C5-Modified 2'-deoxycytidine-5'-triphosphates for DNA sequencing

Thesis submitted for the degree of Doctor of Philosophy

Esther Isobel Allen

Supervisor: Dr D M Williams

The University of Sheffield - Department of Chemistry

March 2016

Declaration

Except where specific references have been made to other sources, the work in this thesis is the original work of the author, and it has not been submitted, wholly or in part, for any other degree.

Esther Isobel Allen

Contents

Acknowledgements

Firstly, my thanks go to my supervisor David Williams for his support and guidance throughout my PhD. Thanks also to group members both past and present who have always made the lab an enjoyable place to work and provided help, support and tea when needed.

I am grateful to BBRSC and QuantuMDx for providing the funding for this research. My special thanks to the team at QuantuMDx for their infectious passion for research. It has been a privilege being part of your company. Thanks to Heather Murton for her kindness and support throughout my two placements in Newcastle and for sticking with me through my first tentative PAGE gels. Also for her much appreciated contribution of additional biological data, including the information on the Illumina compound shown in chapter four.

To my friends within the department, it is no exaggeration that this thesis would not have been completed without the many cups of tea and glasses of wine you have provided along the way. Many of you have celebrated (and occasionally commiserated!) alongside me throughout the last four years and I hope that it will continue for many years to come.

To my extended Sheffield family, you have left me with many amazing memories over the years; from a late night swim in Western park to spending far too much money on sushi and cocktails. Sheffield would not have been the same without you.

Ash, thank you. For always managing to make me smile, your support throughout the last four years and the endless belief you have had in me.

Finally to my small but perfectly formed family: Martha, you are everything a big sister should be. You have always been there when I have needed you and for that I am, as always, grateful. To my parents, I'm sorry. I don't imagine when you sent your daughter off to university 8 years ago you'd expect her to still be there! However I know that you are proud of me, and I hope that I can continue to make you so.

Abstract

Seven C5-modified 2'-deoxycytidine-5'-triphosphates have been synthesised as proof-of-concept molecules for use in a hand-held DNA sequencing by synthesis (SBS) device in development by QuantuMDx. SBS is a continuous sequencing method where 3' reversibly blocked base-modified 2'-deoxynucleoside-5'-triphosphates (dNTPs) are incorporated into the complementary DNA strand by a DNA polymerase. The dNTPs are modified at the C5 position of the pyrimidines, or the C7 position of the 7-deazapurines with a reporter group unique to each dNTP. Following incorporation the reporter group and blocking group are removed and the next modified dNTP incorporated to continue the SBS cycle.

The SBS device in development will sequence DNA on a nanowire, recognising each dNTP by its unique anionic reporter group causing a characteristic change in the surrounding electric field. This thesis describes the synthesis of 2'-deoxycytidine-5'-triphosphates (dCTPs) with anionic reporter groups attached by means of polyethylene glycol (PEG) linkers of differing lengths. Five reporter groups were investigated; trimesic acid, PAMAM dendrimers 1-D2 and 1-D3 and two oligodeoxyribonucleotides (ODNs), dT_6 and dT_{24}. The reporter group was attached to the dCTP through a *O*-(*N*-succinimidyl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TSTU) coupling or Cu(I) catalysed click reaction. A dCTP analogue containing a photocleavable, 2-nitrobenzyl linker is also described.

All modified dCTPs were tested as substrates for DNA polymerases using primer extension reactions. All dCTPs were successfully incorporated by Vent (exo-) and Therminator DNA polymerases as evidenced by polyacrylamide gel electrophoresis (PAGE). This result was then confirmed using arrayed primer extension reactions performed on a pre-functionalised glass surface.

Abbreviations

A	Adenine
A.U.	Arbitrary units
ADP	Adenosine-5'-diphosphate
Aq.	Aqueous
ATP	Adenosine-5'-triphosphate
C	Cytosine
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
DCC	*N*,*N*-Dicyclohexylcarbodiimide
DCM	Dichloromethane
dCTP	2'-deoxycytidine-5'-triphosphate
ddCTP	2',3'-dideoxycytidine-5'-triphosphate
ddNTP	2',3'-dideoxynucleoside-5'-triphosphate
dG	2'-deoxyguanosine
DIPEA	*N*,*N*-diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNMP	2'-deoxynucleoside-5'-monophosphate
dNTP	2'-deoxynucleoside-5'-triphosphate
dT	2'-deoxythymidine
dU	2'-deoxyuridine
dUTP	2'-deoxyuridine-5'-triphosphate
EDC	*N*-(3-dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
Eq.	Equivalents
ESI	Electron spray ionisation
Et	Ethyl
EtOAc	Ethyl acetate

FAM	5-carboxyfluorescein
FMOC	Fluorenylmethyloxycarbonyl
G	Guanine
h	Hours
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ISFET	Ion sensitive field effect transistor
Me	Methyl
MeCN	Acetonitrile
MeOH	Methanol
Mins	Minutes
NHS	*N*-hydroxysuccinimide
NMR	Nuclear magnetic resonance spectroscopy
NW	Nanowire
ODN	Oligodeoxynucleotides
PAGE	Polyacrylamide gel electrophoresis
PAMAM	Polyamido amine
PEG	Polyethylene glycol
PNA	Peptide nucleic acid
PPi	Pyrophosphate
Prep-RP-HPLC	Preparative reversed phase high performance liquid chromatography
RNA	Ribonucleic acid
Sat.	Saturated
SBS	Sequencing by synthesis
SiNW	Silicon nanowire
SMRT	Single molecule real time sequencing
SNP	Single nucleotide polymorphisms
ssDNA	Single stranded DNA
T	Thymine
TBAF	Tetra-*n*-butylammonium fluoride
TBDMS	*Tert*-butyldimethylsilyl
TBDMSCl	*Tert*-butyldimethylsilyl chloride
^{t}Bu	*Tert*-butyl
TCEP	(tris(2-carboxyethyl)phosphine)

TEAB	Triethylammonium bicarbonate buffer
TEAF	Tetra-*n*-ethylammonium fluoride
TFA	Trifluoroacetyl
TFAA	Trifluoroacetic anhydride
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TSTU	*O*-(*N*-succinimidyl)-1,1,3,3-tetramethyluronium tetrafluoroborate

Chapter 1 –Introduction

1. Introduction

1.1 DNA Structure

The components of DNA

Deoxyribonucleic acid (DNA) is a polymeric molecule containing all the genetic information required for the function and development of living organisms. DNA is comprised of monomer units called nucleotides (figure 1.1). Each nucleotide monomer is composed of one of four heterocyclic bases (adenine, cytosine, guanine or thymine), a pentose sugar (2-deoxyribose) and a phosphate diester group which connects the nucleotide monomers to form a single strand of DNA. Ribonucleic acid (RNA) has an analogous primary structure to DNA but has a ribose sugar in place of 2-deoxyribose and the heterocyclic base uracil in place of thymine.

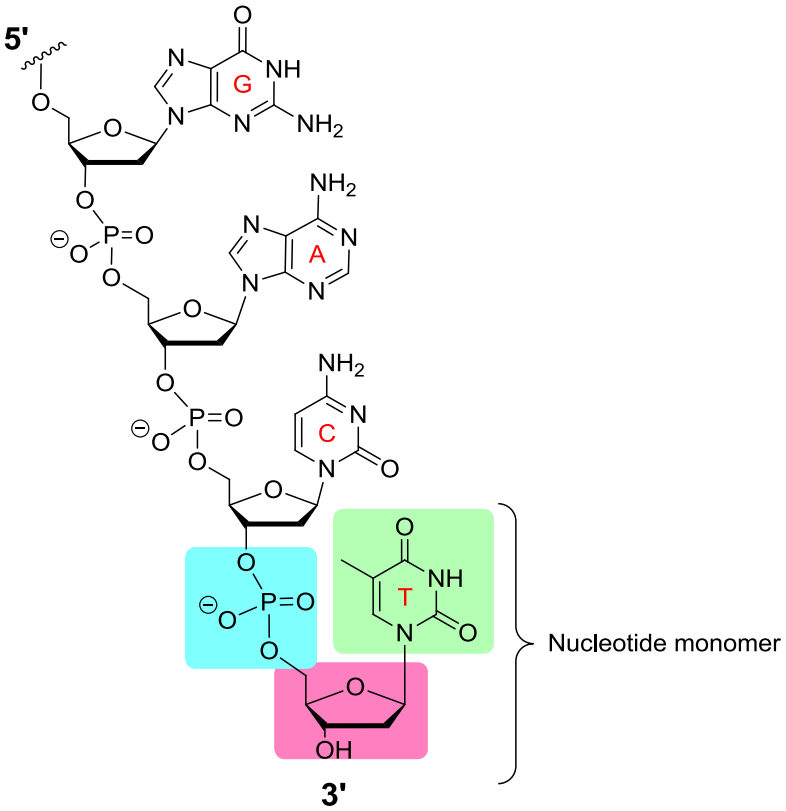


Figure 1.1: Primary structure of DNA. Phosphate diester (blue), 2-deoxyribose sugar (pink) and heterocyclic base (green) highlighted for one nucleotide monomer unit.

The heterocyclic bases of DNA are either purines (adenine and guanine) or pyrimidines (cytosine and thymine) as shown in figure 1.2.[1].

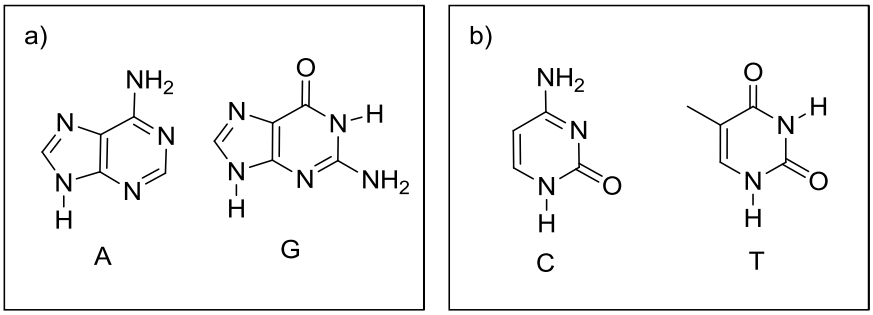


Figure 1.2: (a) The purine bases. (b) The pyrimidine bases.

The heterocyclic base is attached to the anomeric carbon of the sugar 2-deoxyribose via a glycosidic bond. This attachment is at the N9 position of the purines and the N1 position of the pyrimidines. The IUPAC numbering system for 2'-deoxyribonucleosides (unit comprising of base and 2-deoxyribose) is shown in figure 1.3. The numbering system of the sugar ring is distinguished from that of the base by the suffix prime (').

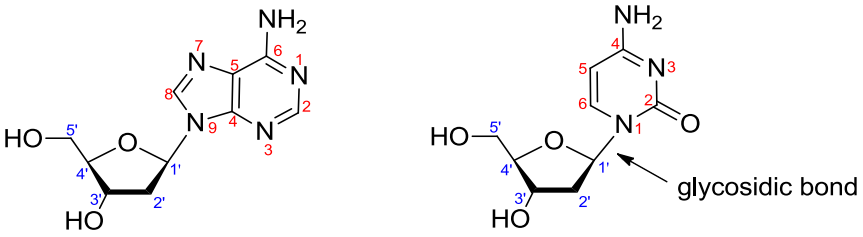


Figure 1.3: IUPAC numbering for the 2'-deoxyribose sugar and a purine (2'-deoxyadenosine) and pyrimidine (2'-deoxycytidine) base. Glycosidic linkage highlighted between 2-deoxyribose and the heterocyclic base.

A nucleotide is defined as a nucleoside which has at least one phosphate group (PO_4^{3-}) present. For example, nucleoside-5'-monophosphates have a single phosphate group attached to the 5'-hydroxyl of the 2-deoxyribose sugar. The most common nucleotides are monophosphates, diphosphates and triphosphates (figure 1.4). Some higher phosphate units are known but will not be discussed here. DNA has a phosphate diester backbone

holding the nucleotide monomers as a strand however each monomer must be introduced into the DNA strand via a 2'-deoxynucleoside 5'-triphosphate (dNTP).

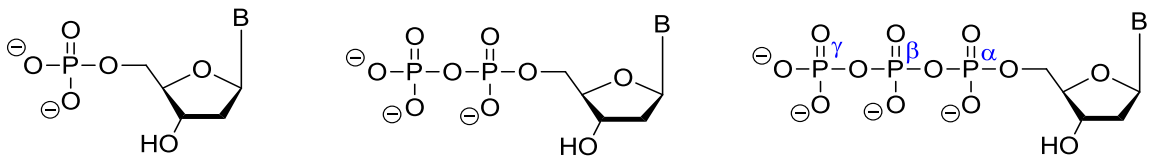


Figure 1.4: 2'-Deoxyribonucleoside 5′-mono-, di- and triphosphates. Where B represents one of the four heterocyclic bases; A, G, C or T.

The triphosphate moiety can exist in multiple ionisation states. For example the first three ionisations at the α, β and γ phosphates of adenosine-5'-triphosphate occur with a pK_a of 4.5 or less. The final ionisation on the γ-phosphate has a pK_a of 7 giving 2'-deoxyadenosine-5'-triphosphate an overall negative charge of 3.5 at pH 7 (figure 1.5).[2, 3]

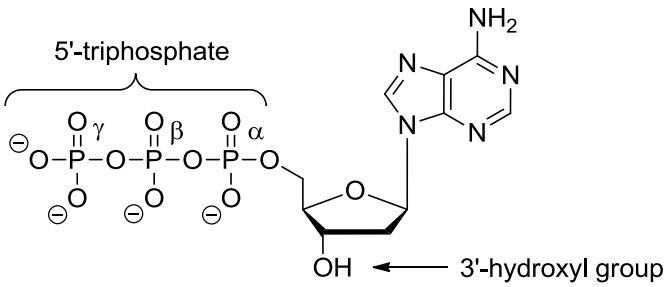


Figure 1.5: 2'-Deoxyadenosine-5'-triphosphate.

The secondary structure of DNA

Complementary base pairing was discovered by James Watson and Francis Crick in the early 1950s and as such, is commonly referred to as Watson and Crick base pairing.[4] Complementary base pairing occurs by selective hydrogen bonding between adenine to thymidine and guanine to cytosine respectively (figure 1.6).

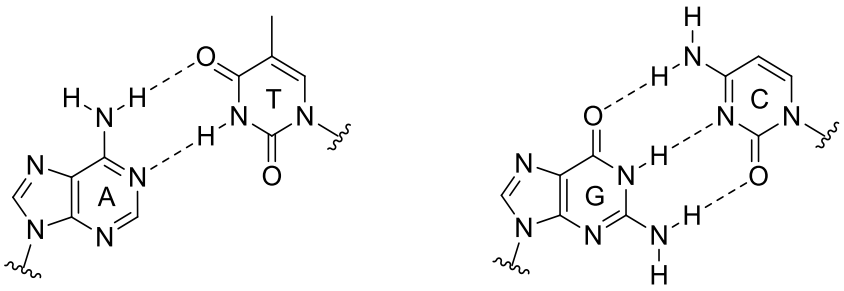


Figure 1.6: Watson and Crick base pairing shown between A – T and G – C.

The complementary base pairing holds two DNA strands together; every adenine (A) on one strand is bonded to a thymine (T) on the other DNA strand and similarly every guanine (G) is bonded to a cytosine (C). This forms the secondary structure of DNA. For this complementary base pairing to occur the two DNA strands run anti-parallel to one another, with one strand running in a 5'→3' direction and its complementary strand in a 3'→5' direction (figure 1.7). The two strands then wrap around each other to form a right-handed double helix.[4, 5]

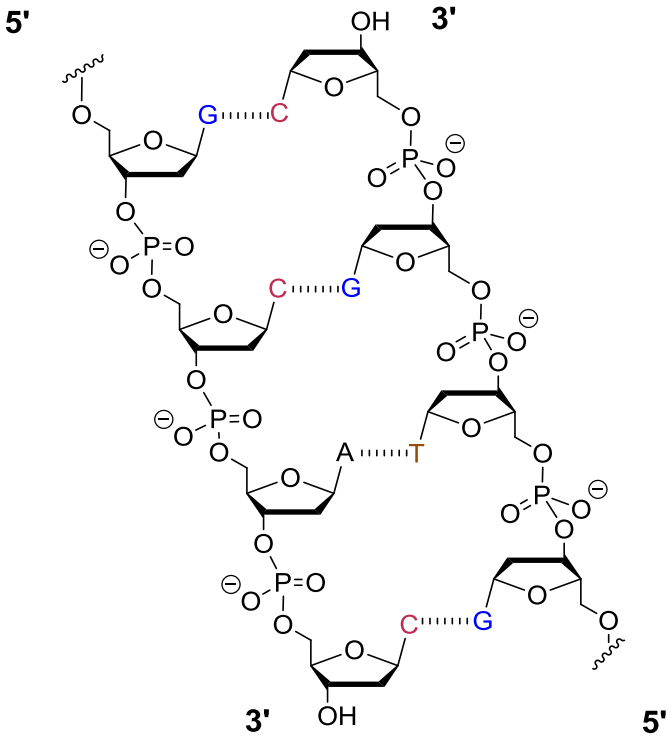


Figure 1.7: Secondary structure of DNA. Complementary strands of DNA run anti-parallel to one another.

The tertiary structure of DNA

The double helical DNA structure modelled by Watson and Crick, known as B-DNA is the most abundant DNA conformation in living organisms. B-DNA is a right handed helical structure with a major and minor groove, stemming from the twisting of the phosphate backbones around each other (figure 1.8). In addition to the hydrogen bonding of the base pairs, π-stacking between bases parallel to the DNA structure further stabilises this structure (figure 1.9).

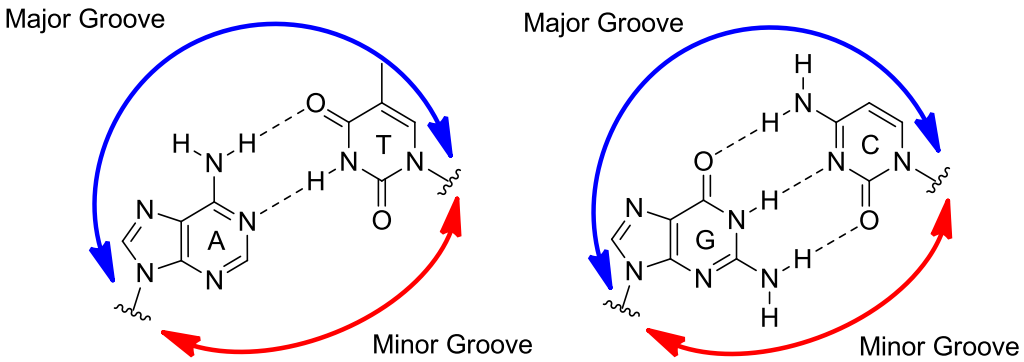


Figure 1.8: Origin of major and minor groove from complementary base pairing.

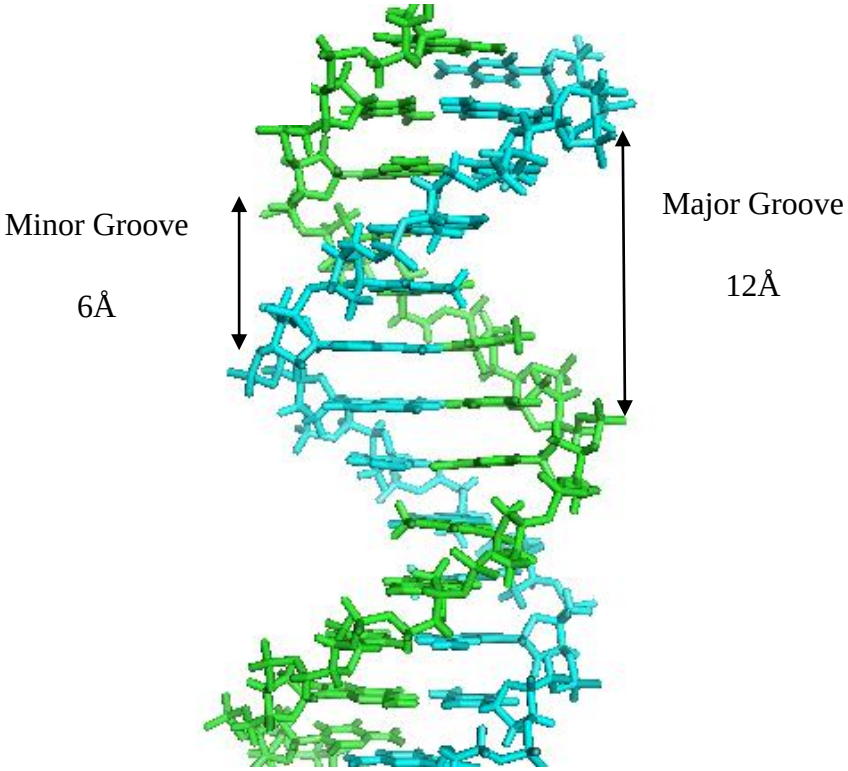


Figure 1.9: The tertiary structure of DNA. (PDB ref: 2M2C).

Alternative tertiary structures such as A-DNA and Z-DNA have been found. A-DNA has a higher number of base pairs (bp) per turn of the helix, with a repeat of 11 bp per turn compared to the 10 bp of B-DNA, leading to a shorter and broader helix. Z-DNA has a left-handed conformation. Alternating G and C rich sequences are known to induce the Z-DNA conformation, whereas non alternating G and C rich sequences favour A-DNA.[6,7]

1.2 DNA Replication and Amplification

DNA replication

DNA replication occurs in all living organisms resulting in the production of new, replica DNA material from the original DNA; this is required for cell reproduction and the passing on of the genetic code to new cells. The genetic code consists of triplets of nucleotides known as codons, each codon translates to one of the 20 amino acids. Therefore the sequence of nucleotides in DNA translates into the amino acid sequence in proteins.

DNA replication starts with the separation of double stranded DNA into two single strands. A nick is created in the double helix using a DNA gyrase enzyme, DNA helicase then unwinds the double stranded DNA to leave two pieces of single stranded DNA which act as templates for replication (figure 1.10). Replication is known as a semi-conservative process due to the new DNA helix formed consisting of one strand of original or parent DNA and a new daughter strand.[8] A primer (a short piece of DNA which binds to a complementary region on the DNA) or a primase (an enzyme which creates a primer) must be present for the DNA polymerase enzyme to be able to initiate replication. The DNA polymerase then ‘walks’ down the DNA strands, in the 5' to 3' direction, inserting the complementary nucleotides into the new DNA strand, replicating the original DNA (figure 1.11).

DNA replication is more complex in the case of the lagging strand. Unlike the leading strand, synthesis is discontinuous, leading to fragments known as Okazaki fragments (figure 1.10). These are later joined together using a DNA ligase enzyme. The specificity of base pairing ensures faithful replication of the original DNA. Additionally many DNA

polymerases have proof reading activity in the form of a 3'-5'-exonuclease to reduce the number of errors made through replication by removing incorrectly inserted nucleotides.

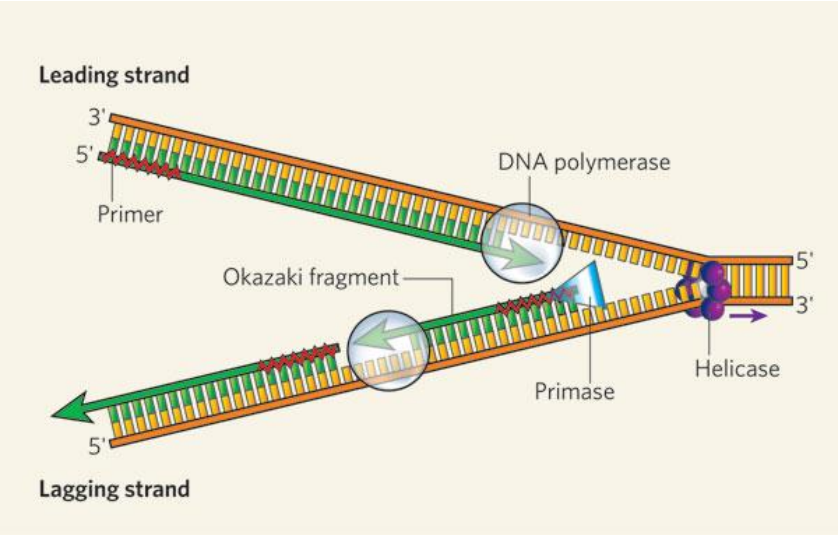


Figure 1.10: A figure showing the replication fork formed during DNA replication, with leading and lagging strands highlighted. (http://www.nature.com/scitable/content/dna-replication-of-the-leading-and-lagging-14668888 accessed 15/07/2015)

The fundamental reaction of DNA replication is the enzymatically catalysed formation of phosphodiester bonds to synthesise new DNA. A phosphate diester linkage is formed between the hydroxyl group at the 3'-terminus of the growing DNA strand and the α-phosphate of the complementary dNTP being incorporated (figure 1.11). The phosphate diester linkage between the α and β-phosphate of the dNTP is broken by nucleophilic attack of the 3' hydroxyl and pyrophosphate (from the dNTP) and a proton (from the 3'-hydroxyl) are released as the phosphodiester bond is formed.[9] DNA chain growth only occurs in the 5′→3′ direction. The reaction is catalysed by a DNA polymerase enzyme in the presence of magnesium ions which ensures the incorporation of the associated complementary dNTP.

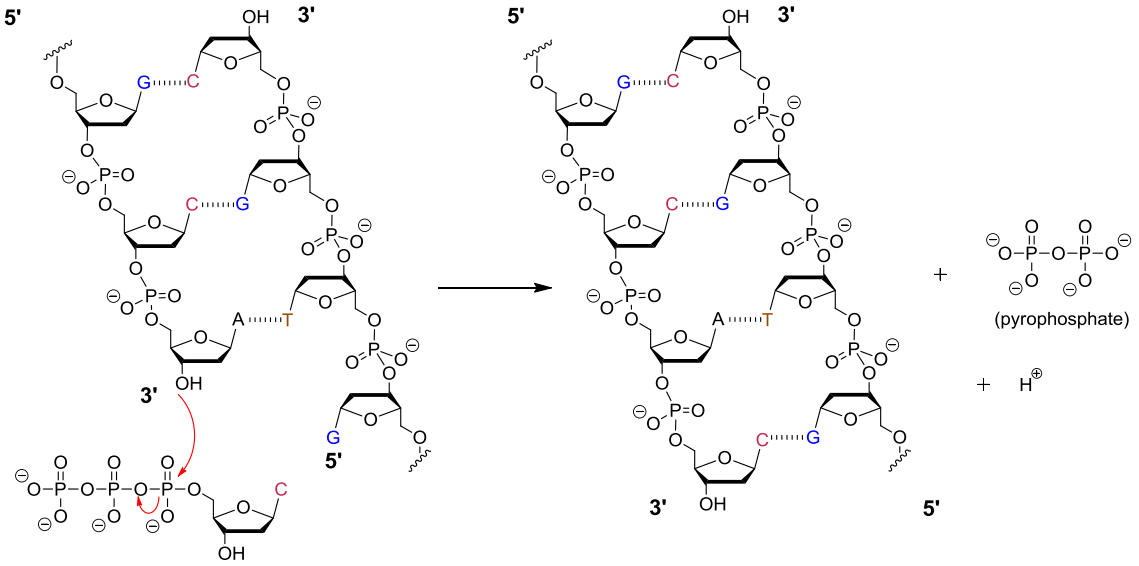


Figure 1.11: Formation of a phosphodiester bond during DNA synthesis.

DNA amplification

Amplification of DNA is of specific importance to many biotechnology applications such as DNA fingerprinting and DNA sequencing. To study an individual's genes or investigate a DNA region of interest it is necessary to obtain sufficient copies of the DNA often from very small amounts of the original DNA. DNA amplification allows the generation of multiple copies of the DNA from one original molecule of DNA.

The polymerase chain reaction (PCR) allows the amplification of a DNA sequence and multiplies it by several orders of magnitude. PCR begins by denaturing double stranded DNA by heating to form single stranded DNA. Cooling in the presence of a complementary DNA primer allows annealing of the primer to a specific complementary region on the template. The attached primer then acts as an initiation point for DNA synthesis resulting in the reproduction of the sequence of the original DNA (figure 1.12). The use of a complementary reverse primer on the opposite DNA strand allows DNA replication to take place on both strands.[10, 11]

The process can be repeated or cycled to generate double-stranded copies of a section of the original DNA until the required amount of DNA is present. At this point a final step

of elongation ensures there is no single stranded DNA remaining. The process is very accurate; the reported error rates for a DNA polymerase incorporating an incorrect nucleotide for the main families of polymerases (A, B and C) are between 1 in 10^6 and 1 in 10^8 nucleotides during DNA synthesis.[12, 13]

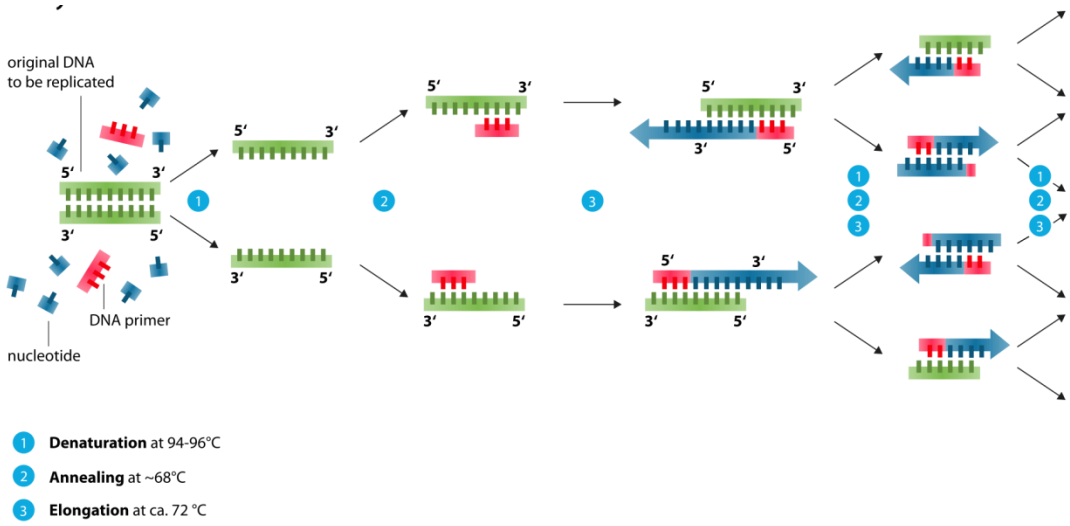


Figure 1.12: PCR schematic showing the PCR cycle with both forward and reverse primers. (http://www.socmucimm.org/category/publications/protocols/laboratory-techniques/page/2/ accessed 05/08/2015)

1.3 Introduction to DNA sequencing

The ability to sequence and analyse DNA has become a key tool in many fields from medical diagnostics to forensic biology and genetic research. The synthesis of modified dNTPs has been pivotal to first and second generation sequencing techniques and remains the key synthetic strategy of many sequencing devices.

The key to sequencing is the ability to extract the genetic code of the DNA and analyse it accurately. All genetic information is held by the order of the four nucleotides in a strand of DNA and sequencing allows access to this genetic information. One of the most important fields for DNA sequencing is disease diagnostics. Often the identity of a disease, or the risk of developing a disease, can be determined by sequencing the DNA of the patient and identifying variations (usually mutations) which are known markers for a certain disease or condition. Usually for disease to be present there will be multiple

mutations within the DNA sequence but single nucleotide polymorphisms (SNPs) where the DNA sequence varies by one nucleotide can also be the cause. An example of an adverse SNP is deep vein thrombosis (DVT) where a G to A substitution at nucleotide position 1,691 in the factor V gene increases the individual's risk of developing DVT sevenfold.[14]

It is the ability to use DNA sequencing to gain an insight into genetics and its potential for use in disease diagnostics that inspired the initial development of sequencing techniques in the late 1970s. The Human Genome Project drove further development and interest in DNA sequencing; officially founded in 1990 the project aimed to sequence the full human genome within 15 years. Due to technological advances, importantly including those in the field of computer technology, the human genome was officially completed in April 2003.[15] This was a worldwide effort spanning many countries and disciplines highlighting the importance perceived in sequencing the genome. It was hoped that by sequencing the full genome scientists would be able to identify the genes responsible for many of the diseases that affect humankind and in turn find ways to treat, cure, or prevent them in the future.

Recent global efforts have focussed on the '$1000 genome'. It is believed that by lowering the cost of sequencing a human genome to $1000, DNA sequencing will be able to be used on a more frequent basis and in more applications; arguably the most important possibility is the use of DNA sequencing to develop personalised medicine. Personalised medicine is the concept that a patient could enter a medical facility, have their DNA sequenced and the genetic information from this used to diagnose any genetic conditions or diseases present. Crucially their genetic code could then be referenced against that of others to find the most effective treatment plan or dosage of medication based on their genetic information.[16]

1.4 First generation DNA sequencing techniques

The first DNA sequencing technologies were developed in the 1970s (Sanger[17], Maxam and Gilbert[18]). Maxam and Gilbert proposed sequencing by chemical cleavage[19] whereas Sanger sequencing is based on chain termination during DNA replication.

Early sequencing methods depended on two methods of radioactively labelling DNA in order to visualise the DNA fragments, formed through the sequencing methods described, by polyacrylamide gel electrophoresis (PAGE). A radioactive label can be introduced onto the 5'-hydroxyl of a 5'-dephospho-DNA primer using a ^{32}P radioactively labelled adenosine-5'-triphosphate (ATP). The reaction is catalysed by a polynucleotide kinase and results in a 5'-radioactively labelled primer (figure 1.13). Alternatively a radioactive label can be added to DNA by using a ^{32}P labelled phosphorus at the α-position of the incoming triphosphate moiety (figure 1.14).

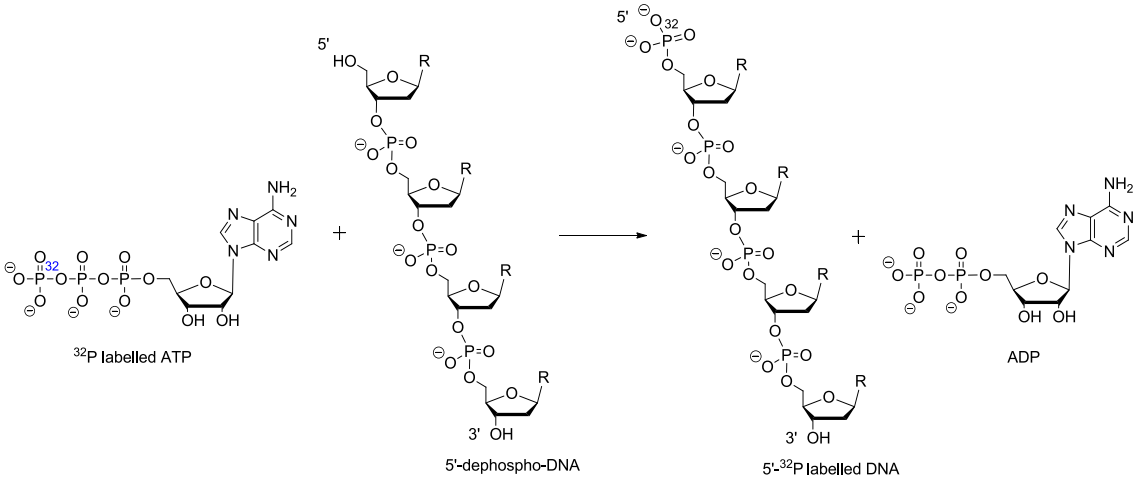


Figure 1.13: Synthesis of a ^{32}P labelled primer.

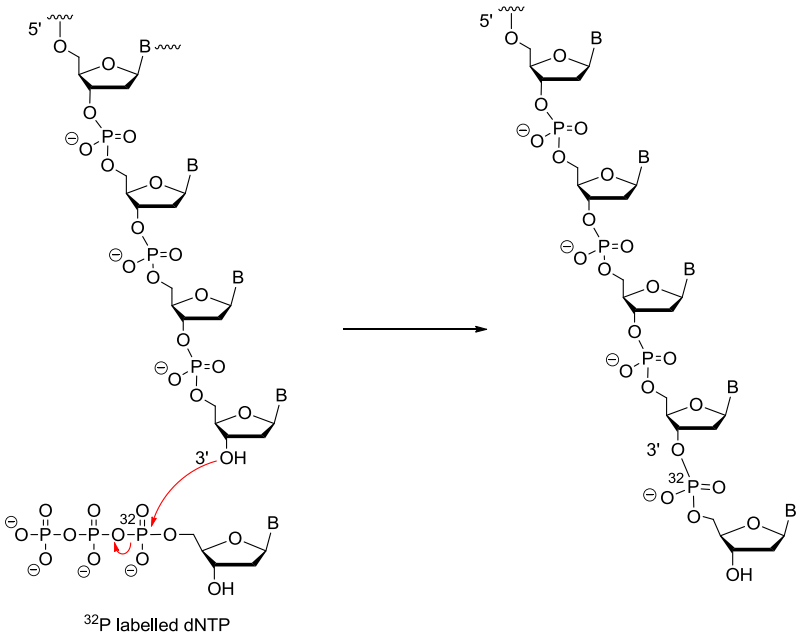


Figure 1.14: ^{32}P radioactively labelled dNTP incorporation into DNA.

Maxam and Gilbert DNA sequencing

Maxam-Gilbert sequencing was first reported in 1977.[18] It takes ^{32}P radioactively-labelled DNA and chemically modifies the purines or pyrimidines (A, G or C, T) to create damage within the labelled DNA strand. This leads to loss of the base from the 2-deoxyribose sugar creating an abasic site, a weak point in the backbone, which can be broken via a β-elimination reaction in the presence of piperidine. The four bases can be identified by using four different cleavage conditions which preferentially cleave at a known base or bases. This results in a series of DNA fragments of varying lengths which can be analysed by PAGE.

The purine specific reagent for cleavage of the base from the sugar phosphate backbone is dimethyl sulphate (DMS). DMS methylates guanine at the N7 position and adenine at the N3 position (figure 1.15).

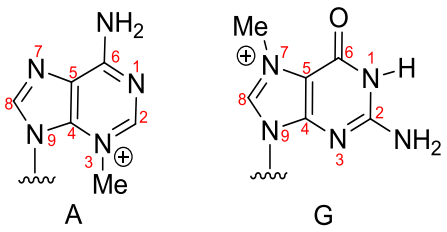


Figure 1.15: Methylation of N3 adenine and N7 guanine.

Methylation produces a positive charge on the purine ring causing an unstable glycosidic bond which can be broken, leaving the sugar free to be cleaved from the DNA strand. The glycosidic bond of methylated adenine is less stable than that of methylated guanine, therefore by treating methylated adenine with dilute acid it is possible to induce preferential adenine cleavage and to distinguish between A and G.

Cytosine and thymine are cleaved using hydrazine (figure 1.16). Cytosine can be distinguished from thymine by the addition of 2M NaCl which suppresses the initial reaction between thymine and hydrazine and allows cytosine to be cleaved preferentially.[18]

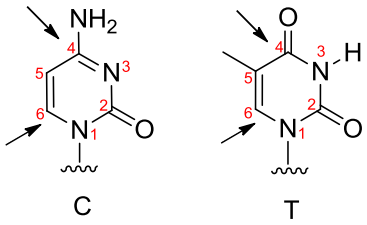


Figure 1.16: Positions of attack by hydrazine during the T and C specific chemical cleavage reaction.

The reactions involved in the Maxam-Gilbert chemical cleavage method only cause damage to 1 in 50 to 100 base pairs. This allows different length fragments to be made and the sequence of the target DNA to be identified by combining results of all the fragments to create the sequence of the whole strand.

The DNA fragments produced are analysed by PAGE which separates the fragments dependent on their charge and size, effectively acting like a sieve allowing single nucleotide resolution between fragment lengths. The smallest fragments, having the least

resistance, move quickest from the top of the gel towards the anode at the bottom of the gel. The larger DNA fragments move slower and reside nearer the top. The gel is imaged by autoradiography through the inclusion of a radioactive label as previously described. As a radioactive label is used in all experiments each reaction cannot be distinguished from another. Therefore the four experiments described must be analysed in separate lanes on the PAGE gel to determine between the four nucleotides. The sequence of the target complementary DNA can be identified by interpreting the PAGE gel, relying on the knowledge of complementary base pairing to determine the sequence of the target DNA strand (figure 1.17).

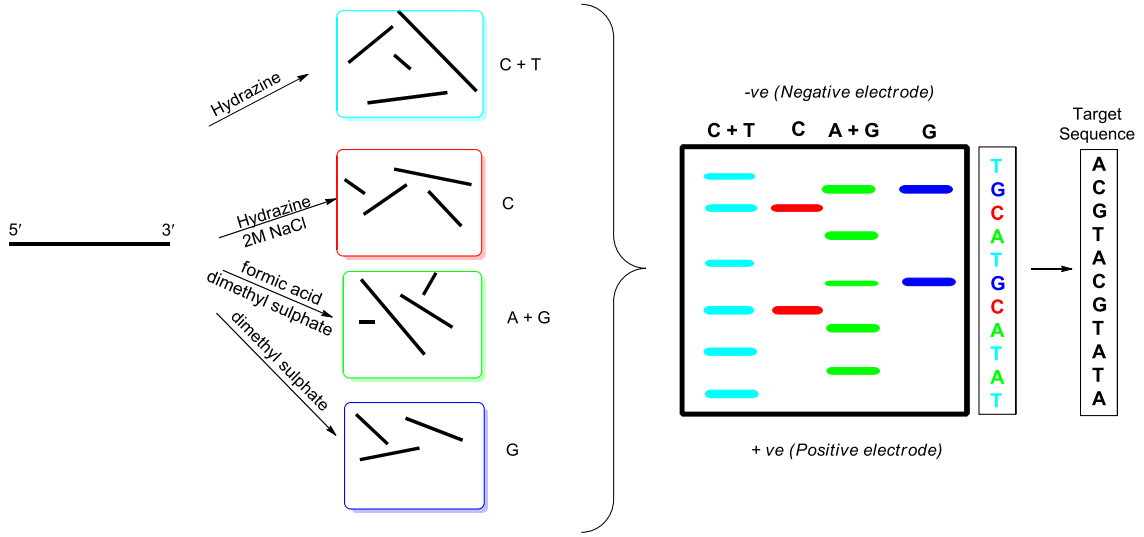


Figure 1.17: Polyacrylamide gel electrophoresis (PAGE) diagram for the separation of fragments created through Maxam and Gilbert sequencing.

Although PAGE is very effective at analysing DNA fragments it is limited in the length of base pairs that it can successfully analyse with sufficient resolution to achieve the accuracy needed in DNA sequencing.[18, 20]

Sanger DNA sequencing using ^{32}P labelled DNA

In 1977 Sanger[21] published his dideoxynucleotide chain-terminating DNA sequencing method which was to become the most commonly used sequencing method of the 20th century. The principle is that a small amount of 2',3'-dideoxyribonucleoside triphosphate (ddNTP, figure 1.18) is added to a sequencing reaction alongside the natural dNTPs.

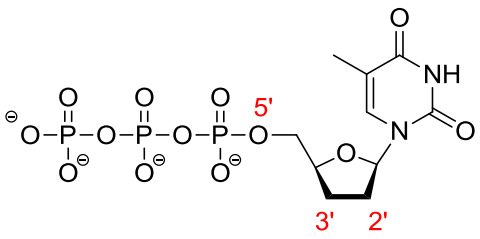


Figure 1.18: Example of a dideoxynucleotide: 2', 3'-dideoxythymidine-5'-triphosphate.

When the ddNTP is incorporated into the DNA strand it terminates the sequencing reaction due to the lack of a 3'-hydroxyl for formation of the next phosphate diester linkage (figure 1.19b). As the amount of ddNTP added is considerably lower (~1% typically) than the amount of natural dNTP present the ddNTP will not be incorporated every time the complementary base arises. This results in a distribution of DNA fragments of varying lengths.

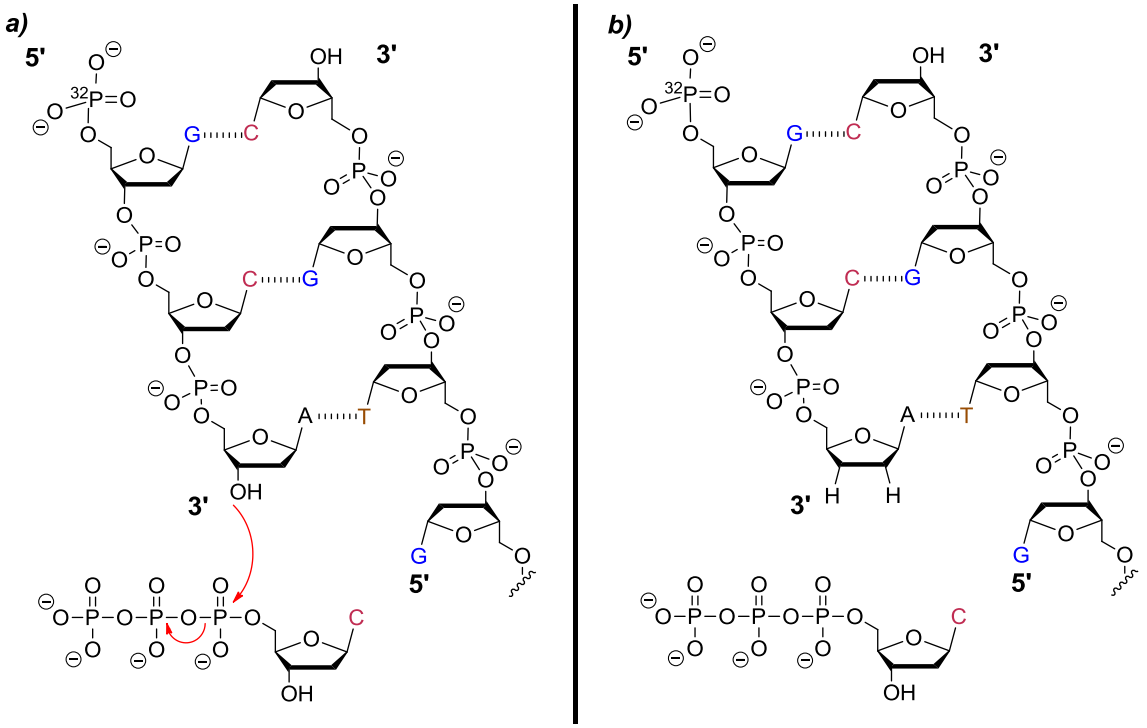


Figure 1.19: (a) Formation of a phosphate diester linkage from the nucleophillic attack of 3'-hydroxyl on the DNA strand onto the 5'-triphosphate of the next complementary nucleotide (b) The lack of a 3'-hydroxyl prevents formation of a phosphate diester linkage, terminating DNA chain extension. ^{32}P label shown on the primer.

As the detection of the DNA fragments formed during sequencing is through the inclusion of a common ^{32}P label, each sequencing reaction must be performed in a separate vessel, one for each ddNTP. The reactions are analysed by PAGE as described for Maxam and Gilbert sequencing, with a different lane being used for each ddNTP reaction. After imaging through autoradiography the sequence of the complementary DNA can be determined by analysis of the gel.

Although it remains one of the most commonly used sequencing methods the original Sanger sequencing method is not without disadvantages: it is a time consuming process that cannot be performed as a 'one pot reaction' and requires PAGE followed by the manual interpretation of the resulting autoradiogram.

Sanger DNA sequencing using fluorescently labelled primers

As Sanger sequencing became more widely used variations to the method, such as introducing fluorescent reporter groups in order to automate sequencing, were explored. This significantly decreased the time it took to analyse results from sequencing reactions.[22–24]

Connell *et al.*[23] proposed automation by the use of a fluorophore covalently attached to the primer. They described four primers each labelled with a distinct fluorophore, one for each base-specific sequencing reaction. The introduction of the dideoxyribonucleotides must be performed as a four reaction pot process but the resulting DNA fragments can be combined and analysed in a single lane on a polyacrylamide gel. However variations in the electrophoretic mobility and cross-over of emission signals from the four dyes were often found to affect the accuracy of interpreting the sequence. Ansorge *et al.*[25] suggested that to overcome this problem a fluorescent dye could be introduced onto a primer as shown in figure 1.20 and the same primer used for all DNA sequencing reactions. This meant a return to a four lane analysis system, but eliminated any concerns over the varying mobility of different dyes during electrophoresis. It also allowed for automation of results to be continued.

A further advantage of labelling the primer as described is the convenience of using unmodified dideoxynucleotides, as the fluorescent modification remains on the primer. This avoids the requirement for synthesising fluorescently modified dideoxynucleotides.

It can also be assumed that this process will be suitable for use with many more DNA polymerases, as the process no longer relies on modified dideoxynucleotides being substrates for a DNA polymerase.

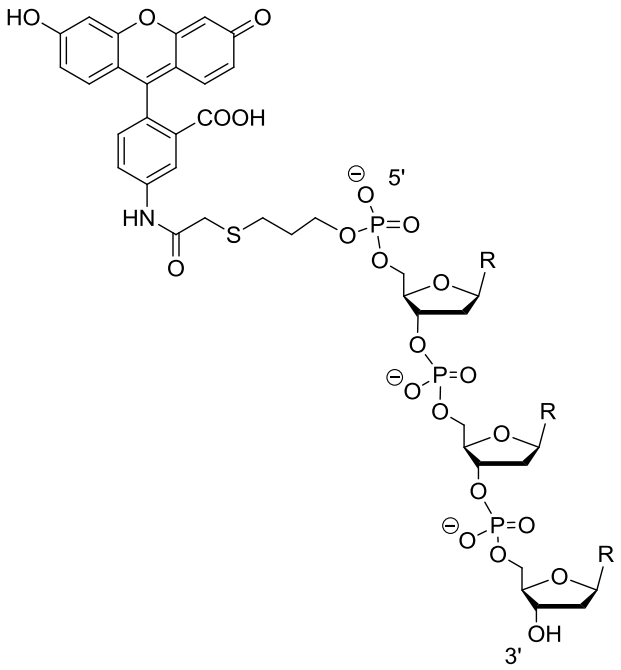


Figure 1.20: 5'-Fluorescently labelled primer.

