, 360–364 (2001).

234. Sazani, P. *et al.* Systemically delivered antisense oligomers upregulate gene

expression in mouse tissues. *Nature Biotechnology* **20**, 1228–1233 (2002).

235. Koppelhus, U. & Nielsen, P. E. Cellular delivery of peptide nucleic acid (PNA). *Advanced Drug Delivery Reviews* **55**, 267–280 (2003).

236. Ball, R. J., Green, P. S., Gale, N., Langley, G. J. & Brown, T. Peptide nucleic acid probes with charged photocleavable mass markers: Towards PNA-based MALDI-TOF MS genetic analysis. *Artificial DNA: PNA and XNA* **1**, 27–35 (2010).

237. Feagin, T. A., Shah, N. I. & Heemstra, J. M. Convenient and scalable synthesis of Fmoc-protected peptide nucleic acid backbone. *Journal of Nucleic Acids* **2012**, (2012).

238. Egholm, M. *et al.* PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993).

239. Wojciechowski, F. & Hudson, R. H. E. A convenient route to N-[2-(Fmoc)aminoethyl]glycine esters and PNA oligomerization using a bis-N-Boc nucleobase protecting group strategy. *Journal of Organic Chemistry* **73**, 3807–3816 (2008).

240. Thomson, S. A. *et al.* Fmoc mediated synthesis of Peptide Nucleic Acids. *Tetrahedron* **51**, 6179–6194 (1995).

241. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. Peptide Nucleic Acids

(PNA). Oligonucleotide Analogues with an Achiral Peptide Backbone. *Journal of the American Chemical Society* **114**, 1895–1897 (1992).

242. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254**, 1497–1500 (1991).

243. Fukuyama, T., Jow, C. K. & Cheung, M. 2- and 4-Nitrobenzenesulfonamides: Exceptionally versatile means for preparation of secondary amines and protection of amines. *Tetrahedron Letters* **36**, 6373–6374 (1995).

244. Pipkorn, R. *et al.* Improved synthesis strategy for peptide nucleic acids (PNA) appropriate for cell-specific fluorescence imaging. *International Journal of Medical Sciences* **9**, 1–10 (2012).

245. Matysiak, S. *et al.* Searching for avidity by chemical ligation of combinatorially self-assembled DNA-encoded ligand libraries. *Organic and Biomolecular Chemistry* **16**, 48–52 (2017).

246. El-Sagheer, A. H. & Brown, T. Click nucleic acid ligation: Applications in biology and nanotechnology. *Accounts of Chemical Research* **45**, 1258–1267 (2012).

247. El-Sagheer, A. H. & Brown, T. Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *Journal of*

*the American Chemical Society* **131**, 3958–3964 (2009).

248. Kim, Y., Cao, Z. & Tan, W. Molecular assembly for high-performance bivalent nucleic acid inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5664–5669 (2008).

249. Birts, C. N. *et al.* Transcription of click-linked DNA in human cells. *Angewandte Chemie - International Edition* **53**, 2362–2365 (2014).

250. Banga, R. J., Chernyak, N., Narayan, S. P., Nguyen, S. T. & Mirkin, C. A. Liposomal spherical nucleic acids. *Journal of the American Chemical Society* **136**, 9866–9869 (2014).

251. Ustinov, A. V. *et al.* Modification of nucleic acids using [3 + 2]-dipolar cycloaddition of azides and alkynes. *Russian Journal of Bioorganic Chemistry* **36**, 401–445 (2010).

252. Keefe, A. D., Clark, M. A., Hupp, C. D., Litovchick, A. & Zhang, Y. Chemical ligation methods for the tagging of DNA-encoded chemical libraries. *Current Opinion in Chemical Biology* **26**, 80–88 (2015).

253. 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. *Proceedings of the National Academy of Sciences*

*of the United States of America* **108**, 11338–11343 (2011).

254. Matysiak, S. *et al.* Searching for avidity by chemical ligation of combinatorially self-assembled DNA-encoded ligand libraries. *Organic and Biomolecular Chemistry* **16**, 48–52 (2017).

255. Sayed, E. M. *et al.* Nitrophenyl-Group-Containing Heterocycles. I. Synthesis, Characterization, Crystal Structure, Anticancer Activity, and Antioxidant Properties of Some New 5,6,7,8-Tetrahydroisoquinolines Bearing 3(4)-Nitrophenyl Group. *ACS Omega* **7**, 8767–8776 (2022).

256. Nepali, K., Lee, H.-Y. & Liou, J.-P. Nitro-Group-Containing Drugs. *J. Med. Chem.* **62**, 2851–2893 (2019).

257. Metrangolo, P., Meyer, F., Pilati, T., Resnati, G. & Terraneo, G. Halogen bonding in supramolecular chemistry. *Angewandte Chemie - International Edition* vol. 47 6114–6127 Preprint at https://doi.org/10.1002/anie.200800128 (2008).

258. Gilday, L. C. *et al.* Halogen Bonding in Supramolecular Chemistry. *Chemical Reviews* vol. 115 7118–7195 Preprint at https://doi.org/10.1021/cr500674c (2015).

259. Metrangolo, P., Neukirch, H., Pilati, T. & Resnati, G. Halogen bonding based recognition processes: A world parallel to hydrogen bonding. *Accounts of Chemical Research* **38**, 386–395 (2005).

260. 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. *Nature Chemistry* **6**, 1039–1043 (2014).

261. Brown, A. & Beer, P. D. Halogen bonding anion recognition. *Chemical Communications* **52**, 8645–8658 (2016).

262. Breipohl, G., Will, D. W., Peyman, A. & Uhlmann, E. Novel synthetic routes to PNA monomers and PNA-DNA linker molecules. *Tetrahedron* **53**, 14671–14686 (1997).

263. Krapcho, A. P. & Kuell, C. S. Mono-protected diamines. n-tert-butoxycarbonyl-α,ω- alkanediamines from α,ω-alkanediamines. *Synthetic Communications* **20**, 2559–2564 (1990).

264. Fader, L. D., Boyd, M. & Tsantrizos, Y. S. Backbone modifications of aromatic peptide nucleic acid (APNA) monomers and their hybridization properties with DNA and RNA. *Journal of Organic Chemistry* **66**, 3372–3379 (2001).

265. Aldrian-Herrada, G., Rabié, A., Wintersteiger, R. & Brugidou, J. Solid-phase synthesis of peptide nucleic acid (PNA) monomers and their oligomerization using disulphide anchoring linkers. *Journal of Peptide Science* **4**, 266–281 (1998).

266. Reagents, P. Coupling Reagents. *Tetrahedron* 5603–5606 (1996).

267. Dicyclohexylcarbodiimide, C. *et al.* Coupling Reagents - Novabiochem. *Tetrahedron* 1–4 (1996).

268. El-Faham, A. & Albericio, F. Peptide coupling reagents, more than a letter soup. *Chemical Reviews* **111**, 6557–6602 (2011).