Sanger DNA sequencing using fluorescently labelled dideoxyribonucleotides

The aim of 'one pot' Sanger sequencing was re-addressed by introducing succinylfluorescein fluorescent dyes to the dideoxynucleotides.[22] The fluorescent dye is attached to the 2',3'-dideoxyribonucleoside by a linker at the C5-position of the pyrimidines and the C7-position of 7-deazapurines (figure 1.21). The purines bases cannot be modified on their natural heterocyclic base as this would create a positive charge at the N7 position (further discussed in figure 1.34), therefore the heteroatom is replaced with a carbon.

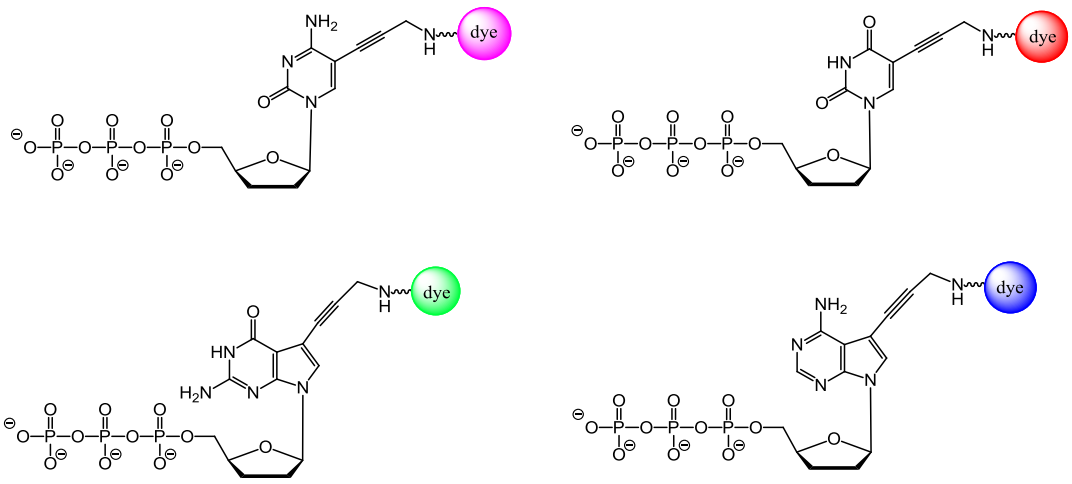


Figure 1.21: Example structures of ddNTPs labelled with four fluorescent dyes with unique emission profiles.

Each dideoxynucleotides has a unique fluorescent dye with its own emission characteristics; it is therefore possible to determine which dideoxyribonucleotide terminated the DNA synthesis by its discrete emission profile. The method of sequencing by using four fluorescently labelled dideoxynucleotides allows for one pot sequencing and simplified automated analysis. Sanger sequencing using fluorescently labelled nucleotides was the main method of sequencing used to complete the Human Genome Project.[26]

1.5 Second (Next) generation sequencing

Next generation sequencing (NGS) was developed to meet the demand for faster and more affordable DNA sequencing methods.[26] Second generation techniques can be broadly classified as sequencing methods which can be performed in a massively parallel fashion or that require amplification of DNA prior to sequencing. Numerous techniques are now available which can be classed as NGS technologies.

One significant way to increase the speed of sequencing is to use a "sequencing by synthesis" (SBS) method. SBS is a continuous process where DNA synthesis does not terminate after a modified nucleotide has been added to the DNA chain. Instead a

reversible 3'-hydroxyl blocking group briefly pauses elongation while detection of the reporter group determines the nucleotides identity. The blocking group and reporter group are then cleaved and sequencing continues (figure 1.22).

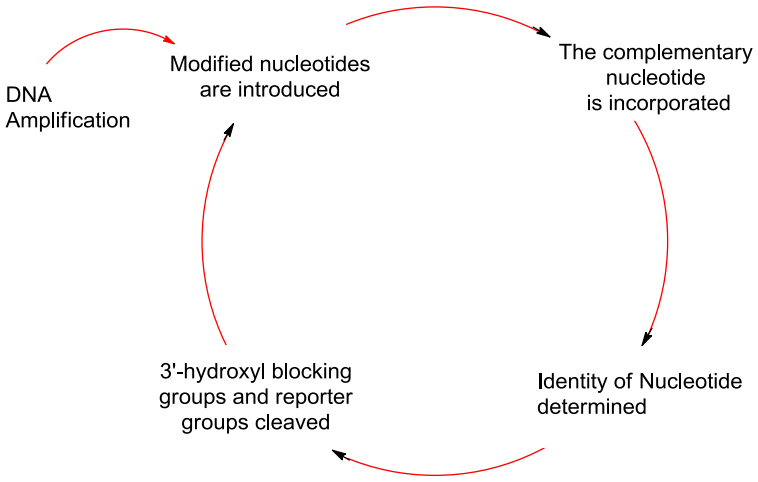


Figure 1.22: Sequencing by synthesis cycle.

Alternatively the need for a 3'-blocking group can be avoided by a cycle of introducing one nucleotide at a time with washing between introductions. When a signal is detected, indicating the complementary nucleotide has been incorporated, the reporter group can then be cleaved and washed away allowing for the next nucleotide to be incorporated. This process continues until the sequence of the target DNA is determined.

Pyrosequencing

Roche's 454 sequencer was the first next generation sequencing technique to achieve commercialisation, becoming available in 2004.[27] Roche confirmed the sequencing of the full genome of James Watson using pyrosequencing in a letter to the journal Nature in 2008.[28] Although not the first full genome to be published (that accolade belongs to J. Craig Venter whose genome was sequenced using Sanger sequencing technology and published in 2007[29]) it was the first genome to be sequenced using massively parallel next generation sequencing.

Pyrosequencing technology detects the release of pyrophosphate ($P_2O_7^{4-}$ or 'PPi') during DNA synthesis (see figure 1.23). Following a cascade of enzymatic reactions the release of pyrophosphate is detected as visible light. Oxyluciferin generates light in proportion to the amount of ATP present which is directly related to the production of pyrophosphate. Therefore, if three nucleotides are incorporated into the template strand then three times the amount of light will be generated and measured.

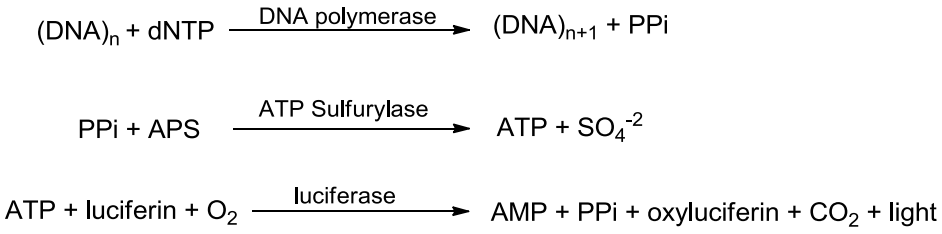


Figure 1.23: The enzymatic reaction scheme of Pyrosequencing (PPI = pyrophosphate, APS = adenosine-5′-phosphosulfate, ATP = adenosine-5′-triphosphate, AMP = adenosine-5′-monophosphate).

Although this phenomenon was first reported in 1985 it was significantly later that that the concept as a sequencing method became a reality.[30,31] Initial work using pyrosequencing for DNA sequencing was not without problems; it required each dNTP to be introduced separately, extensive washing to remove unincorporated nucleotides and was unable to correctly analyse homopolymeric regions (repeating sections of the same nucleotide) in the template. To address the problem of resolution of homopolymeric regions a cleavable 3'-hydroxyl blocking group was investigated by Turro *et al.*[32] who reported the use of a 3'-*O*-allyl group as a chemically cleavable 3'-hydroxyl blocking group. The use of a 3'-blocking group temporarily halts the sequencing reaction after each dNTP incorporation so that only one dNTP can be incorporated at a time. Problems with signal noise have been partially corrected by the addition of a further enzyme (apyrase) that breaks down excess nucleotide material at the end of each extension step. This ensures that excessive nucleotide material can be washed away. This combined with a combination of software algorithms and increased sensitivity in light detection allows for sections of up to 6 repeating nucleotides to be accurately read.[27, 33]

Results from pyrosequencing are analysed using a pyrogram™ (figure 1.24). The pyrogram shown does not use a 3'-blocking group therefore the height of the signal indicates how many nucleotides have been incorporated. Where no signal is seen, no incorporation has occurred.[34]

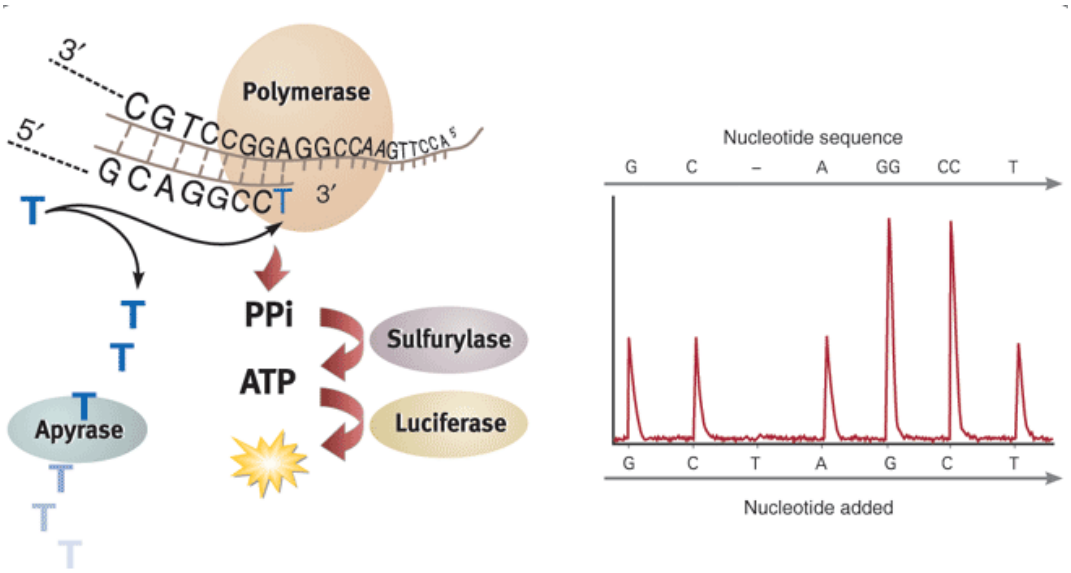


Figure 1.24: Pyrogram™ showing results of pyrosequencing.[34]

Reversible terminator sequencing – Illumina sequencing

One of the leading companies in next generation sequencing is Illumina (previously Solexa), whose technology is based on SBS and the use of fluorescent reporter groups.

As in Sanger sequencing using fluorescently labelled 2',3'-dideoxyribonucleotides (section 1.4), Illumina technology modifies each dNTP with a discrete fluorescent dye as a reporter group (figure 1.25). These fluorescent dyes allow for each nucleotide to be individually identified by the different colour they emit when excited by a laser. The Illumina modified nucleotides used for sequencing have been carefully designed to overcome the issues seen with electrophoretic mobility and to allow for 'one pot' sequencing.

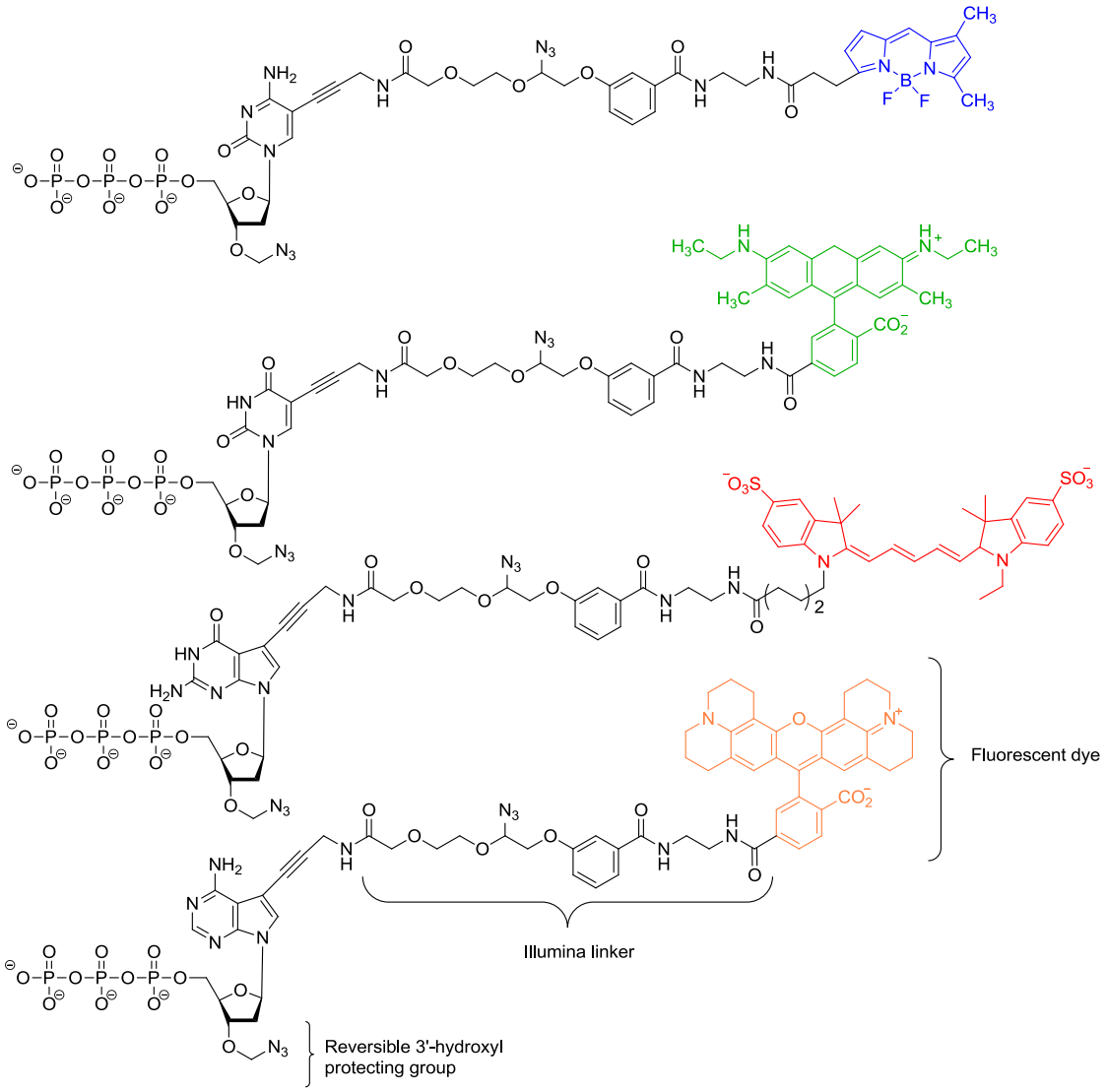


Figure 1.25: The Illumina modified dNTPs, dC, dU, dG and dA (top to bottom).

During Illumina sequencing all four modified dNTPs are simultaneously present in the sequencing reaction. When the complementary modified dNTP is incorporated into the growing DNA strand the sequencing reaction is temporarily halted due to the presence of the 3'-blocking group. The dNTP is then identified through its unique fluorescence emission. Once the identity of the nucleotide has been determined, the 3'-hydroxyl blocking group and linker are cleaved, allowing the DNA sequencing reaction to continue. The use of a 3'-blocking group is of particular importance for resolving homopolymeric regions on the DNA template with high accuracy. The blocking group also needs to be cleaved under mild conditions with high efficiency to allow formation of the next phosphate diester linkage. The choice of the 3'-hydroxyl blocking group is

critical as during nucleotide incorporation DNA polymerase enzymes are particularly sensitive to changes at the 3'-hydroxyl position of dNTPs.[35] Illumina sequencing uses a small azidomethyl group as the 3'-blocking group.

An important feature of SBS is that cleavage of the linker/reporter group on the heterocyclic base returns the 3'-terminus of the growing DNA chain to a structure as close to its natural form as possible, this is key for continued fidelity of the polymerase. Illumina exploit an azidomethyl moiety to allow cleavage of both the linker and the 3'-blocking group in a single step. The azidomethyl group is reduced using a phosphine such as TCEP in the Staudinger reaction (figure 1.26) to form a labile hemiaminal ether that undergoes rapid hydrolysis and cleavage of the reporter and blocking groups.[36] This removes the bulk of the modification from the C5-position of the base and leaves a free 3'-hydroxyl for sequencing to continue.

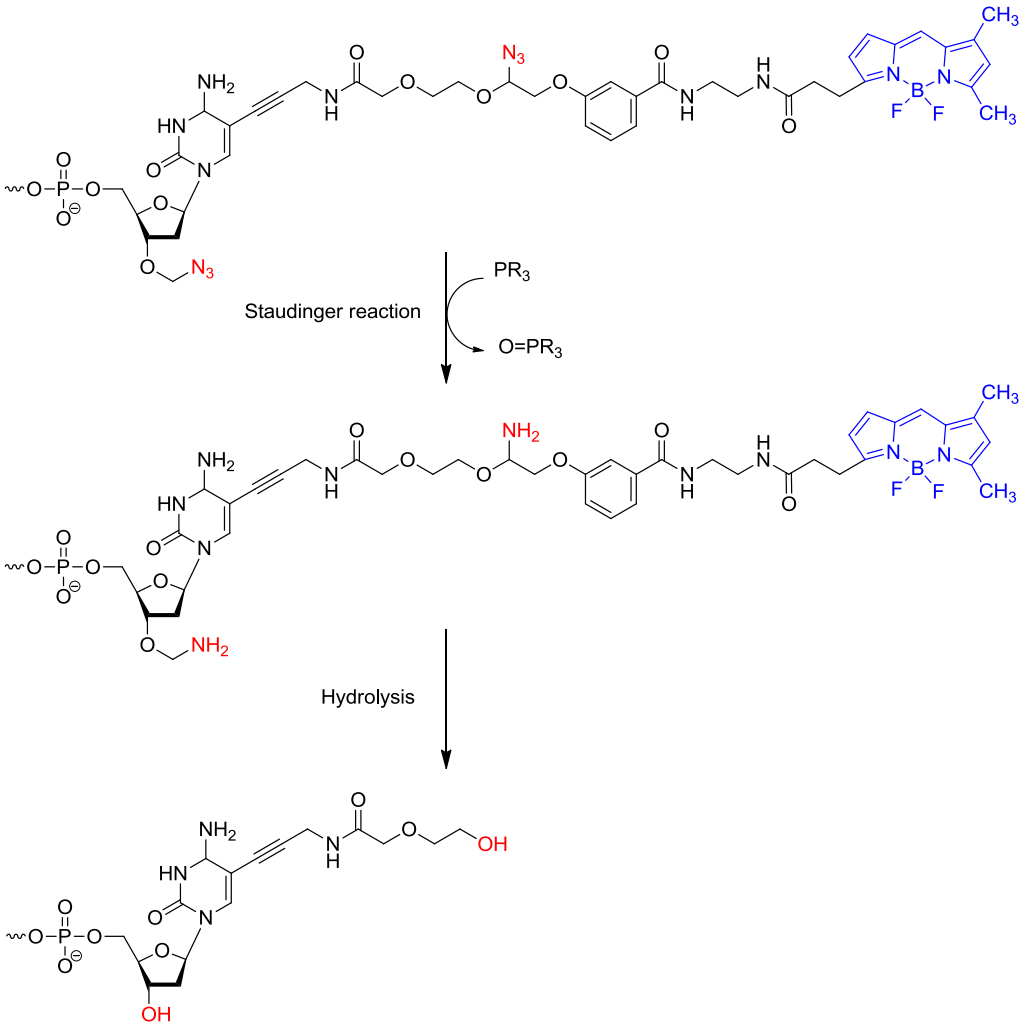


Figure 1.26: Staudinger reaction for the cleavage of the azidomethyl 3'-hydroxyl protecting group and reporter group.

1.6 Third generation sequencing

As the DNA sequencing market continues to expand a third generation of DNA sequencers have started to emerge. Third generation sequencing methods are predominately based on single molecule sequencing performed in real time (SMRT). Real time sequencing varies from other sequencing methods discussed because it does not halt the process of DNA synthesis to detect the incorporation of a nucleotide and does not require amplification of the DNA prior to sequencing. Instead the sequence of a single DNA molecule is detected in real time as each dNTP is incorporated.

BioPacific single molecule real time sequencing

BioPacific have developed a sequencer that detects fluorescently modified dNTPs during DNA synthesis in real time.[37,38] The modified dNTPs have the fluorophore on the terminal phosphate group of the 5'-polyphosphate, allowing for the fluorophore to be cleaved during the formation of the phosphodiester bond (figure 1.27). This is advantageous as once removed the nucleotide is returned to its natural form.

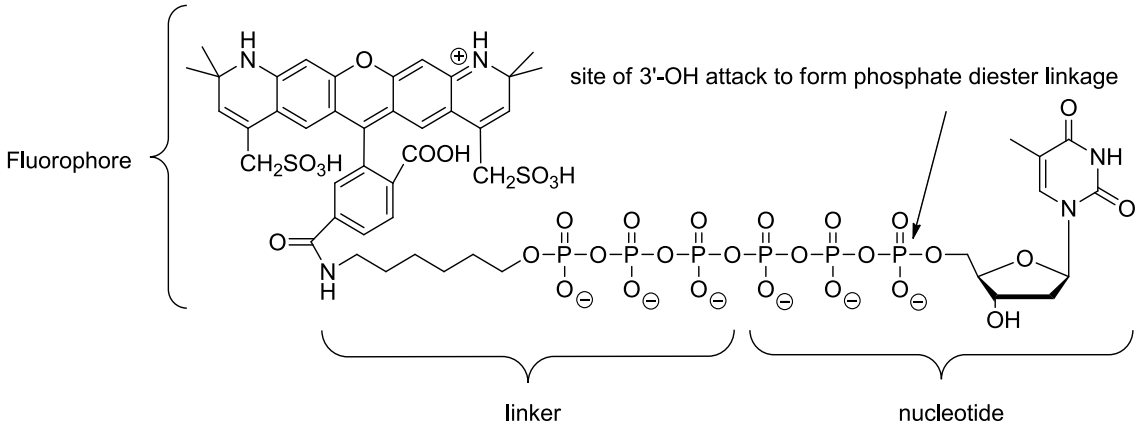


Figure 1.27: BioPacific 5'-polyphosphate fluorescently labelled modified 2'-deoxythymidine.[38]

The identity of the base is detected by the colour of light emitted when the cleaved fluorophore is excited by a laser (figure 1.28).

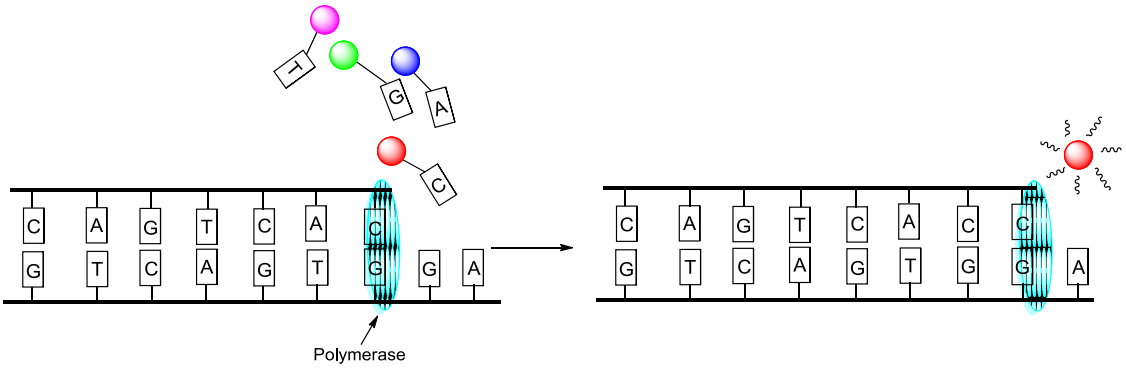


Figure 1.28: Incorporation and detection of a 5'-fluorescently labelled BioPacific nucleotide.

Initial problems were signal to noise ratios and the development of a machine capable of distinguishing between very quick pulses of light (~200 ms between pulses) consistently and accurately. Error rates reported for DNA sequencing results have also been generally higher than those of Illumina or Ion torrent sequencing.[39, 40]

Ion torrent sequencing

Ion torrent sequencing translates the incorporation of a nucleotide and the subsequent proton release into digital information. No modification is required to the dNTPs as the proton detected is a by-product of the formation of a phosphate diester linkage (figure 1.11).

The template strand to be sequenced is embedded on a bead within a well on a semi-conductive surface with a small amount of DNA polymerase present. Each dNTP is then passed over the well separately and when the complementary dNTP is present the polymerase adds the dNTP to the growing DNA chain. This releases pyrophosphate from the triphosphate group of the incorporated nucleotide and a proton from the 3'-hydroxyl of the growing DNA strand. The hydrogen ion is detected by an ion sensitive field effect transistor (ISFET). This is extremely sensitive to very small changes in pH with each base altering the pH of the surrounding solution by approximately 0.02 pH units.[41] The change in pH leads to a change in the surface potential of the metal oxide sensing layer and an overall change in the voltage of the field effect transistor (figure 1.29).

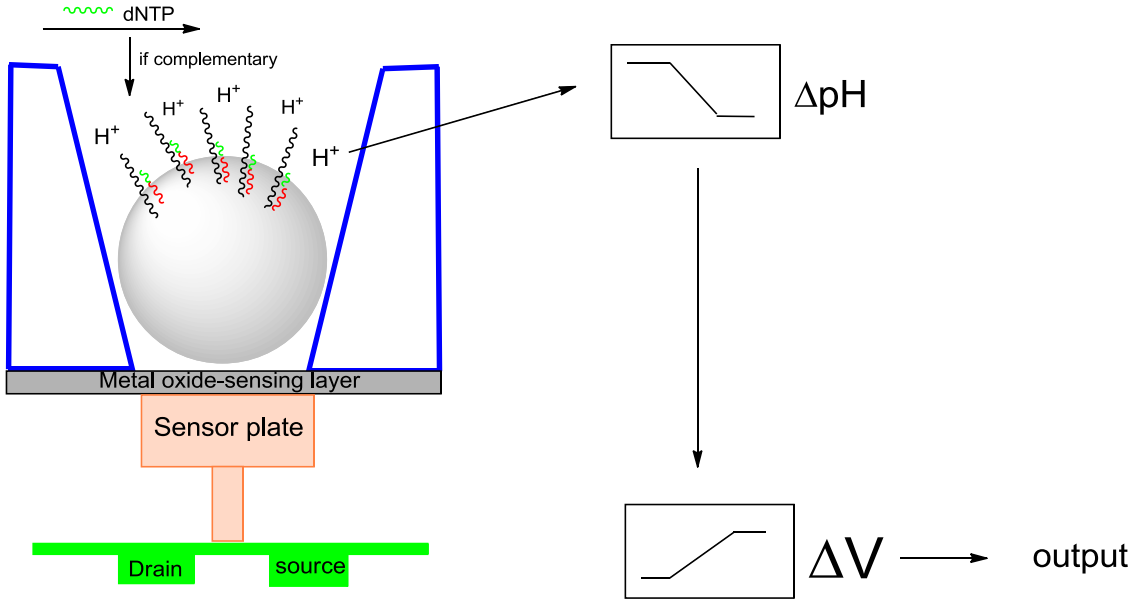


Figure 1.29: Detection of the release of a proton and output.

The addition of a dNTP can therefore be recorded as a difference in voltage and digitalised. Homopolymeric regions are accurately reported by distinguishing the voltage difference between one or multiple base incorporations (figure 1.30).[42,43] Ion Torrent sequencing is capable of recording data from a large number of simultaneous sequencing reactions without the need of further instrumentation to decode the data such as optical systems.

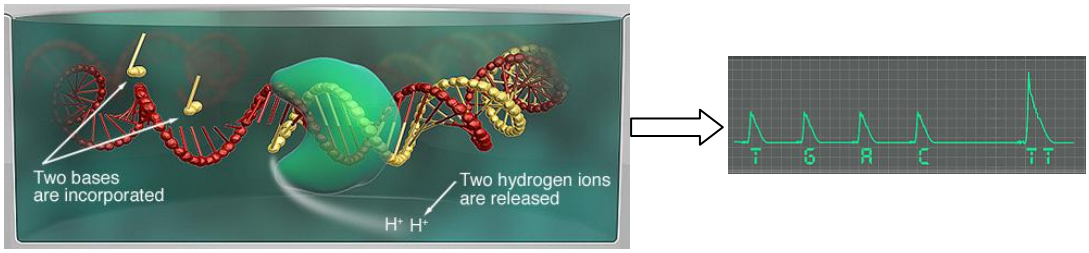


Figure 1.30: Ion torrent sequencing and read-out result (http://www.lifetechnologies.com/uk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html accessed 16/07/2015)

Overall ion torrent sequencing is still in its infancy but could be a promising sequencing platform to significantly speed up the process of sequencing the human genome and to meet the increased demand for sequencing.

Nanopore sequencing

Oxford Nanopore announced the release of two nanopore based sequencing devices in 2014, the GridION and the MinION.[44] Nanopore sequencing involves detecting dNTPs as they pass through a nanopore.

Figure 1.31: Nanopore release MiniIon.[44]

Initial DNA sequencing studies struggled to sufficiently slow the passage of the DNA strand through the nanopore to be able to detect individual nucleotides as they passed through.[45] After extensive research into different types and designs of nanopore, two distinct areas of research have emerged; those using biological nanopores such as a lipid bilayer and those using synthetic solid state nanopores such as graphene.[46] The nanopore is inserted into a membrane and a potential is applied which creates a current to force the polynucleotide DNA strand to run through the nanopore (figure 1.32).

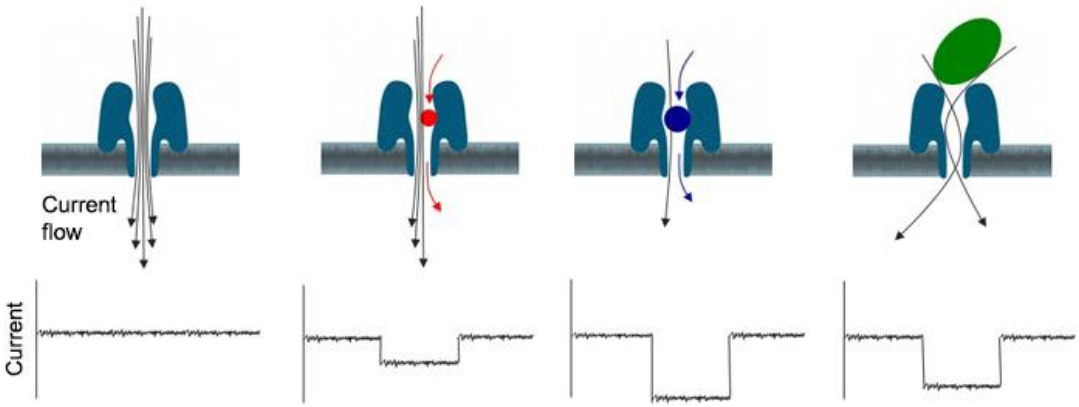


Figure 1.32: Schematic of data output from Nanopore sequencing for three analytes. (https://www.nanoporetech.com/science-technology/introduction-to-nanopore-sensing/introduction-to-nanopore-sensing accessed 16/07/2015)

As individual nucleotides flow through the nanopore they cause characteristic disruptions in the current which allow for nucleotide identification (figure 1.33).[47],[48] This information can then be used to determine the sequence of the DNA which has passed through the nanopore.

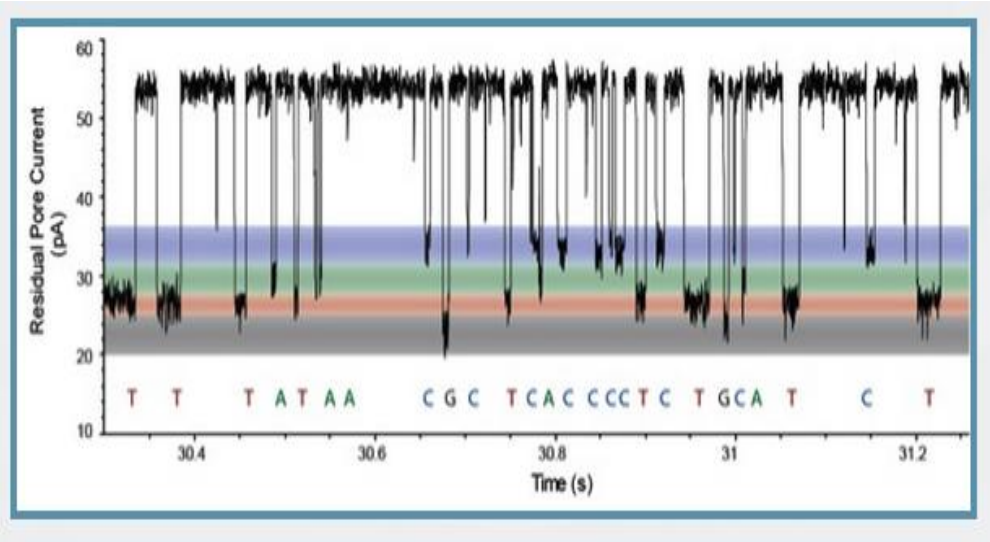


Figure 1.33: Output from nanopore sequencing showing the increased resistance associated with passing a nucleotide through the nanopore membrane and the unique character of the peaks. *http://www.nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/dna-exonuclease-sequencing- Accessed 07/04/13*

1.7 Modified dNTPs as DNA polymerase substrates

Many established and developing sequencing methodologies rely on the use of modified dNTPs. The research undertaken in this thesis has been in collaboration with QuantuMDx, a biotechnology company developing an SBS method which will detect the incorporation of modified dNTPs during sequencing through charge recognition on a silicon nanowire.

SBS sequencing methods are dependent on a modified dNTP bearing a reporter group being incorporating into the growing DNA strand. For the incorporation of a modified dNTP to occur it must be a substrate for a DNA polymerase (section 1.2). DNA polymerases have evolved to be highly sensitive to the incorporation of nucleotides onto the growing 3'-end of DNA. They are capable of distinguishing between RNA and DNA

and in detecting small variations in nucleotide structure.[35] It is therefore challenging to find appropriate DNA polymerases for use with modified nucleotides. Modifications at the C5 position of the pyrimidines and the C7 position of the 7-deazapurines have been shown to be best tolerated by DNA polymerases since such modifications are held away from Watson and Crick base pairing faces.[49, 50] Therefore modifications of nucleotides for use in DNA sequencing predominately occur at the C5-position of the pyrimidines and the C7-position of deazapurines (figure 1.34). Additionally C5 and C7 modifications are positioned within the major groove of the DNA helix, which has been shown to have a stabilising effect upon the DNA.[49, 50]

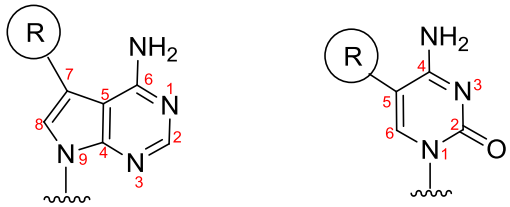


Figure 1.34: C7 and C5 modifications of pyrimidines (2′-deoxyribose ring not shown).

The purines bases cannot be modified as their natural nucleotide structure as this would create a positive charge at the N7 position, therefore the heteroatom is replaced with a carbon (figure 1.35).

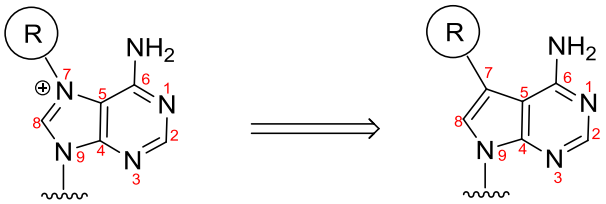


Figure 1.35: The use of 7-deazapurines removes the positive charge at the N7 position

In some cases a DNA polymerase must be engineered to accept a desired modified nucleotide or to increase efficiency of incorporation.[51] This is particularly the case for nucleotides with 3'-hydroxyl modifications. Although many sequencing methods call for the use of 3'-hydroxyl blocking groups, finding suitable modifications that do not adversely affect the polymerase substrate properties has proved difficult.[35] For those that have been found to be substrates (3′-*O*-allyl[32], 3′-*O*-NH_2[35] and 3′-*O*-

azidomethyl,[52] amongst others) the moiety must be small and this can be limiting in the design of new 3'-blocking groups.

Directed evolution is one technique which can be implemented to engineer DNA polymerases better capable of accepting unnatural nucleotides or to increase efficiency of incorporation of modified dNTPs. For example, *Taq* DNA polymerase was found to incorporate the four dideoxynucleotides at varying rates during Sanger sequencing, with 2',3'-dideoxyguanosine-5'-triphosphate (ddGTP) being incorporated ten times faster than other ddNTPs present. This presented a problem with the premature ddGTP termination of DNA synthesis. However Li *et al.*[53] reported that by modifying the amino acid residues of the *Taq* polymerase (specifically Arg-660) this could be rectified.

Many DNA polymerases have an inbuilt proof reading mechanism by which incorrectly inserted nucleotides can be removed by a 3'-5' exonuclease. This is useful in nature as it allows the polymerase to check DNA for errors and subsequently remove the incorrect sequence. However with the use of modified nucleotides exonuclease activity can be undesirable as the polymerase may remove the modified nucleotide from the DNA sequence. This would prevent the sequencing reaction from occurring and also cause the DNA synthesis to terminate. Due to this, many DNA polymerases have had their 3'-5' exonuclease activity removed for use in DNA sequencing.

C5-modified 2'-deoxypyrimidine nucleotides as substrates for DNA polymerases

As this thesis is focussed on the synthesis of modified 2'-deoxycytidine-5'-triphosphates (dCTPs) we are particularly interested in examples within the literature of C5-modified 2'-deoxypyrimidines as DNA polymerase substrates.

Previous studies of C5-modified 2'-deoxyuridine-5'-triphosphates (dUTP's) have shown that C5 modifications based on rigid linkers, where the modification is held away from the Waston and Crick base pairing face are best tolerate by DNA polymerases. This finding was reported by Lee *et al.*[54] who reported that modifications at the C5 position of dUTP must be based on either alkene or alkyne moieties to be substrates for common DNA polymerases. It was also reported that only the trans alkene modified dUTP isomer (5-(*E*)-[3-aminopropenyl]-2'-deoxyuridine-5'-triphosphate) was found to be a substrate,

not the cis (5-(*Z*)-(3-aminopropenyl)-2'-deoxyuridine-5'-triphosphate) configuration (figure 1.36). This was believed to be due to the stereochemistry of the cis-isomer holding the modification too close to the base for polymerase activity. Alkane-based linkers were found to be unsuitable for modified dNTPS where the dNTP is required to be a substrate for DNA polymerases; presumably due to the flexibility at the C5 position interfering with the Watson and Crick base paring.

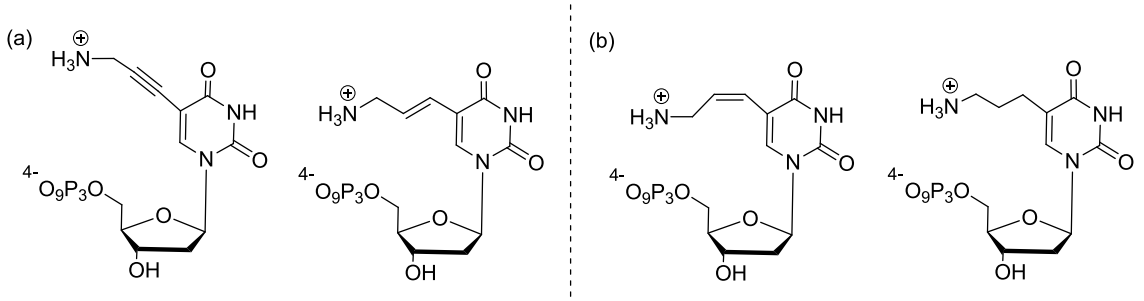


Figure 1.36: (a) Alkyne (3-aminopropynyl) and alkene (trans-(5-(3-aminopropenyl)) derived C5 modifications found to be substrates for a *taq* and Vent or *taq* and *rTh* polymerases respectively. (b) alkene (cis (5-(3-aminopropenyl)) and alkane (propylamine) based linkers found to not be DNA polymerase substrates.

There are a number of examples of highly functionalised C5-modified 2'-deoxypyrimidines which have been found to be substrates for DNA polymerases within the literature. For example, Marx *et al.*[50] published the successful incorporation of a dendrimer modified dUTP (figure 1.37) into an oligodeoxynucleotide (ODN) using KlenTaq DNA polymerase. They observed that the rigid C5-anchor group was sufficient to hold the more flexible dendrimer branches outside of the active site of the polymerase.

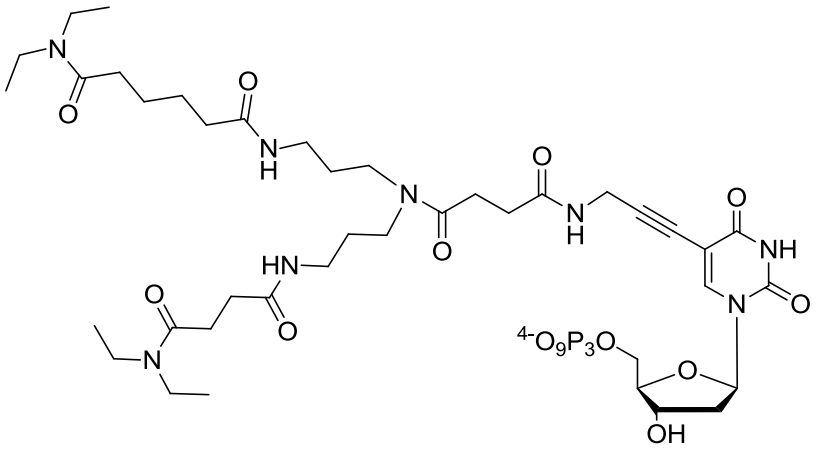


Figure 1.37: Dendrimer functionalised dUTP synthesised by Marx et al[50].

Thoresen *et al.*[55] showed the successful incorporation of C5-modified dUTPs functionalised with fluorescent linkers (figure 1.38). They demonstrated that their fluorescently modified dUTPs were substrates for *Taq*FS DNA polymerase, a commonly used polymerase in DNA sequencing. It is of interest to note that they recorded a positive correlation between the distance the fluorescent group was held from the base and the efficiency of its enzymatic incorporation into DNA.

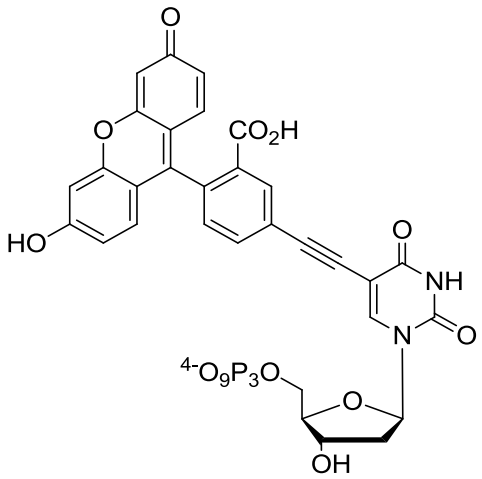


Figure 1.38: An example dUTP with a C5fluorescent modification synthesised by Thoresen *et al*.[55]

Most studies on modified pyrimidines focus on dUTP analogues however there are a number of C5-modified dCTP analogues which have been investigated as substrates for DNA polymerases. Kuwahara and Sawai[56] reported 2'-deoxycytidine-5'-triphosphates (dCTPs) modified with short linkers based on alkynes or trans-alkenes to be substrates for *taq* and Vent (exo-) DNA polymerases (figure 1.39).

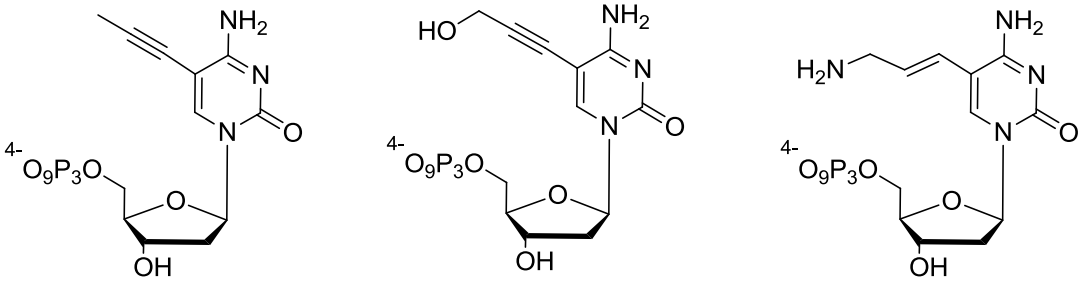


Figure 1.39: C5- modified dCTP analogues found to be substrates for *taq* and Vent (exo-) DNA polymerase.[56]

dNTPs functionalised with charged species

Charged modifications are of particular interest to this project. Few examples of negatively charged reporter groups or linkers are available in the literature and this is further restricted when examining dCTP analogues.

Famulok[57] and Marx[58] have reported two dUTP analogues with carboxylic acid side chains at the C5 position to be substrates for Vent (exo-) and *Pwo* DNA polymerases respectively (figure 1.40).

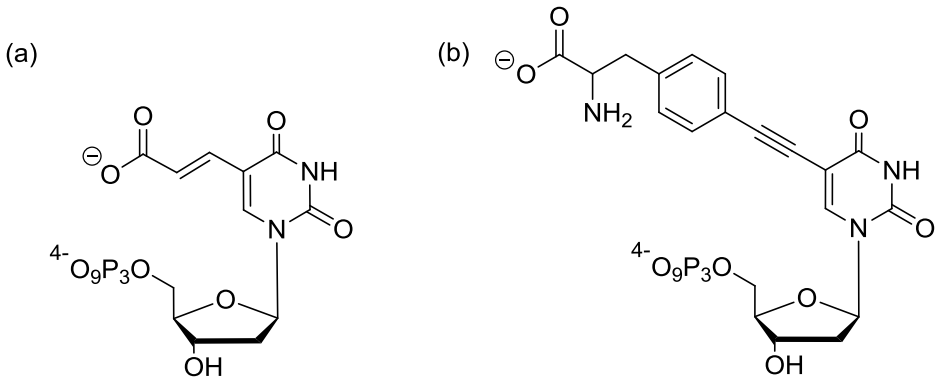


Figure 1.40: Negatively charged C5 modified 2'-deoxyuridine-5'-triphosphates as synthesised by (a) Famlouk and (b) Marx.[57, 58]

A dCTP analogue with a carboxylate linker (5-valeric acid-dCTP, figure 1.41) was published by Hollenstein[59] for investigation as a DNA polymerase substrate. He reported the successful incorporation during primer extension reactions by Vent (exo-), *Pwo* and 9°N_m DNA polymerases up to three consecutive insertions. Successful amplification during PCR was also demonstrated.

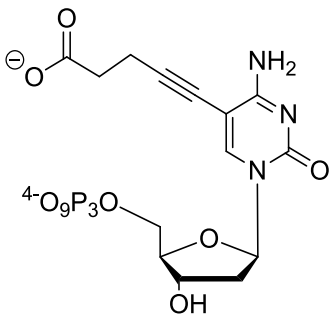


Figure 1.41: 5-valeric acid dCTP as synthesised by Hollenstein.[59]

There are also reports within the literature of positively charged ddNTPs as DNA polymerase substrates. Kumar *et al.*[60] demonstated the incorporation of a positively charged ddNTPs (figure 1.42) for use in Sanger sequencing using ThermoSequeanase II. It is worth noting that Kumar *et al.*[61] also synthesised an α-sulpho-β-alanine linker to produce negatively charged 2',3'-dideoxycytidine-5-triphosphate (ddCTP) analogues, again for use in Sanger sequencing.

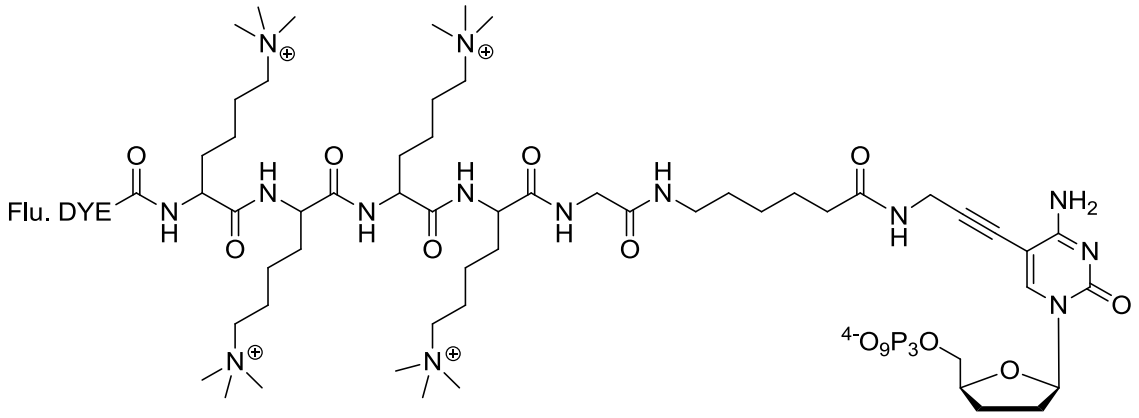


Figure 1.42: Positively charged ddCTP for use in Sanger sequencing (Flu. Dye = ROX fluorescent dye) [60, 61]

1.8 Nanowires for the detection of biological species

Nanowires have recently emerged as an area of interest for use as a sequencing platform. In 2001 Lieber *et al.*[62] reported the use of boron-doped silicon nanowires (SiNWs) as highly sensitive detectors for biological species. Recently this technology has been further developed by Zhang *et al.*[63] as a platform for DNA sensing. Peptide nucleic acid (PNA) is a neutral polymer with a similar structure to DNA or RNA. In order to show the potential of nanowires to detect DNA an engineered strand of PNA was immobilised onto the silicon surface of the nanowire and a DNA strand with a complementary sequence introduced. The binding of the DNA was detected as a change in the electric field surrounding the nanowire.

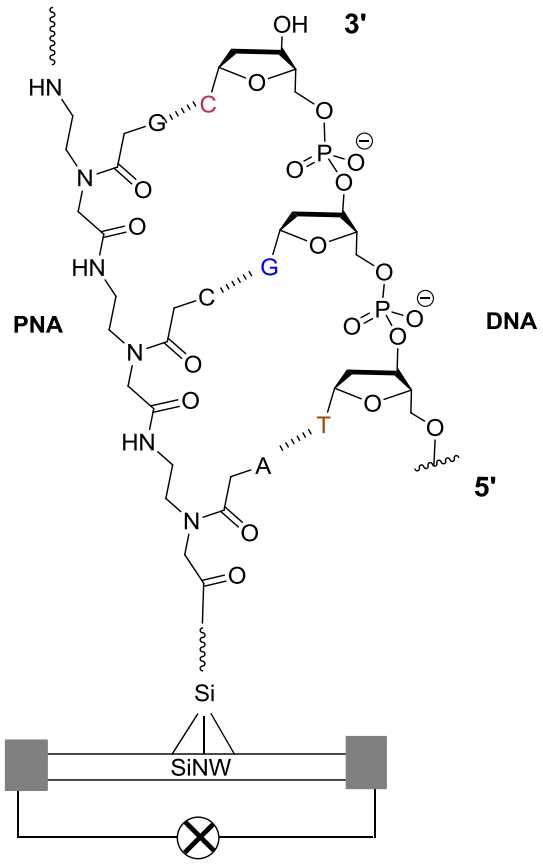


Figure 1.43: PNA attached to a silicon nanowire for the detection of DNA.

The nanowire is acting as a field effect transistor (FET), exhibiting a change in conductivity in response to variations in the electrical field; such as detection of the anionic DNA backbone. Due to the sensitivity of the nanowire small variations in the amount of charge in the electrical field should be detectable and discernible from each other.[64] QuantuMDx aim to exploit the sensitivity of nanowires to charge variations by developing a nanowire based SBS device to detect modified dNTPs through charged reporter groups.

1.9 Aims of the project

The aim of this project is to synthesise C5-modified dCTPs as proof-of-concept molecules for use in the DNA SBS device in development by the biotechnology company QuantuMDx.

QuantuMDx DNA sequencing

QuantuMDx are developing a nanowire based handheld SBS device. Handheld devices for DNA sequencing is an area which has not yet been fully exploited. QuantuMDx wish to develop a completely portable device, small enough to be handheld and capable of rapid analysis of short sequences of DNA. The overall aim is to achieve sample to result turnaround in 15 minutes.

The device will work using SBS methodology as previously described in section 1.5. As a 3'-blocking group is not currently being explored it is imagined that initially the four modified-dNTPs will be introduced sequentially. The detection of dNTPs during DNA sequencing will be through an anionic reporter group attached via a linker and rigid anchor to the dNTP at the C5 position of the pyrimidines and the C7 position of the 7-deazapurines. As described in section 1.8 nanowires show a change in conductivity in response to changes in the electrical field around them. Therefore if the complementary dNTP is incorporated into the growing DNA chain, the nanowire will detect the anionic reporter group in close proximity to itself and the resulting change in conductance recorded. By associating the change in conductance with the order of introduction of the dNTPs the identity of the nucleotide can be determined.

QuantuMDx's sequencing device is still in its infancy however in the final device a cleavable linker and 3'-blocking group will be used during sequencing (figure 1.44). With the use of a cleavable linker each dNTP will be functionalised with a discrete reporter group. When all four dNTPs are placed in the reaction mixture together the identity of the complementary dNTP can be determined from its unique reporter group. The cleavable linker facilitates the removal of the reporter group from the growing DNA strand. The 3'-blocking group will prevent the incorporation of more than one dNTP at a time. This is advantageous for the resolution of homopolymeric regions on the template

DNA. After the successful incorporation and detection of the complementary dNTP by the nanowire the linker is cleaved to remove the reporter group and the 3'-blocking group simultaneously removed to allow DNA synthesis to continue. Ideally conditions of the cleavage of both should be complementary to each other.

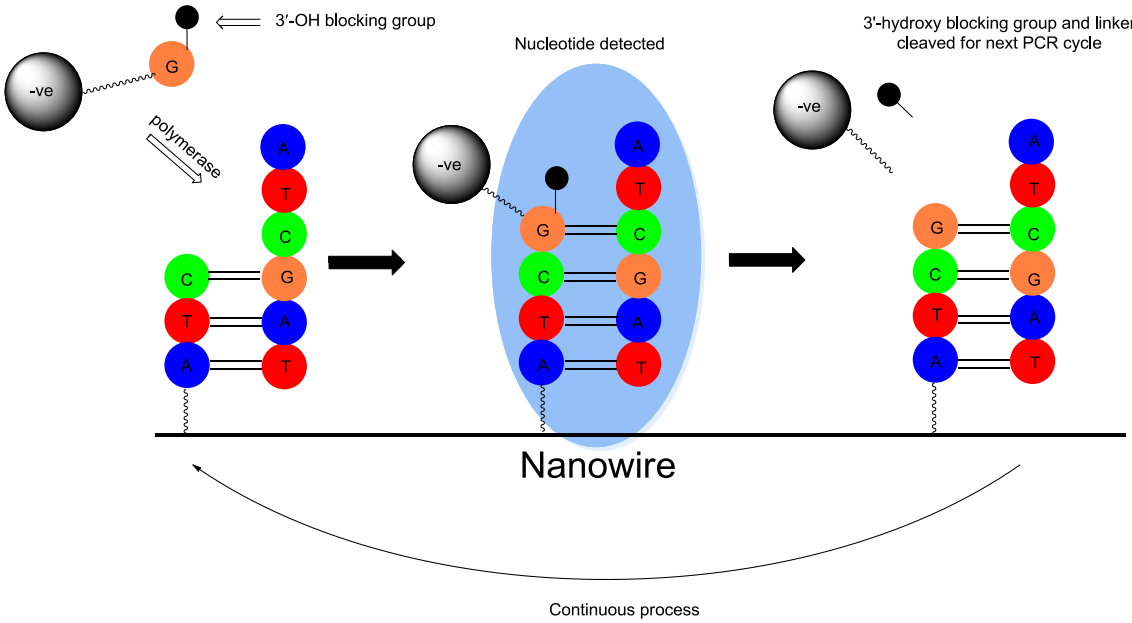


Figure 1.44: The principle design of QuantuMDx's nanowire based sequencing device.

The target modified dNTPs for use in SBS sequencing

The focus of this thesis is to synthesise and investigate C5-modified dCTPs as proof-of-principle compounds for QuantuMDx's charge-detection based SBS. The modified dCTPs synthesised will be follow the generalised structure shown in figure 1.45.

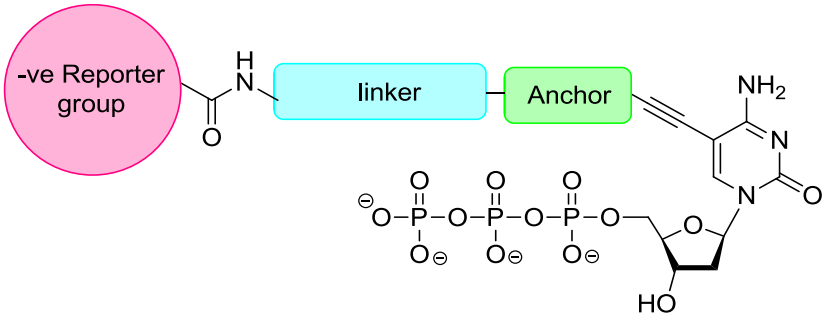


Figure 1.45: Template for modified 2'-deoxycytidine analogues for QuantuMDx DNA sequencing.

Four features require introduction the 2'-deoxycytidine base; An anchor, linker, anionic reporter group and the triphosphate moiety. As the design of the sequencing device is still underway, little is known about the prerequisites needed for the modified dCTP analogues to be suitable. However three areas for interest have been focussed upon: Synthesising dCTPs with differing length linkers; the synthesis and introduction of new anionic reporter groups; and the synthesis of a photocleavable linker. These will be discussed in chapter 3.

Chapter 2 –*Phosphorylation Chemistry*

2. Phosphorylation chemistry – Chemical synthesis of dNTPs

2.1 Introduction to Phosphorylation Chemistry

The first chemical synthesis of a 2'-deoxyribonucleoside 5'-triphosphate (specifically dATP) from the reaction of adenosine-5'-dibenzyl pyrophosphate with dibenzyl chlorophosphonate was described over 6 decades ago by Baddiley *et al.*[65] Since their initial report numerous advances have been made in synthetic methods to phosphorylate both natural and modified 2'-deoxyribonucleosides.

Synthesis of 5'-triphosphates through the activation of 5'-monophosphates

The synthesis of dATP described by Baddiley *et al.* is a lengthy procedure with multiple time consuming purifications steps culminating in low amounts of isolated dATP.[65] Due to this an alternative route to synthesising 5'-triphosphates through the activation of 5'-monophosphates was developed.

The first widely used synthesis of dNTPs from 5'-monophosphates was reported by Smith and Khorana.[66] Pyrophosphoric acid and dicyclohexylcarbodiimide (DCC) were used and proved a significant improvement upon previous syntheses. The reaction proceeds through an active intermediate (**1**) formed through the condensation of DCC with 5'-monophosphate (figure 2.1). The intermediate is then displaced by the nucleophilic pyrophosphoric acid to form the 5'-triphosphate.

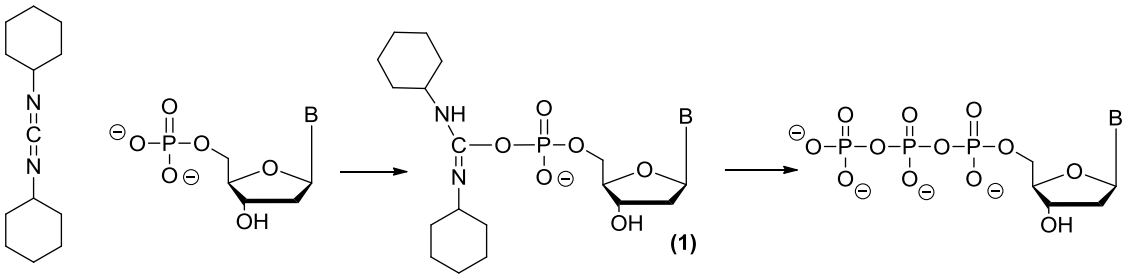


Figure 2.1: Reaction of DCC with 5'-monophosphate and subsequent conversion to the 5'-triphosphate (where B represents one of the four heterocyclic bases; A, G, T or C).

The process did however have problems, predominately the challenge of the low solubility of the reagents and starting materials and the formation of diphosphates as a side reaction. This was later addressed by Moffatt and Khorana[67] who reported the synthesis of adenosine-5'-phosphoromorpholidate (**2**) as a more convenient and stable activated 5'-monophosphate. The synthesis proceeded with an initial condensation reaction between a 5'-monophosphate and DCC, as previously described, however the displacement using the nucleophilic morpholine led to the 5'-phosphoromorpholidate being isolated in good yields. The phosphoromorpholidate can be reacted with pyrophosphate to give the corresponding 5'-triphosphate in up to 90% yields (figure 2.2).[68]

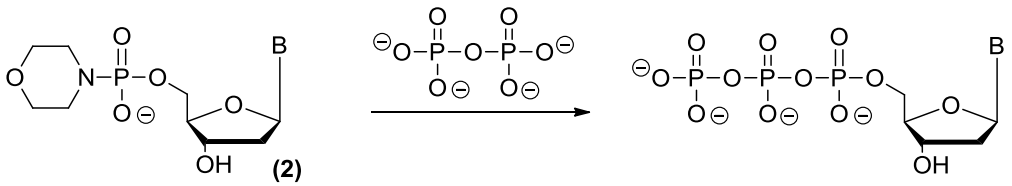


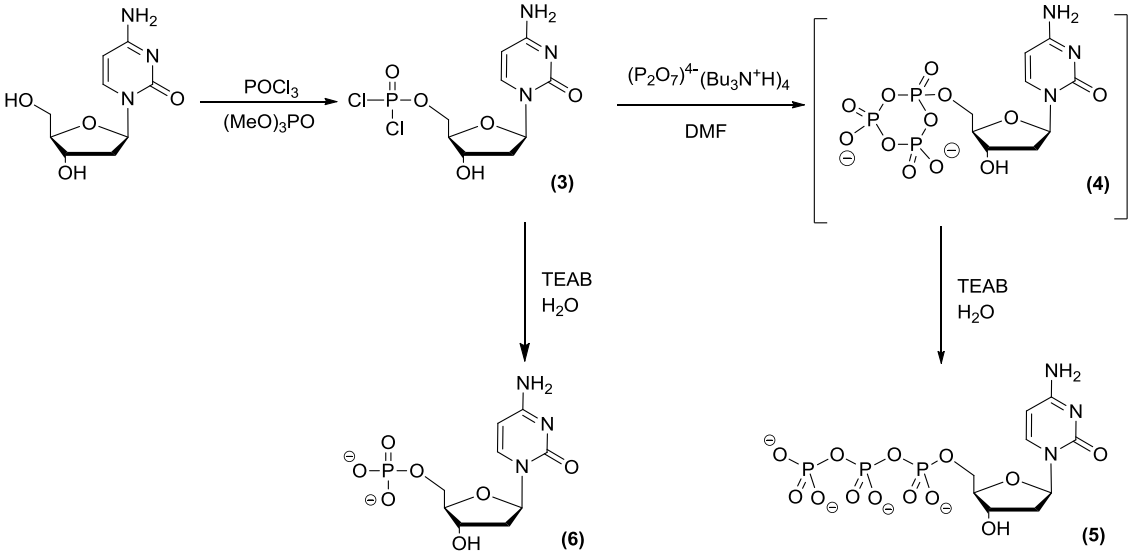
Figure 2.2: Synthesis of 5'-triphosphates from the reaction of 5'-phosphoromorpholidate (**2**) with pyrophosphate. (where B = A, G, T or C).

As earlier methods relied upon the deoxynucleoside-5'-monophosphates for conversion to triphosphates the selective synthesis of 5'-monophosphates in the presence of 2' or 3'-hydroxyl groups became an important challenge. At the same time new phosphorylation methods were explored with a focus being placed on those which could synthesise 5'-triphosphates in a ‘one pot’ fashion. The methods most commonly used today can be

separated into two categories- P(V) or P(III) phosphorylations- and will be discussed in the following sections.

P(V) phosphorylation methodology

Yoshikawa *et al.*[69] described the reaction of 2'-deoxyribonucleosides with phosphoryl chloride ($POCl_3$) to form the highly reactive phosphorodichloridate intermediate (**3**) (scheme 2.1). This phosphorodichloridate intermediate can then be hydrolysed, for example using triethylammonium bicarbonate buffer (TEAB), to form the nucleoside 5'-monophosphate (**6**). This method is described as a P(V) phosphorylation due to the oxidation state of the phosphorus atom of $POCl_3$. Ludwig[70] later proposed a simple modification to Yoshikawa's method which allowed the syntheses of dNTPs. This can be achieved by the addition of tributylammonium pyrophosphate to the phosphorodichloridate intermediate (**3**) to form the cyclic intermediate (**4**). Following hydrolysis the linear dNTP product (**5**) is formed. Ludwig's modification of Yoshikawa's P(V) method was a significant breakthrough in producing dNTPs in a 'one pot' procedure without the requirement for protecting groups on the exocyclic amines of the bases or the 3'-hydroxyl group. This was key for simplifying and increasing reliability of the synthesis of dNTPs.[69]



Scheme 2.1: P(V) phosphorylation of unprotected 2'-deoxycytidine.

P(V) phosphorylations are unfortunately prone to number of side reactions; should the phosphorodichloridate intermediate (**3**) come into contact with water before the addition of pyrophosphate, the 5'-monophosphate by-product (**6**) is formed. Additionally, by using an unprotected 2'-deoxyribonucleoside, phosphorylation of the nucleophilic 3'-hydroxyl can occur forming the 3',5'-bisphosphate (**7**) (figure 2.3).

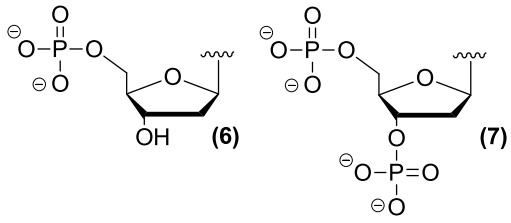


Figure 2.3: 5'-Monophosphate (**6**) and 3',5'-diphosphate 2'-deoxynucleotides (**7**) can both be formed in P(V) phosphorylations.

Yoshikawa *et al.*[71] subsequently reported a number of alterations to the method in order to reduce side products and increase regioselectivity towards synthesising 5'-phosphates. Decreased temperatures were suggested, with phosphorylation being carried out between 0 to -5°C showing increased selectivity to the 5'-hydroxyl. They also report that the use of trialkyl phosphate solvents, such as trimethyl phosphate, increased the yield of 5'-nucleotides by up to 70% depending on the nucleoside. Although the increased yield is likely due to the increased solubility of the nucleosides in the phosphate solvents, they postulated that an interaction between trialkylphosphate and $POCl_3$ was possible and could form an ionised, active phosphorylating agent (**8**) (figure 2.4).

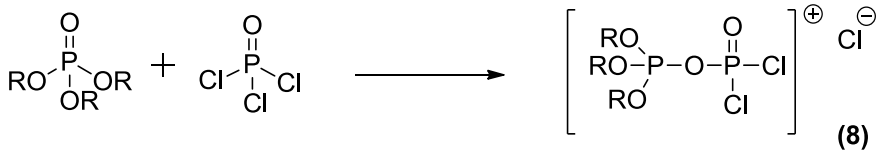


Figure 2.4: Yoshikawa *et al*. proposed phosphorylating agent formed through the reaction of trialkyl phosphate solvents and $POCl_3$.[71]

Other modifications have also been proposed in recent years such as the use of a proton sponge® (1,8-bis(dimethylamino)naphthalene) where alkyne and alkene modified nucleosides are phosphorylated. Without the use of a proton sponge the HCl generated

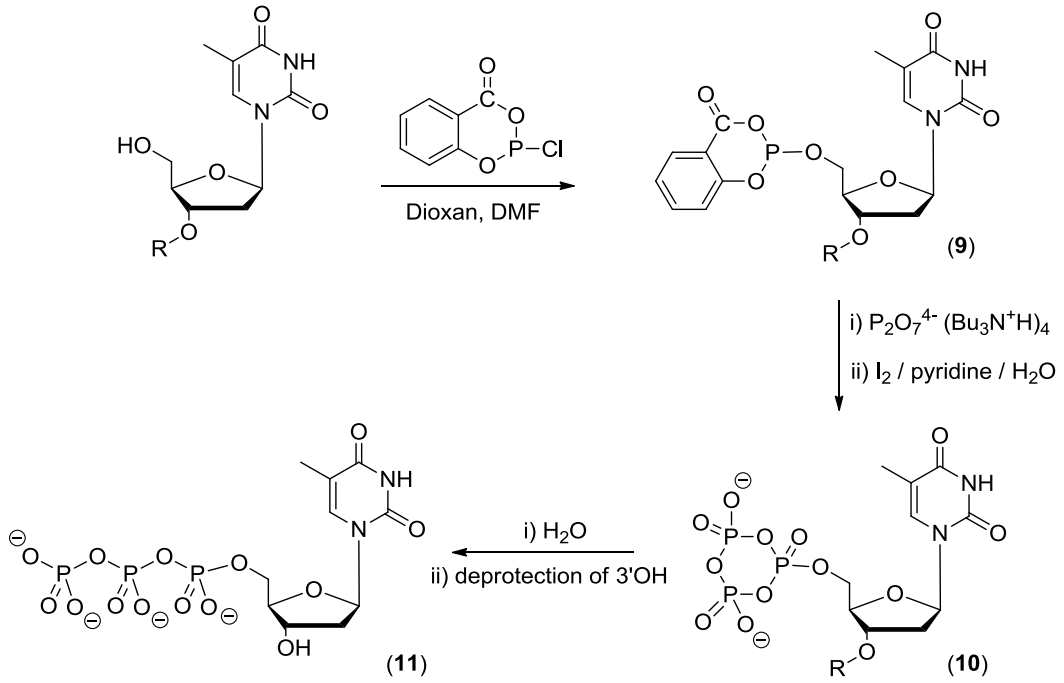
during the phosphorylation reaction was reported to undergo an addition reaction to the unsaturated bond present in modified nucleosides.[72] The proton sponge acts as a strong base to neutralise the HCl generated but has very little nucleophilic behaviour due to steric effects.[73] The use of proton sponge was found to lead to significantly shortened reaction times making its use popular in all P(V) phosphorylations.

Given the recent rise in the need to phosphorylate modified nucleosides which are often time-consuming to make, the drawbacks of P(V) phosphorylations discussed have led to new phosphorylation strategies. Specifically 2'-deoxyribonucleoside protection methodologies can be used to increase 5'-triphosphate yields by reducing the formation of by-products and facilitating the use of highly reactive phosphorylating agents.

P(III) phosphorylation methodology

Ludwig and Eckstein proposed the use of salicyl chlorophosphite as a P(III) phosphorylating agent.[74] P(III) reagents have a trigonal pyramidal structure with a lone pair of electrons on the phosphorus atom making them more reactive than pentavalent P(V) alternatives but also more susceptible to hydrolysis and oxidation.[75]

P(III) phosphorylation proceeds through a reactive phosphite intermediate (**9**), which reacts with pyrophosphate through two nucleophilic substitutions before being oxidised to give the 5'-cyclic triphosphate (**10**). Finally the cyclic triphosphate is hydrolysed to give the desired linear dNTP (**11**) (scheme 2.2). Due to the 3'-hydroxyl position being protected, no by-products derived from reaction at this position are formed. Usually the 3'-hydroxyl group is protected with a base labile protecting group, such as acetyl, as the removal is compatible with the stability of the triphosphate.



Scheme 2.2: P(III) phosphorylation of a 3'-hydroxyl protected 2'-deoxythymidine where R = base-labile protecting group.

There are some disadvantages to the use of P(III) reagents; due to their reactivity they lack the regioselectivity of P(V) reagents and require sites of unwanted phosphorylation to be protected before use. It is universally agreed that the 3'-hydroxyl group must be protected prior to phosphorylation but the need for blocking groups at exocyclic amino functionalities of the bases is debated. In Ludwig and Eckstein's initial report they phosphorylated 2'-deoxy-adenosine, guanosine and cytidine with the exocyclic amines either protected or as the unprotected amines and determined that they did not believe that protection chemistry of the heterocyclic bases was required.[74] However it has also been reported that although the exocyclic amino groups of adenine and guanine are poor nucleophiles when compared to the hydroxyl groups of the sugar, the exocyclic amino group of cytosine was found to display a similar reactivity.[76] In addition, since many sugar protection strategies lead to additional protection of the amino groups of the base, this procedure can often be achieved without the addition of a further synthetic step. Therefore protection of the exocyclic amino functionality is often performed for P(III) phosphorylations.

2.2 Synthetic strategy towards modified dNTPs

The phosphorylation of modified 2'-deoxyribonucleosides requires some consideration towards the synthetic pathway for introducing the desired C5-modification and the triphosphate moiety. As previously discussed this thesis is focussed on the synthesis of dCTPs for use in SBS. There are four features to be introduced to 2'-deoxycytidine in order to make it suitable for use in SBS; an anchor, linker, reporter group and triphosphate moiety (figure 2.5). However the order of synthesis can be varied.

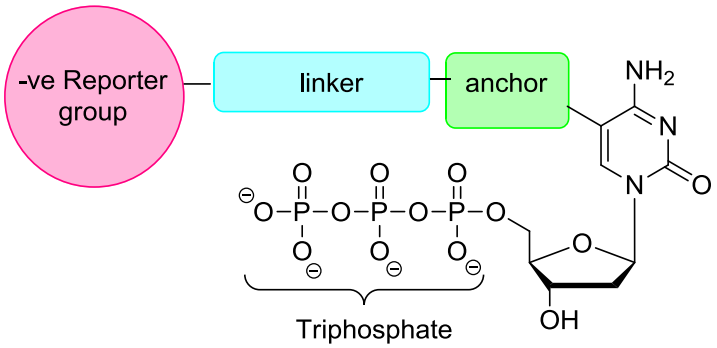
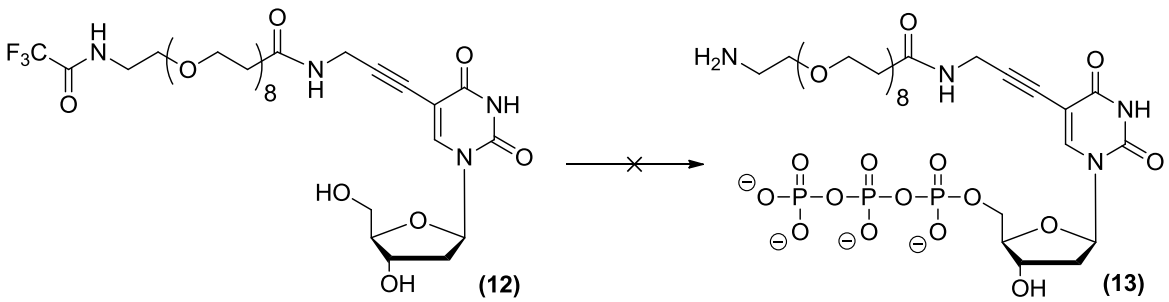


Figure 2.5: SBS dCTP design.

Synthetic pathway towards modified dCTPs

When deciding on a synthetic route towards modified dCTP analogues introducing the anchor first (out of the four components required, figure 2.5) is arguably the most rational choice as it will be present in all modified dCTPs. The introduction of an anchor is pivotal to providing the rigidity to help ensure the dCTPs are accepted as substrates for DNA polymerases (section 1.7). A Sonogashira cross-coupling between protected propargylamine allows the introduction of the rigid propargylamino anchor onto 5-iodo-2'-deoxycytidine.[77] From this point a choice between functionalising the anchor or functionalising the sugar with the triphosphate moiety needed to be made. We initially chose to functionalise the anchor with a linker (performed through a TSTU coupling reaction) and subsequently attempted to phosphorylate. A trial reaction for expediency was performed using PEG_8-modified 2'-deoxyuridine (**12**) (Scheme 2.3). The PEG_8-2'-deoxyuridine nucleoside had been previously synthesised by a co-worker in the Williams

group. Due to the lack of protection chemistry at the 3'-hydroxyl the phosphorylation of **12** was attempted using P(V) chemistry.



Scheme 2.3: P(V) phosphorylation. Reagents and conditions: (i) $POCl_3$ (1.5eq), $PO(OMe)_3$, proton sponge, -10°C (ii) PPi, Bu_3N, DMF (iii) 0.1M TEAB (iv) 35% NH_4OH.

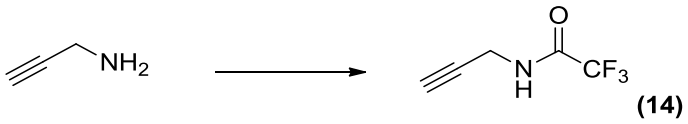
The synthesis of the triphosphate (**13)** proved unsuccessful. This was attributed to the steric bulk of the linker hindering phosphorylation of the 5'-hydroxyl. Similar issues have been noted in the literature, for example Thoresen *et al.*[55] reported significant difficulties when phosphorylating 2'-deoxyuridine bearing a fluorescent dye at the C5 position. Attempts to protect the base in order to perform P(III) phosphorylation chemistry were also unsuccessful.

Due to these findings it was decided that the triphosphate moiety should be introduced prior to the addition of the linker but after introduction of the propargylamino anchor. We therefore began synthetic efforts to synthesise the anchor-modified dCTP analogue using P(III) phosphorylation methods.

2.3 Synthesis of anchor modified dCTP

Synthesis of the anchor

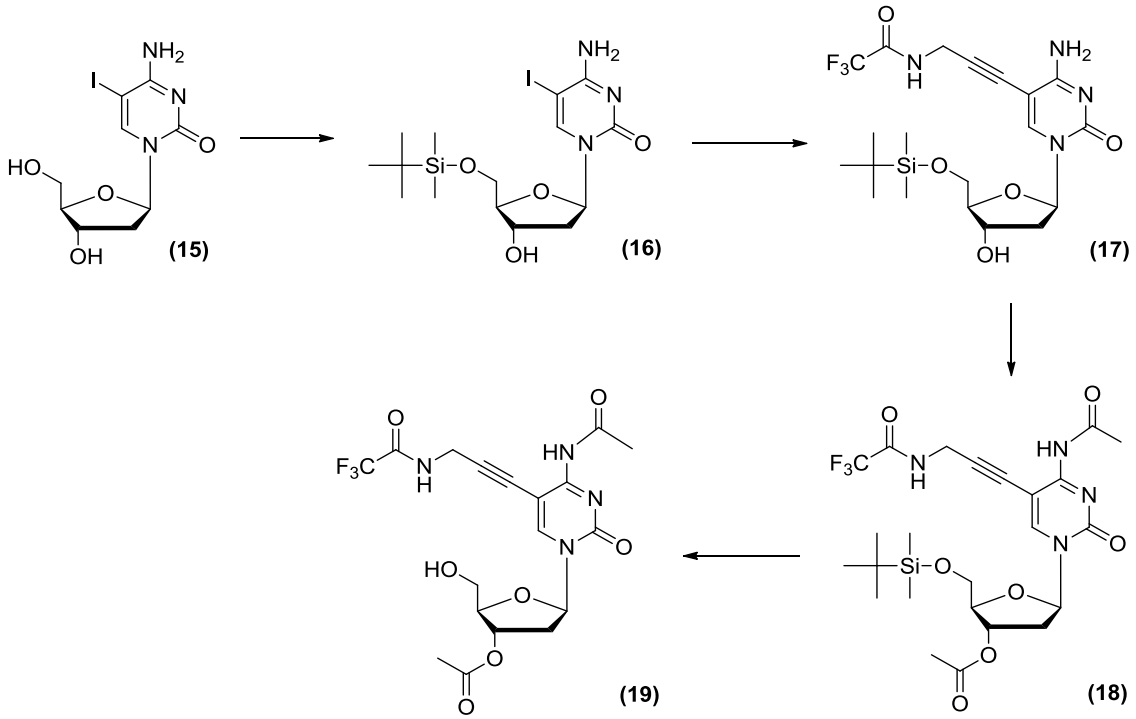
Propargylamine was protected with trifluoroacetamide in a good yield (76%, (**14)**, scheme 2.4). Trifluoroacetamide was chosen as the protecting group as it can be easily removed with aqueous (aq.) ammonia at room temperature, which is compatible with the stability of triphosphates.



Scheme 2.4: Trifluoroacetylation of propargylamine to give **14.** Reagents and conditions**:** methyl trifluoroacetate, DCM, 0°C, 48 h, 76 % yield.

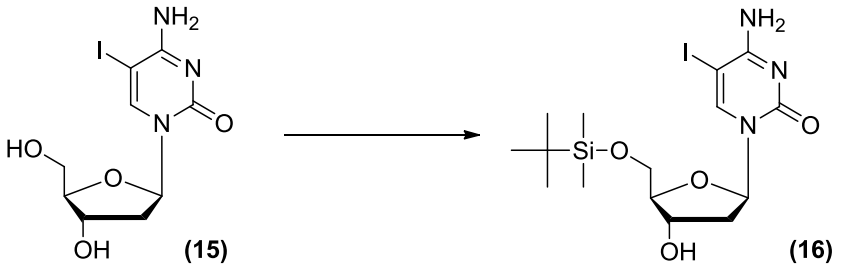
Protection chemistry towards P(III) phosphorylation

Phosphorylation was envisaged via P(III) methodology, therefore the synthesis of 3'-protected 2'-deoxycytidine was required. The proposed synthetic route (scheme 2.5) was based on previous trial protections performed on 2'-deoxythymidine and 5-iodo-2'-deoxyuridine. Due to the N4-amino functionality of 2'-deoxycytidine additional protection chemistry of the heterocyclic base was employed. The reaction scheme is a four step pathway requiring the use of two alternative hydroxyl protecting groups. The 5'-hydroxyl protecting group must be removed in the presence of the 3'-hydroxyl and N4 protecting groups; Silyl and acetyl protecting groups were chosen respectively.



Scheme 2.5: Proposed synthetic route to 3'-protected-2'-deoxycytidine (**19**).

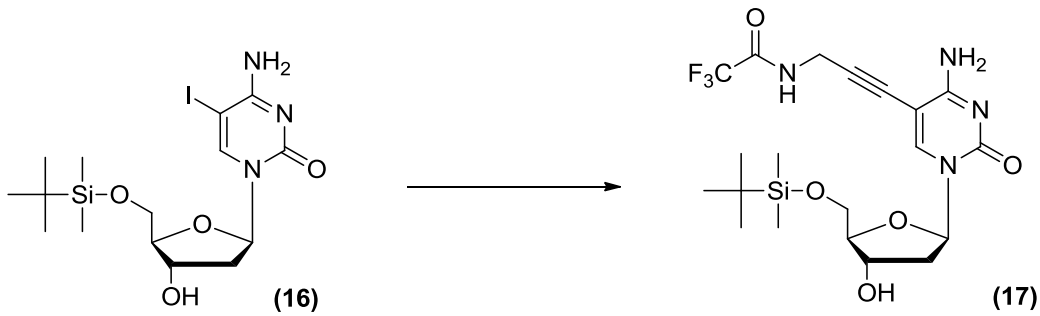
In the first synthetic step, *tert*-butyldimethylsilyl chloride (TBDMSCl) was used to protect the 5'-hydroxyl group of the commercially available 5-iodo-2'-deoxycytidine (**15**). The addition of the silyl group before the Sonogashira coupling reduced the polarity of the compound in order to allow for easier purification by flash chromatography.



Scheme 2.6: TBDMSCl protection to give 5'-protected 2'-deoxycytidine **16**. Reagents and conditions**:** TBDMSCl, Imidazole in DMF. Overnight, RT, 74% yield.

Literature precedent implies that controlling the equivalents of imidazole to TBDMSCl and exploiting the additional reactivity of the primary alcohol over the secondary alcohol are key to reducing the formation of the 3',5'-*O*-TBDMS-2'-deoxycytidine species.[78] We found that 2.2 : 1.1 equivalents imidazole to TBDMSCl gave the desired 5'-*O*-*tert*butyldimethylsilyl-5-iodo-2'-deoxycytidine (**16**) in a 74% yield after purification (scheme 2.6).

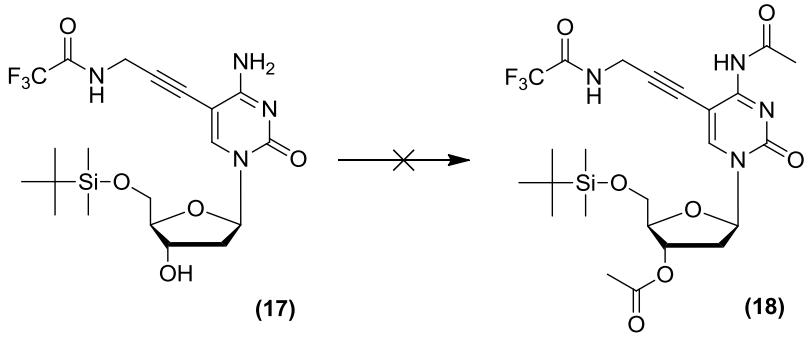
The anchor (**14**), was then attached to the C5-position by a palladium catalysed Sonogashira cross coupling to 5'-*O*-(*tert*-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (**17**) in a 75% yield (scheme 2.7).



Scheme 2.7: Sonogashira coupling of anchor **(14)** to 5'-protected 2'-deoxycytidine **(16)**. Reagents and conditions: (i) Et_3N, $Pd(PPh_3)_3$, **(14)**, DMF, absence of light, 0°C (ii) overnight at RT, 75% yield.

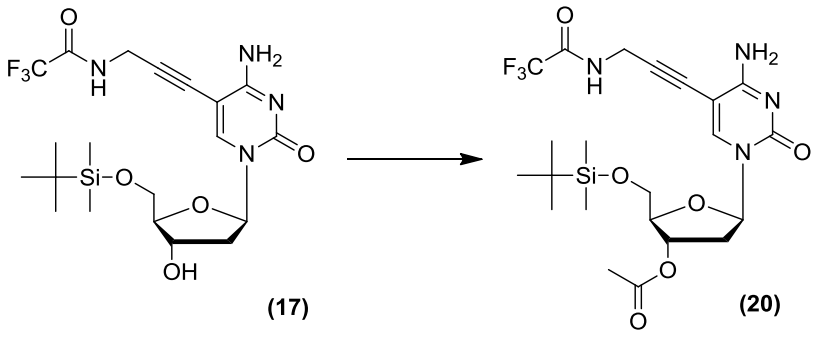
Prior to phosphorylation, we envisaged acetyl protection of both the 3'-hydroxyl group and the N4-amino group due to its ease of removal and its compatibility with cleavage of the silyl group from the 5'-hydroxyl. In addition the actyl group can be removed during cleavage of the trifluoroacetyl group from the anchor.

Unfortunately the synthesis of the *bis*-acetyl protected 2'-deoxycytidine (**18**) proved to be challenging (scheme 2.8). Multiple undesired acetyl products could be detected by mass spectrometry. Repeated attempts to vary the synthetic conditions proved to be ineffective. Due to this obtaining a pure sample of **18** in sufficient quantity for continued synthesis could not be achieved. The proposed protection route (scheme 2.5) was therefore not deemed viable.



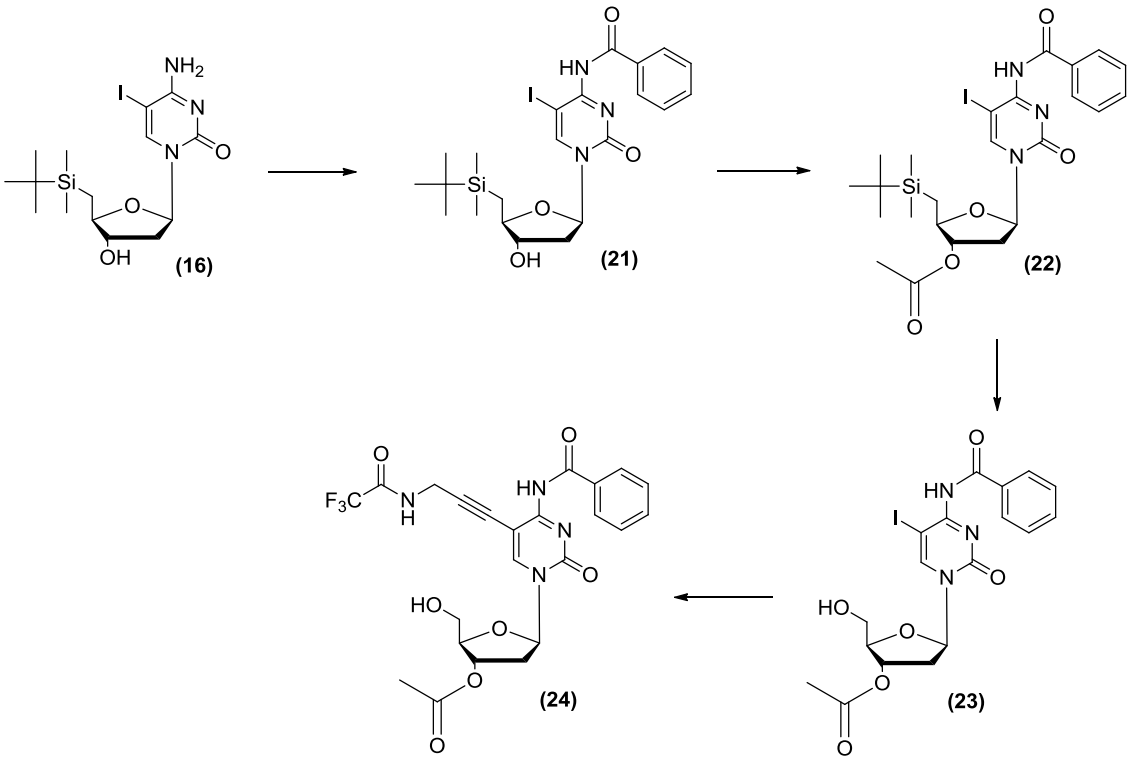
Scheme 2.8: Failed synthetic attempt at the synthesis of **18.** Reagents and conditions: Ac_2O, Et_3N, DMAP, Py.

Attempts to produce 3'-monoacetylated material (**20**) (Scheme 2.9) were more successful however remained low yielding (54%). Attempts to increase the yield resulted in the same challenges seen with the attempted synthesis of (**18**).



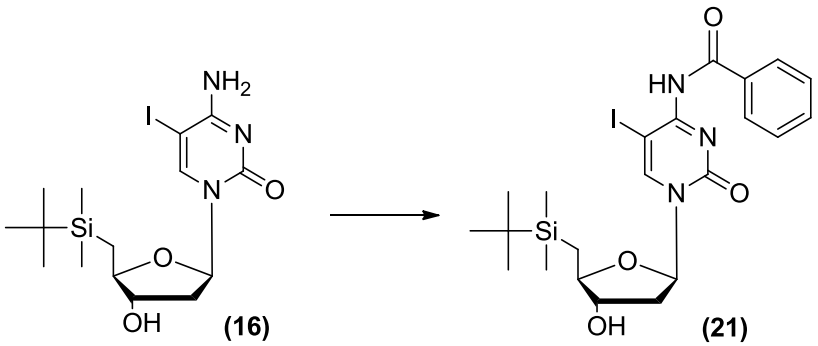
Scheme 2.9: Acetylation of **17.** Reagents and conditions**:** Ac_2O Et_3N, DMAP, Py, overnight, RT, 54% yield.

As an alternative we considered the use of the more stable benzoyl protecting group for the N4 amino group that would allow the preparation of a 3'-hydroxyl protected nucleoside (scheme 2.10).



Scheme 2.10: Proposed alternative synthetic route to 3'-OH protected-2'-deoxycytidine (**24**).

Benzoylation of (**16**) using transient protection[79] employing trimethylsilyl chloride (TMSCl) followed by benzoylation and mild hydrolysis of the 3'-TMS ether gave the desired compound (**21**) in 61% yield (scheme 2.11).



Scheme 2.11: Benzoylation of **16**. Reagents and conditions**:** (i) TMSCl, Py, 3 h (ii) BzCl, 2 h (iii) H_2O, 5 h, 61% yield.

Once the successful synthesis of (**21**) had been confirmed, three small scale acetylations were performed, in each case varying the equivalents (eq.) of acetic anhydride used (figure 2.6). Each reaction was analysed by analytical reversed phase high pressure liquid chromatography (RP-HPLC) after 12 hours of stirring at ambient temperature. The starting material (**21**), being more polar elutes at ~ 25 minutes compared to the ~ 27 minutes of the product (**22**). The latter peak was isolated and confirmed as the desired (**22**) by NMR and mass spectrometry.

When increasing the acetic anhydride from 2 to 3 eq., little difference in conversion of starting material to product was seen. In comparison, 12 hours with 5% v:v acetic anhydride in pyridine produced (**22**) in an 94% conversion as determined by RP-HPLC.

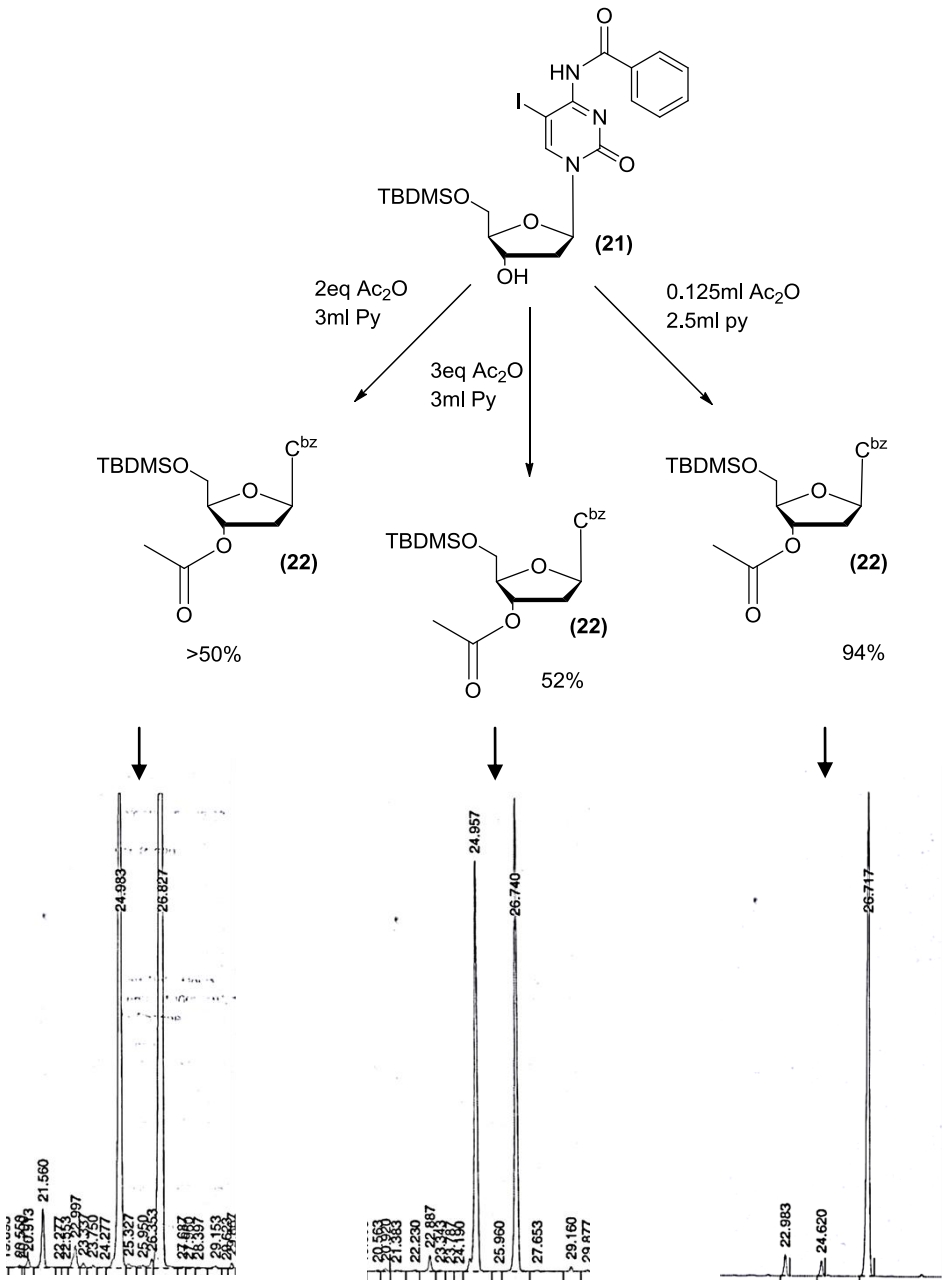
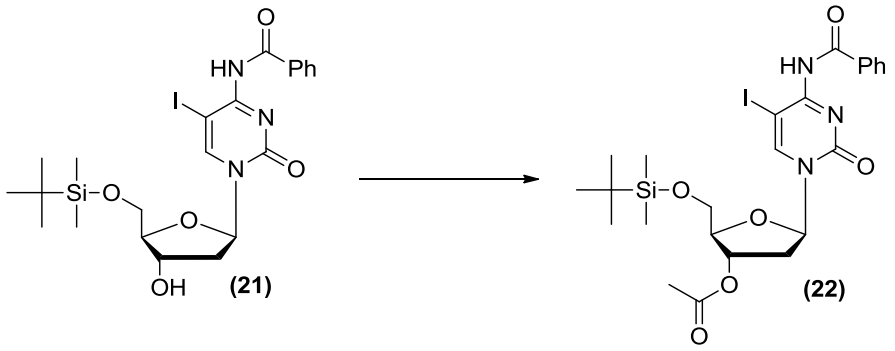


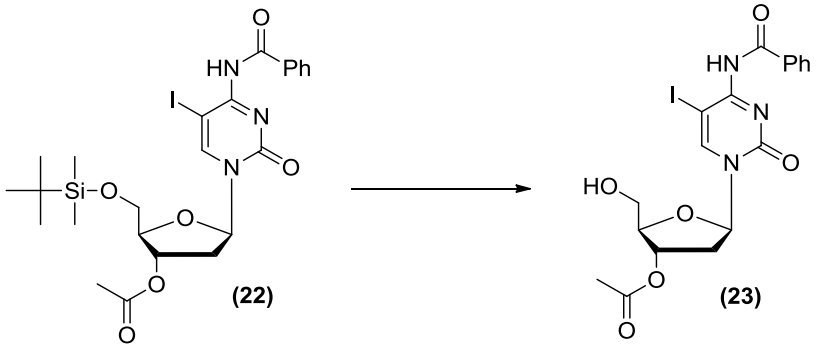
Figure 2.6: A comparison of conditions for the acetylation of (**21**)**.** *N*4-benzoyl protected base represented as C^{bz}. Reaction conditions stated and HPLC chromatographs shown. HPLC conditions: 5 to 100 % B over 30 mins where A = H_2O and B = MeCN, UV detection = 295 nm, flow rate = 1 mL/min.

No side reactions were seen in any of the acetylations and as conditions of 5% v.v acetic anhydride in pyridine gave the best conversion of starting material to (**22**) this was scaled up to produce a larger quantity of the 3'-acetylated nucleoside (scheme 2.12).



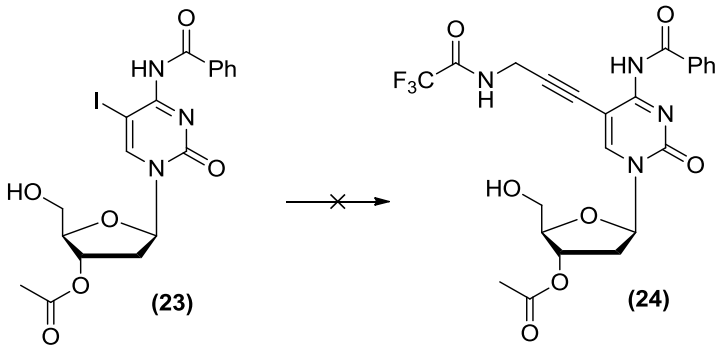
Scheme 2.12: Acetylation of (**21**). Reagents and conditions**:** 5% v:v Ac_2O/Py, overnight, RT, 76% yield

The silyl group of **22** was then removed using tetra-*n*-butylammonium fluoride (TBAF) as a source of fluoride ions in THF. Silicon-fluorine bonds have a much higher bond strength than silicon-oxygen bonds (810 $kJmol^{-1}$ vs 530 $kJmol^{-1}$).[80] This results in the cleavage of the Si-O bond to give the unprotected 5'-hydroxyl (**23**) (scheme 2.13).



Scheme 2.13: Desilyation of (**22**) Reagents and conditions**:** 1 M TBAF in THF, THF, RT, overnight. 86% yield.

After the protection chemistry to synthesise the 3'-protected-deoxycytidine (**23**) had been successfully achieved, the next synthetic step was the introduction of the anchor by a Sonogashira cross coupling with the previously synthesised (**14**).



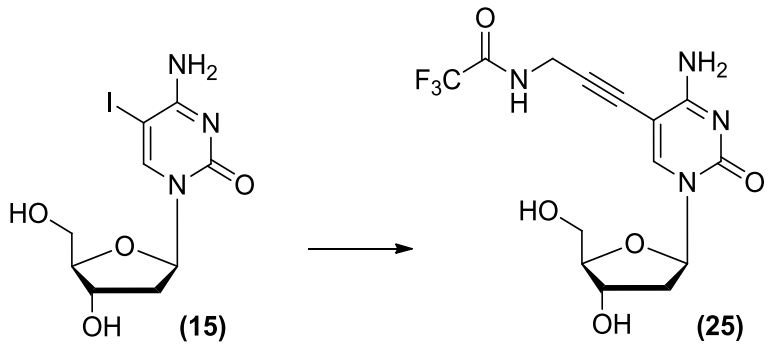
Scheme 2.14: Sonogashira cross coupling of anchor, **14**, to **23** Reagents and conditions**:** (i) Et_3N, $Pd(PPh_3)_3$, **(14)**, DMF, 0°C, absence of light (ii) overnight at RT.

Unfortunately, with the benzoyl group in place the desired Sonogashira coupling could not be achieved (scheme 2.14). This was postulated as being due to the bulky benzoyl group being in close proximity to the 5-iodo-functionality preventing its participation in the coupling.

Due to difficulties in synthesising the 3'-protected 5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine analogues (**19**) and (**24**), P(V) chemistry was investigated as an alternative due to its suitability for phosphorylation of unprotected nucleosides.

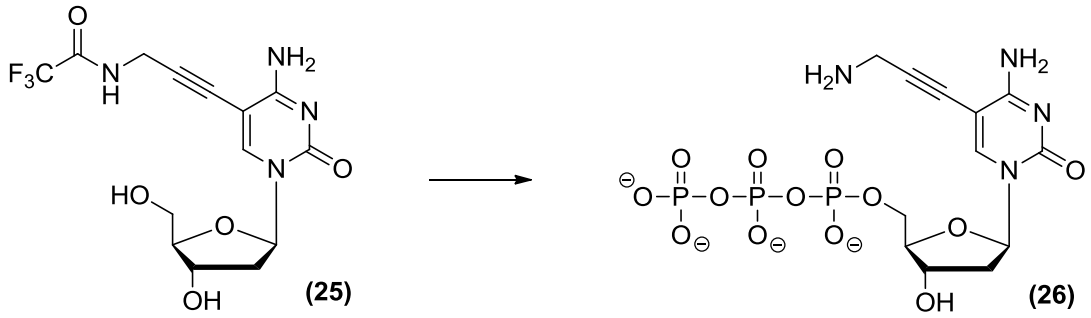
Unprotected P(V) phosphorylation

The anchor group (**14**) was coupled to 5-iodo-2'-deoxycytidine (**15**) by a Sonogashira palladium catalysed cross coupling reaction. This gave the desired 5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (**25**) in a 90% yield after purification by flash chromatography (scheme 2.15).



Scheme 2.15: Sonogashira cross coupling reaction of anchor, **14**, with 5-iodo-2'-deoxyctidine **(15).** Reagents and conditions**:** i) Et_3N, $Pd(PPh_3)_3$, (**14**), DMF, 0°C, absence of light (ii) overnight at RT, 90% yield.

5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (**25**) was then phosphorylated following Ludwig's procedure[70] without any protection chemistry (scheme 2.16).



Scheme 2.16: P(V) phosphorylation of **25**. Reagents and conditions: (i) $POCl_3$, $PO(OMe)_3$, (ii) PPi, Bu_3N, DMF, 0^oC (iii) 0.1M TEAB (iv) 35% NH_4OH, 20% yield.

As the P(V) phosphorylating agent ($POCl_3$) is less reactive than a P(III) agent, selectivity relies on phosphorylation occurring at the primary alcohol, producing the 5'-triphosphate without phosphorylating the less reactive secondary 3' alcohol. To help aid the selectivity towards the 5'-hydroxyl the reaction was run at a lowered temperature (0°C). Finally, ammonia was added at the end of the phosphorylation reaction to remove the trifluoroacetamide protecting group of the anchor.

Once the phosphorylation was complete the crude mixture was purified by medium pressure liquid chromatography (MPLC) using anionic exchange media (DEAE,

Sephadex A25) to separate the components by charge (figure 2.7). By using a TEAB salt gradient lower charge compounds elute first, enabling the mono, - di and triphosphates to be resolved (figure 2.7). This is advantageous in removing nucleoside or nucleotide side products from the crude mixture.

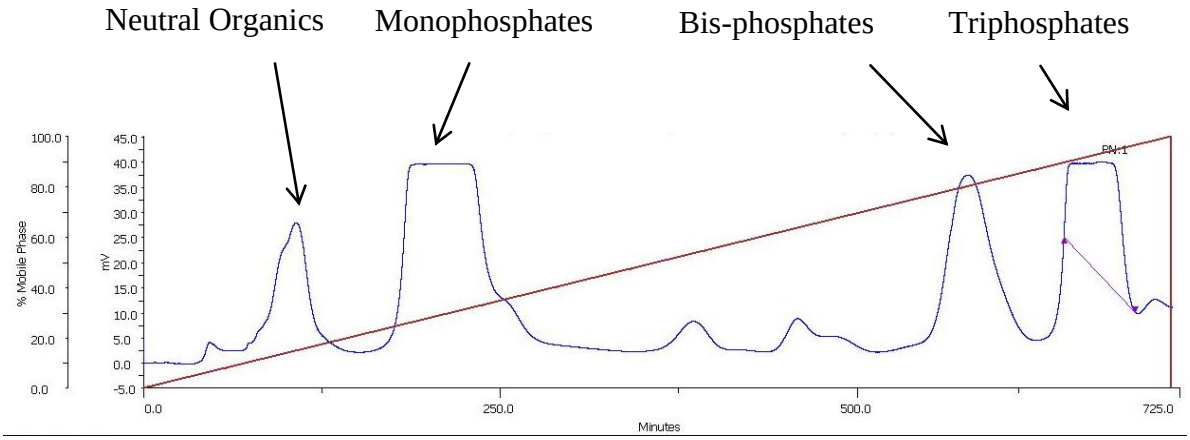


Figure 2.7: MPLC trace of the purification of (**26**) (DEAE Sephadex A25 Conditions: 50 – 900 mM TEAB, 8 mL / min, 12 h)

Due to the need for excess pyrophosphate in the reaction, which has an analogous charge to the dCTP, further purification was required to isolate pure triphosphate material. The dCTP was therefore subsequently purified by preparative reversed phase high pressure liquid chromatography (prep-RP-HPLC) using a mobile phase comprising of a gradient of acetonitrile (MeCN) in aqueous TEAB (figure 2.8).

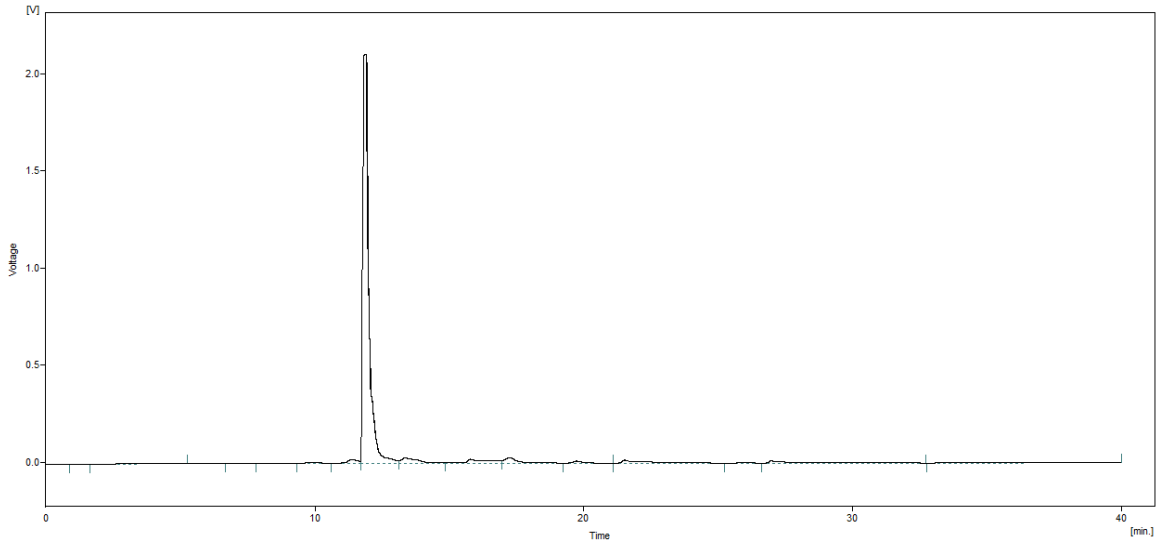


Figure 2.8: HPLC trace of (**26)** (5-65% B over 30 minutes, A= 0.1 M TEAB, B= 30 % MeCN/ 0.1 M TEAB, 1ml/min, UV detection = 295 nm).

Concerns with lack of regiospecificity of P(V) chemistry leading to low yields of the 5'-triphosphate were proved unfounded. By using 1.1 eq. of $POCl_3$, the isolated yield of purified 5'-*O*-triphosphate (**26**) was recorded at 20%. This is comparable to other published work.[54,57,81] Although the yield is lower than those regularly recorded with the use of P(III) chemistry,[74,82,83] the overall procedure is a two-step synthesis compared to the five-step synthesis required for P(III) chemistry. The yield is therefore highly competitive in addition to saving time. The triphosphate moiety can be seen by ^{31}P NMR, with the γ and α phosphorus signals being observed as doublets at ~ -10 and -12 ppm respectively and the β phosphorus as the expected triplet at ~ -23 ppm (figure 2.9).

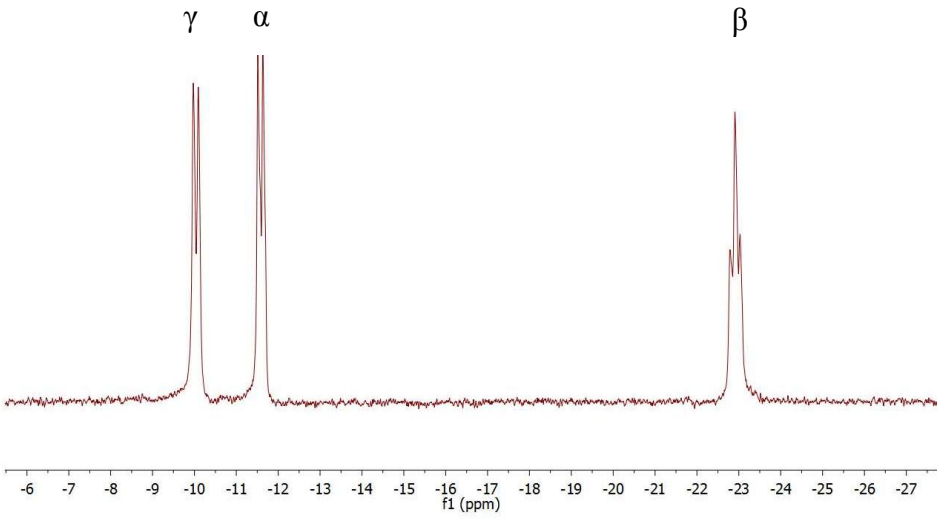


Figure 2.9: ^{31}P NMR of (**26**) purified by MPLC and HPLC.

Due to this success all methods attempting to protect 5-iodo-2'-deoxycytidine were discontinued and our focus turned to various phosphorylation methods suitable for unprotected nucleosides.

It is worth noting that although P(V) chemistry was successful in synthesising the desired dCTP (**26**), care must be taken when hydrolysing the cyclic triphosphate intermediate (**4**) to the linear triphosphate. In our initial syntheses, ammonia was introduced before complete hydrolysis (in aq. TEAB) resulting in the formation of the triphosphoramidate in a 1:4 ratio, (**27)** to (**26**) (figure 2.10).

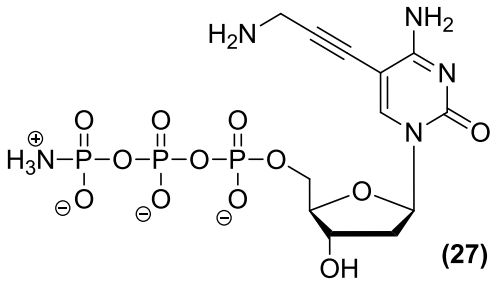


Figure 2.10: Triphosphamidate (**27**) synthesised by the addition of NH_4OH to the intermediate (**4**) prior to the complete hydrolysis**.**

The presence of the triphosphoramidate (**27**) can identified using phosphorus NMR (figure 2.11) by a shift in the γ-phosphate from ~ -10 ppm to ~ -1 ppm. Modification of

the experimental procedure through extension of the hydrolysis time from 45 minutes to three hours in all future phosphorylations resulted in the formation of the desired dCTP (**26**) only.

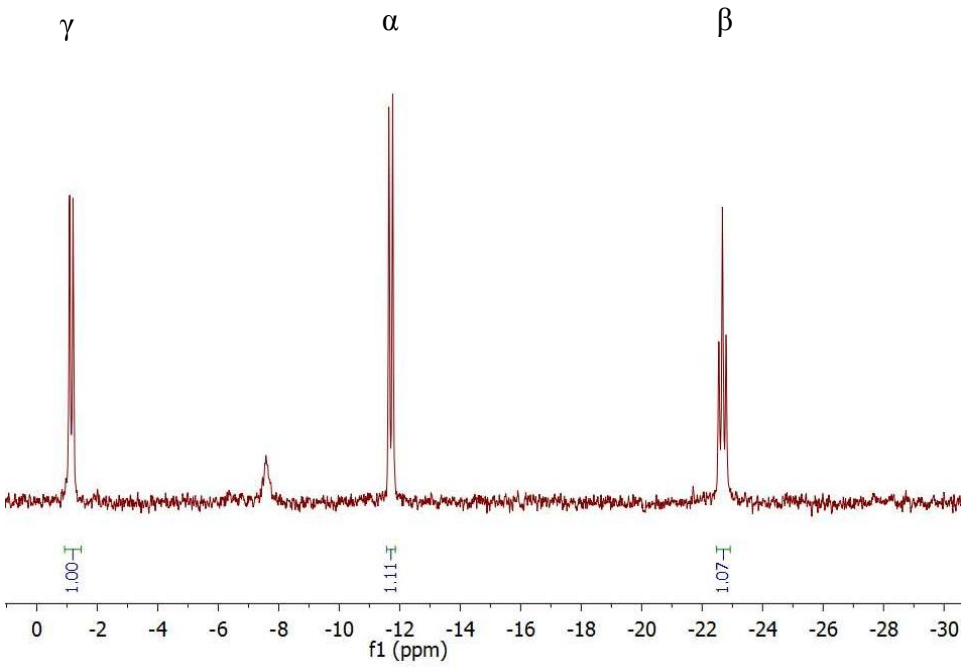


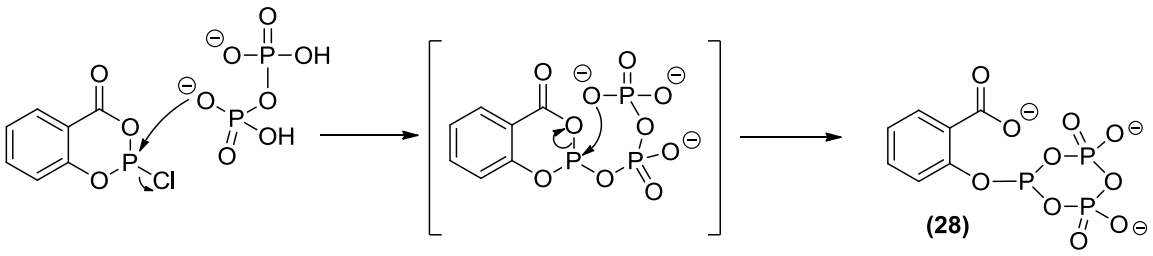
Figure 2.11: ^{31}P NMR of **27**.

P(III) phosphorylation

Following from the success of using P(V) methods on the unprotected starting nucleoside (**25**), it was decided to investigate the applicability of P(III) chemistry on the same system. Although P(III) phosphorylations are most commonly performed with 3'-hydroxyl protecting groups, due to the use of the more reactive phosphorylating agent, we wished to compare the yield of 5'-triphosphates synthesised through both P(V) and P(III) phosphorylations on unprotected nucleosides. Almost no work has been published in this area, with the notable exception of Huang *et al.*,[84, 85] however we believed that it was worthy of investigation due to the laborious and impractical process of 3'-hydroxyl protection chemistry.

Modified unprotected P(III) phosphorylation

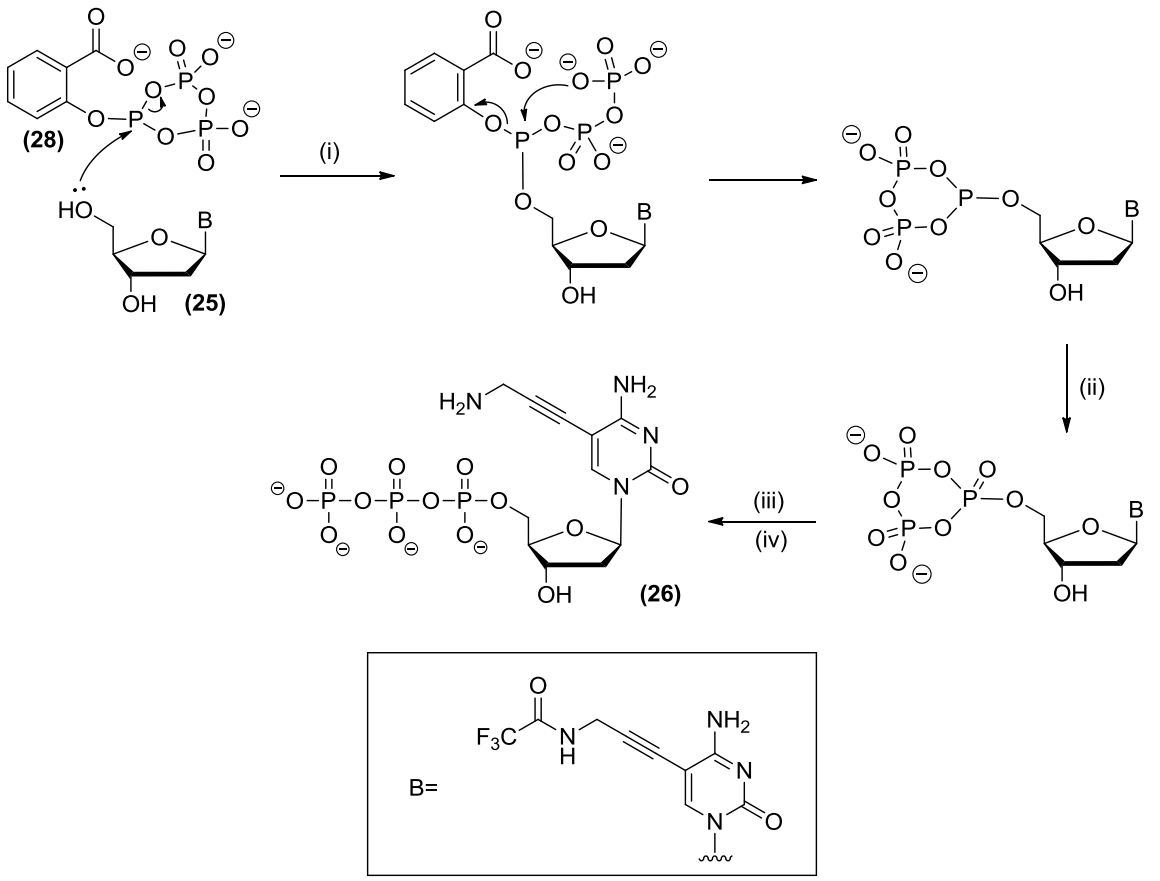
Following work published by Huang *et al.* in 2011[84, 86] we investigated the use of a new P(III) phosphorylating agent. Rather than using salicyl chlorophosphite as described by Ludwig and Eckstein[74] an alternative phosphitylating reagent (**28**) was synthesised from salicyl chlorophosphite and tributylammonium pyrophosphate, as shown in scheme 2.17.[84, 85]



Scheme 2.17: Synthesis of the P(III) phosphitylating agent (**28**)**.** Reagents and conditions: Salicyl chlorophosphite, PPi, Bu_3N, DMF, 1 h, RT. Intermediate not isolated.

Huang *et al.*[84, 86] describe this as a mild regioselective phosphitylating agent for preferentially phosphorylating all four bases at the 5'-hydroxyl group. No phosphorylation at the exocyclic amines is reported and the undesired 3'-triphosphate is isolated as the major by-product in a 5-10% yield. The regioselectivity is reported to be due to the bulky nature of (**28**) favouring the primary 5'-hydroxyl over the more hindered and less nucleophilic 3'-hydroxyl and exocyclic amines. This was therefore ideal for attempting phosphorylation on the unprotected 2'-deoxycytidine (**25**).

The phosphorylation was performed using 1.2 eq. of the phosphitylating agent (**28**) and was oxidised using iodine to give the desired triphosphate (**26**) (scheme 2.18).



Scheme 2.18: Synthesis of (**26**) following Huang *et al.* proposed mechanism for phosphorylation of 2'-deoxynucleotides using the phosphitylation reagent (**28**).[84, 86] Reagents and conditions: (i) (**28**), DMF, -13°C, 3 h (ii) 3 % I_2 Py/H_2O (1:9) (iii) 0.1 M TEAB, 3 h (iv) NH_4OH, 3 h. 18% yield.

It had been hoped that due to pyrophosphate being consumed in making the phosphorylating intermediate, MPLC purification would be sufficient for the purification of (**26**) without the need for further purification by HPLC. However ^{31}P NMR, post MPLC, showed no pyrophosphate as expected (usually seen ~ -10 ppm) but a large amount of the inorganic cyclic triphosphate by-product remained (~ -22 ppm, figure 2.12 (a)). The triphosphate was therefore purified by prep-RP-HPLC as in previous P(V) syntheses to give the pure dCTP (**26**) (figure 2.12, (b)) in a 18% yield.

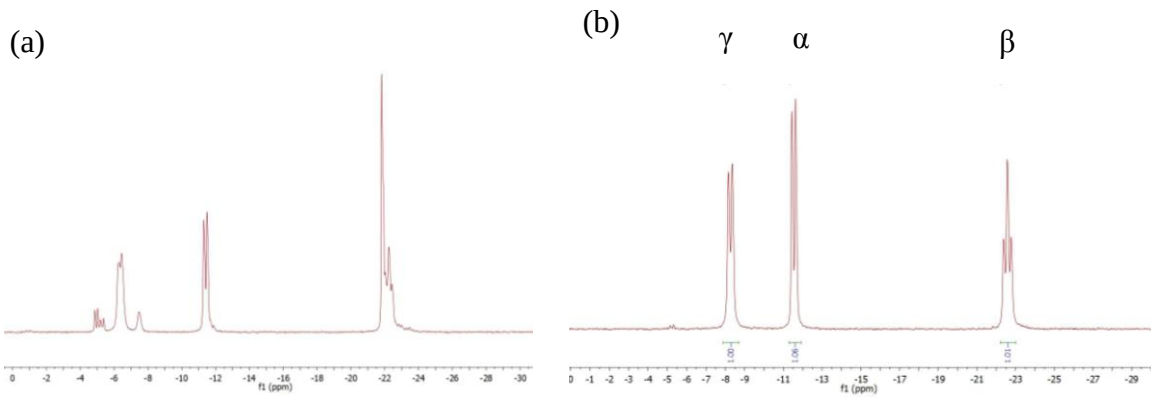


Figure 2.12: ^{31}P NMR of (**26**) after P(III) phosphorylation (a) post MPLC purification and (b) post MPLC and HPLC purification.

The isolated yield (18%) was lower than expected given the previous yield seen using P(V) methods. This was presumed to be due to ineffectiveness of the phosphitylating agent. This particular method also proved to be inconsistent upon repeat, which would provide circumstantial evidence for potential issues with the formation of the phosphorylating intermediate.

2.4 Conclusions

Although Ludwig and Eckstein's P(III) methodology[74] was our desired phosphorylation method due to the higher yield of 5'-triphosphates and lack of side products formed during the phosphorylation, the required protection chemistry upon 5-iodo-2'-deoxycytidine proved to be challenging and eventually deemed impractical. We therefore sought an alternative phosphorylation method and found P(V) chemistry[70] to be a reliable method for producing the C5-modified dCTP, (**26**). Attempts to increase the yield of (**26**) using a modified P(III) method proposed by Huang *et al.*[84, 85] resulted in disappointing yields and issues with reliability upon repeat, further supporting the use of P(V) phosphorylation techniques for our work.

Phosphorylation with a linker present (scheme 2.3) proved to be unfeasible, however having just the anchor present as in the synthesis of (**26**), does not seem to have a detrimental effect upon the phosphorylation. Therefore (**26**) is an important compound to have successfully synthesised as it provides a stock nucleotide from which a large variety of modified dCTPs can result. Importantly the required linker and consequently the reporter group, can be attached by the primary amine at the terminus of the anchor via the nucleotide (**26**).

Chapter 3 – *Introducing new functionality to 2'-deoxycytidine 5'-triphosphates*

3. Introducing new functionality to dCTPs

3.1 Introduction

To be suitable as proof-of-concept molecules in QuantuMDx's SBS device the anchor modified 2'-deoxycytidine-5'-triphosphate (**26**) requires the introduction of a linker and an anionic reporter group. When synthesising modified dCTPs it is important to ensure that the modifications present (the anchor, linker and reporter group) do not prevent the acceptance of the modified dCTPs as DNA polymerase substrates. Therefore the effect of varying linker length, introducing new anionic reporter groups and the introduction of a cleavable linker on polymerase substrate properties of the dCTPs will be investigated through the synthesis of novel C5-modified dCTPs.

As both a linker and reporter group need to be present in all modified dCTPs the general pathway to functionalising the anchor modified dCTP (**26)** needs to be determined. Two feasible routes were envisaged; coupling the linker and reporter together prior to attachment to dCTP or introducing the linker onto the dCTP first followed by the reporter group in a two-step methodology. To save time-consuming multiple purifications of dCTP analogues, it is preferential to couple the linker and reporter group together and subsequently attach the combined linker-reporter unit to the anchor modified dCTP (**26**). The chosen pathway for synthesising modified dCTPs was therefore decided upon as illustrated in figure 3.1. This route is chosen as it limits the synthetic and more importantly the purification steps taking place on the nucleotide, but also by introducing the linker and reporter group to the anchor modified dCTP avoids phosphorylation of the fully functionalised nucleoside, a procedure which has been shown to be problematic (section 2.2).

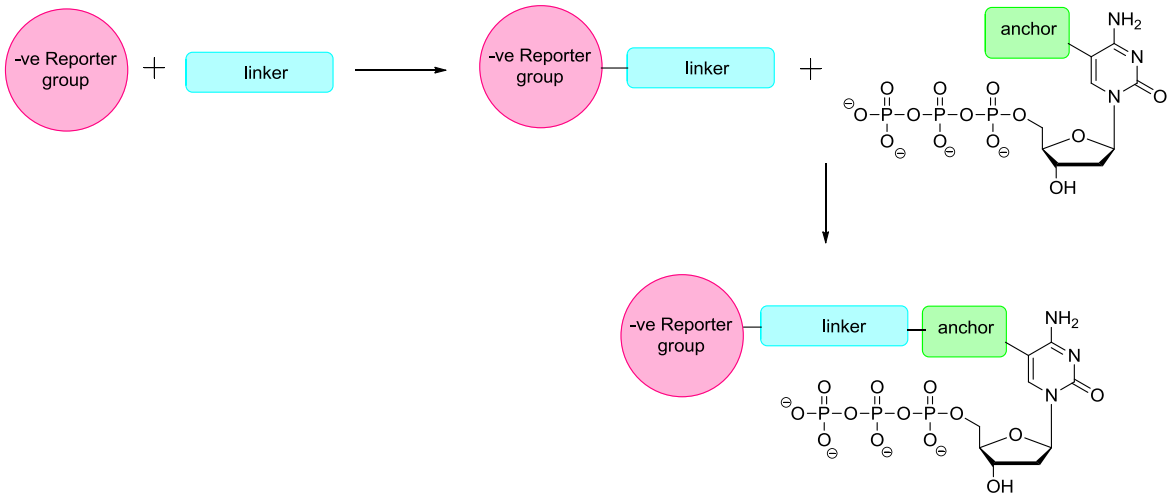


Figure 3.1: Synthetic pathway for the synthesis of modified dCTP analogues.

Section A: Investigating linker length

There are few examples in the literature of studies specifically looking at the relationship between the length of a linker and its effect on the substrate properties of the modified dNTP. A study by Waggoner *et al.*[87] investigated the incorporation of fluorescent nucleotides with varying linker lengths and describes an increase in efficiency of incorporation during PCR when the linker length was increased (figure 3.2). This finding was repeated by Quake *et al.*[88] who investigated the effect of linker length on the incorporation of fluorescently labelled nucleotides by thermophilic DNA polymerases during primer extension reactions.

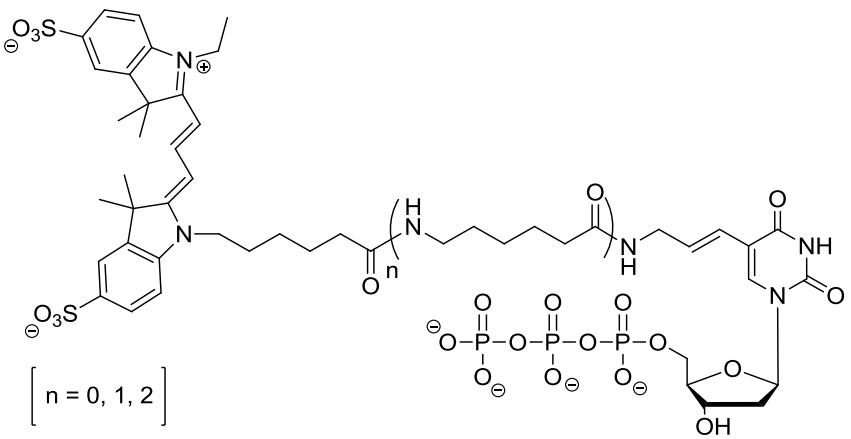


Figure 3.2: Cy3-labelled dUTP analogues with varying linker lengths as investigated by Waggoner *et al.* as PCR substrates.[87]

QuantuMDx's nanowire-based sequencing device requires modified dNTPs with linkers which must be long and flexible enough to position the reporter group within the range of detection by the nanowire. The nanowire detection range will be subject to Debye's length which dictates the distance that electrostatic effects persist in solution. As such, literature suggests that the charged reporter group will need to be in close proximity to the nanowire to be recognised (figure 3.3).[63]

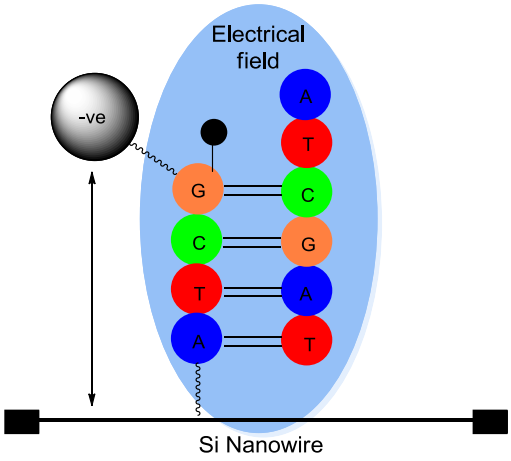


Figure 3.3: Schematic of the detection of the reporter group by a Silicon nanowire.

It is imagined that as the DNA sequencing reaction proceeds and the modified dNTPs are incorporated at increasingly greater distances from the nanowire, progressively longer linkers may be needed for the recognition of the reporter group to occur. Therefore the development of a method of synthesising increasingly longer linkers from a shorter basic linker was of particular interest. By varying the linker length we will also determine how far the bulkiest part of the modification (the reporter group) must be held from the nucleobase for the modified dCTP to be a DNA polymerase substrate. Consequentially we planned to synthesise C5-modified dCTPs with the trimesic acid reporter group supported by PEG linkers of varying lengths. The effect of linker length on the substrate properties of the modified dCTPs is to be investigated in chapter 4.

3.2 Synthesis of $(PEG)_8$ modified 2'-deoxycytidine-5'-triphosphate

The first proof-of-concept compound to be synthesised used the commercially available carboxy-amine polyethylene glycol ($CA(PEG)_8$) linker (**29**). $CA(PEG)_8$ is a bifunctional PEG unit which has both carboxyl and amine functionality hence the prefix 'CA'. The length of $CA(PEG)_8$ (approx. 3.6 nm) is equivalent to 12 base pairs and the PEG backbone provides flexibility to the linker.

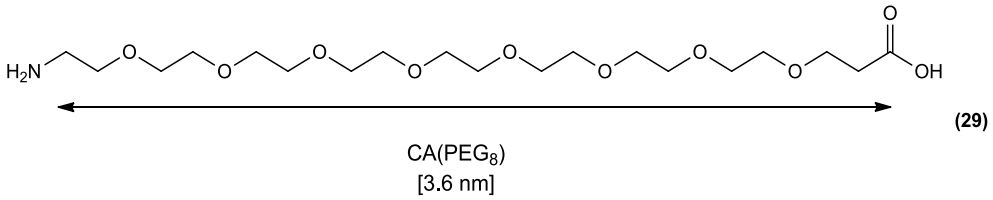


Figure 3.4: $CA(PEG_8)$ linker length.[89]

Due to its commercial availability as the diethyl protected carboxylate (**30**) (figure 3.5) trimesic acid was chosen as the first reporter group.

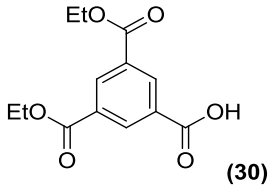
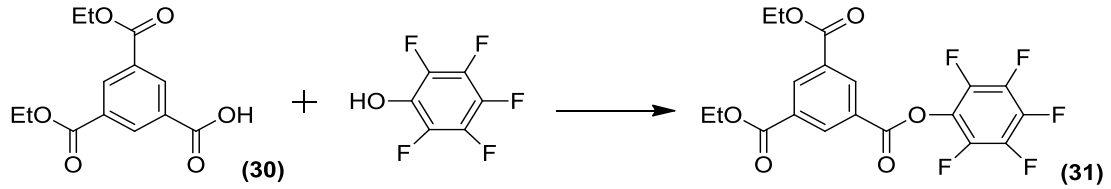


Figure 3.5: Protected trimesic acid reporter group - diethyl 1,3,5 benzenetricarboxylate (**30**).

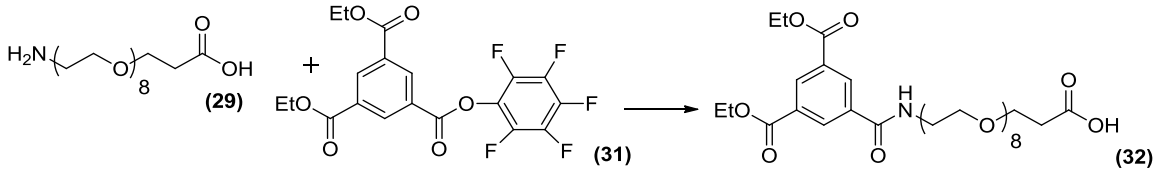
By purchasing the pre-protected trimesic acid the remaining unprotected carboxylate can be coupled to the $CA(PEG)_8$ amine without the need for any prior protection. Due to the cost of (**29**) it was imperative that the coupling reaction proceeded in a high yield to avoid potential loss of the $CA(PEG)_8$ material. Therefore the carboxylate of the reporter group (**30**) was first converted to an activated-ester to increase its reactivity towards the $CA(PEG)_8$ amine. Although many reagents exists for the synthesis of activated esters for amide synthesis (HATU, HBTU, NHS *etc*.) a stable activated ester which could be

isolated was favoured. For this purpose the diethyl 1,3,5 benzenetricarboxylate was converted to the activated PFP-ester (**31**) with the use of pentafluorophenol (PFP) and the coupling agent 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) in dry DMF. The desired product (**31**) was purified by flash chromatography and obtained in 61% yield (scheme 3.1).



Scheme 3.1: Pentafluorophenol activation of diethyl protected trimesic acid (**30**). Reagents and conditions: Pentafluorophenol, EDC, DMF, 1.5 h, 61%.

The linker, CA(PEG_8) (**29**), can then be reacted with the PFP-ester of diethyl 1,3,5 benzenetricarboxylate (**31**) to form the linker-reporter unit (**32**) as described in figure 3.1. This reaction was performed by another group member in anhydrous methanol (MeOH) and the product (**32**) purified prep-RP-HPLC.



Scheme 3.2: Coupling of PFP-activated diethyl-trimesic acid (**31**) to CA(PEG_8) (**29**).

The next synthetic step required is the coupling of the combined linker-reporter unit (**32**) to the anchor-modified dCTP (**26**). *O*-(*N*-succinimidyl)-*N,N,N′,N′*-tetramethyluronium tetrafluoroborate (TSTU, figure 3.6) was chosen as the coupling agent as it is stable under aqueous conditions; this makes it an appropriate choice for use with triphosphates which are commonly stored in aqueous conditions.

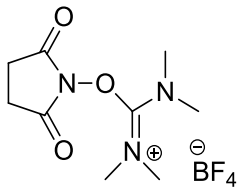
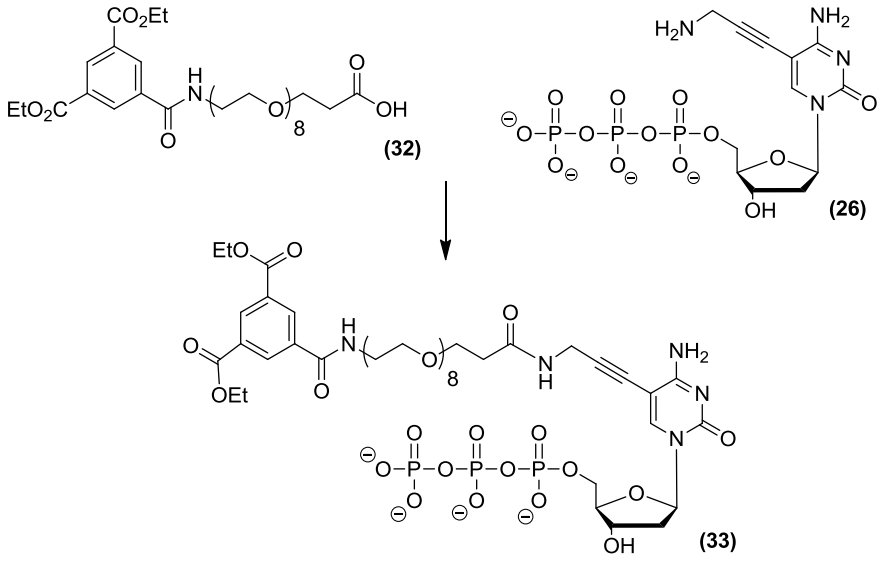


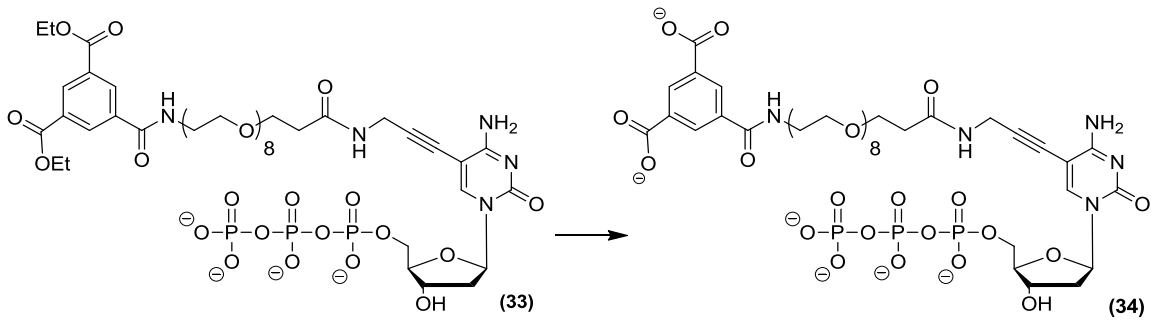
Figure 3.6: Peptide coupling reagent, TSTU.

The coupling reaction produced the precursor (**33**) for the first proof-of-concept compound with a suitable anchor, linker and protected reporter group attached to the C5 position of dCTP. The reaction was analysed by RP-HPLC and the main product (**33**) subsequently isolated by prep-RP-HPLC in a 57% yield (scheme 3.3)



Scheme 3.3: TSTU coupling of reporter-linker unit **32** to anchor modified dCTP (26). Reagents and conditions: (i) **32**, TSTU, DIPEA, DMF, 2 h (ii) **26**, 0.1 M sodium borate buffer, 24 h, 57% yield.

The final synthetic step performed was hydrolysis of the ethyl esters with 1 M aqueous sodium hydroxide solution (NaOH) at room temperature for 2 hours. This generates the dicarboxylate dCTP (**34**) (scheme 3.4) and provides a suitable dCTP analogue for testing as a DNA polymerase substrate.



Scheme 3.4: NaOH deprotection of (**33)**. Reagents and conditions: 1 M NaOH, 0°C, 2 h, 82% yield.

The triphosphate (**34**) was analysed by RP-HPLC (figure 3.7) and subsequently purified by prep-RP-HPLC with an 82% yield for the deprotection and subsequent conversion to Na^+ salt by ion exchange resin (dowex Na^+ form).

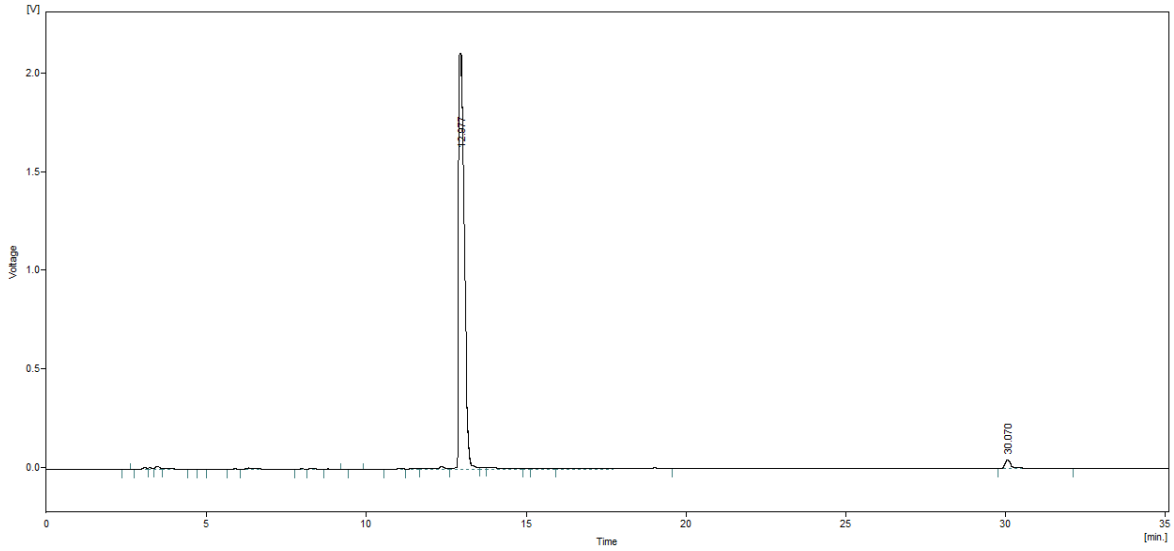


Figure 3.7: RP-HPLC trace of (**34**). Conditions: 5-95 % B over 45 minutes where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min. Retention time of product = 13.0 mins.

3.3 Synthesis of a $(PEG)_4$ modified 2'-deoxycytidine-5'-triphosphate

As long linkers such as $CA(PEG)_8$ are typically expensive to purchase, the synthesis of a shorter linker which could be used as a linker in its own right or coupled together to form longer linkers was explored. For example, $CA(PEG)_8$ is an expensive reagent with 1g retailing at over £700, whilst tetraethylene glycol (PEG_4) is readily available for the significantly reduced cost of 3p per gram. The functionalisation of shorter more affordable PEG units is therefore highly desirable. The decision to remain focussed on a PEG linker was based on their known stability and biocompatibility.[90]

Following the developed procedures for $CA(PEG)_8$ (section 3.2) it was consequently decided to synthesise the equivalent linker, carboxy-amine PEG_4 ($CA(PEG)_4$), from tetraethylene glycol and to attach this to the diethyl trimesic acid reporter group (**30**).

The proposed $CA(PEG_4)$ linker (**38**) has an approximate length of 1.7 nm which is equivalent to 6 base pairs (figure 3.8).

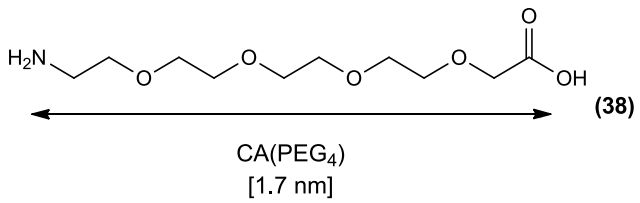
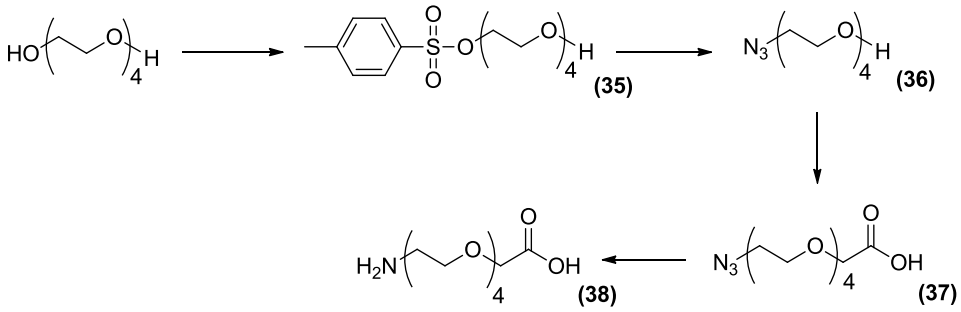


Figure 3.8: Carboxy-amine tetraethylene glycol linker, $CA(PEG)_4$.[89]

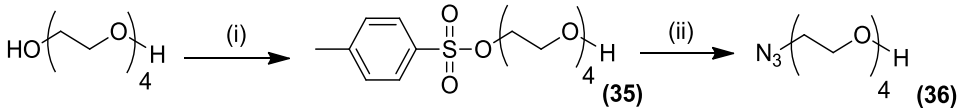
$CA(PEG)_4$ linker synthesis

The synthetic route to the $CA(PEG)_4$ linker (**38**) is shown in scheme 3.5. Starting from the mono-functional PEG_4 starting material it is a four step route to the desired linker.



Scheme 3.5: Synthetic overview to the synthesis of $CA(PEG)_{4.}$

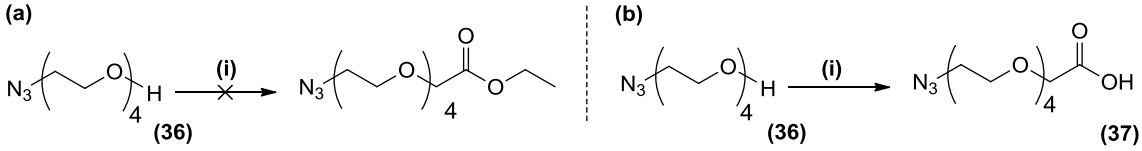
The first synthetic step is the mono-tosolyation of PEG_4. To avoid potential problems with *bis*-tosylation an excess of tetraethylene glycol was used; at 10-20 mol. % tosyl chloride no *bis*-tosylated product was observed. The tetraethylene glycol monotosylate product (**35**) (scheme 3.6) was isolated by solvent extraction in good yield and high purity. The tosylate moiety can then be displaced using sodium azide to give the mono-azide functionalised PEG_4 (**36**) (scheme 3.6). This was again isolated by solvent extraction and required no further purification. The yield over two steps was 73%. However it is noteworthy that this yield is based upon the use of tosyl chloride as a limiting reagent therefore a large proportion of the starting material was not functionalised and therefore lost. Given the low cost of PEG_4 and the simple extraction techniques used to remove unreacted starting material this was not of significant concern in this instance.



Scheme 3.6: Synthesis of tetraethylene glycol monotosylate (**35**) and subsequent displacement with NaN_3 to give mono-azide tetraethylene glycol (**36**). Reagents and conditions: (i) Et_3N, 10% mol. 4-toluenesulfonyl chloride in DCM, overnight, 92%, (ii) NaN_3, EtOH, 70 °C, 20 h, 79%.

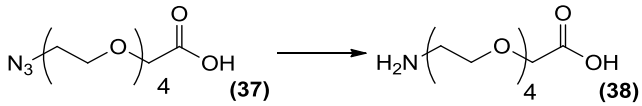
In the first instance ethyl bromoacetate was used to introduce the carboxylate function onto the monoazide-PEG_4 (**36**) (scheme 3.7 (a)) However, yields were low and as multiple by-products were seen by thin layer chromatography (TLC) flash chromatography purification was required. This proved to be difficult as the compound

was not UV active and streaked heavily on silica, unfortunately no pure material was isolated after flash chromatography. Due to the scale of synthesis purification by HPLC was not a viable option. An alternative method of introducing the carboxylate functionality to the PEG linker was therefore required. Using bromoacetic acid in place of ethyl bromoacetate proved to be a much cleaner reaction with only one compound, the desired (**37**), being seen (scheme 3.7 (b)). By using bromoacetic acid there is also the advantage of eliminating the need for subsequent hydrolysis of the ester, thus removing a synthetic step. The carboxy-azide PEG_4 linker ($CAz(PEG)_4$) (**37**) was then isolated by solvent extraction and did not require flash chromatography purification.



Scheme 3.7: (a) Attempted esterification of 36 using ethyl bromoacetate (b) Successful carboxylation of 36 using bromoacetic acid. Reagents and conditions: (a) NaH, ethyl bromoacetate, THF, 0°C; (b) NaH, Bromoacetic acid, THF, 0°C, overnight at RT, 88%.

After carboxylation had been successfully achieved and the purity of (**37**) confirmed by ^{1}H, ^{13}C NMR and liquid chromatography – mass spectrometry (LC-MS), the final synthetic step was to reduce the azide to the amine (scheme 3.8). The reduction of (**37**) was achieved cleanly by hydrogenation using a ThalesNano H-Cube as a source of hydrogen with a 10% palladium on carbon catalyst contained within a cartridge (CatCart®).

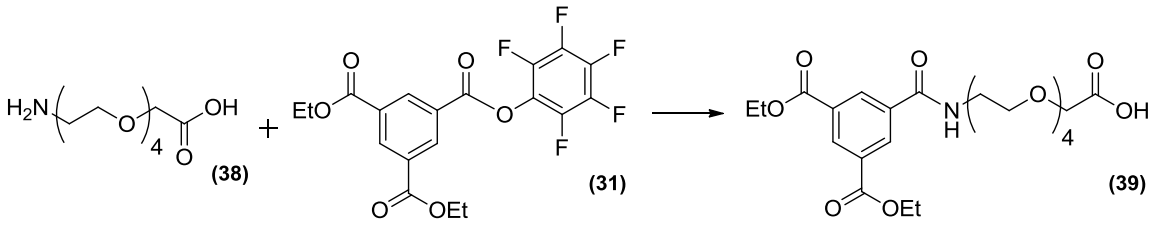


Scheme 3.8: Hydrogenation of $CAz(PEG)_4$ (**37**) to $CA(PEG)_4$ (**38**) using a ThalesNano H-cube. Reagents and conditions: H_2, 10% Pd on C, MeOH, 50°C, 30 bar, 90%.

The hydrogenation was initially attempted at room temperature under atmospheric pressure however the reaction was incomplete after one pass over the CatCart. By increasing the temperature to 50°C and reducing the pressure to 30 bar subsequent attempts to hydrogenate the azide were successful after one pass over the catalyst

(0.14 M solution, 1 mL/min). The desired product (**38**) was then isolated in a 90% yield by evaporating the solution to dryness. The light yellow oil obtained was characterised by ^{1}H, ^{13}C NMR and LC-MS and no further purification was required.

Once the $CA(PEG)_4$ linker (38) had been successfully synthesised it was coupled with the PFP-ester (**31**) (scheme 3.9).

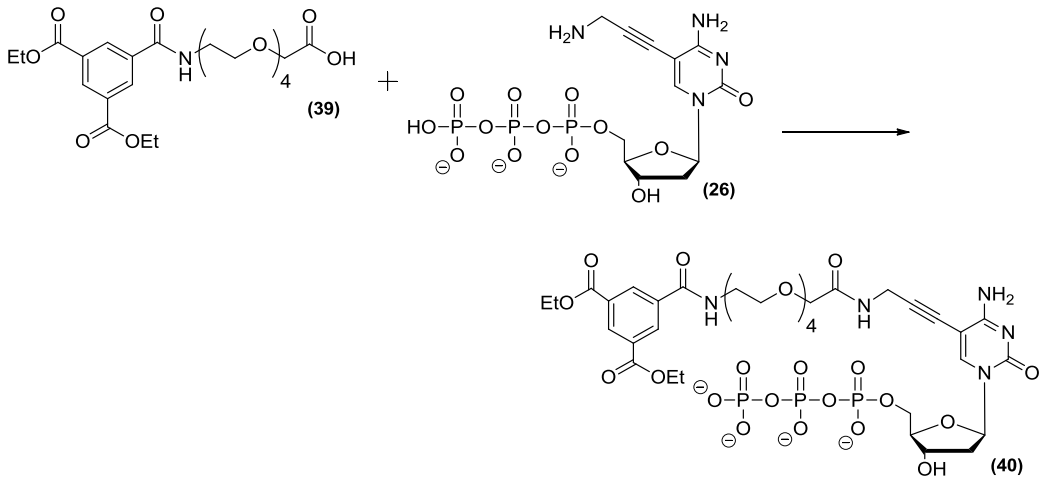


Scheme 3.9: Coupling of PFP-activated trimesic acid reporter group (**31**) to $CA(PEG)_4$ linker (**38**). Reagents and conditions: DMF, DIPEA, overnight, 22%.

The attachment of the reporter group to $CA(PEG)_4$ was initially attempted with MeOH in the absence of a base, however, problems with solubility hindered the reaction. Changing the solvent to DMF and adding 1 eq. of *N*,*N*-diisopropylethylamine (DIPEA) increased the reaction yield, unfortunately the yield was still relatively low (22%). Alternative routes such as making the *N*-hydroxysuccinimide (NHS) ester of the 1,3-diethyl-trimesic acid (**30**) could be investigated to increase the yield of coupling if required.

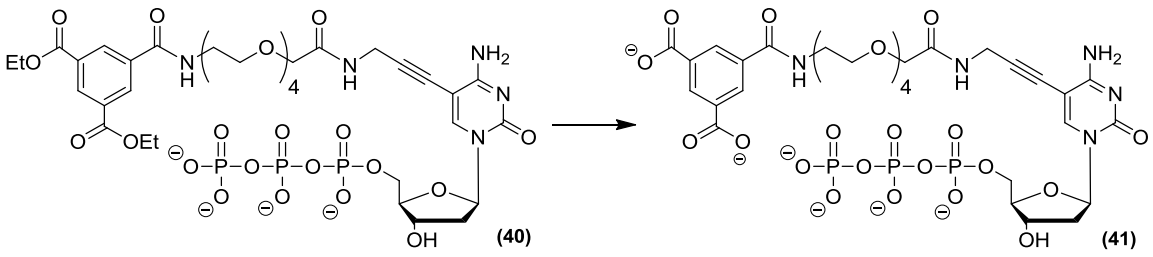
Attachment of PEG_4 linker-reporter unit to a modified dCTP

The PEG_4 linker-reporter unit (**39**) was then coupled with the anchor modified-dCTP (**26**) via a TSTU coupling. Purification was performed by prep-RP-HPLC to give (**40**) in a 74% yield (scheme 3.10).



Scheme 3.10: TSTU coupling of linker-reporter unit (**39**) to C5-propargylamino-dCTP (**26**). Reagents and conditions: (i) (**39**), TSTU, DIPEA, DMF, 2 h, (ii) (**26**), 0.1 M sodium borate buffer, 0°C, 24 h at RT, 74%.

Ester hydrolysis using 1 M NaOH gave the PEG_4 modified-dCTP analogue (**41**) in 82% yield following prep-RP-HPLC (scheme 3.11). By using the same reporter group, (**41**) and (**34**) vary only by linker length and therefore can be used to examine the effect of linker length on their substrate properties.



Scheme 3.11: NaOH deprotection of (**40**). Reagents and conditions: 1 M NaOH, 0°C, 2 h, 82%.

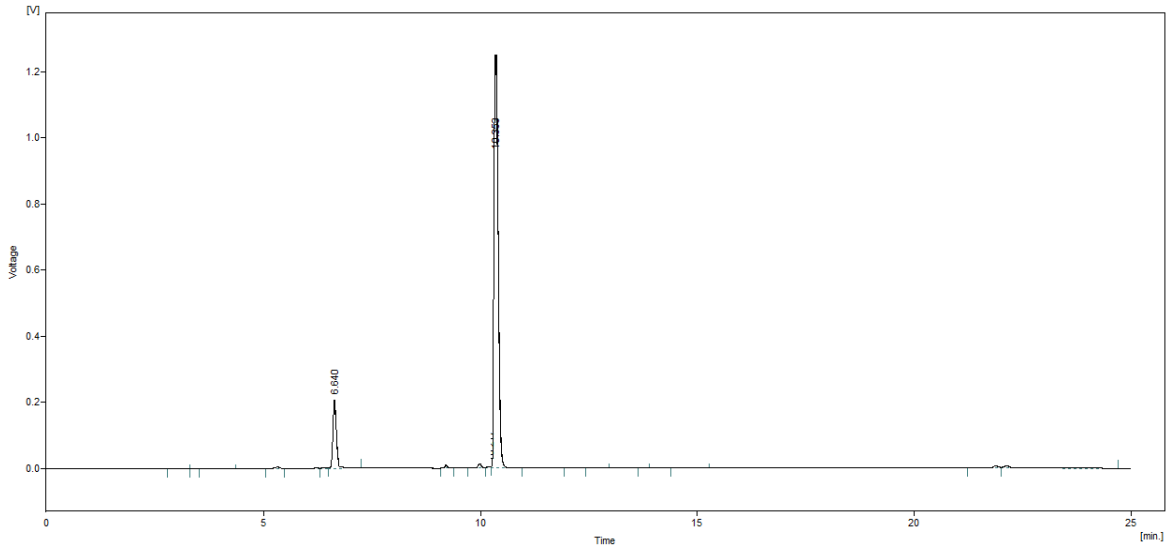
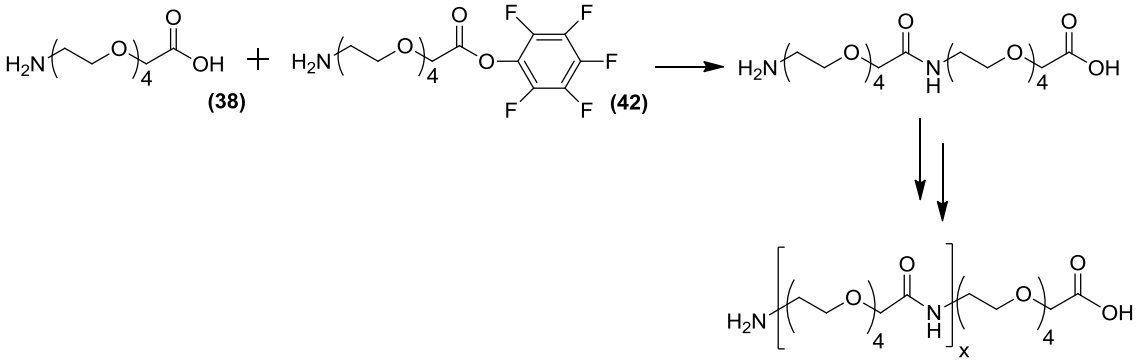


Figure 3.9: Analytical HPLC trace of crude (**41**). Conditions: 5 – 95% B over 30 minutes where A= 0.1 M TEAB and B= 50% MeCN/ 0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min, Retention time of product = 10.3 minutes.

3.4 A route to longer linkers from short, PEG_4 linkers

The use of the $CA(PEG)_4$ linker (**38**) as a monomer to allow a variety of different length linkers to be synthesised by coupling units together to form double, triple, quadruple etc. length linkers was envisaged. This would avoid the need for longer, expensive linkers and gives a high degree of control in producing variable-length linkers.

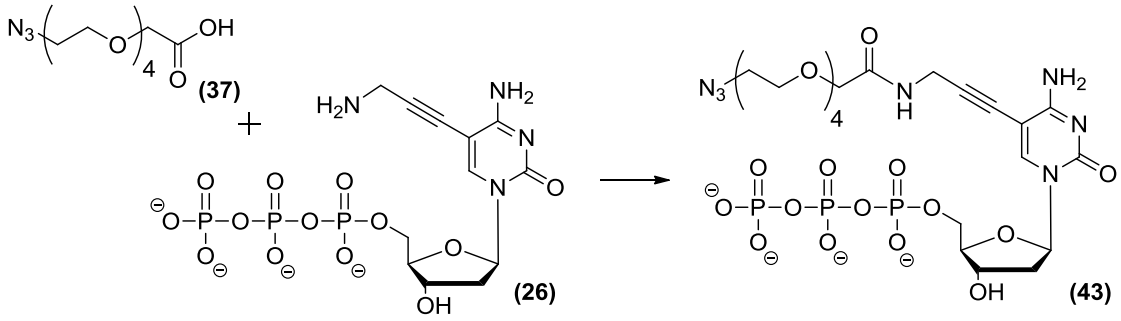
Unfortunately the PFP-activation of $CA(PEG)_4$ linker and its purification (attempted by a MChem student in the group) proved challenging and although possible, the overall yield of the product made the route unfeasible as a strategy towards longer linkers.



Scheme 3.12: Potential synthetic route to synthesising long linkers of the form $CA(PEG_4)_x$ from the $CA(PEG)_4$ linker (**38**) and the related activated PFP-ester (**42**).

We therefore attempted an alternative route based on coupling the linkers together *in situ* on the triphosphate. This was achieved by coupling the $CA(PEG)_4$ precursor $CAz(PEG_4)$ (**37**) to the anchor modified dCTP (**26**). The $CAz(PEG)_4$ linker must be used rather than the $CA(PEG)_4$ unit to avoid polymerisation of the $CA(PEG)_4$ monomers together during the activation with TSTU.

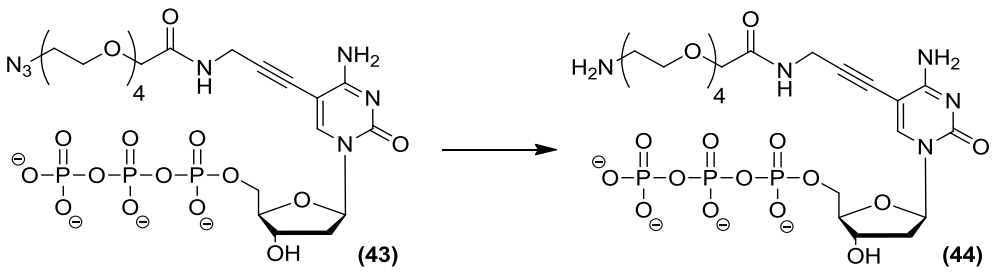
The first addition of $CAz(PEG_4)$ (**37**) to C5-propargylamino-dCTP (**26**) was accomplished in a 78% yield after prep-RP-HPLC purification (scheme 3.13).



Scheme 3.13: TSTU coupling of $CAz(PEG)_4$ to the anchor modified dCTP (**26**). Reagents and conditions: (i) (**37**), TSTU, DIPEA, DMF, 3h, (ii) (**26**), 0.1 M sodium borate buffer, 0°C, 24 h at RT, 78%.

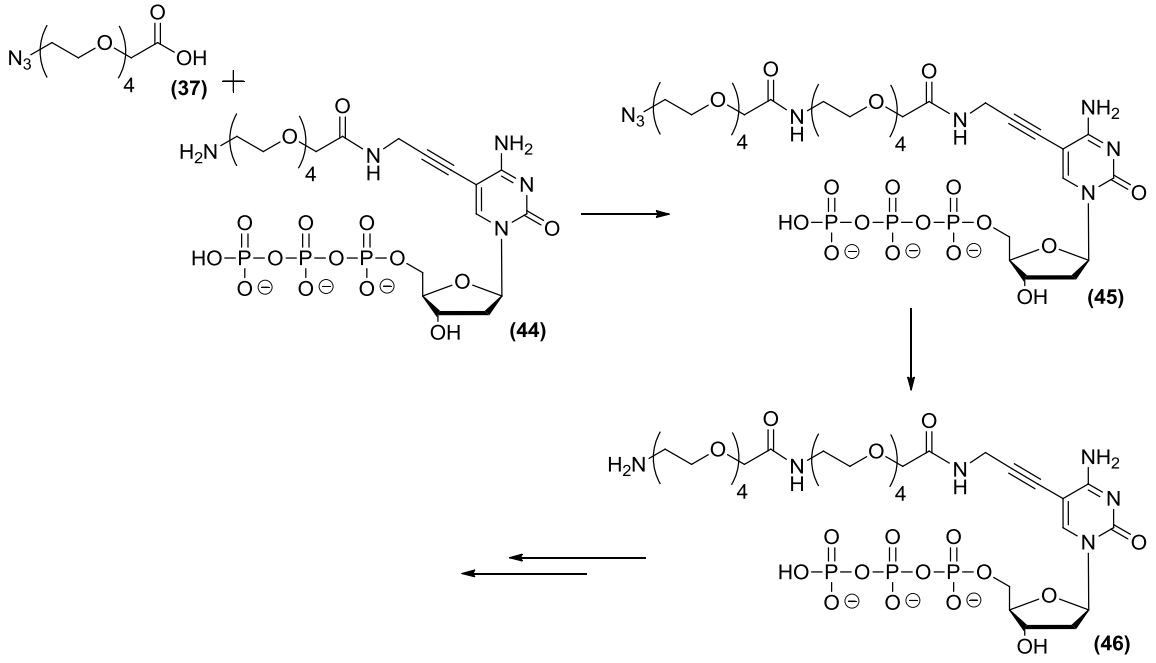
The azido moiety can then be reduced via the Staudinger reaction to give the corresponding amine. Hydrogenation cannot be used in this case due the unsaturated alkyne present. Accordingly (**43**) was reduced to (**44**) using triphenyl phosphine (PPh_3)

in pyridine to afford (**44**) in 59% yield after purification by ion-exhange MPLC (scheme 3.14).



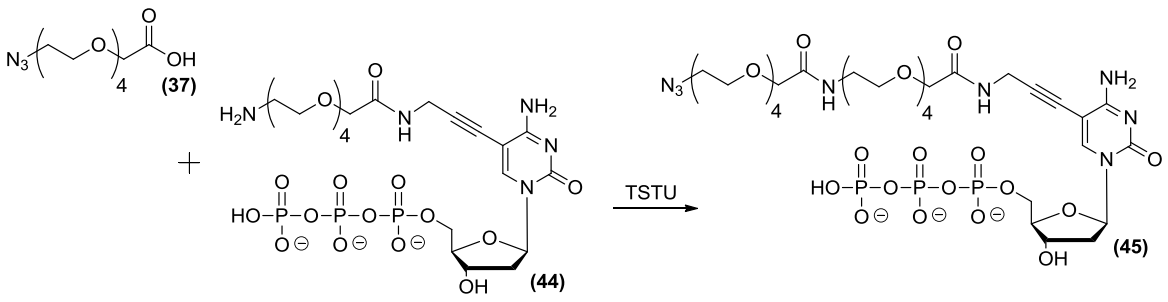
Scheme 3.14: Reduction of (**43**) to (**44**). Reagents and conditions: (i) PPh_3, Pyridine, 7 h (ii) 5% NH_3 solution, RT, overnight, 59%.

In principle the TSTU coupling/reduction process can be repeated as required to gain double, triple etc. linker length modified dCTPs (Scheme 3.15).



Scheme 3.15: Proposed TSTU coupling/reduction cycle to longer linkers. Demonstrated by the coupling of an additional $CAz(PEG)_4$ linker (**37**) to (**44**) and subsequent reduction to form the amino-terminated (**46**). The process could then be repeated, extending the linker in PEG_4 repeating units.

The feasibility of this route was demonstrated by attaching a further CAz(PEG)$_4$ linker to **44** as shown in scheme 3.16.



Scheme 3.16: TSTU coupling of an additional CAz(PEG)$_4$ linker (**37**) to (**44**).
Reagents and conditions: (i) (**37**), TSTU, DIPEA, DMF, 3 h, (ii) (**44**), 0.1 M sodium borate buffer, 0°C, 24 h at RT, 65%.

The TSTU coupling proceeded in a good yield (65%) with (**45**) being purified by prep-RP-HPLC. However, given the relatively high purity of the crude compound (figure 3.10) it was pleasantly discovered that HPLC purification does not appear necessary and therefore introducing each extra linker would require only one purification step. However as the azido terminated dCTP (**45**) was sought for another area of interest (section 3.6) it was not reduced for further extension (scheme 3.15).

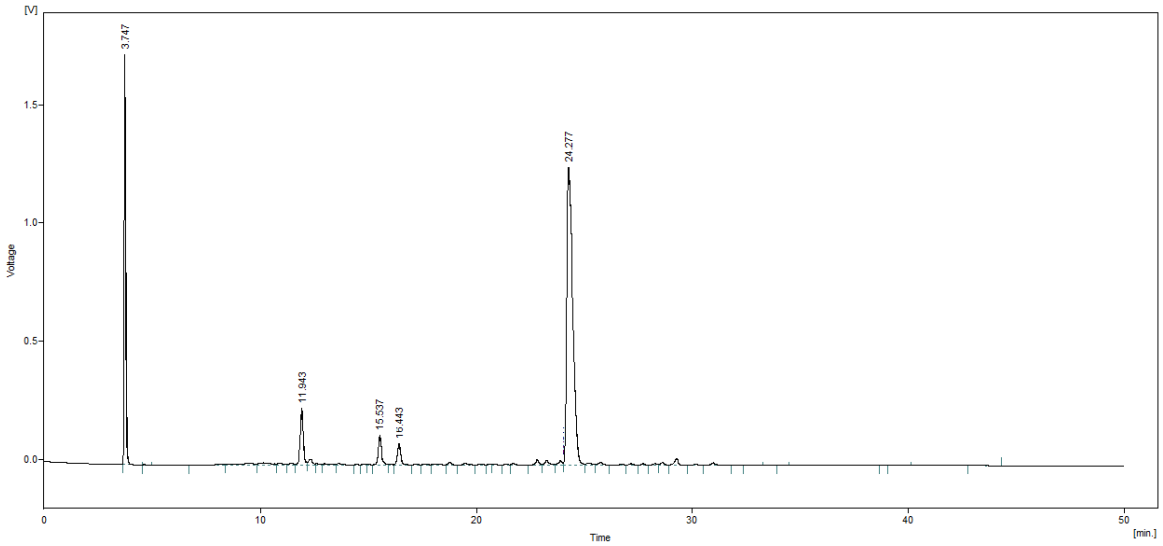


Figure 3.10: Analytical HPLC trace of crude (**45**). Conditions 5-100% B over 40 minutes where A= 0.1 M TEAB and B= 50% MeCN / 0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min, Retention time of product = 24.2 minutes.

Section B: Reporter groups

The SBS sequencing device in development by QuantuMDx will use modified dNTPs bearing negatively charged reporter groups (section 1.9). There are however relatively few examples of base-modified dNTPs functionalised with negatively charged groups that have been investigated as substrates for DNA polymerases: to our knowledge, there are two such examples of modified dCTPs as DNA substrates.[59, 61]

Famulok[57](**1**) and Marx[58](**2**) have reported dUTP analogues with carboxylic acid side chains at the C5 position to be substrates for vent (exo-) and *pwo* DNA polymerases respectively. A dCTP analogue with a C5-carboxylate modification (5-valeric acid-dCTP, (**49**)) has been reported by Hollenstein[59] and is substrate for vent (exo-), *Pwo* and 9°N$_m$ DNA polymerase (figure 3.11).

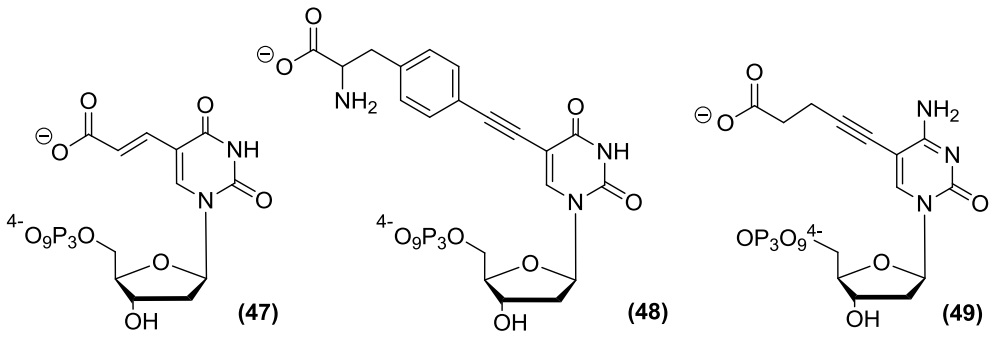


Figure 3.11: Negatively charged C5 modified dUTPs (**47** and **48**) found to be substrates for two DNA polymerases, ((vent (exo-)*)* and (*pwo*)) and 5-valeric acid-dCTP (**49**) found to be a substrate for vent (exo-), *Pwo* and 9°N$_m$ DNA polymerases.[57–59]

The previously highlighted (section 1.7) α-sulpho-β-alanine-modified ddCTP for use in adapted Sanger DNA sequencing has been shown to be a substrate for Thermo Sequenase™ II and a modified *taq* DNA polymerase by Kumar *et al.* (figure 3.12).[61] Interestingly they also discussed a correlation between increased negative charge upon the modified ddNTPs and decreased incorporation efficiency but state the effect can be diminished by extending the linker arm.

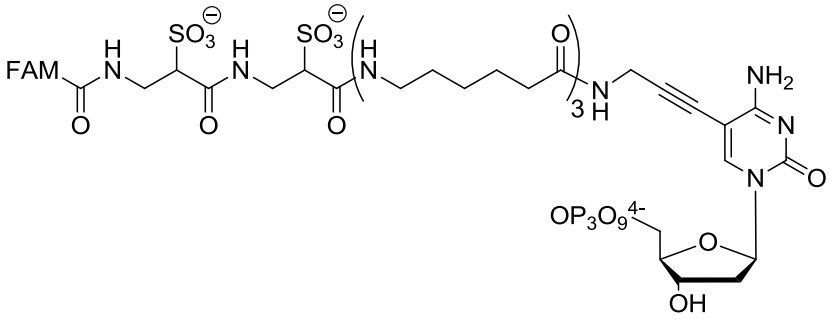
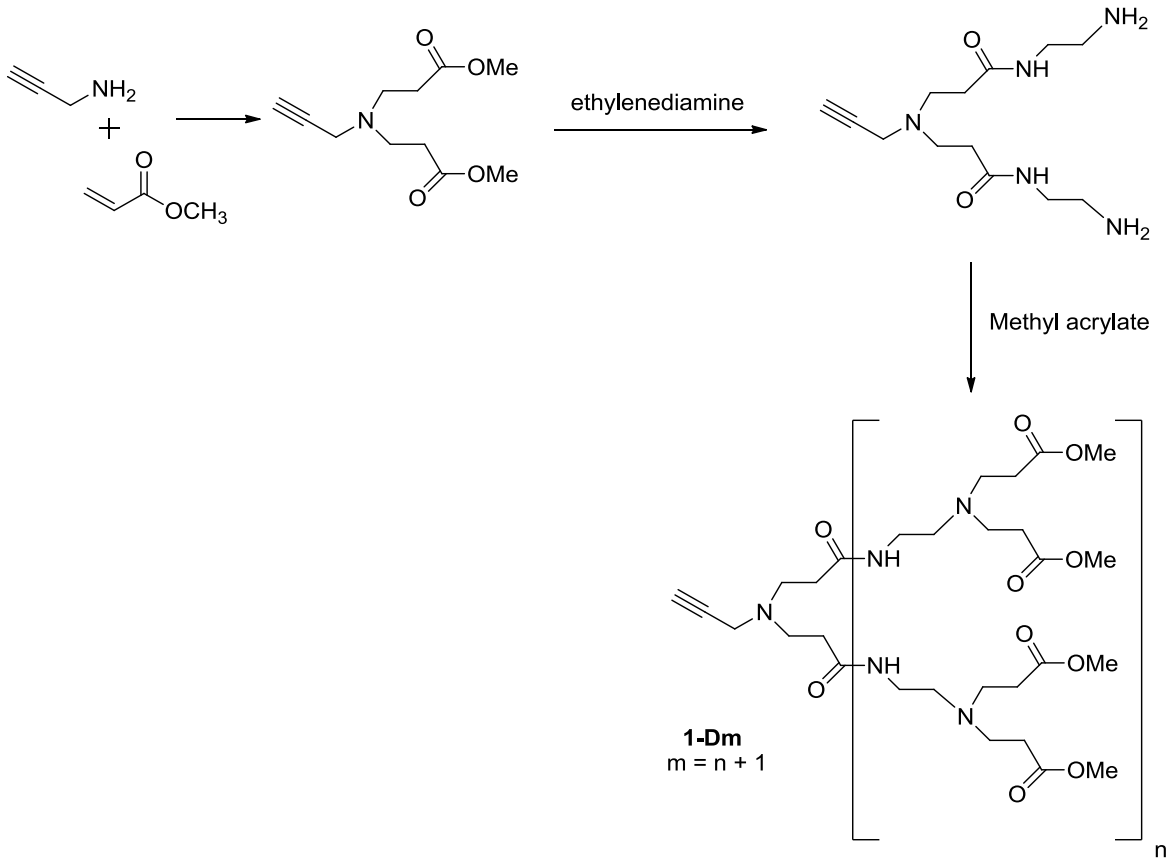


Figure 3.12: An example α-sulpho-β-alanine-modified ddCTP as synthesised by Kumar *et al.*[61] which was found to be a substrate for Thermo Sequenase™ II and a modified *taq* DNA polymerase. (Where FAM = 5-carboxyfluorescein, a fluorescent dye).

The synthesis of two dCTP analogues bearing an anionic trimesic acid reporter group has already been discussed (section A) and their substrate properties will be investigated in Chapter 4. Additionally the trimesic acid reporter group has also been attached to the remaining dNTPs (dATP, dGTP and dUTP) by other research members of QuantuMDx and all have been found to be substrates for at least one DNA polymerase; however the synthesis of these analogues will not be discussed here.

3.5 PAMAM dendrimers as reporter groups

After the successful synthesis of modified dCTPs (**34**) and (**41**) the synthesis of dCTPs functionalised with highly charged reporter groups was undertaken. Polyamido amine (PAMAM) dendrimers are a family of dendrimers which have a propargylamine core unit which is extended through a cycle of reactions with methyl acrylate and ethylenediamine into a branched dendrimer.[91] Each methyl acrylate and ethylenediamine addition extends the dendrimer into a new class; D1, D2, D3 etc. (scheme 3.17). If terminated after the addition of methyl acrylate a dendrimer with multiple methyl esters is synthesised. The methyl esters can be subsequently hydrolysed to the charged carboxylates to give a polyanionic reporter group. By this technique we envisaged the synthesis of D2 and D3 dendrimers as reporter groups which could be attached through their terminal alkyne to the $CAz(PEG)_4$ linker or the azido-functionalised dCTP (**45**) using click chemistry.

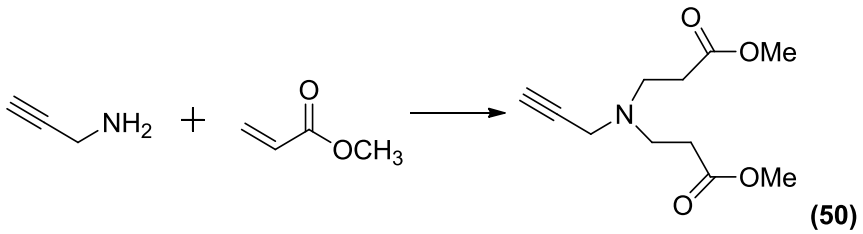


Scheme 3.17: Synthetic overview of the synthesis of PAMAM dendrimers.

PAMAM dendrimers offer the advantage that they can be synthesised relatively simply and the synthesis adjusted to access reporter groups with varying anionic charges. So far only the 1-D1 to 1-D3 dendrimers have been synthesised (**50, 52, 53**) however the ethylenediamine/methyl acrylate cycle could be continued to form larger and more anionic dendrimers if required in the future (scheme 3.17).

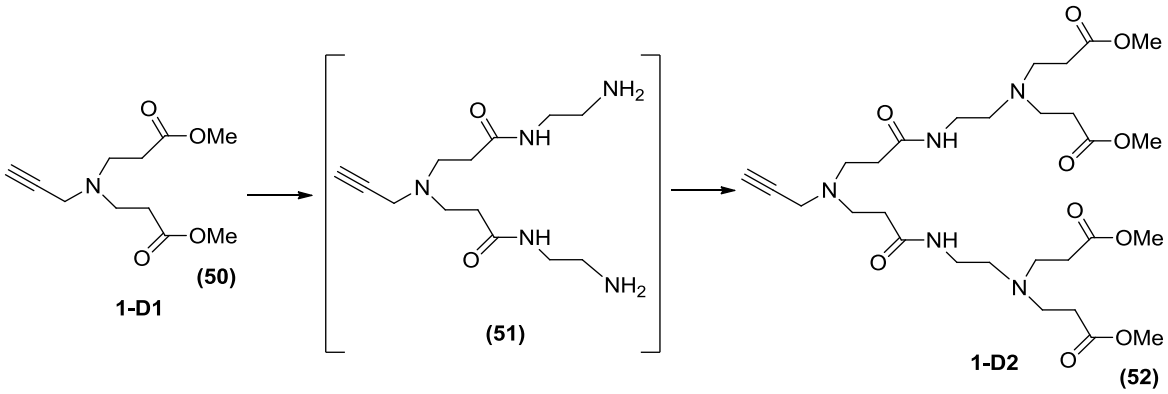
Synthesis of PAMAM dendrimers 1-D1 to 1-D3

Propargylamine was reacted with methyl acrylate in MeOH through a Michael addition to form the PAMAM 1-D1 dendrimer (**50**).[92] The reaction was purified by flash chromatography and product, PAMAM 1-D1, isolated in a 97% yield (scheme 3.18).



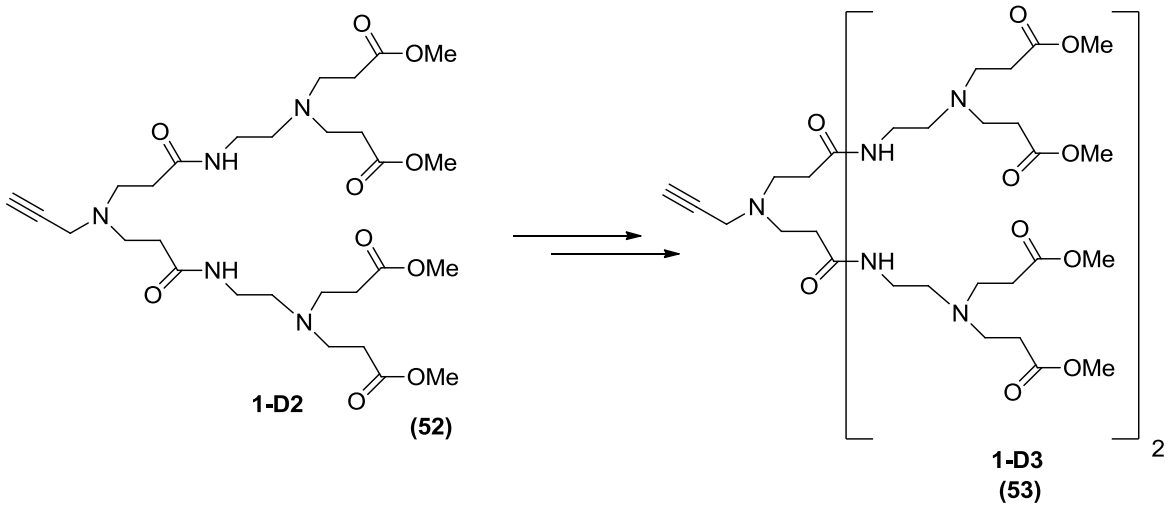
Scheme 3.18: Synthesis of PAMAM 1-D1 dendrimer. Reagents and conditions: Methyl acrylate, MeOH, 30°C, 48 h, 97%.

The 1-D1 dendrimer (**50**) was then extended to the 1-D2 dendrimer (**52**) through the ethylenediamine/methylacrylate two-step synthesis. Initially ethylenediamine was added to PAMAM 1-D1 through an amide condensation reaction and the amine-functionalised product (**51**) isolated, but not characterised. The intermediate was then reacted with methyl acrylate to form the 1-D2 dendrimer (**52**) (scheme 3.19).



Scheme 3.19: Synthesis of PAMAM 1-D2 dendrimer. Reagents and conditions: (i) Ethylenediamine, MeOH, 30°C, 48h, (ii) (**51**)**,** Methyl acrylate, MeOH, 72 h, 52%.

The ethylenediamine/methyl acrylate cycle was repeated to extend the 1-D2 dendrimer to the 1-D3 dendrimer (**53**) (scheme 3.20). This doubled the charge of the reporter group from a four carboxylate reporter to an eight carboxylate reporter group.

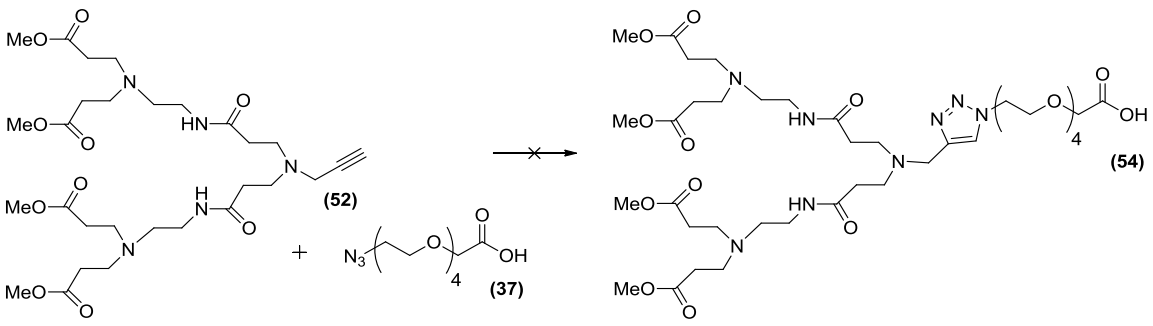


Scheme 3.20: Synthesis of PAMAM 1-D3 dendrimer (**53**). Reagents and conditions: Ethylenediamine, MeOH, 48h, (ii) Methyl acrylate, MeOH, 72 h, 52%.

Attachment of PAMAM dendrimers to a PEG_4 linker

The term 'click chemistry' was coined by K. Barry Sharpless and co-workers in 2001 to describe a method of reacting two compounds together quickly, cleanly and in high yields (among other requirements).[93] Although the term does not apply to one specific reaction it is often associated with the copper catalysed reaction between an alkyne and an azide to form a triazole linkage. Therefore, utilising the alkyne present in the PAMAM dendrimers and an azide terminated PEG linker, the synthetic route to synthesising novel dCTP analogues with PAMAM reporter groups is described using click chemistry.

Our preliminary efforts were focussed on the click reaction between the PAMAM dendrimer (**52**) and the CAz$(PEG)_4$ linker (**37**) using copper iodide in anhydrous DMF (scheme 3.21).[94] However this reaction was difficult to monitor due to the absence of a chromophore in either starting compound. Purification was attempted by prep-RP-HPLC but unfortunately only the starting material (PAMAM 1-D2) (**52**) was isolated, indicating the reaction was unsuccessful. An alternative synthetic technique of clicking the two compounds together using copper (II) sulphate ($CuSO_4$) and sodium ascorbate was trialled however this was also unsuccessful and again, only starting material was isolated.[95]

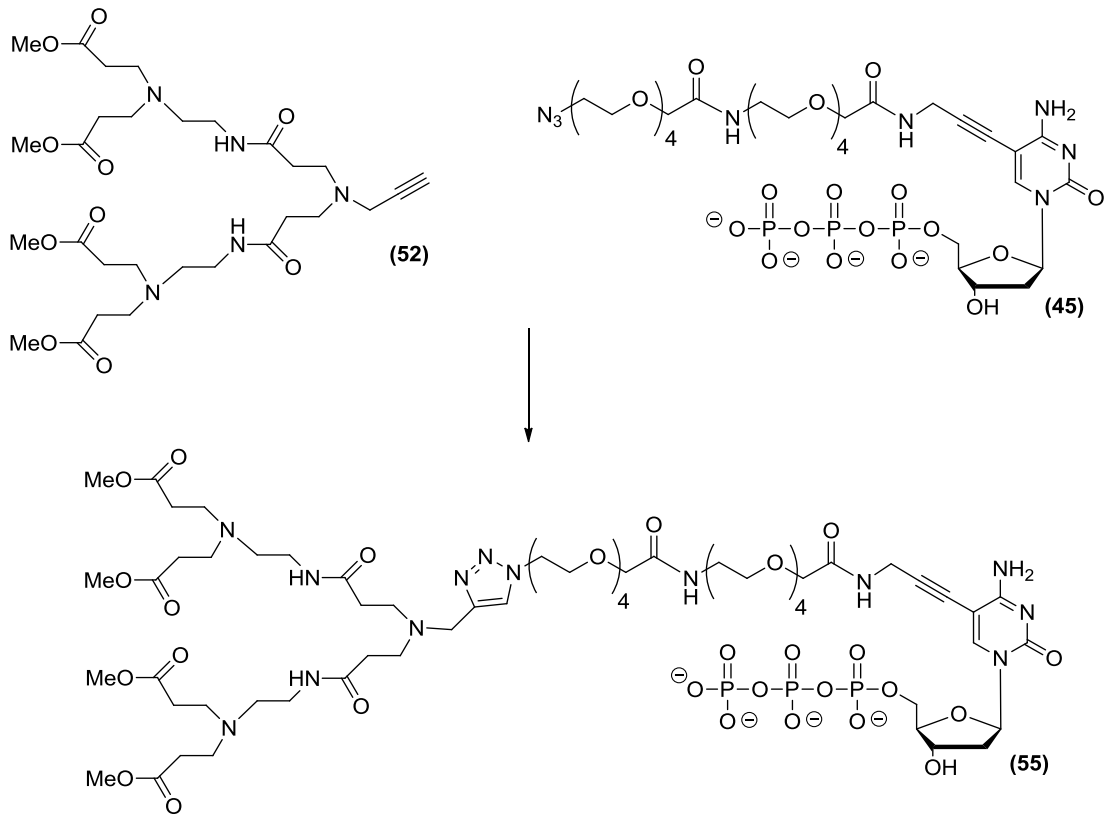


Scheme 3.21: Attempted click reaction between CAz(PEG)$_4$ (**37**) and PAMAM 1-D2 (**52**). Reagents and conditions: Attempted synthesis 1) CuI, DMF, 72 h. Attempted synthesis 2) $CuSO_4.5H_2O$, sodium ascorbate, 1:1 THF:H_2O, 8 h.

3.6 Click reactions performed with modified dCTPs

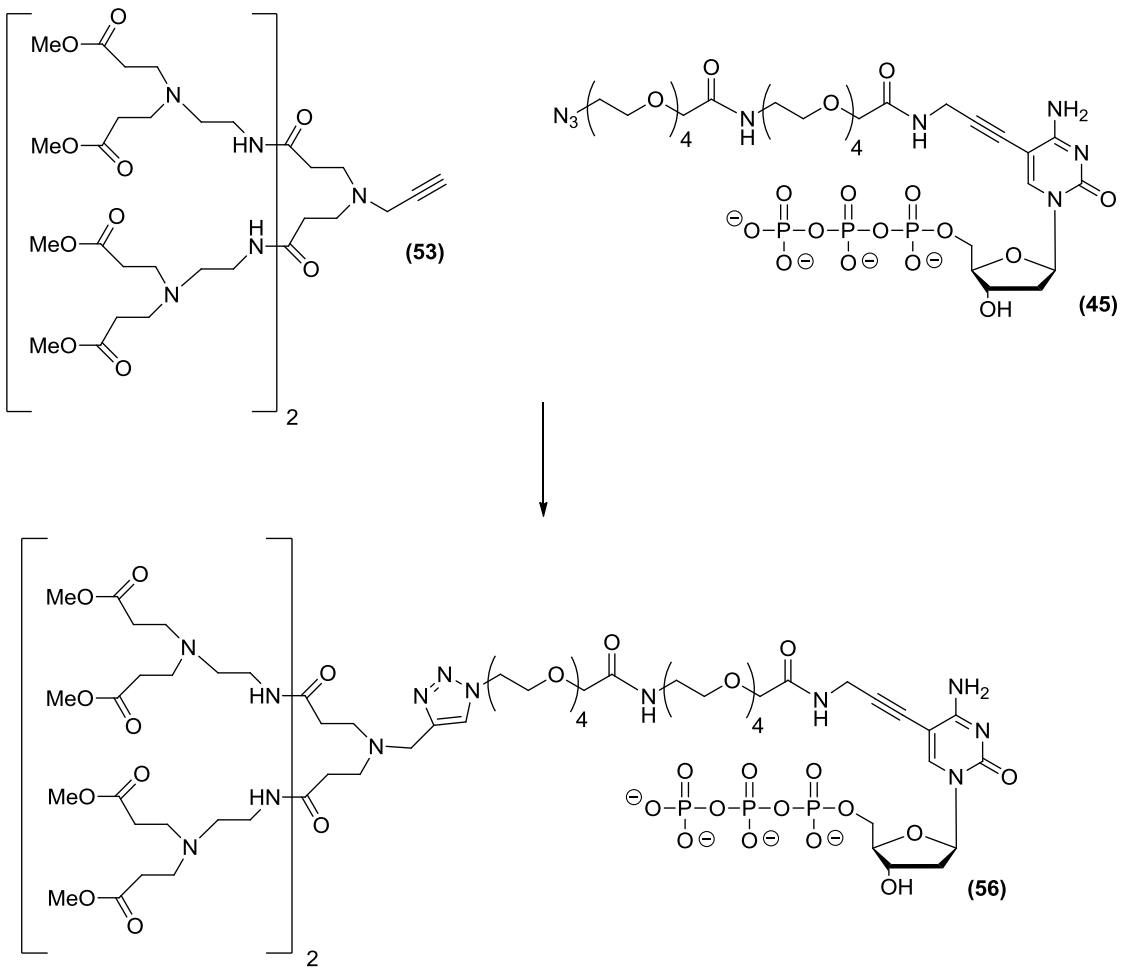
Published work by Seela and Ingale[96] reported the successful use of click chemistry to connect alkyne functionalised modified nucleosides and ODNs together through a *bis*-azide linker. Given this precedent subsequent attempts at click chemistry were performed directly on the dCTP analogue. The click reaction was initially performed using the classical $CuSO_4$/sodium ascorbate approach.[95] Given the previously discussed examples within the literature which suggest that a nucleotide modified with a longer linker will be more favourable as a DNA polymerase substrate, it was decided that the azido-$(PEG_4)_2$-dCTP (**45**) would be the most suitable dCTP analogue for functionalisation with a PAMAM dendrimer.[87, 88] This may be particularly prudent given the additional size of the dendrimer compared to the previously tested reporter group. Azido-$(PEG_4)_2$-dCTP (**45**) was synthesised as described in scheme 3.16.

The subsequent Cu(I) catalysed click reaction was performed with 0.1 eq. $CuSO_4$ and 4 eq. of sodium ascorbate. Sodium ascorbate is a reducing agent which is required to reduce Cu(II) to Cu(I) in the reaction. The synthesis of (**55**) was successful through the click reaction of (**52**) and (**45**), with the desired triphosphate isolated in a 35% yield after prep-RP-HPLC purification (scheme 3.22).



Scheme 3.22: Click reaction between PAMAM 1-D2 (**52**) and (**45**). Reagents and conditions: $CuSO_4$, sodium ascorbate, THF, 2h, 35%.

The click reaction was then repeated using the larger, PAMAM 1-D3, dendrimer (**53**). A higher loading of $CuSO_4$ (0.2 eq.) was used upon repeat to see if the yield could be increased (scheme 3.23) and the reaction was monitored by analytical HPLC (figure 3.13).



Scheme 3.23: Click reaction between PAMAM 1-D3 (**53**) and (**45**). Reagents and conditions: $CuSO_4$, sodium ascorbate, THF, 2h, 37%.

After 1 hour it was observed by RP-HPLC that approximately 40% of the starting material (retention time of 15 mins) has been consumed, with a product peak developing at 23.5 mins (figure 3.13). The peak height for the dCTP starting material (**45**) gradually decreased until after 4 hours the reaction was deemed complete. The reaction was then purified by prep-RP-HPLC and (**56**) isolated in a 37% yield.

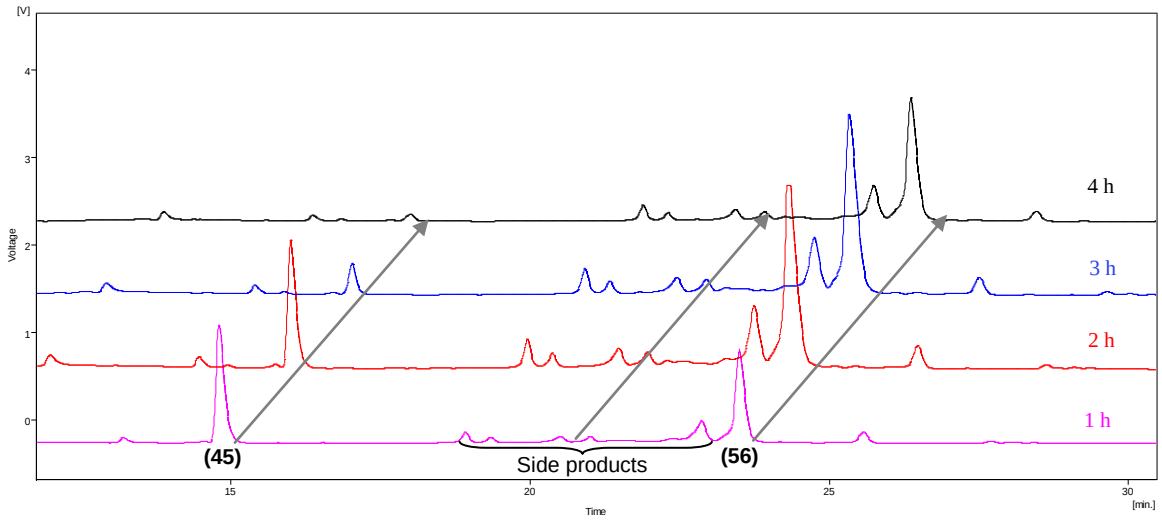


Figure 3.13: Analytical HPLC to monitor the coupling reaction between (**53**) and (**45**) to form the modified dCTP (**56**). Conditions: 5 – 100% B over 30 minutes where A= 0.1 M TEAB and B= 50 % MeCN / 0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min, Retention time of product = 23.4 mins Retention time of starting material = 15.0 mins.

One theory for the relatively low yields of the click reaction products was the instability of the Cu(I) ion needed for the click reaction to proceed. To test this assumption 0.2 eq Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) was added to the click reaction. TBTA forms a complex with the copper ion, stabilising the 1+ oxidation state (figure 3.14).[97]

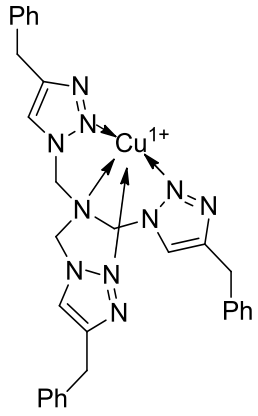
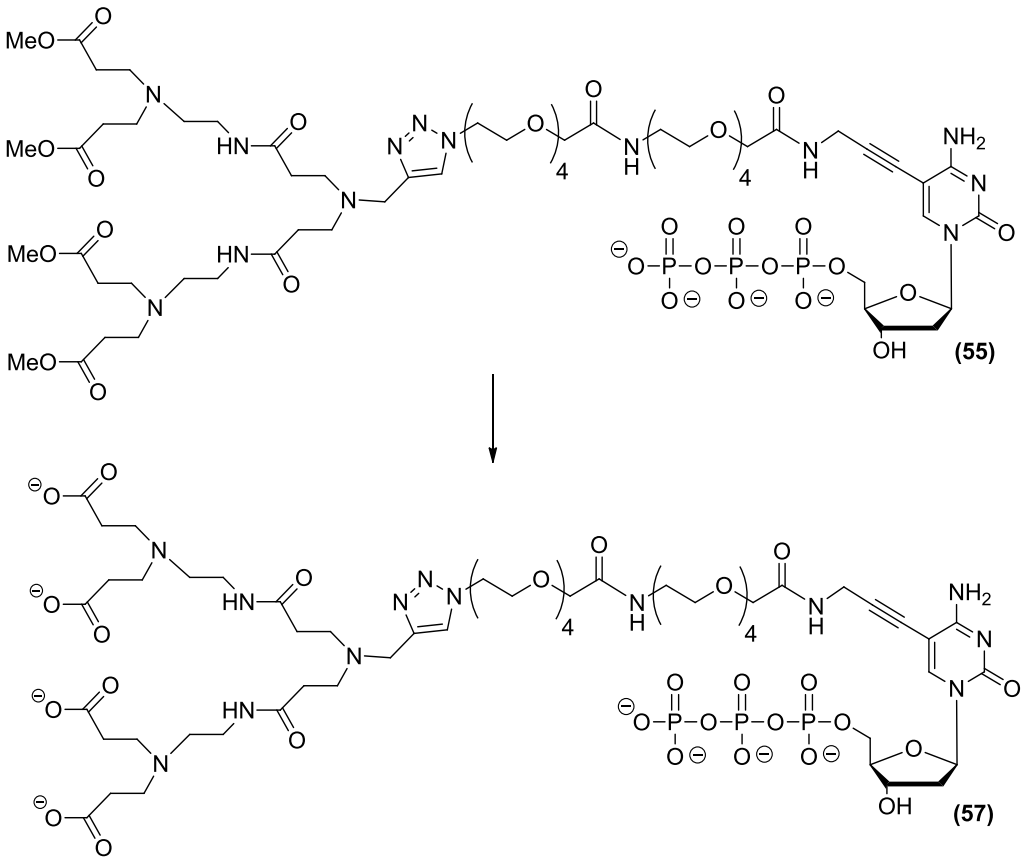


Figure 3.14: Cu (I) complex with TBTA.

The click reaction was therefore repeated using the smaller 1-D2 PAMAM dendrimer (**52**) with the addition of TBTA. For the copper complex to form (figure 3.14) $CuSO_4$ (i.e. Cu(II)) is premixed with TBTA. The TBTA-Cu complex is added to the reaction and reduced *in situ* by sodium ascorbate.[98] The reaction was then performed in an analogous manner to the previous synthesis. After purification by prep-RP-HPLC a two-fold increase in yield of (**55**) (70% isolated yield) was observed.

To provide a modified dCTP suitable for testing as a polymerase substrate the methyl esters required hydrolysis. The deprotection of (**55**) was performed with 2 M NaOH for 2 hours, by analogy of the deprotection of the trimesic acid reporter group. Unfortunately, no triphosphate-containing material was isolated following prep-RP-HPLC purification. However stirring (**55**) in a 0.5 M NaOH solution for 2 hours at room temperature resulted in complete hydrolysis to the desired carboxylate dCTP (**57**) (scheme 3.24) as determined by HPLC analysis of the crude reaction mixture.



Scheme 3.24: Deprotection of the dendrimer reporter group of (**55**). Reagents and conditions: 0.5 M NaOH, 2 h, 68%.

The analytical RP-HPLC trace of the crude material from the deprotection of (**55**) is shown (figure 3.15). The major peak (retention time 10.7 mins) was subsequently isolated by prep-RP-HPLC to give (**57**) in a 68% yield.

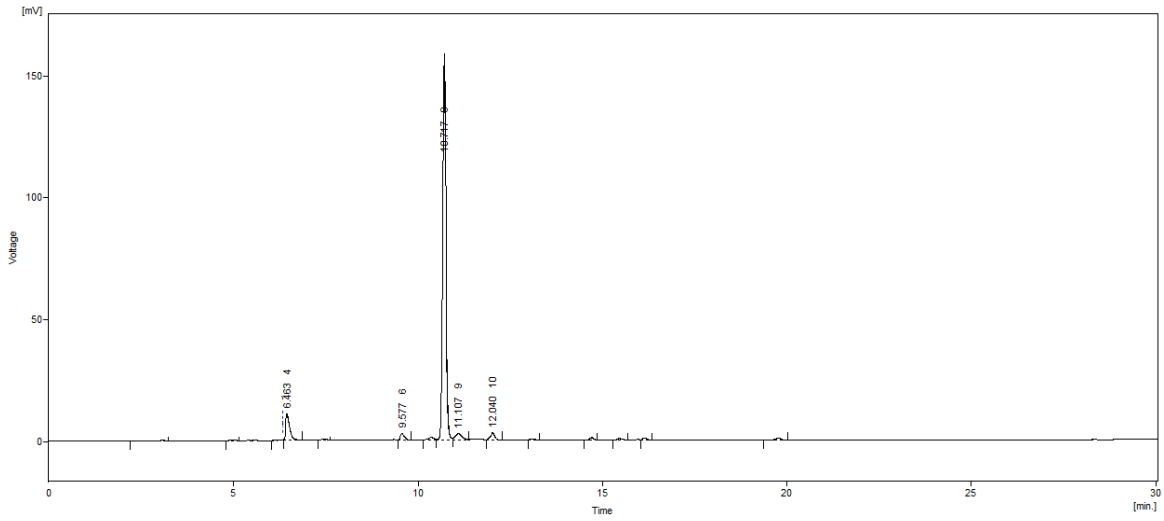
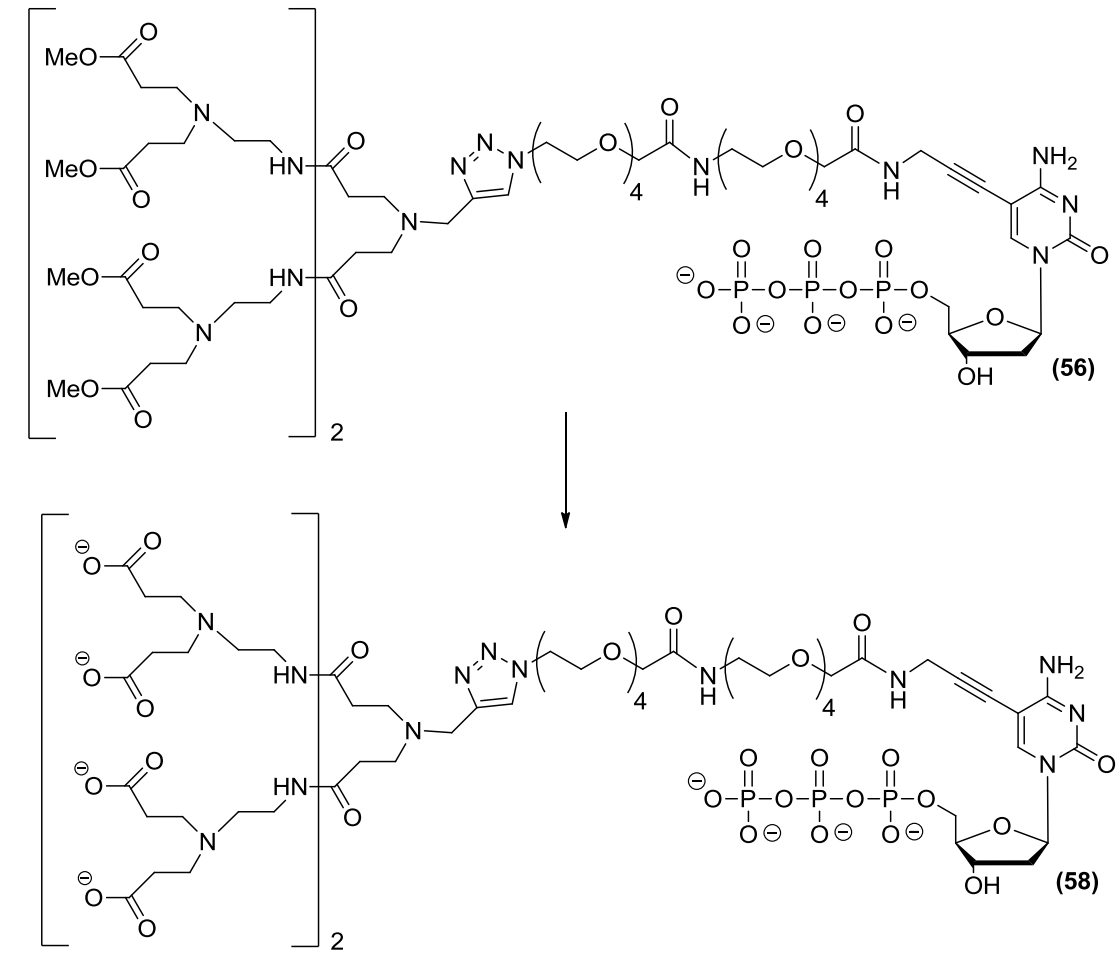


Figure 3.15: Analytical HPLC of crude (**57**). Conditions: 5 – 100% B over 30 minutes where A= 0.1 M TEAB and B= 50% MeCN / 0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min, Retention time of product = 10.7 minutes.

With appropriate conditions for the hydrolysis of (**55**) in hand the hydrolysis of the methyl esters of the larger 1-D3 dendrimer modified dCTP (**56**) was then performed (scheme 3.25). Again a lower concentration of NaOH (0.5 M) proved sufficient for deprotection within 2 h. The reaction was analysed by RP-HPLC and the major peak (figure 3.16, retention time 6.5 mins) isolated by prep-RP-HPLC in a 63% yield.



Scheme 3.25: Deprotection of the 5-(PAMAM-1-D3-$(PEG_4)_2$)-dCTP (**56**). Reagents and conditions: 0.5 M NaOH, 2 h, 63%.

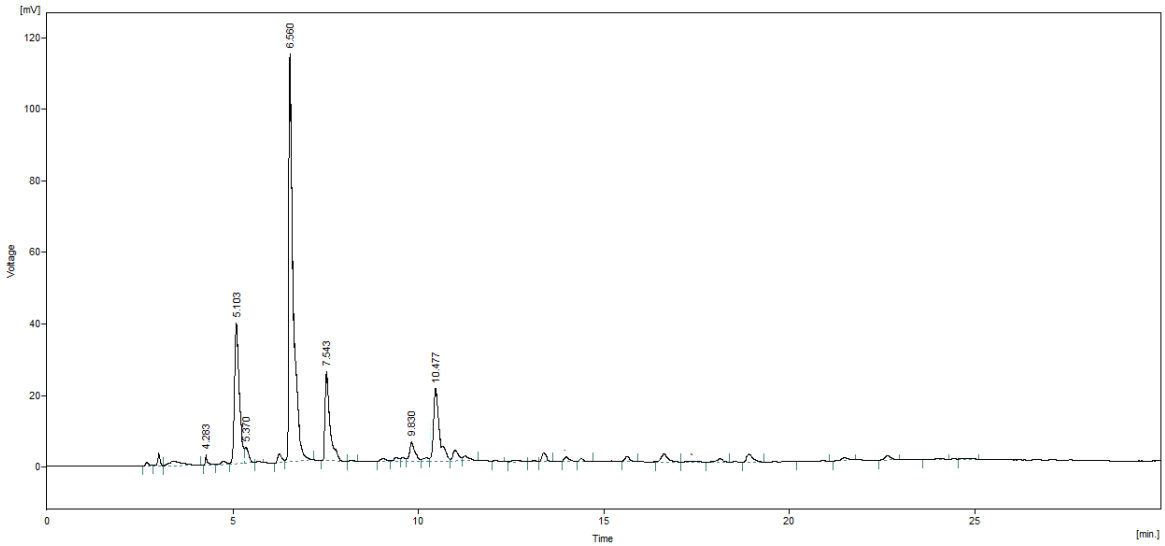


Figure 3.16: Analytical HPLC trace of crude (**58**). Conditions: 5 – 100% B over 30 minutes where A= 0.1 M TEAB and B= 50% MeCN / 0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min, Retention time of product = 6.5 minutes.

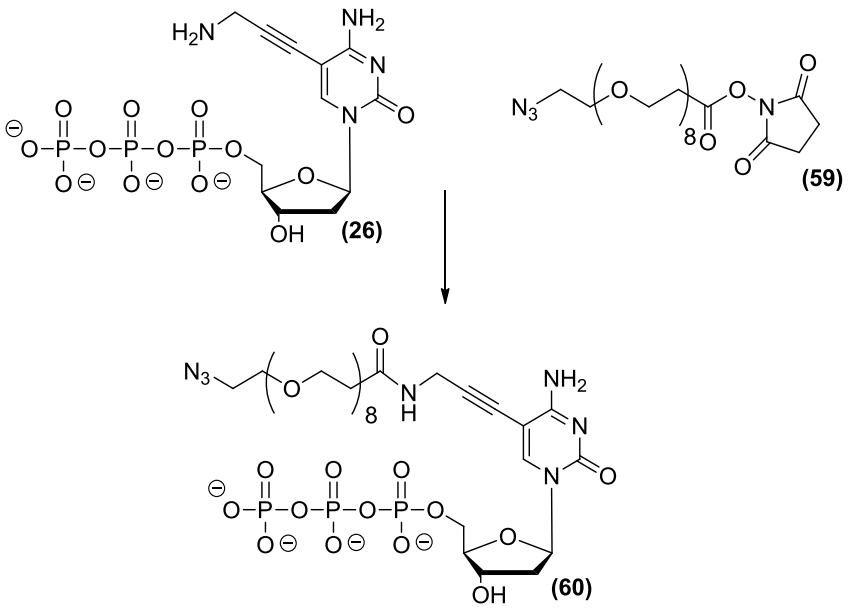
With the synthesis of four novel dCTPs complete it was of interest to synthesise a dCTP analogue with a significantly higher charge. This is of importance as in the final SBS device there may need to be a significant difference between the charges of the reporter groups depending on the sensitivity of the nanowire.

3.7 Using single stranded DNA as a reporter group

The use of single stranded DNA (ssDNA) as a linear anionic reporter group was investigated as ssDNA possesses an inherent negative charge due to its phosphate diester backbone. A significant advantage of using ssDNA as a reporter group is that the anionic charge can be easily adjusted by synthesising different length DNA strands. Marx *et al.*[99] have previously reported a C5-modified dTTP with an ssDNA modification of up to 40 bases in length to be a substrate for Therminator DNA polymerase.

We expected that by introducing a 5'-alkyne modification to ssDNA the same click chemistry described in section 3.6 could be utilised to attach the ssDNA to an azido-modified dCTP analogue. This would give a C5-modified dCTP with a linear reporter group bearing a higher anionic charge than the reporter groups previously investigated.

A suitable modified dCTP analogue was synthesised by coupling the commercially available azido-PEG_8-NHS ester (**59**) to the previously synthesised anchor modified-dCTP (**26**). As the linker was purchased as the NHS-ester this could be simply achieved with the use of aq. acetonitrile in (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 8). The desired product was isolated in a 70% yield after purification by prep-RP-HPLC.



Scheme 3.26: Coupling of the azido-PEG_8-NHS ester (**59**) to (**26**). Reagents and conditions: MeCN, HEPES buffer, 70%.

Two alkyne modified poly(dT) ssDNA compounds were purchased from DNA Technology, Denmark. The single stranded 2'-deoxythymidine ODNS (ssdT) were 6 and 24 nucleotides in length, pre-functionalised at the 5'-hydroxyl with an alkyne linker (figure 3.17).

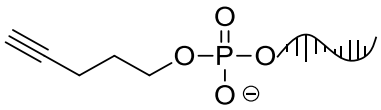
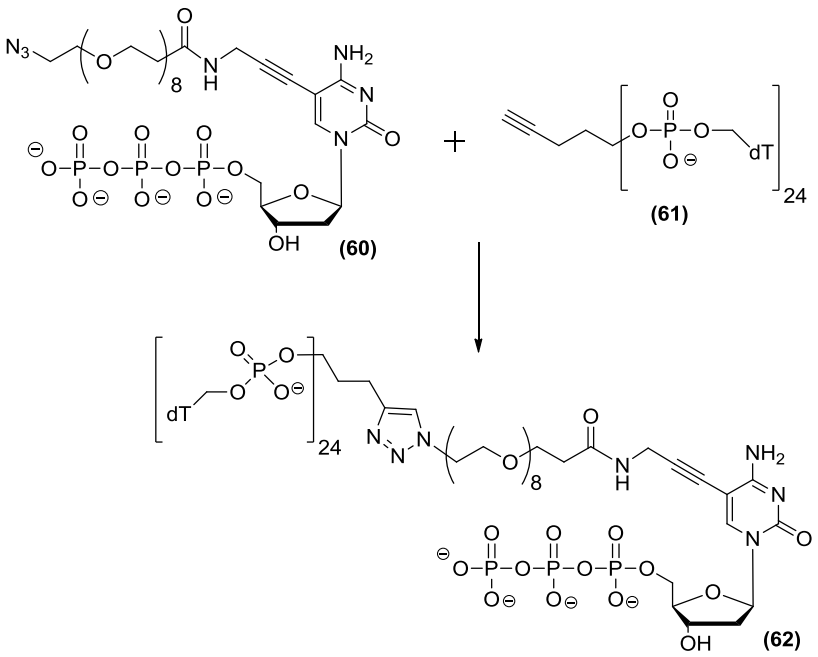


Figure 3.17: Schematic representation of the 1-pentyne linker used to modified ssDNA for use in click reactions.

A variation of the Sharpless click reaction[93] was once again used with the addition of TBTA as a Cu(I) stabilising agent. The Cu(I) source was varied from copper sulphate ($CuSO_4$) to copper acetate ($Cu(OAc)_2$) to mimic the conditions described by Seela *et al.*[96] for clicking alkyne modified ODNs to a *bis*-azide.

The initial click reaction was trialled on a 100 nmol scale; the alkyne modified dT_{24} ssDNA (**61**) reacting with the dCTP (**60**) (4 eq.) in the presence of a Cu-TBTA complex and sodium ascorbate in aq. THF (scheme 3.27).



Scheme 3.27: Click reaction of dT_{24} ssDNA (**61**) to dCTP (**60**). Reagents and conditions: $Cu(OAC)_2$, TBTA, sodium ascorbate, DMSO, H_2O, THF, *t*BuOH, sodium bicarbonate (0.2 M), 8 h, 62%.

The reaction was analysed by HPLC (figure 3.18). The HPLC trace of the starting material is overlaid, shown in red, with the trace of the product from the click reaction shown in blue. A new peak can be seen at 12.3 mins which does not correlate to either starting material (**60**) or (**61**). The retention time of the starting material for the alkyne $ssdT_{24}$ (**61**) is not shown for clarity, as its retention time (15.6 mins) partially overlaps with that of the azido-PEG_8-dCTP (**60**). It is worth noting for C5-alkyne modified dCTP analogues, λ_{max} = 295 nm therefore as the HPLC detection is at 270 nm to detect the ssDNA this is not a faithful representation of the composition of the crude reaction mixture. However the retention time of the starting material did not vary when UV detection was performed at 295 nm or 270 nm. Therefore the identification of the residual material being starting material is still valid.

The click product (**62**) (12.3 mins) was subsequently isolated by RP-HPLC in a 62 % yield.

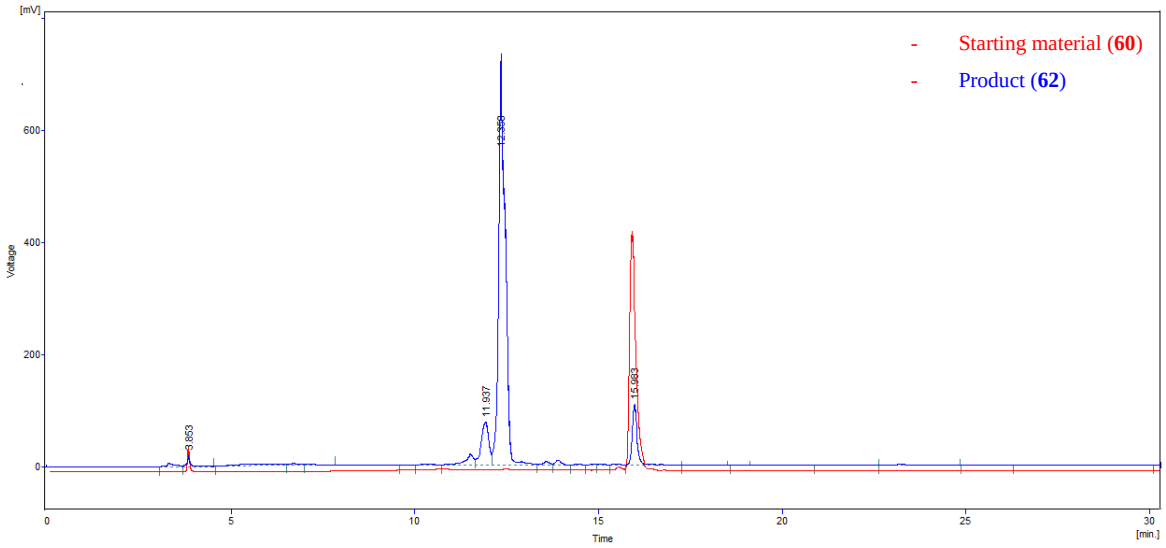
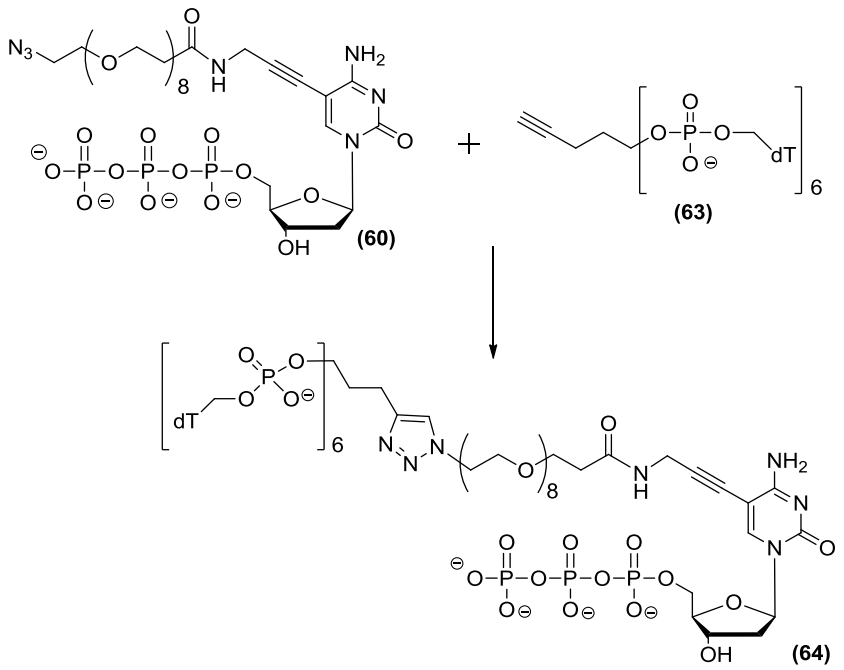


Figure 3.18: HPLC trace showing the click product (**62**) with the starting material (**60**) shown for clarity. Conditions: 5 – 100 % B over 30 minutes where A= 0.1 M TEAB and B= 50 % MeCN/0.1 M TEAB. UV = 270 nm, flow rate = 1 ml / min. Retention time of product = 12.3 minutes, starting material = 15.9 minutes.

After the successful synthesis of (**62**) the click reaction was repeated with the shorter dT_6 (**63**) (scheme 3.28). The click reaction was analysed and subsequently purified by RP-HPLC to isolate (**64**) in a 68% yield. A comparison of retention times between the purified product (**64**) to starting material (**60**) is shown in figure 3.19 with the RP-HPLC trace of the starting material (**60**) overlaid (in red) on the RP-HPLC trace of the purified (**64**) shown in black.



Scheme 3.28.: Click reaction of dT_{24} oligonucleotide (**63**) to dCTP (**60**). Reagents and conditions: $Cu(OAC)_2$, TBTA, sodium ascorbate, DMSO, H_2O, THF, *t*BuOH, sodium bicarbonate (0.2 M), 8 h, 68 %.

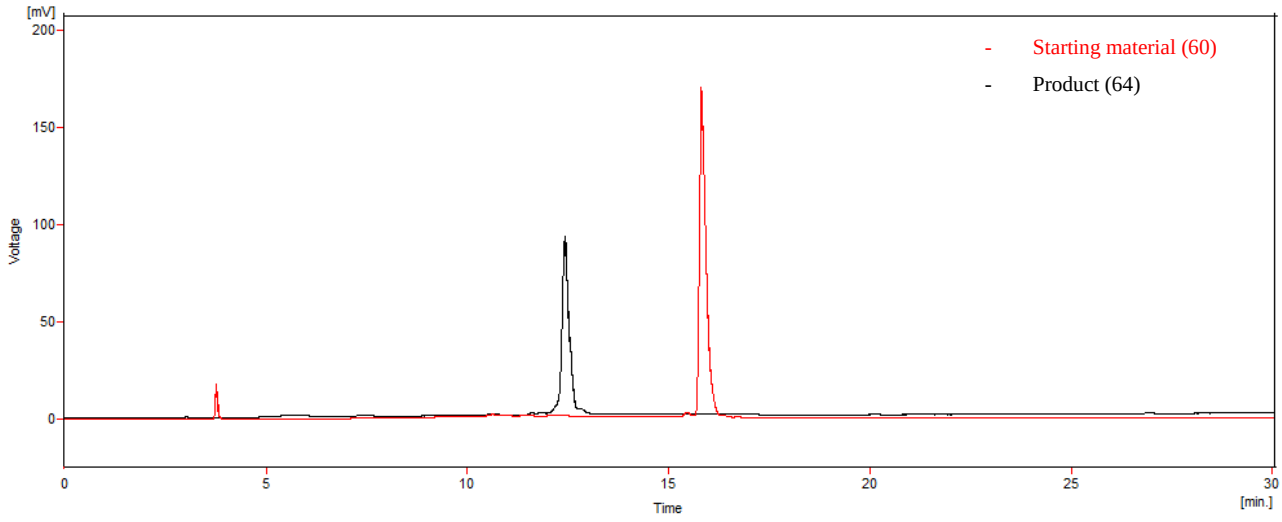


Figure 3.19: HPLC trace showing the click product (**64**) with the starting material (**60**) shown for clarity. Conditions: 5-100% B over 30 minutes where A= 0.1 M TEAB and B= 50 % MeCN / 0.1 M TEAB. UV = 270 nm, flow rate = 1 ml / min. Retention time of product = 12.4 minutes, starting material = 15.9 minutes.

Section C: Cleavable reporter groups

As the SBS cycle (figure 1.22) continues, more and more nucleotides bearing reporter groups are incorporated into the growing DNA. As QuantuMDx are developing modified dNTPs with anionic reporter groups this would result in a build-up of charge on the nanowire. This could affect the nanowires ability to accurately detect the identity of the incoming nucleotide from its reporter group. To prevent this from occurring, the reporter group must be cleaved after each dNTP incorporation. Cleavage can be achieved in many ways, for example: chemically, enzymatically or photolytically.

Illumina SBS sequencing is one of the most utilised SBS methodologies. It relies upon the chemical cleavage of fluorescent reporter groups and the 3'-blocking group between dNTP incorporations. The Illumina modified dNTPs have a cleavable azidomethyl group within the linker and at the 3'-hydroxyl. The linker and 3'-blocking group are cleaved through the Staudinger reaction using tris(2-carboxyethyl)phosphine (TCEP). This removes the fluorescent reporter group which is tethered to the dNTPs through the linker and provides an unprotected 3'-hydroxyl for the formation of the next phosphodiester bond (figure 3.20).

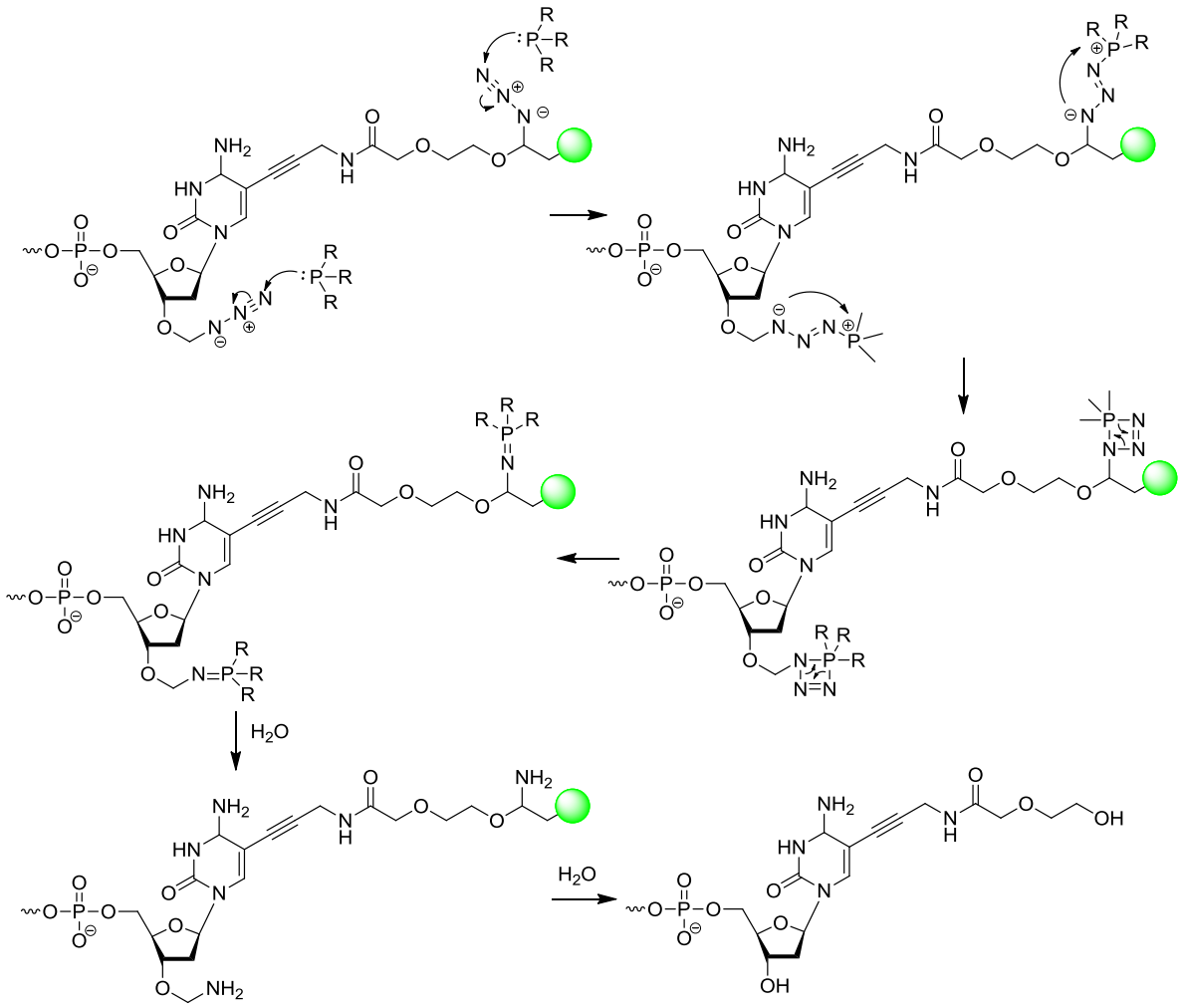


Figure 3.20: Chemical cleavage of the Illumina's C5-modified dCTP through the Staudinger reaction.

Photocleavable linkers have also been proposed for use in SBS with the synthesis of a photocleavable dNTP described by Bai *et al.*[100] They describe a fluorescent reporter group attached via a photocleavable, 2-nitrobenzyl linker to the C5 position of dUTP (figure 3.21). The 2-nitrobenzyl linker can be cleaved by UV irradiation at 340 nm to release the fluorescent group attached to a 1,2-nitrosophenyl-ethanone fragment and leave the unlabelled 3-aminopropenyl-dUTP.

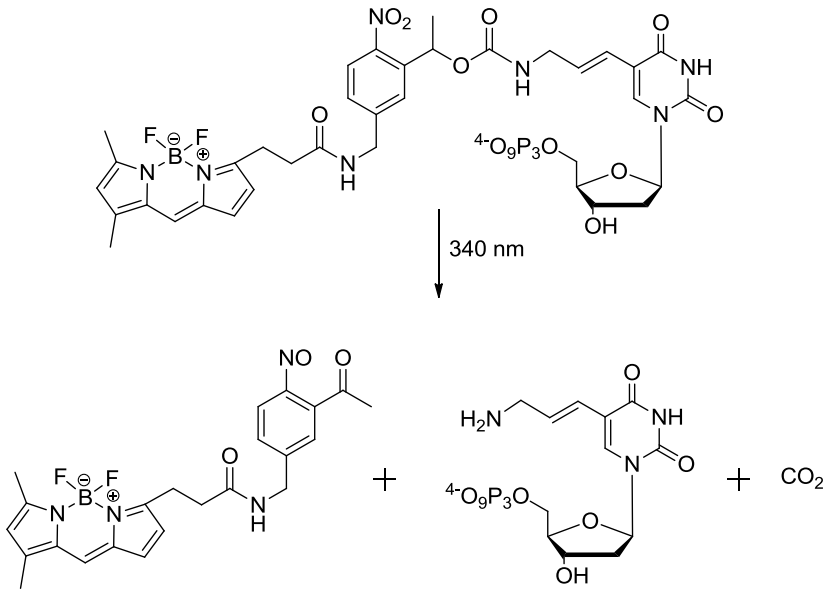


Figure 3.21: The photocleavage of the modified dUTP analogue synthesised by Bai and co-workers.[100]

Ramos *et al.*[101] subsequently investigated a variation of the photocleavable linker used by Bai *et al.*[102] where they showed the photocleavable unit could vary within the constraints shown in figure 3.22.

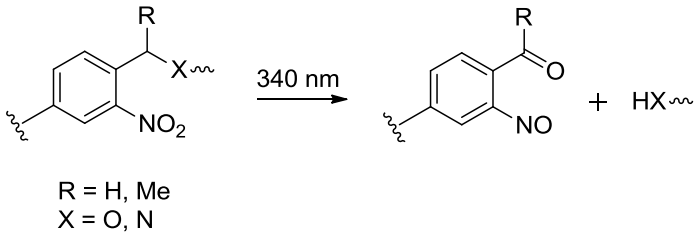


Figure 3.22: Structural variations allowed for an active photocleavable unit.

The amido derivative of the *O*-nitrobenzyl photocleavable unit was used by Abramova and Silnikov[103] to synthesise a series of oligonucleotides containing nucleotides modified with photocleavable linkers. An example of a ribonucleoside (cytidine) bearing such a modification is shown in figure 3.23.

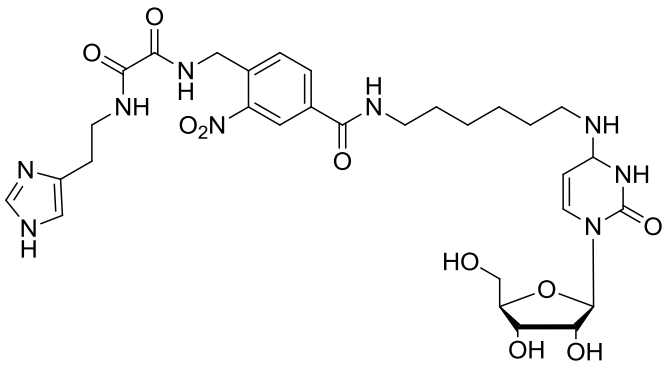


Figure 3.23: Cytidine bearing a photocleavable linker at the N4 position of the nucleobase as synthesised by Abramova and Silnikov.[103]

Photocleavable groups have also been established as suitable 3'-blocking groups. Metzker *et al.*[104] described the synthesis of 2'-deoxy-3'-*O*-(2-nitrobenzyl)adenosine-5'-triphosphate (figure 3.24) for use in SBS. They subsequently demonstrated its application in a SBS cycle through successful DNA synthesis termination, cleavage through photolysis and the successful re-initiation of DNA synthesis.

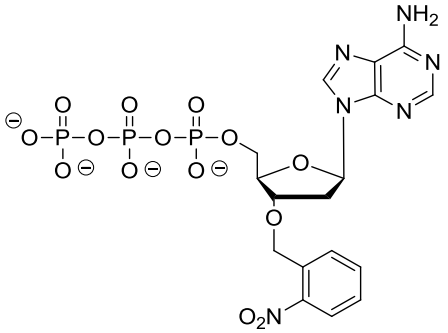


Figure 3.24: 2-Nitrobenzyl as a 3'-blocking group as described by Metzker *et al.*[104]

3.8 Synthesis of a cleavable linker

Previous efforts to synthesis dNTPs possessing cleavable functionality have focussed on the synthesis of chemically cleavable linkers. For this endeavour the synthesis of a linker based upon benzyl azide was attempted by a previous group member (figure 3.25).

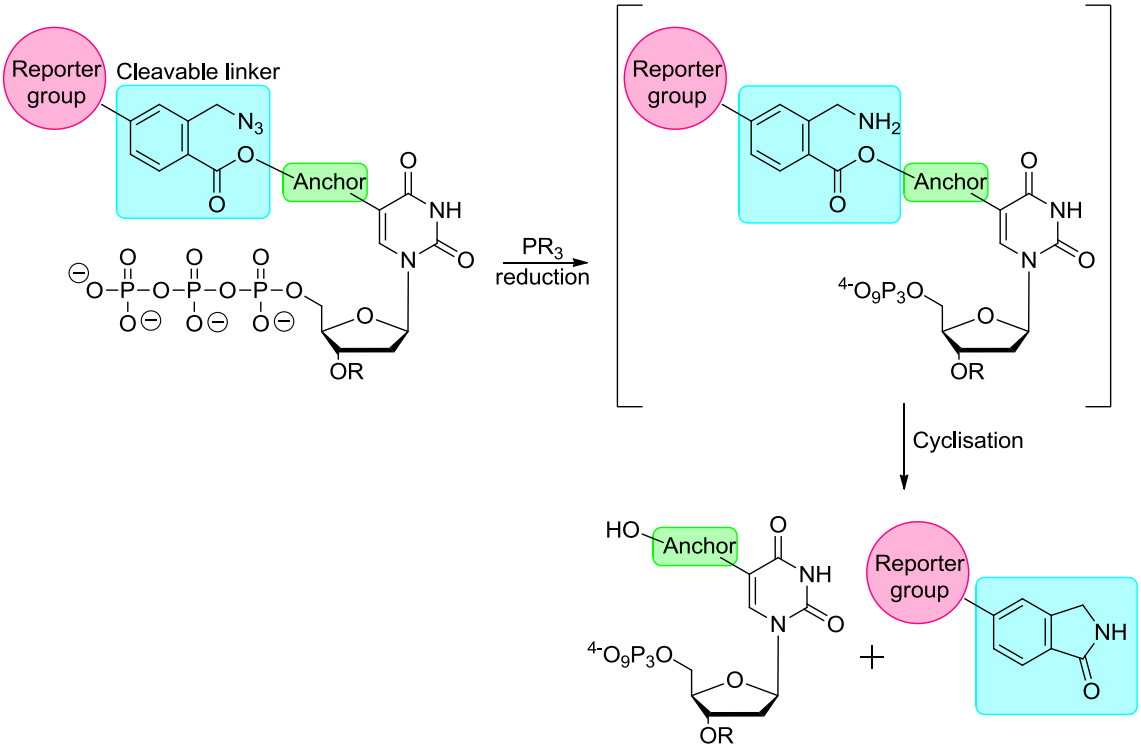
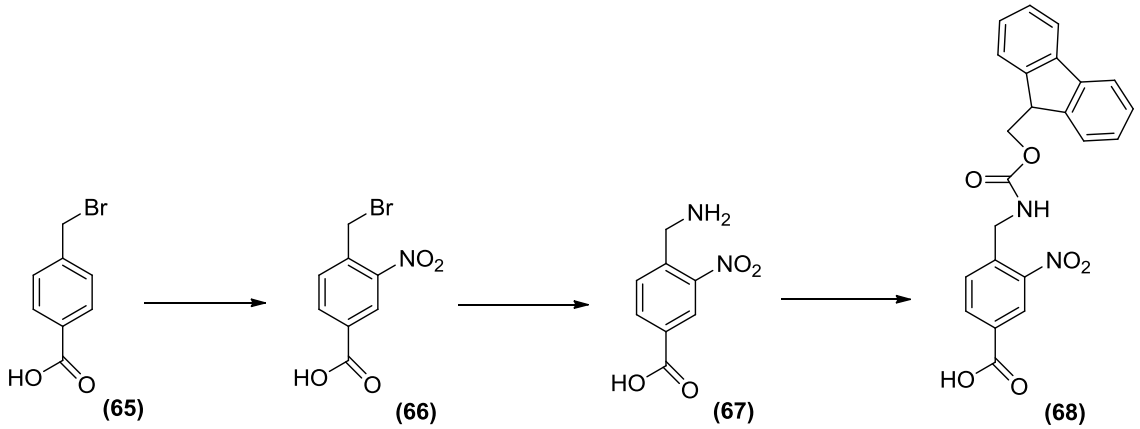


Figure 3.25: Proposed design of a chemically cleavable linker by a previous group member.[105]

The proposed cleavage was through the reduction of the benzyl azide to the corresponding amine. This would then react with the ester, cleaving the reporter group through a cyclisation reaction (figure 3.25). Unfortunately although the method of cleavage was shown to be valid by test reactions performed on the cleavable linker alone, the linker could not be successfully functionalised to allow its incorporation into a dNTP and the attachment of a reporter group. These problems have been addressed with the synthesis of a photocleavable linker which provides appropriate functionality to allow for the attachment of a reporter group and the tethering to a dNTP.

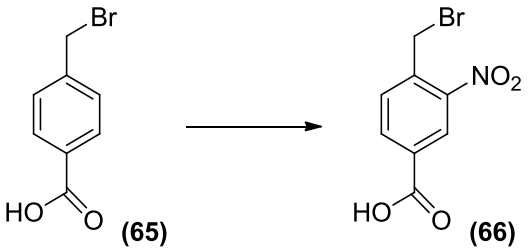
Synthesis of a photocleavable linker

Following a variation of the synthetic route proposed by Ramos *et al.*[101] we synthesised the photocleavable, 2-nitrobenzyl derivative (**68**) (scheme 3.29).



Scheme 3.29: Synthetic overview for the synthesis of the photocleavable linker (**68**).

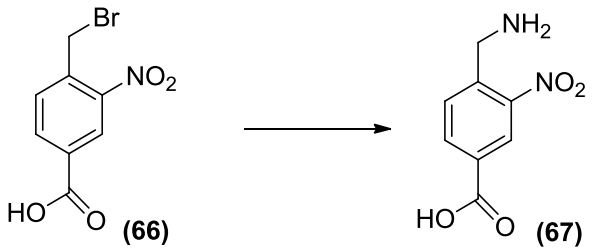
Nitration of the commercially available 4-(bromomethyl)benzoic acid (**65**) was attempted using concentrated sulphuric and nitric acid with refluxing overnight. Unfortunately, a mixture of bis- and mono-nitrated products were formed and could not be separated by flash chromatography. Nitration was reattempted with the use of fuming nitric acid at -10°C and the desired mono-nitrated product (**66**) isolated by precipitation in a 63% yield (scheme 3.30).



Scheme 3.30: Nitration of 4-(bromomethyl)benzoic acid (**65**). Reagents and conditions: Fuming HNO_3, -10°C, 1 h, 63%.

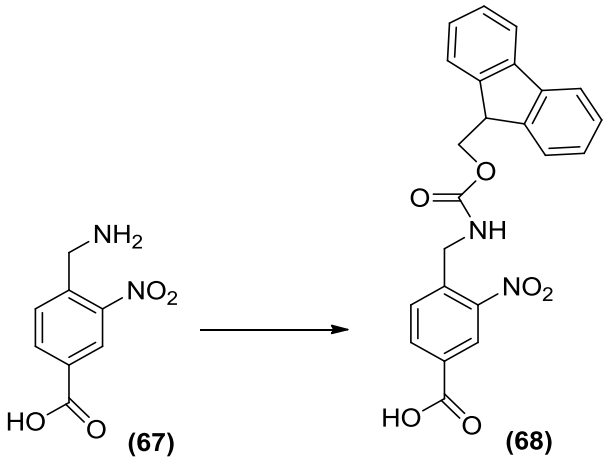
The bromoalkane must then undergo an S_N2 nucleophilic substitution with ammonia to generate the desired 4-(aminomethyl)benzoic acid (**67**). Amination was achieved by

stirring in 8% ammonia in ethanol solution for 20 h at room temperature (scheme 3.31). 4-aminomethyl-3-nitrobenzoic acid (**67**) was isolated by precipitation in a 34% yield.



Scheme 3.31: Amination of (**66**). Reagents and conditions: 8% NH_3 in EtOH, 20 h, 34%.

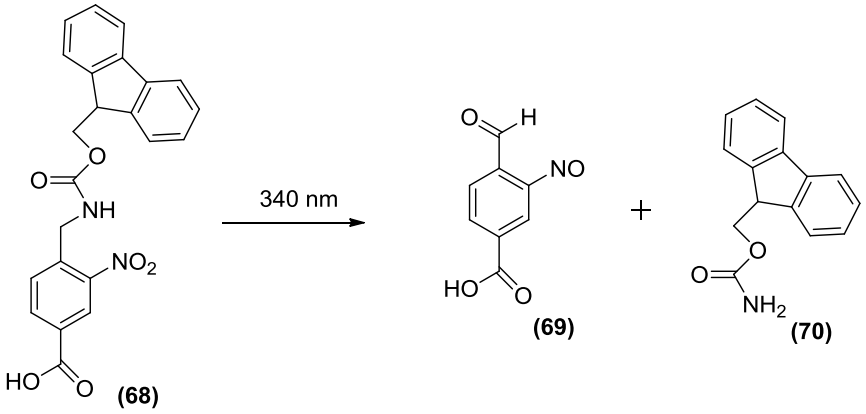
The amine then required protection to block its nucleophilicity and avoid its participation in the subsequent TSTU coupling. This was achieved with a fluorenylmethyloxycarbonyl chloride (FMOC) protection. FMOC was chosen as the protecting group as it is base labile and therefore the conditions of its deprotection are suitable for use with triphosphates. Compound (**68)** was isolated by precipitation through drop-wise addition of 1 M hydrochloric acid in a 28% yield (scheme 3.32).



Scheme 3.32: FMOC-Cl protection of (**67**). Reagents and conditions: 10% $NaCO_3$, FMOC-Cl, 0°C, overnight, 28%.

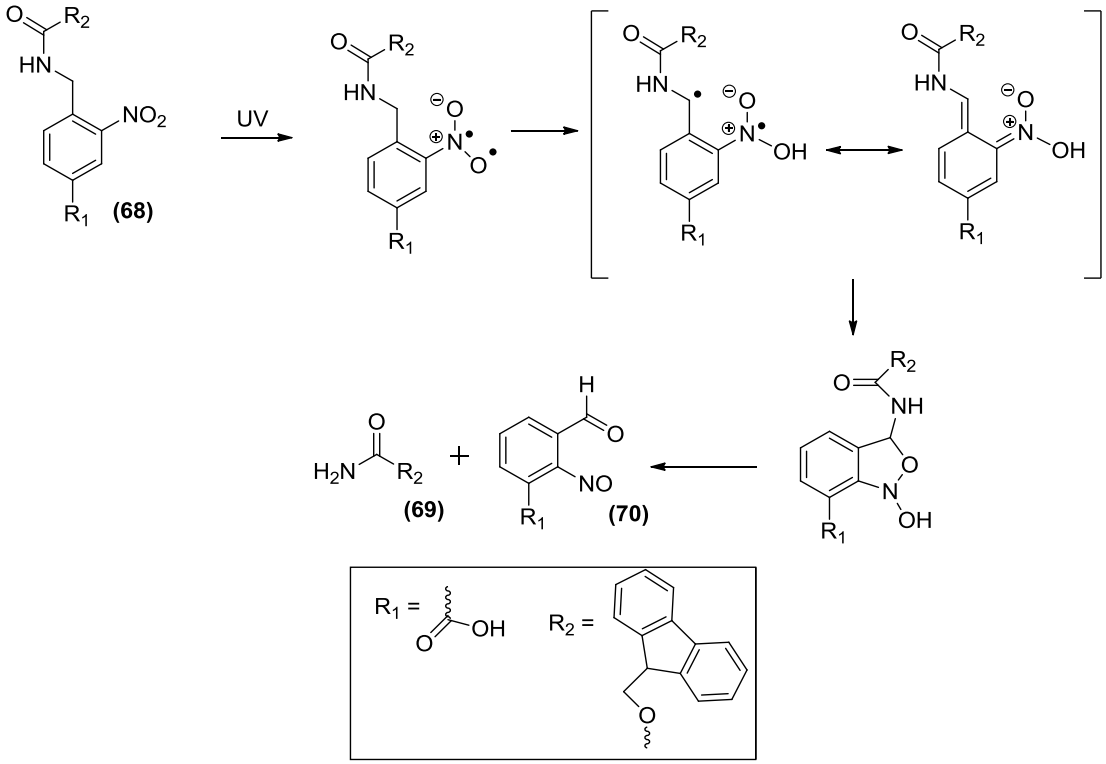
Test photolysis reactions

To ensure the cleavage of the proposed photocleavable moiety (**67**), test photolysis reactions were performed to ensure the linker was 'fit for purpose'. Although the photocleavable linker (**67**) (scheme 3.32) will be the linker performing the cleavage in the dCTP, compound (**68**) provided an early opportunity to test the photolysis due to the fluorescence of the FMOC group (scheme 3.33).



Scheme 3.33: Test photolysis reaction of (**68**) and projected products.

The cleavage mechanism of 2-nitrobenzyl derivatives is generally considered a photoisomerisation mechanism which proceeds *via* a 1,4-biradical intermediate (scheme 3.34). Due to the formation of the 1,4-biradical the photolysis is sometimes referred to as Norrish-type II photoreaction.[106] Once formed through UV irradiation at 340 nm the radical intermediate rearranges to the cyclic isoxazolidine. A final rearrangement results in the cleavage of 9-fluorenyl carbamate (**70**) and formation of the 4-formyl-3-nitroso-benzoic acid (**69**) by-product.[107]



Scheme 3.34: Photolysis of the test compound **68**

Due to the fluorescence of 9-fluorenylmethyl carbamates, successful cleavage will be determined by analysis of the change of the fluorescent profile as the cleavage reaction occurs. The cleavage reaction was irradiated with UV light (340 nm) using a fluorometer and the emission profile recorded every 150 seconds (figure 3.26).

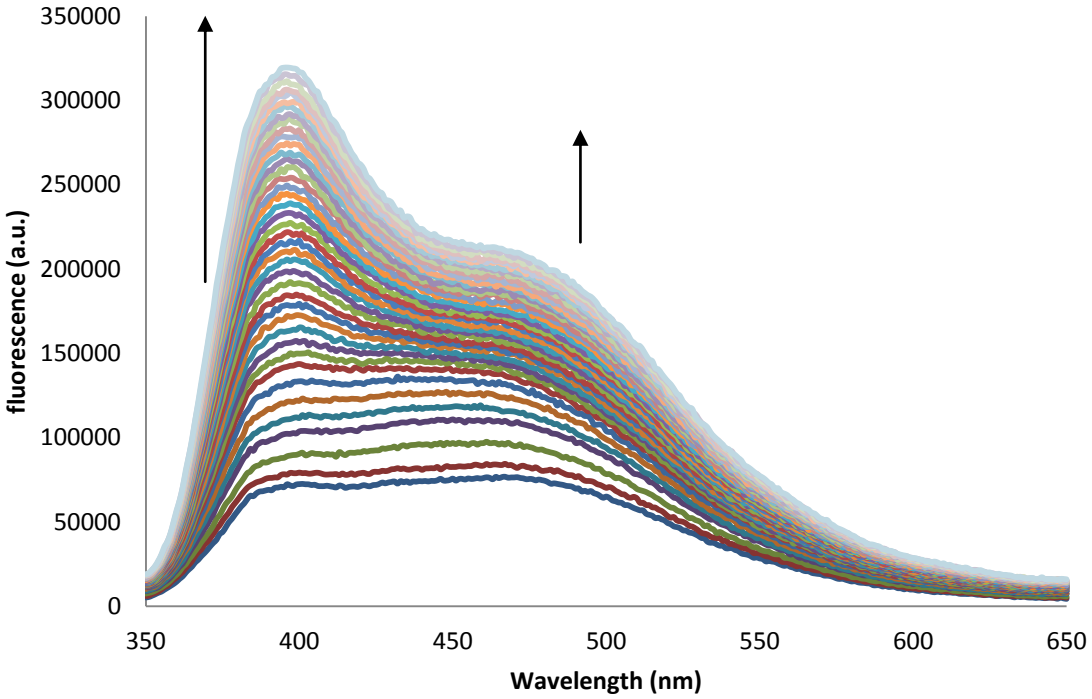


Figure 3.26: Fluorescence emission spectra showing the increasing fluorescence maxima over time with the photocleavage of (**68**) (data shown for every 5 scans).

The reaction was monitored for 500 minutes. However the fluorescence emission continued to show an increasing maxima which indicates the reaction had not yet reached completion after 500 minutes. This was further confirmed by examining the fluorescence maxima at 400 nm over time (figure 3.27).

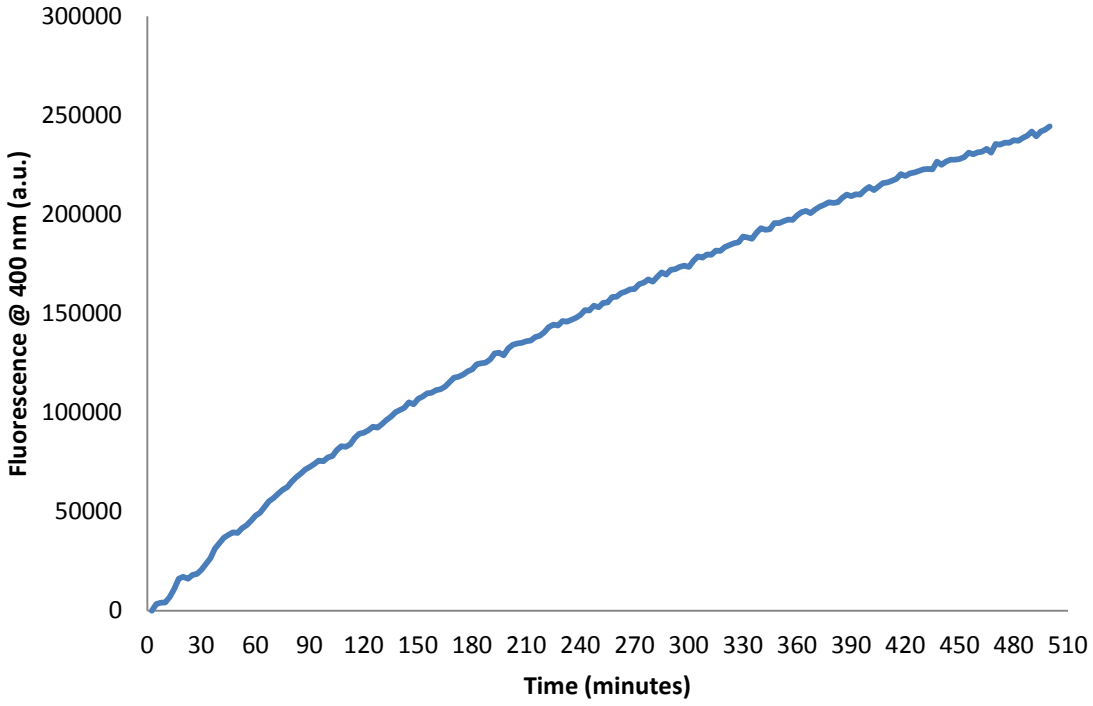


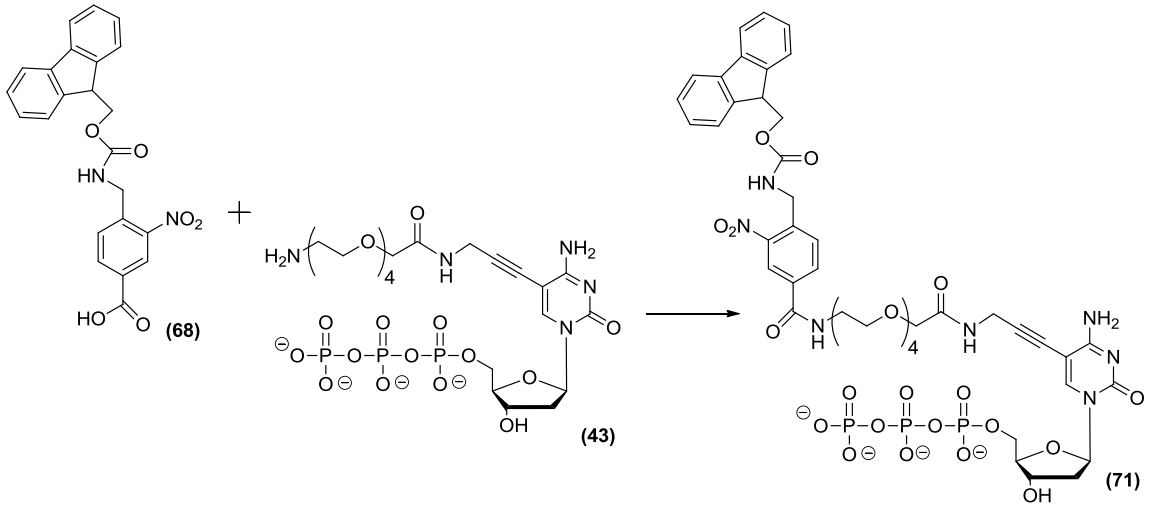
Figure 3.27: Fluorescence emission maxima at 400 nm over time for the photocleavage test reaction of (**68**).

If the cleavage reaction had been complete a levelling of the curve would have been expected as no further change would occur in the fluorescence emission when all starting material has been converted to the product.

In conclusion, although the reaction has not gone to completion the clear change in the emission over time indicates that cleavage is occurring. For future photolysis experiments the conditions of cleavage will require optimisation. Importantly the source of light for UV irradiation will need to be considered as this will influence the speed of photolysis.

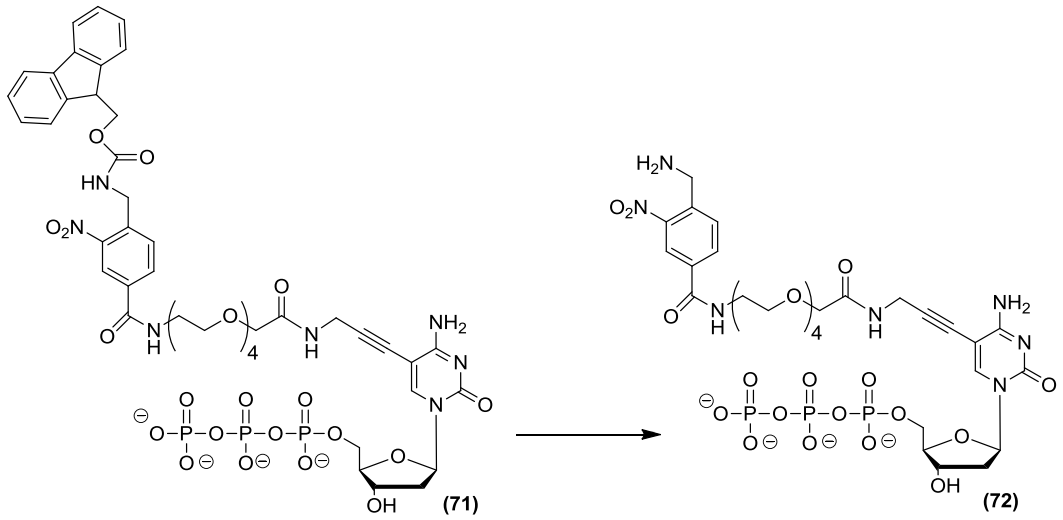
3.9 Synthesis of a cleavable dCTP

With the synthesis of the photocleavable unit complete, the amide coupling of (**68**) with the previously synthesised dCTP analogue (**43**) was undertaken. This was performed through a TSTU coupling and (**71**) isolated by prep-RP-HPLC (scheme 3.25).



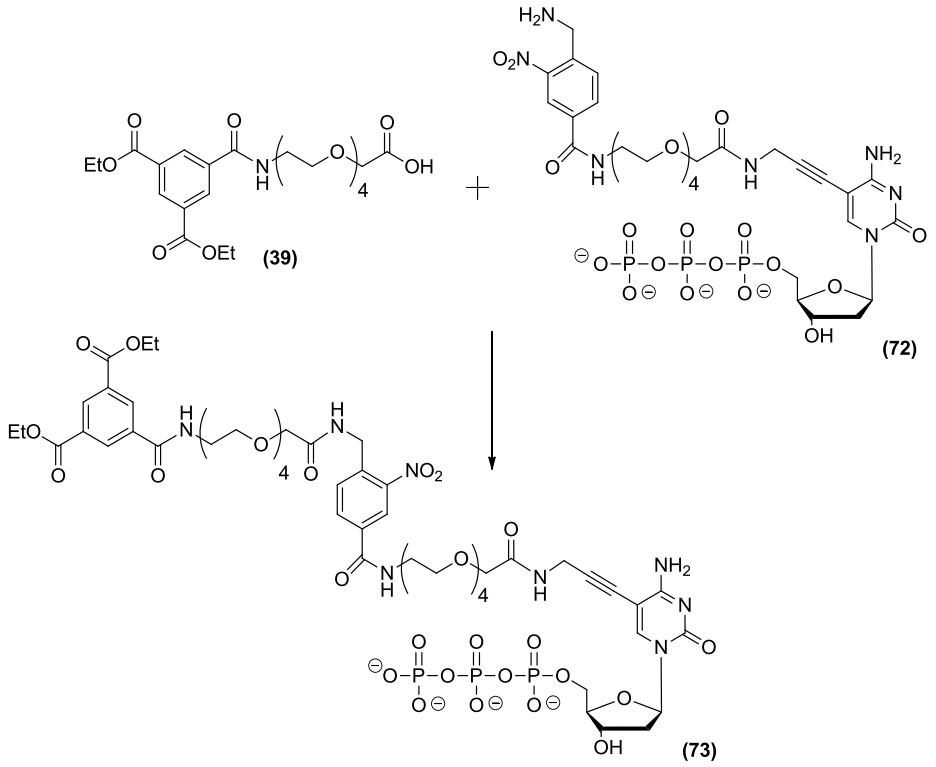
Scheme 3.35: TSTU coupling of the photocleavable (**68**) to dCTP (**43**). Reagents and conditions: (i) (**68**), TSTU, DIPEA, DMF, 3h, (ii) (**43**), 0.1 M sodium borate buffer (yield reported over two steps).

The modified dCTP (**71)** requires the addition of a reporter group to possess all appropriate functionality for use in SBS. Deprotection of the amine by the base-mediated removal of the FMOC protecting group allows subsequent functionalisation of the resulting primary amine with a reporter group. The FMOC deprotection was achieved by stirring (**71)** in 17.5% NH_4OH overnight. The desired product, (**72)**, was subsequently isolated by prep-RP-HPLC in a 35% yield over the TSTU coupling and FMOC deprotection (scheme 3.36).



Scheme 3.36: FMOC deprotection of (**71**). Reagents and conditions: 17.5% NH_4OH, overnight, 35% (over TSTU coupling and deprotection).

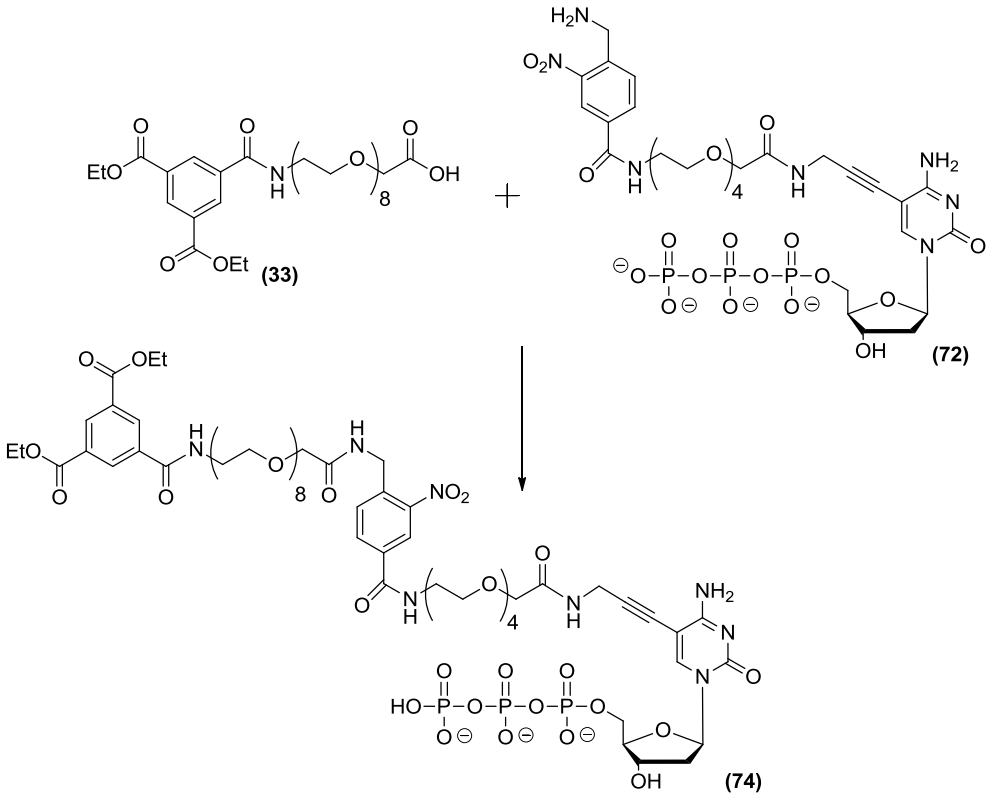
The previously synthesised (**39**) provided a suitable compound to introduce a reporter group and a further PEG_4 linker through a TSTU coupling (scheme 3.37). The modified-dCTP (**72)** was purified by prep-RP-HPLC with an isolated yield of 81%.



Scheme 3.37: TSTU coupling of the PEG_4 linker-reporter unit (**39**) to dCTP (**72)**. Reagents and conditions: (i) (**39**), TSTU, DIPEA, DMF, 3 h, (ii) (**72**), 0.1 M sodium borate buffer, 24 h, 81%.

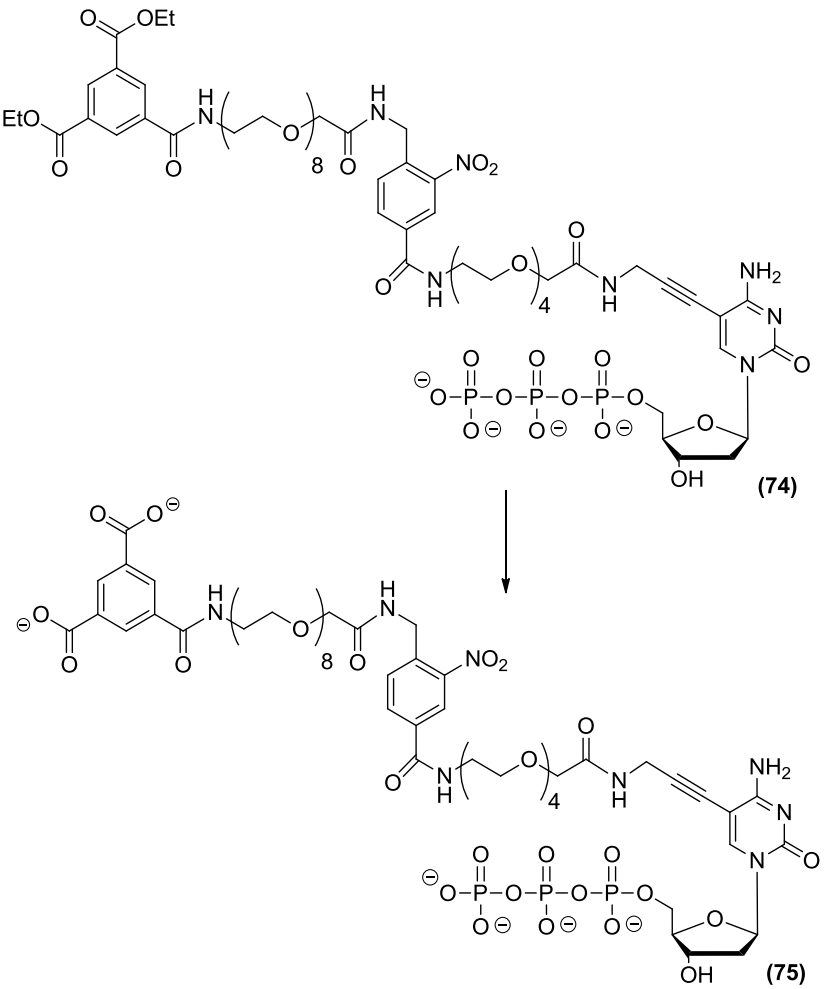
Unfortunately when ester hydrolysis of (**73**) was attempted (using 1 M NaOH for 2 h at room temperature) no desired material could be isolated after attempted HPLC purification. ^{1}H and ^{31}P NMR analysis of the isolated compound showed an aromatic compound with no phosphorus groups present. Mass spectrometry (ESI) analysis was unable to identify a mass relating to any of the likely molecule fragments caused by photolysis or decomposition.

Due to complications seen in the synthesis of (**73**) the amide coupling to attach the linker-reporter unit required repeating. As the entirety of the PEG_4-trimesic acid compound (**39**) had been used in previous coupling reactions the reaction was repeated using the comparative PEG_8-trimesic acid compound, (**33)** (scheme 3.38).



Scheme 3.38: TSTU coupling of PEG_8 linker-reporter unit (**33**) to dCTP (**72**). Reagents and conditions: (i) (**33**), TSTU, DIPEA, DMF, 3 h, (ii) (**72**), 0.1 M sodium borate buffer, 24 h, 4 %.

Compound (**74**) was isolated by prep-RP-HPLC as with the PEG_4 analogue (**73**), however the subsequent hydrolysis of the ethyl esters was performed with a decreased concentration of NaOH of 0.25 M (scheme 3.39).



Scheme 3.39: NaOH deprotection of (**74**). Reagents and conditions: 0.25 M NaOH, 1 h, 89%).

The reaction was monitored by RP-HPLC and after one hour of stirring at room temperature the cleavage reaction was complete (figure 3.28). The reaction mixture was then neutralised by the addition of 1 M acetic acid and the product purified by prep-RP-HPLC.

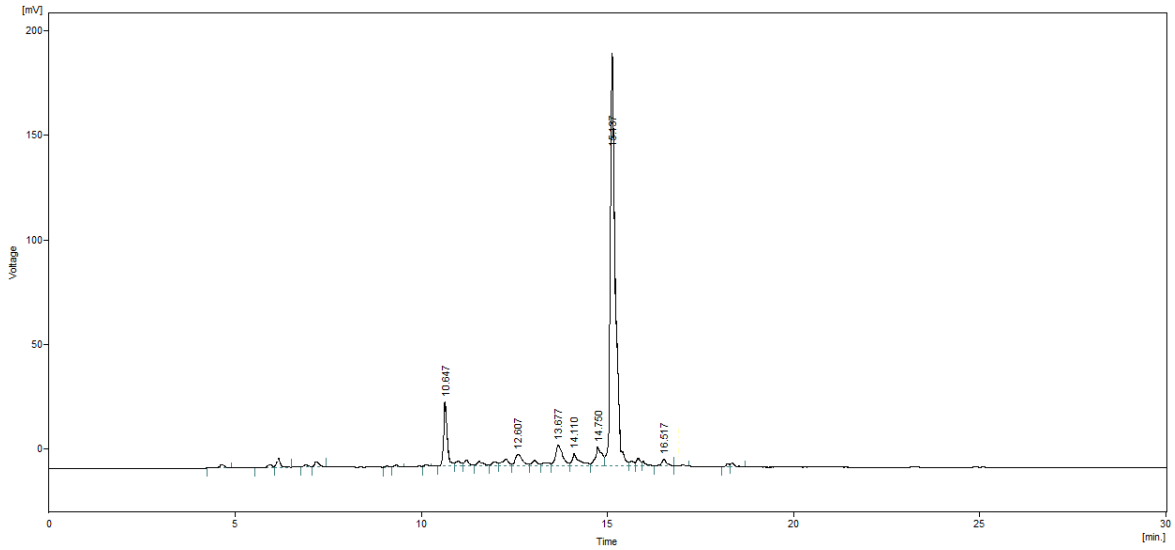


Figure 3.28: HPLC analysis of the crude reaction mixture following the deprotection of (**74**) to the dicarboxylate (**75**). Conditions: 5 – 100% B over 30 minutes where A= 0.1 M TEAB and B= 50 % MeCN/ 0.1 M TEAB. UV = 295 nm, flow rate = 1 ml / min. Retention time of product = 15.1 minutes.

It is noteworthy that due to the low yield of the TSTU coupling of (**33**) to (**74**) (scheme 3.38) it was difficult to obtain ^{31}P NMR data for the two final photocleavable dCTPs (**74**) and (**75**). For the latter compound ^{31}P NMR data was obtained by increasing the number of scans in the NMR experiment from 256 to 70,000. Although the splitting pattern of the three phosphorus's cannot be deduced, the ppm of the three peaks (-7.59, -10.95, -20.44) is typical of a triphosphate. When this data is combined with the electron spray ionisation (ESI) accurate mass, ^{1}H NMR and RP-HPLC data the identity of (**75**) was confidently determined to be the desired triphosphate.

The synthesis of a dCTP with a cleavable linker (**75**) has now been successfully demonstrated. The modified dCTP therefore requires testing as a DNA polymerase substrate to determine the suitability of compounds of its type in SBS.

3.10 Conclusions

The synthesis of seven modified dCTPs with novel C5 modifications has been described. The modifications used have been chosen to investigate the effect of linker length, reporter group size and charge and a photocleavable linker on the DNA polymerase substrate properties.

Two modified dCTPs (**34, 41**) have been synthesised with the trimesic acid reporter group tethered to CA-PEG linkers of differing lengths. The linkers used varied from the equivalent length of 6 base pairs (1.7 nm, $CA(PEG)_4$ linker) to 12 base pairs (3.6 nm, $CA(PEG)_8$ linker). The shorter PEG_4 linker (**29**) was also used to demonstrate the potential to couple short linkers together to make longer linkers. This was exemplified by the synthesis of the double length azido-$(PEG_4)_2$-dCTP (**45**).

Previously only the trimesic acid reporter group has been investigated as a suitable reporter group for use in the nanowire-based sequencing device. If a 3'-blocking group is used in the SBS device and sequencing performed in a 'one pot' fashion (with all four dNTPs present in the reaction mixture at once) four unique reporter groups, each associated to one of the four dNTPs, will be required. Therefore, the study of new reporter groups was of great interest. Reporter groups with a high degree of anionic charge and those where the charge could be easily varied were of particular interest. For this purpose four novel dCTPs were synthesised with reporter groups varying in anionic charge from -4 to -24. Two dendrimer reporter groups were synthesised (PAMAM 1-D2 (**52**) and PAMAM 1-D3 (**53**)) terminated with four or eight carboxylates respectively. Two ssDNA poly-2'-deoxythymidine ODNs (dT_6 (**68**) and dT_{24} (**61**)) have also been utilised as potential reporter groups, exploiting the intrinsically anionic phosphate diester backbone present. The use of ssDNA as a reporter group is of particular interest as the length, and therefore charge, can be easily adapted as required by changing the DNA length. Both the dendrimer reporter groups and the ssDNA were attached to an azide terminated linker at the C5 position of dCTP using Cu(I) catalysed click chemistry.

In the final SBS device the modified dNTPs will need to possess a cleavable moiety capable of removing the reporter group after dNTP incorporation into the growing DNA strand. The exploration of using a photocleavable linker positioned in-between the reporter group and DNA base has been investigated for this purpose. The synthesis of the

photocleavable, 2-nitrobenzyl linker (**67**) has been described and its subsequent addition to a modified dCTP successfully demonstrated. The coupling of the trimesic acid reporter group proved more challenging than expected but was achieved with the successful synthesis of (**75**).

All seven modified dCTP analogues will now be tested as substrates for a selection of DNA polymerases using primer extension reactions to ensure the suitability of the modifications present for potential use in QuantuMDxs SBS sequencing device.

Chapter 4 –*Biological studies*

4. Biological studies

4.1 Introduction

For use in SBS the previously synthesised modified dCTP analogues (chapter 3) must be substrates for a DNA polymerase. This is essential as the modified dCTP must be incorporated into the growing DNA strand to identify the presence of guanine on the template DNA strand being sequenced. As discussed, base-modified nucleotides are best tolerated by DNA polymerases when the modification is at the C5-position on the pyrimidines (section 1.7). This chapter will therefore focus on determining whether the C5-modified dCTP analogues described in chapter 3 are substrates for DNA polymerases. This will be assessed using primer extension reactions with a variety of polymerases. The other modified dNTPs required to determine the full sequence (modified dUTP, dGTP and dATP analogues) will not be discussed here but are in development by QuantuMDx and have been tested separately as substrates for DNA polymerases.

During the primer extension reaction a primer is bound to its complementary region on an oligonucleotide template and serves as the starting point for the synthesis of the complementary strand to the template. The template is designed in order to investigate the incorporation of the modified dCTP analogues, for example template HM284 requires the insertion of a single dCTP opposite the template guanine in the fifth position (figure 4.1). If the modified dCTP is a substrate for the DNA polymerase being tested it will be incorporated into the growing strand; if it is not the synthesis will stall at the dCTP incorporation, producing a primer product extended by only four nucleotides. It is also anticipated that the elongation of the primer after the incorporation of a modified dNTP may be challenging for a DNA polymerase due to the large C5-modification still being present.

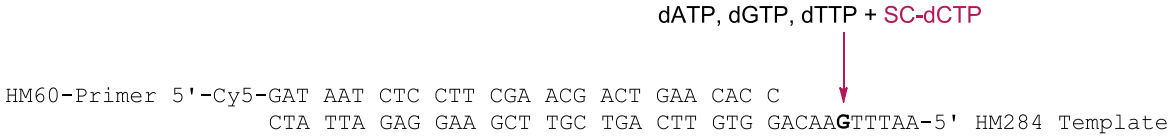


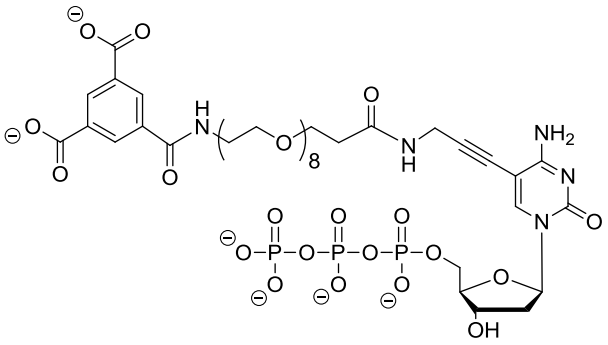
Figure 4.1: HM60 primer and HM284 template sequences

The template can be designed to investigate various potential scenarios that the polymerase may face during DNA sequencing. For example, as an alternative to the design and use of a 3'-blocking group it would be possible to instead introduce each modified dNTP to the sequencing reaction separately (as described in section 1.9); it would therefore be desirable for a polymerase to be able to incorporate multiple modified dNTPs consecutively. This would be required in the case of a homopolymeric region on the template DNA.

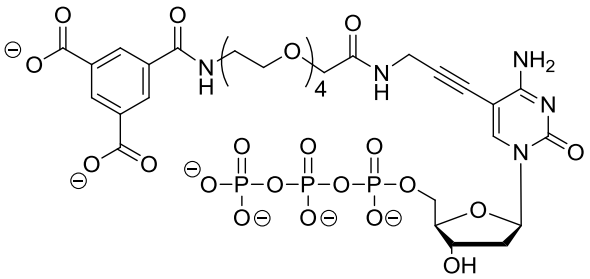
As the current focus is on synthesising proof-of-principle dCTP analogues, not all modified dCTP analogues will be suitable for use in the nanowire SBS sequencing device envisaged by QuantuMDx (section 1.9). This is due to only one dCTP (**75**) possessing a cleavable linker and as 3'-blocking groups have not been explored in our research this functionality is also absent from all modified dCTPs. However the information gathered by testing a variety of linker lengths and reporter groups will help guide the design of future cleavable dCTP analogues for use in the final SBS device. It is worth noting that, as the large C5 modifications will not be removed from the base, extension of the primer after incorporation of the modified dCTP analogue is not a prerequisite. In the final design (with the use of a 3'-blocking group) the bulk of the modification will be cleaved before the addition of the next dNTP, returning the modified dNTP to a structure closer to resembling the natural base.

C5-modified dCTPs for testing as substrates for DNA polymerases

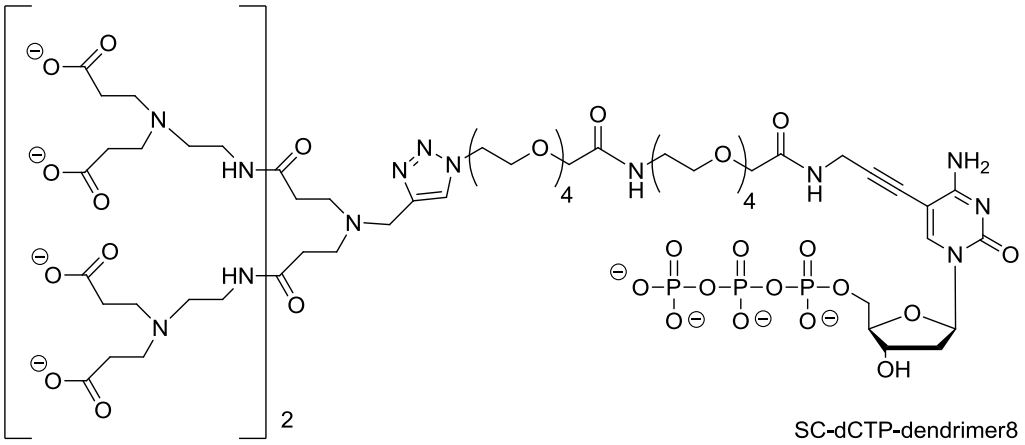
Seven modified dCTP analogues have been synthesised with the aim of investigating a number of variations of modifications: linker length, reporter groups and a cleavable moiety. For future reference modified dCTP analogues will be classified as SC-dCTP-PEG8, SC-dCTP-PEG4, SC-dCTP-dendrimer8, SC-dCTP-dendrimer4, SC-dCTP-ssdT24, SC-dCTP-ssdT6 and SC-dCTP-photocleavable, where the SC prefix relates to ‘super charged’ (figure 4.2).



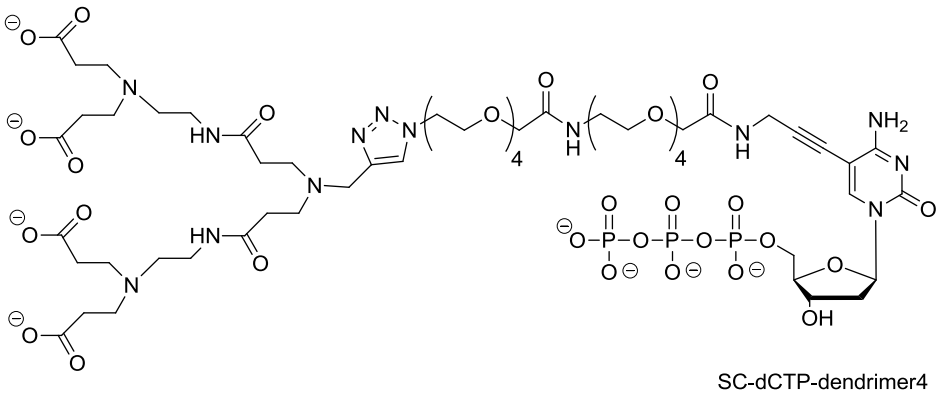
SC-dCTP-PEG8



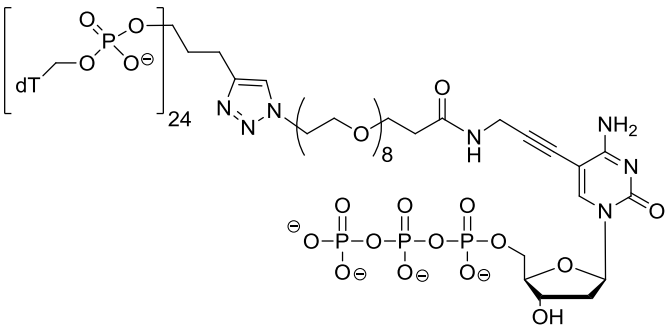
SC-dCTP-PEG4



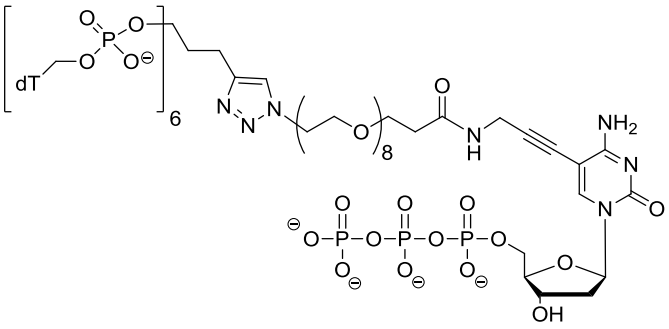
SC-dCTP-dendrimer8



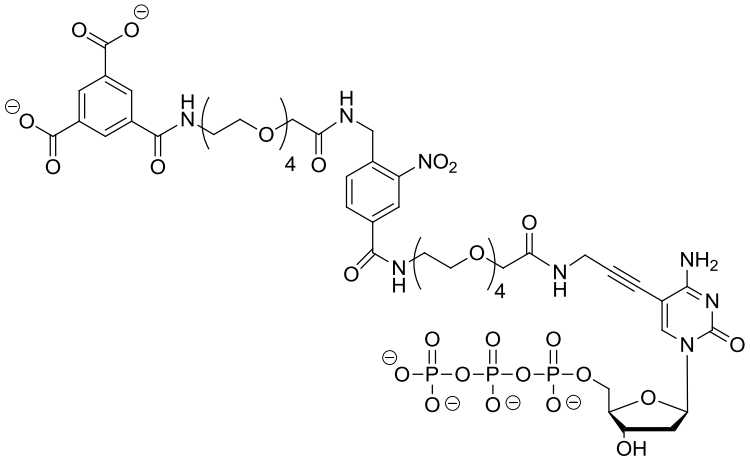
SC-dCTP-dendrimer4



SC-dCTP-ssdT24



SC-dCTP-ssdT6



SC-dCTP-photocleavable

Figure 4.2: The chemical structure of the seven SC-dCTP analogues for biological testing.

4.2 Initial primer extension assays: Polymerase screen for the incorporation of SC-dCTP-PEG8 and SC-dCTP-PEG4

Initial investigations focussed on the single incorporation of two of the simpler modified dCTP analogues synthesised, SC-dCTP-PEG8 and SC-dCTP-PEG4, as potential substrates for seven DNA polymerases. The template HM284 was used which has single guanine at the 5th position (figure 4.3).

```
HM60-Primer 5'-Cy5-GAT AAT CTC CTT CGA ACG ACT GAA CAC CTGTTCAAATT
                    CTA TTA GAG GAA GCT TGC TGA CTT GTG GACAAGTTTAA-5' HM284 Template
```

Figure 4.3: Extended product from the extension of the primer (HM60) using the template HM284. The position of the modified dCTP incorporation is highlighted in red with the complementary dNTPs to be added to the template shown in bold.

Primer extension reactions were carried out with the inclusion of three controls: all four natural dNTPs as a positive control for the synthesis of the fully extended product; the exclusion of dCTP in order to check the fidelity of the polymerase; and a negative control with the exclusion of DNA polymerase.

The regular primer extension reactions had 1 U polymerase in 1x polymerase specific buffer, 20 nM Cy5-5'-labelled primer and 15 nM template in a total volume of 20 µL. To investigate the incorporation of a modified dCTP a SC-dCTP mix (SC-dCTP, dATP, dGTP, dTTP) was included alongside the three controls described above. Although the concentration of a modified nucleotide is often increased in comparison to the natural dNTPs present an equal concentration of modified dCTP analogues to natural dNTPs was used, maintaining a total dNTP concentration of 625 nM. The incubation time and temperature for each primer extension reaction varied and is described for each primer extension reaction; however initial primer extension reactions were incubated at 40°C for 1 hour and 7 hours. After incubation, each reaction was stopped with the addition of complementary sequence to the template in a solution of EDTA and formamide. The reaction mixtures were denatured by heating at 95°C for 3 minutes and analysed by electrophoresis on a 17.5% polyacrylamide gel (PAGE). The gel was imaged by fluorescence visualisation facilitated by the use of the Cy5-labelled primer.

A variety of commercially available DNA polymerases were chosen to identify suitable polymerases for future primer extension assays (table 4.1). All except Phusion and Veraseq 2.0 have had their proof-reading (3'-5'-exonuclease) activity removed.

Polymerase	3'-5' exonuclease activity?	Thermostable?
HotStarTaq	No	Yes
Phusion	Yes	Yes
VeraSeq 2.0	Yes	Yes
Vent (exo-)	No	Yes
Omni- KlenTaq	No	Yes
Therminator	No	Yes
Klenow (exo-)	No	No

Table 4.1: Properties of DNA polymerases chosen for initial polymerase screen.

Our initial polymerase screen with SC-dCTP-PEG8 and SC-dCTP-PEG4 included HotstarTaq, Phusion, VeraSeq 2.0 and Vent (exo-) DNA polymerases (figure 4.4). Samples were incubated at 40°C for 1 hour.

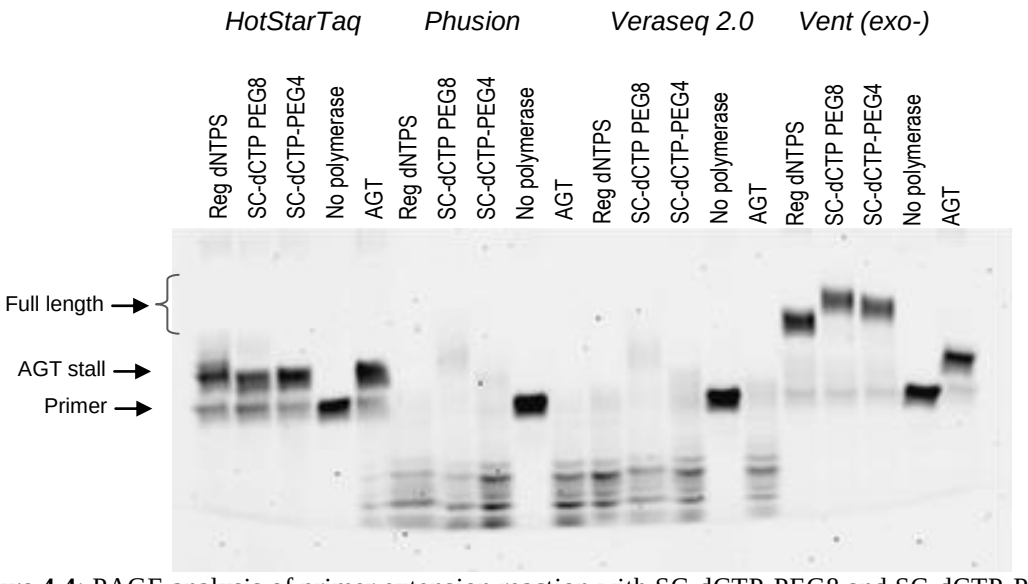


Figure 4.4: PAGE analysis of primer extension reaction with SC-dCTP-PEG8 and SC-dCTP-PEG4 with HotStarTaq (left), Phusion (middle left), VeraSeq 2.0 (middle right), Vent(exo-) (right) DNA polymerases. 40°C, 1h incubation time.

After incubating at 40°C for 1 hour Vent (exo-) appears to be the only suitable polymerase for the incorporation of SC-dCTP-PEG8 and SC-dCTP-PEG4 with good incorporation of both modified dCTPs observed (no stall bands seen). Vent (exo-) also demonstrates high fidelity with no misincorporation in the absence of dCTP. Phusion and VeraSeq 2.0 both have 3'-5' exonuclease activity and degrade the extended product, it is therefore not possible to determine from this result if Phusion or VeraSeq 2.0 incorporated either modified dCTP. HotStarTaq appears to be slow at the synthesis of the fully extended product with no extended product seen in the regular dNTP control and both SC-dCTP mixes stalling prior to the incorporation of the modified dCTP analogues. Again, it is not possible to determine if either modified dCTP analogues are substrates for HotStarTaq polymerase from this result. The incubation time was extended to 7 hours at 40°C to determine if either SC-dCTP analogues were substrates for HotStarTaq DNA polymerase (figure 4.5).

It is noteworthy that a shift in the positioning of the band can be seen in the fully extended products containing a modified dCTP. Where possible brackets have been used to indicate the region where the fully extended product, containing either regular dNTPs or where dCTP has been replaced with modified dCTP, can be found. This variation in mobility is due to the large modifications increasing the overall size of the extended product and causing decreased mobility. The difference in mobility can be translated into the size of the modification present. This is observed in the case of the extended ODN containing either the SC-dCTP-PEG8 or SC-dCTP-PEG4 analogue; with the ODN containing the larger PEG_8 modified dCTP analogue having a lower mobility through the gel than the ODN containing the smaller PEG_4 modified dCTP analogue.

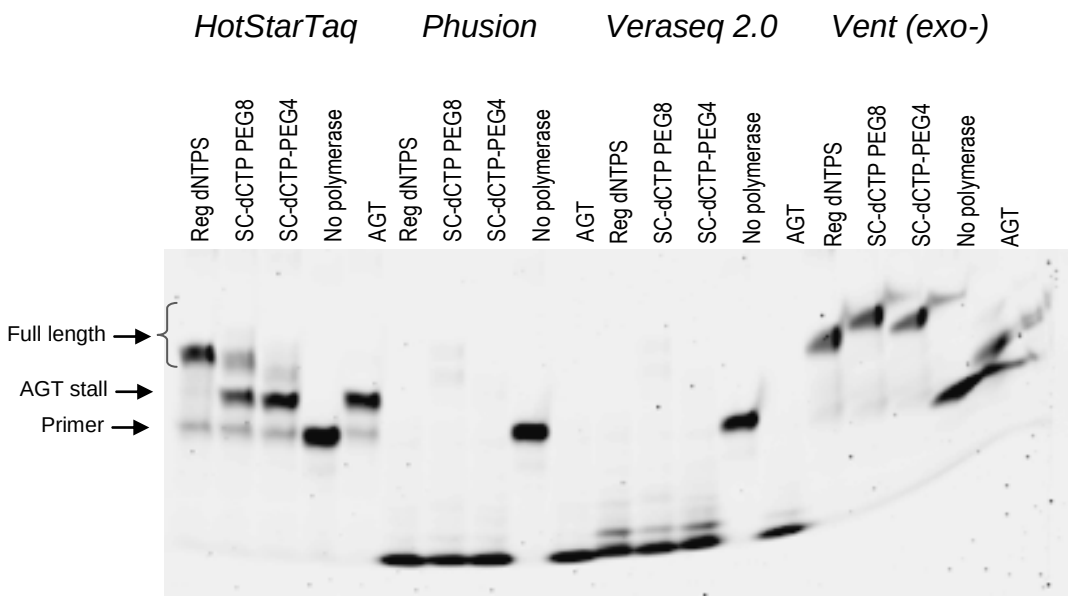


Figure 4.5: PAGE analysis of primer extension reaction with SC-dCTP-PEG8 and SC-dCTP-PEG4 with HotStarTaq (left), Phusion (middle left), VeraSeq 2.0 (middle right), Vent(exo-) (right) DNA polymerases. 40°C, 7 h incubation time.

After extending the incubation time HotStarTaq polymerase has synthesised the fully extended product in the presence of regular dNTPs, however no significant extension product can be seen when natural dCTP has been replaced with SC-dCTP-PEG8 or SC-dCTP-PEG4. This can be determined by a strong stall band in line with the dATP, dGTP, dTTP control. Vent (exo-) remains a good polymerase choice for investigating the incorporation of the other modified dCTP analogues as it shows good incorporation of both modified dCTP analogues with no misincorporation with the dATP, dGTP, dTTP control. With the extended incubation time complete degradation of the primer template sequence can been seen with both Phusion and VeraSeq 2.0 (figure 4.5).

Omni-Klentaq, Therminator and Klenow (exo-) were also investigated as suitable DNA polymerases for use with modified dCTP analogues. Therminator was of particular interest as it is in the 9°N family of polymerases which have been designed specifically for the incorporation of modified nucleotides and it is the polymerase used in Illumina's sequencing methodology.[108]

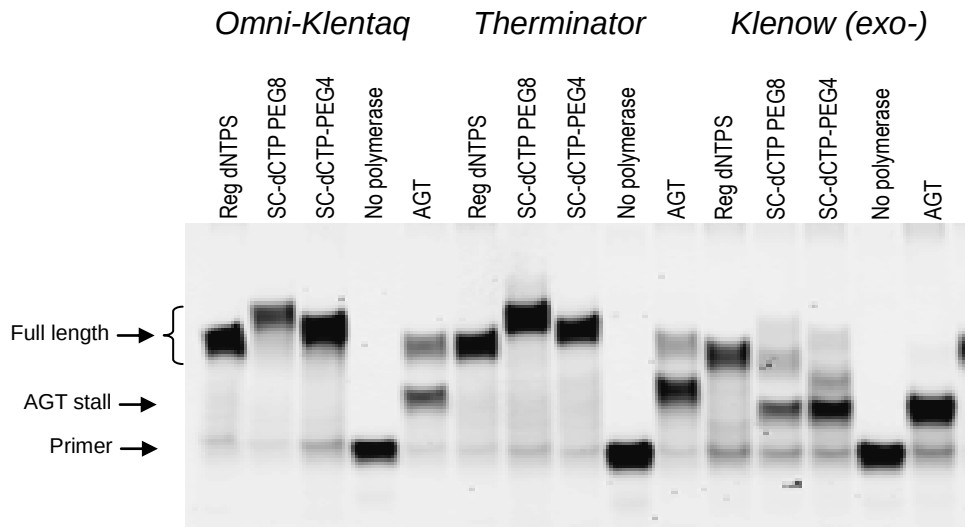


Figure 4.6: PAGE analysis of primer extension reaction with SC-dCTP-PEG8 and SC-dCTP-PEG4 using Omni-Klentaq (left), Therminator (middle) and Klenow(exo-) (right) DNA polymerases. 40°C, 1 h incubation time.

After 1 hour of incubation at 40°C both SC-dCTP-PEG8 and SC-dCTP-PEG4 are successfully incorporated by Omni-Klentaq and Terminator DNA polymerases (figure 4.6). Though a small amount of misincorporation can be seen, more significantly with the use of Omni-Klentaq, this is to be expected due to the lack of 3'-5' exonuclease activity. Klenow (exo-) has not managed to synthesise the fully extended product within the incubation time for either SC-dCTP-PEG8 or SC-dCTP-PEG4, this may be as they are not substrates for Klenow (exo-) or it may be due to an insufficient incubation time or temperature. The incubation was extended to 7 hours in order to determine if either SC-dCTP analogues were substrates for Klenow (exo-) (figure 4.7).

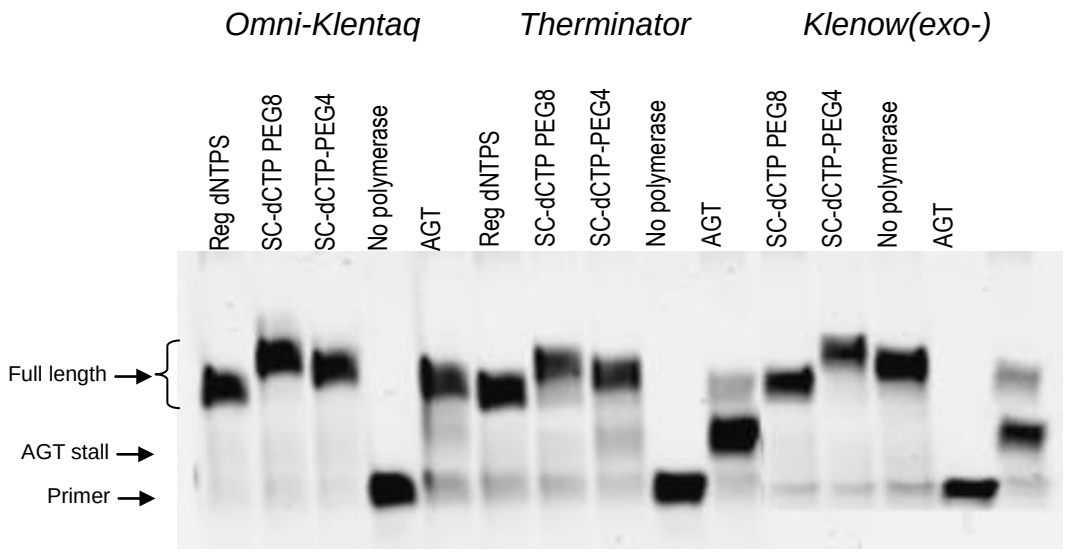


Figure 4.7: PAGE analysis of primer extension reaction with SC-dCTP-PEG8 and SC-dCTP-PEG4 using Omno-Klentaq (left), Therminator (middle) and Klenow(exo-) (right) DNA polymerases. 40°C, 7 h incubation time.

Following an extended incubation time (7 hours) both modified dCTP analogues were incorporated by Klenow(exo-) DNA polymerase. Omni-klentaq however showed disappointing fidelity, with a fully extended product seen in the dATP, dGTP, dTTP control, indicating that another nucleotide has been misincorporated in place of the dCTP. Therminator did not appear to show an increased level of misincorporation from a 1 hour to a 7 hour incubation time (figure 4.7).

From the initial polymerase screenings it was decided to test the remaining five dCTP analogues as substrates for Vent (exo-) and Therminator DNA polymerases. Although both SC-dCTP-PEG analogues were substrates for Omni-Klentaq and Klenow (exo-) DNA polymerases, Omni-Klentaq showed significant levels of misincorporation over time and Klenow (exo-) was slow in the synthesis of the extended product. As fidelity and speed are of key importance to the selection of a suitable DNA polymerase both were excluded from future studies.

4.3 Real-time primer extension assays

To determine if either SC-dCTP-PEG8 or SC-dCTP-PEG4 were incorporated by Phusion or Veraseq 2.0 polymerases prior to the degradation of the product, real time primer extension reactions were performed using an ABI-rotorgene.

To repeat the primer extension reaction SYBR green and an unlabelled primer (HM171) were used in place of the Cy5-labelled primer HM60. The template chosen (HM126) requires the single incorporation of a modified dCTP (figure 4.8), as with template HM284.

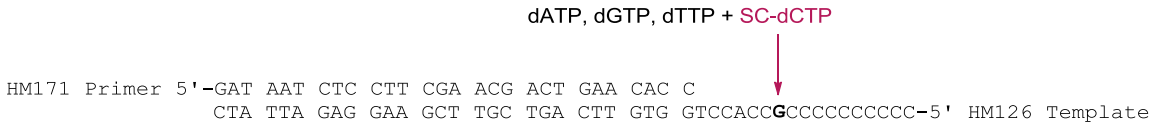


Figure 4.8: HM171 Primer and corresponding HM126 template sequence.

SYBR-green binds selectively to double stranded DNA through intercalation. This results in an increase in fluorescence which can be used to infer the formation of double stranded DNA and therefore successful primer extension. As SYBR-green shows minimal fluorescence when unbound in solution, the addition of SYBR-green is a powerful tool with which to investigate primer extension.

Each experiment was run in an analogous manner to the primer extension assays described in section 4.2, however rather than incubating the reactions and then analysing via PAGE the incubation and analysis stages are combined and a real-time output of fluorescence intensity measured. The reactions were monitored for 4 hours at an incubation temperature of 40°C. The inclusion of three controls was utilised; An AGTC (dATP, dGTP, dTTP and dCTP) positive control, an AGT (dATP, dGTP, dTTP) control to investigate the fidelity of the polymerase and a no dNTP negative control in order to ensure no binding of SYBR-green to the primer-template.

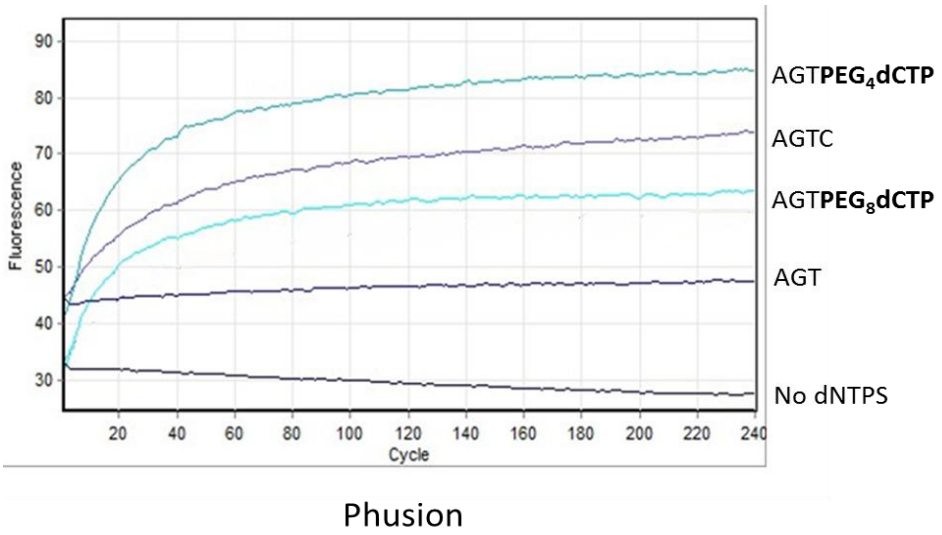


Figure 4.9: Real time primer extension of SC-dCTP-PEG8 and SC-dCTP-PEG4 using Phusion DNA polymerase monitored with the use of an ABI-rotorgene.

From the fluorescence data it can be determined that both SC-dCTP-PEG8 and SC-dCTP-PEG4 are substrates for Phusion polymerase due to the clear increase in fluorescence as the primer is extended (figure 4.9). Phusion also shows high fidelity with no increase in fluorescence in the negative (dATP, dGTP, dTTP) control.

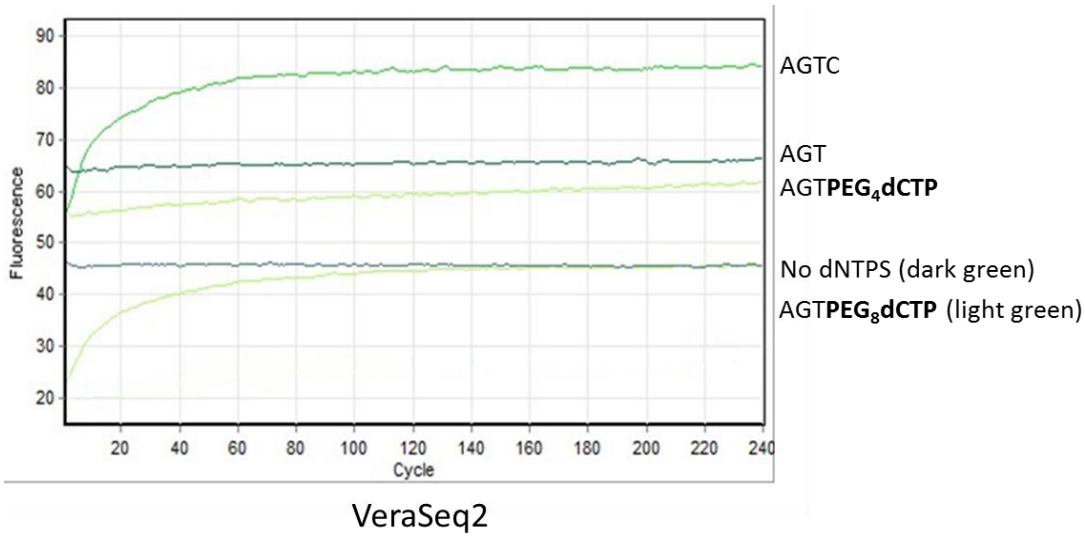


Figure 4.10: Real-time primer extension of SC-dCTP-PEG8 and SC-dCTP-PEG4 using VeraSeq 2.0 polymerase monitored with the use of an ABI-rotorgene.

The data from the real-time primer extension assays using Veraseq 2.0 polymerase implies that only the longer SC-dCTP-PEG8 analogue is a substrate with very little increase in fluorescence seen for incorporation of the SC-dCTP-PEG4 analogue (figure 4.10). Given the findings by Waggonner and Quake[87, 88] previously discussed (chapter 3, section A) this is likely due to the dCTP modified with the longer PEG_8 linker possessing suitable properties to be a polymerase substrate for Veraseq 2.0 but not the dCTP modified with the shorter PEG_4 linker. The high fluorescent value for the fidelity and negative control can be likely attributed to a discrepancy in the amount of SYBR-green present causing a variation in background fluorescence.

4.4 Primer extension reactions: Incorporation of alternative reporter groups

Testing SC-dCTP-dendrimer8 and SC-dCTP-dendrimer4 as DNA polymerase substrates

Initially investigating single dCTP incorporation, the substrate properties of the two dCTP analogues with dendrimer-based reporter groups (SC-dCTP-dendrimer8 and SC-dCTP-dendrimer4, figure 4.2) were examined. These are the first compounds to be investigated for use in the nanowire based sequencing device, not only by ourselves but also by QuantuMDx, that do not contain the trimesic acid reporter group. Dendrimer functionalised nucleotides have been described in the literature by Marx *et al.*[50] who demonstrated the single incorporation of a dendrimer modified dTTP analogue (figure 4.11) by KlenTaq DNA polymerase.

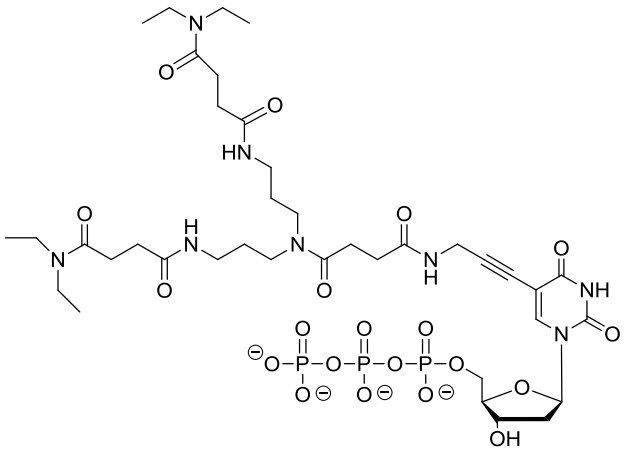


Figure 4.11: dTTP-dendrimer as synthesised by Marx *et al.*[50] found to be a substrate for KlenTaq DNA polymerase.

Marx *et al.*[50] have shown that dendrimer modified dTTPs are substrates for KlenTaq DNA polymerase and therefore it is promising that modifications of this type may be suitable as reporter groups. However, it is worth bearing in mind the dendrimer modified dCTP analogues synthesised (SC-dCTP-dendrimer8 and SC-dCTP-dendrimer4) are larger in size and polyanionic, terminated with either 4 or 8 carboxylates, which could affect the substrate properties.

As with SC-dCTP-PEG4 and SC-dCTP-PEG8 analogues the template HM284 was used to investigate whether SC-dCTP-dendrimer8 and SC-dCTP-dendrimer4 were substrates for Therminator and Vent (exo-) DNA polymerases through a single dCTP incorporation (figure 4.12). As work within the QuantuMDx lab has focussed on sequencing at ambient temperatures (25°C) the temperature of incubation was lowered to 25°C for this data to be comparable.

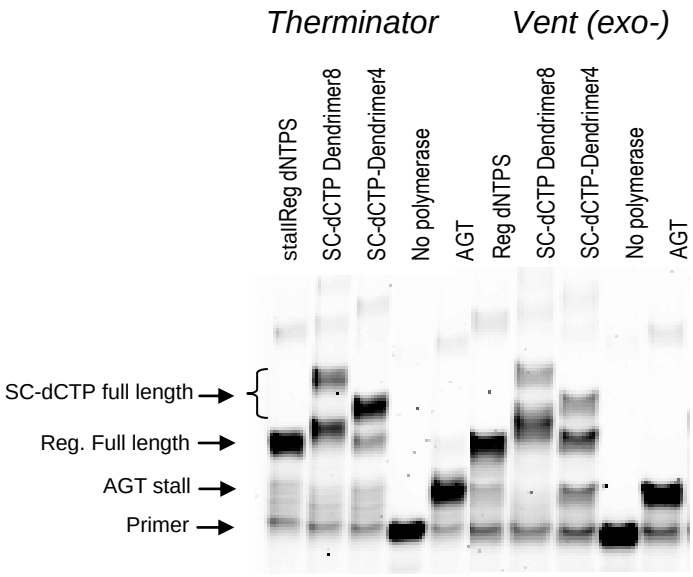


Figure 4.12: PAGE analysis of primer extension reaction with SC-dCTP-Dendrimer8 and SC-dCTP-Dendrimer4 using Therminator (left), Vent (exo-) (right). HM284, 25°C, 1 h incubation time.

Both SC-dCTP-dendrimer8 and SC-dCTP-dendrimer4 were successfully incorporated into the template by Therminator and Vent (exo-) DNA polymerases (figure 4.12). However, two bands are visible with the incorporation of both SC-dCTP-dendrimer analogues, which show the presence of two different length extension products. This can be postulated as the highest band demonstrating the incorporation of the SC-dCTP and subsequent extension and the second, lower band present due to the modified dCTP analogue being incorporated but the primer not being extended afterwards. It is noteworthy that due to the presence of the modified dCTP analogue causing a variation in mobility of the product, the positioning of the stalled, second band appears to be similar in the length to the regular dNTP control but is in fact five nucleotides shorter. Indeed a characteristic shift difference between the two bands, with the larger SC-dCTP-dendrimer8 stall observed higher up the gel and the SC-dCTP-dendrimer4 slightly lower, would support the presence of a SC-dCTP-dendrimer analogue and that this band is indeed due to stalling of the extension.

Therminator appears to extend a higher proportion of the primer to the fully extended product; however both it and Vent (exo-) show the ability to extend the primer fully. A longer extension period or incubating at a higher temperature may lead to higher proportions of the extended product.

Since both SC-dCTP-Dendrimers were substrates for Therminator and Vent (exo-) polymerases we were interested in whether they could be incorporated multiple times. Thus, an alternative template HM245 was used, which has three guanines at the 3rd, 6th and 11th position (figure 4.13).

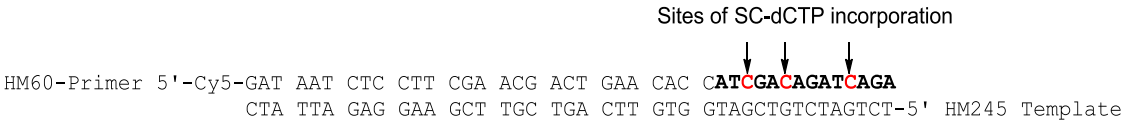


Figure 4.13: Extended product for the primer extension product using primer HM60 and template HM245. With the sites of incorporation of modified dCTP analogues highlighted in red and the complementary dNTPs to be added to the primer shown in bold.

The following primer extension reactions were performed as described previously but with the use of HM245 as the template (figure 4.13). As expected given the results with the template HM284 (figure 4.12) both polymerases struggled to incorporate either SC-dCTP-dendrimer8 or SC-dCTP-dendrimer4 in all three positions, thus synthesising the fully extended product. This can be seen with the incorporation of the SC-dCTP-dendrimer4 using Vent (exo-) where three clear stall bands can be observed (figure 4.14). Although it is clear that both polymerases struggled, multiple incorporations of the modified SC-dCTP dendrimer analogues is possible, with Vent (exo-) performing better than Therminator.

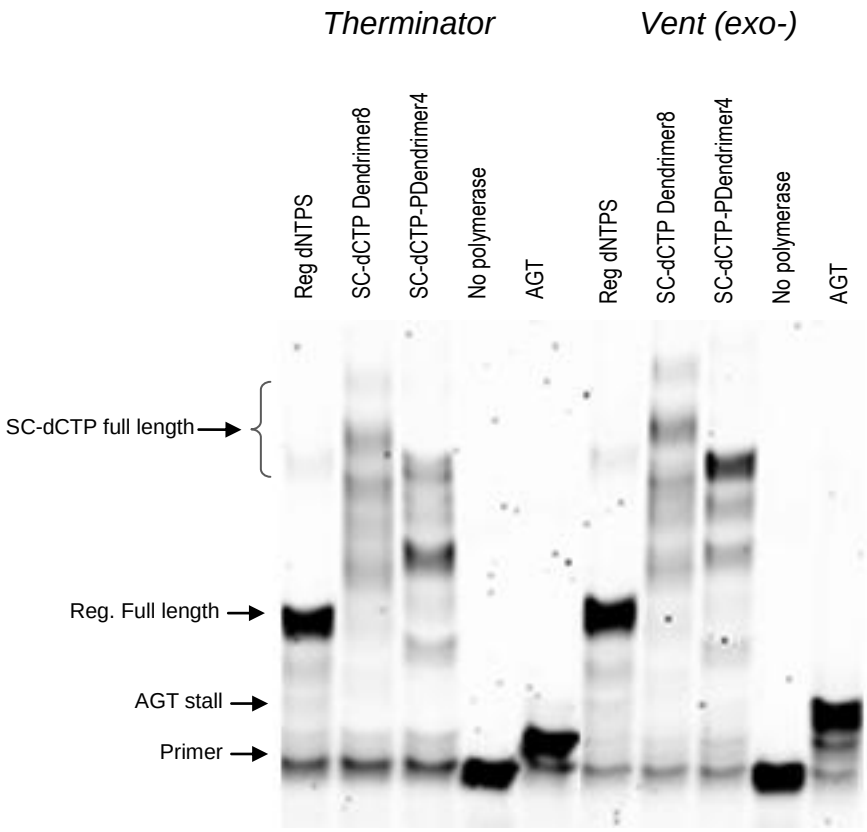


Figure 4.14: PAGE analysis of primer extension reaction with SC-dCTP-Dendrimer8 and SC-dCTP-Dendrimer4 using Therminator (left), Vent (exo-) (right). HM245, 25°C, 1 h incubation time.

Testing SC-dCTP-ssdT24 and SC-dCTP-ssdT6 as DNA polymerase substrates

The use of single stranded DNA as a means of introducing linear reporter groups with varying anionic charge was investigated as this is a modification that can be easily adapted to introduce varying levels of anionic charge. Marx *et al.*[99] described the successful incorporation of C5-ODN-modified dTTP analogues using both Therminator and KlenTaq DNA polymerases. Thus initial primer extension reactions were performed on the larger SC-dCTP-ssT24 analogue synthesised according to their published protocol.[99]

In a variation from the primer extension reactions previously performed the primer was used at a concentration of 3 nM with a template concentration of 5 nM. The nucleotide concentration was adjusted accordingly and decreased to a concentration of 250 nM. The polymerase concentration was varied from the protocol with 0.5 U being used for each

primer extension, a tenfold increase from that used by Marx *et al.*[99] The overall volume was kept the same with a total volume of 20 µL.

SC-dCTP-ssdT24 was initially screened as a substrate for Therminator polymerase. The reaction was incubated at 60°C and stopped after 2 minutes and 1 hour. From figure 4.15 is can be seen that extension is mostly complete after 2 minutes and appears fully extended after 1 hour however, after 1 hour incubation misincorporation becomes significant. The previously discussed (section 4.2) difference in mobility of extended products containing modified dNTPS can be seen with the incorporation of SC-dCTP-ssdT24 causing a significant decrease in the mobility of the product. This results in a large shift between the extended product with the SC-dCTP-ssdT24 incorporated and that of the extended product with natural dCTP (the positive, regular dNTP control). This same observation has been made previously by Marx *et al.* with ODN modified dTTPs.[99]

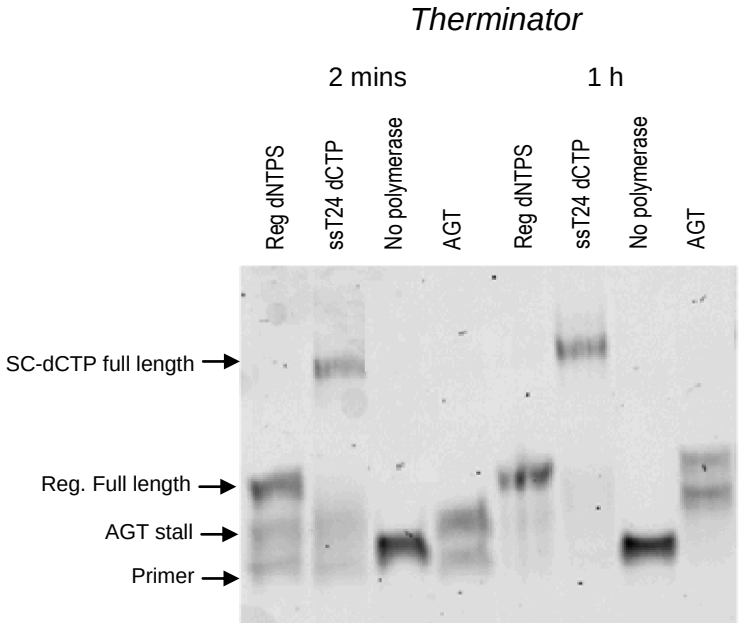


Figure 4.15: PAGE analysis of primer extension reaction with SC-dCTP-ssdT24 using Therminator DNA polymerase and template HM284; 2 mins at 60°C (left) and 1 h at 60°C (right).

It was of interest to see if an increased temperature was required for the incorporation of the modified dCTP analogue. Therefore the primer extension reaction was repeated following the standard primer extension conditions of 1 h incubation at 25°C (section 4.2) with both Vent (exo-) and Therminator DNA polymerases (figure 4.16).

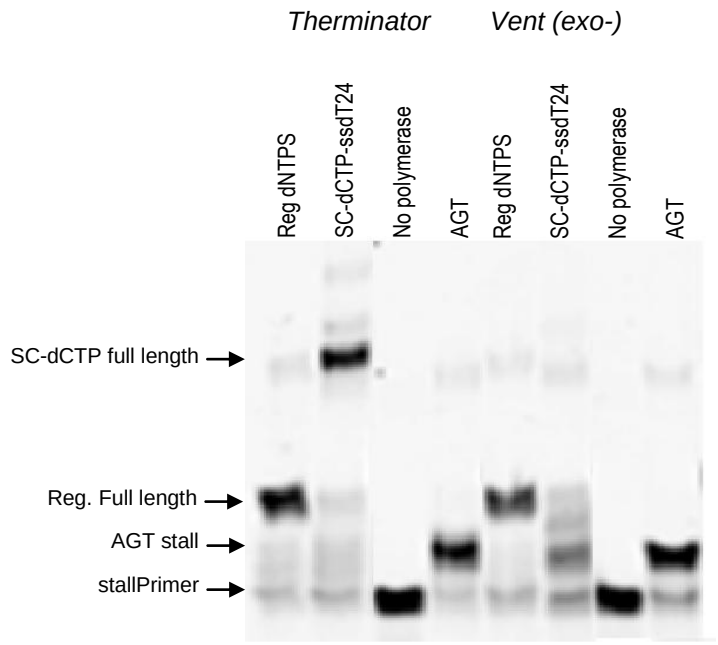


Figure 4.16: PAGE analysis of primer extension reaction with SC-dCTP-ssdT24 using Therminator (left), Vent (exo-) (right). HM284, 25°C, 1 h incubation time.

SC-dCTP-ssdT24 remained a substrate for Therminator DNA polymerase at decreased incubation temperature with full extension being seen after 1 hour at 25°C. However it was found that it is not a substrate for Vent (exo-) polymerase under the same conditions, with no extension being seen (figure 4.16).

We then investigated whether multiple incorporations of the modified SC-dCTP-ssdT24 analogue were possible under the same conditions using the HM245 template (figure 4.17). Due to the large shift in mobility of the SC-dCTP-ssdT24 analogue the results from the elongation of template HM284 are used as a comparison in order to distinguish between single and multiple incorporations. With the use of Therminator DNA polymerase it appears that the SC-dCTP-ssdT24 analogue stalls elongation after a single incorporation (figure 4.17), however further tests will need to be performed in order to determine if the incorporation of SC-dCTP-ssdT24 completely stalls elongation or if the DNA polymerase can extend post-incorporation but cannot incorporate two SC-dCTP-ssdT24 analogues in close proximity.

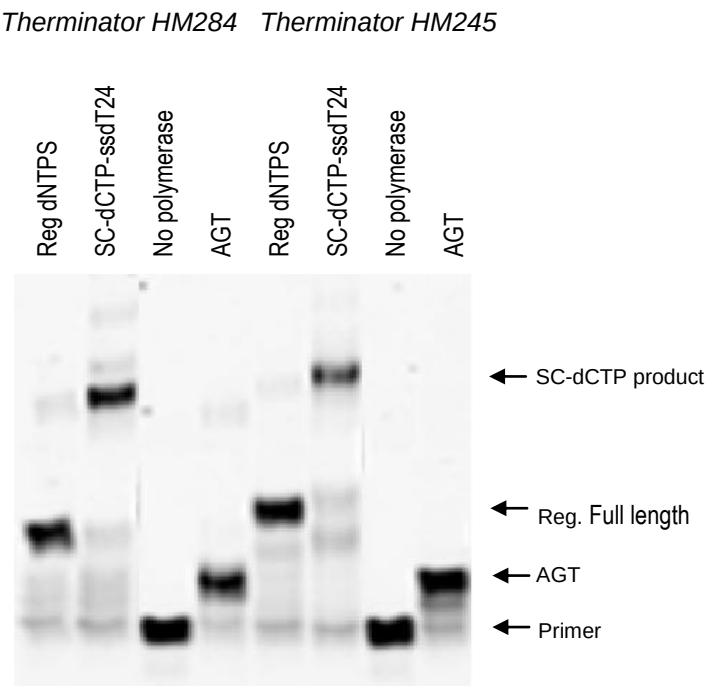


Figure 4.17: PAGE analysis of primer extension reaction with templates HM284 (single incorporation) and HM245 (three SC-dCTP incorporations) in the presence of SC-dCTP-ssdT24 using Therminator DNA polymerase. Incubated at 25°C for 1 h.

With the successful incorporation of SC-dCTP-ssdT24 using Therminator DNA polymerase established an investigation of the substrate properties of the smaller oligonucleotide modified dCTP (SC-dCTP-ssdT6) was examined using Therminator and Vent (exo-) polymerases. Due to the previous lack of incorporation of the related SC-dCTP-ssdT24 with Vent (exo-) after 1 hour at 25°C (figure 4.16) an extension of the incubation period to 18 h at 40°C was performed and SC-dCTP-ssdT24 re-examined as a substrate for both polymerases under these alternative conditions. For these reactions template HM224 was used which has a dCTP incorporation in the first position (figure 4.18).

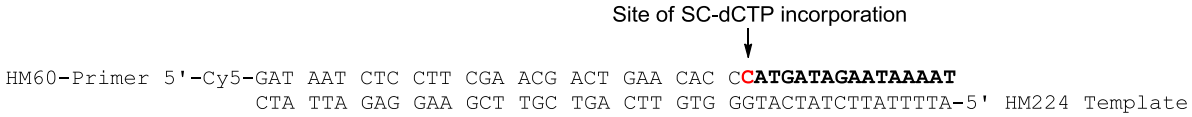


Figure 4.18: Extended product for the primer extension product using primer HM60 and template HM224. With the site of incorporation of modified dCTP analogues highlighted in red and the complementary dNTPs to be added to the primer shown in bold.

Using the extended incubation time at a higher temperature it was found that both SC-dCTP-ssdT analogues were substrates for both polymerases (figure 4.19). This indicates that the previous lack of incorporation of the SC-dCTP-ssdT24 was due to poor kinetics. The pattern of retardation between extended product containing the dCTP with the longer oligonucleotide (SC-dCTP-ssdT24) and the shorter SC-dCTP-ssT6 analogue is as expected, with the shorter modification having greater mobility through the gel.

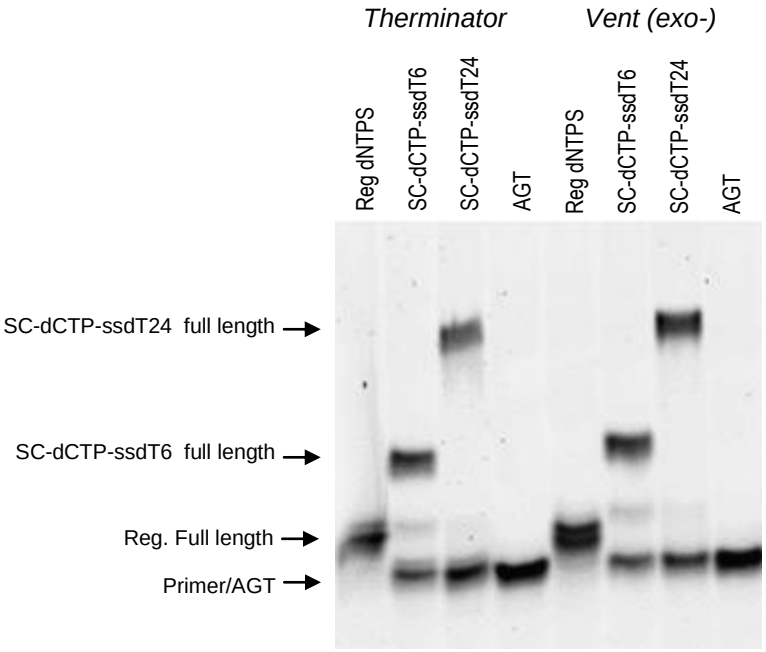


Figure 4.19: PAGE analysis of primer extension reaction using SC-dCTP-ssdT6 and SC-dCTP-ssdT24 with template HM224 using Therminator (left) and Vent (exo-) (right).
. Incubated at 40°C for 18 h.

4.5 Primer extension reactions: Incorporation of a photocleavable dCTP

QuantuMDx have predominately focussed on the use of modified dNTPs containing chemically cleavable reporter groups. They have synthesised the trimesic acid derivative of the Illumina linker and coupled this to the anchor modified dNTPs in order to synthesise cleavable dNTP analogues containing a anionic reporter group (figure 4.20). Cleavage of a compound of this type has been described in figure 3.19.

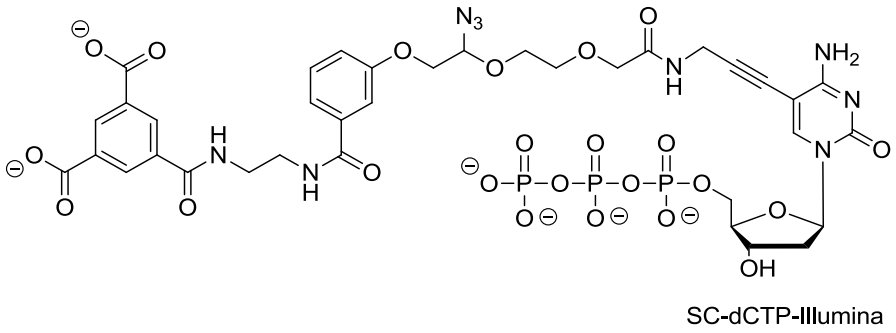


Figure 4.20: Chemically cleavable SC-dCTP-illumina as synthesised by QuantuMDx.

The dCTP derivative, SC-dCTP-Illumina (shown in figure 4.20), has been shown to be a substrate for Vent (exo-). After 30 minutes incubation at 25°C extension is predominately complete with full extension achieved within 4 hours (figure 4.21).

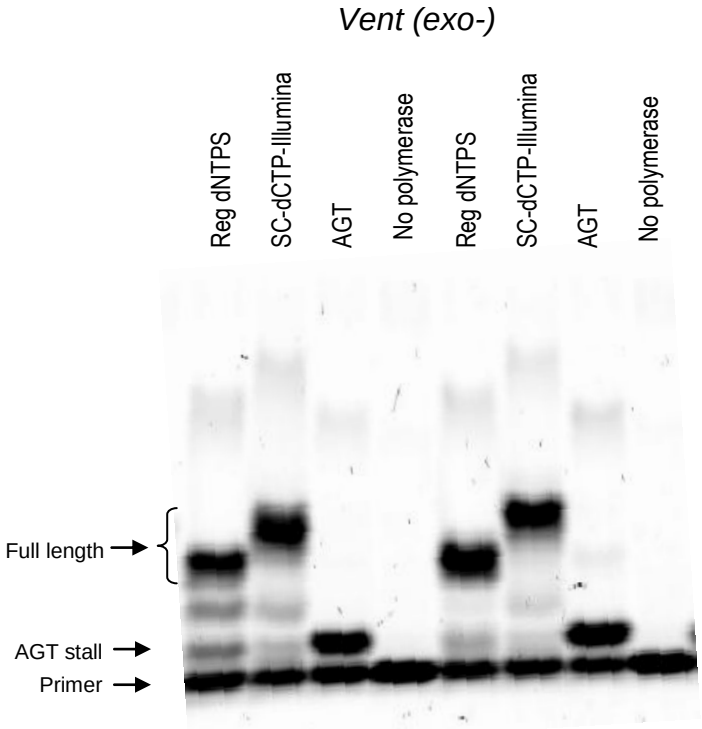


Figure 4.21: PAGE analysis of primer extension reaction with SC-dCTP-Illumina using Vent (exo-) and the template HM284: 25°C, 30min (left), 25°C, 4 h (right). Performed By Heather Murton, QuantuMDx.

Synthesis of a cleavable linker which can be positioned in-between the reporter group and the DNA base is of particular interest as it will allow a route to simple addition of the required cleavage functionality. For this purpose the synthesis of a 2-nitrobenzyl moiety placed between PEG_4 linkers to synthesise the SC-dCTP-photocleavable analogue (figure 4.2, referred to within figures as SC-dCTP-PC) has been described (chapter 3, section C). This was subsequently tested to ensure the modified dCTP was a polymerase substrate.

By investigating single dCTP incorporation using template HM284 (figure 4.3) with the use of Therminator and Vent (exo-) polymerases it was found that Therminator successfully incorporated the SC-dCTP-photocleavable analogue however Vent (exo-) struggled to extend either the natural dNTP control or incorporate the SC-dCTP-photocleavable analogue within 1 hour at 25°C (figure 4.22).

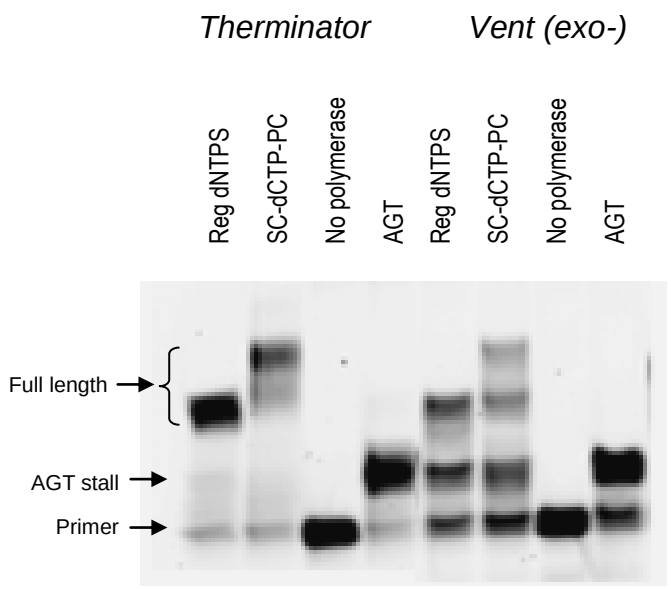


Figure 4.22: PAGE analysis of primer extension reaction with SC-dCTP-photocleavable using Therminator (left), Vent (exo-) (right). HM284, 25°C, 1 h incubation time.

Due to the lack of full extension seen using Vent (exo-) polymerase after 1 hour at 25°C the same temperature was maintained however the incubation time extended to 7 hours. After this extended reaction time Vent (exo-) successfully incorporated the modified dCTP and fully extended the positive control to give the both full length products (figure 4.23).

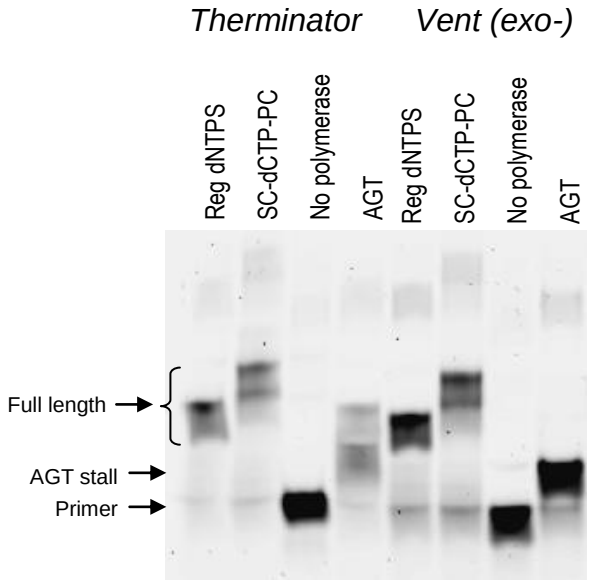


Figure 4.23: PAGE analysis of primer extension reaction with SC-dCTP-photocleavable using Therminator (left), Vent (exo-) (right). HM284, 25°C, 7 h incubation time.

The ability of both polymerases to incorporate multiple SC-dCTP-photocleavable analogues using the template HM245 after 1 h at 25°C (figure 4.13) was subsequently investigated (figure 4.24).

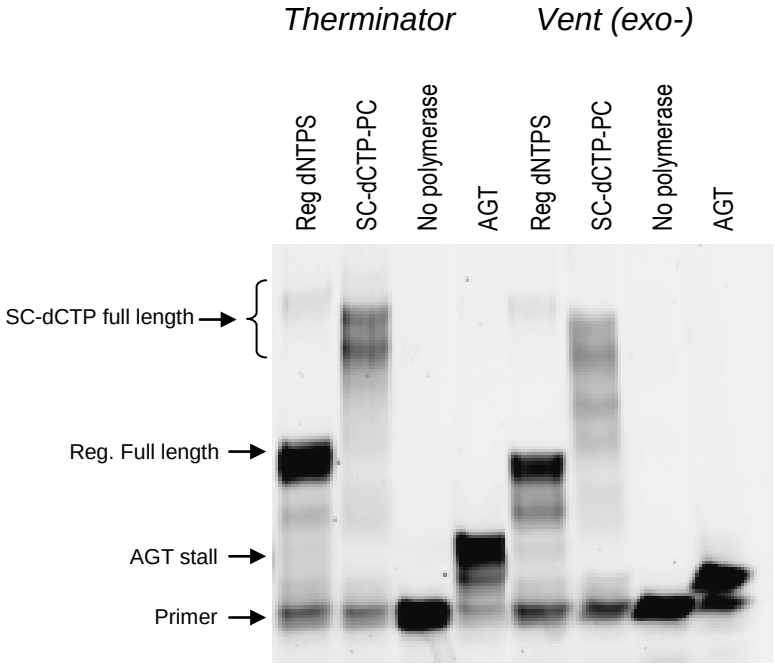


Figure 4.24: PAGE analysis of primer extension reaction with SC-dCTP-photocleavable using Therminator (left), Vent (exo-) (right). HM245, 25°C, 1 h incubation time.

Both polymerases display the capacity to incorporate multiple modified dCTPs but struggle to fully extend the template, as seen by the lack of a strong band correlating to a fully extended product. To further examine this the incubation time was extended to 7 hours (figure 4.25). In the case of Vent (exo-) by extending the incubation period a higher degree of fully extended product was achieved, with a sharper band seen with fewer stalls. However a comparison between figure 4.24 and figure 4.25 with the use of Therminator polymerase does not appear to show a higher degree of extension after the extended incubation.

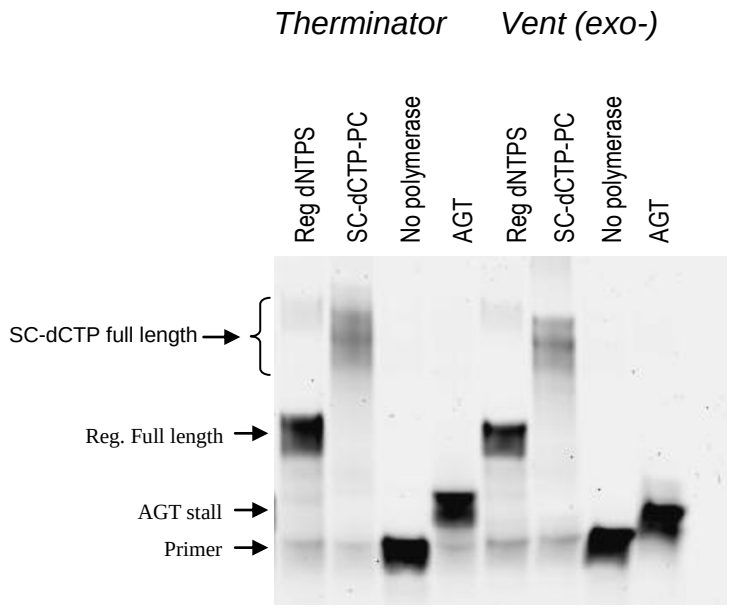


Figure 4.25: PAGE analysis of primer extension reaction with SC-dCTP-photocleavable using Therminator (left), Vent (exo-) (right). HM245, 25°C, 7 h incubation time.

4.6 Arrayed primer extension

QuantuMDx have recently developed primer extension reactions performed on a pre-functionalised glass surface, in a move towards being able to perform sequencing on nanowires. Successful primer extension is inferred by the incorporation of a Cy3-labelled dUTP after the dCTP incorporation.

For this purpose an aldehyde-functionalised glass slide has been functionalised by covalent attachment of 5'-amino modified ODN primer (P 1.3, figure 4.26) through heating in a humid environment before baking. The glass slide has been functionalised in 16 square areas.

```
P 1.3 5′ NH2 CAT ATT CAT CAA CCG CAC TCT GTT TAT CTC
             GTA TAA GTA GTT GGC GTG AGA CAA ATA GAG GACCC 5′ HM484
```

Figure 4.26: Primer (P1.3) and Template (HM484) sequence for arrayed primer extension assays.

Subsequent washing after immobilising the primer removes any unbound material and the slide is briefly dried before placing an adhesive mould containing 16 chambers on the slide (figure 4.27). This allows the user to perform multiple primer extension reactions upon each glass slide.

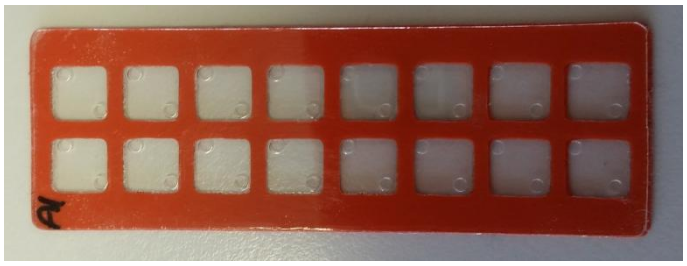


Figure 4.27: Aldehyde functionalised glass slide with chambers for primer extension reactions attached.

The reaction set up varied from the previous primer extension reactions described. As the primer is already bound to the surface the template needs to be annealed to the surface-bound primer.

A 40 µL reaction mixture containing 80 nM template, 2.5 µM dNTP mixes and 8 U polymerase in a 1x polymerase specific buffer was added to each chamber. As with previous primer extension assays three controls are used: A regular mix (dATP, dGTP, Cy3-dUTP, dCTP) as a positive control; an AGU mix (dATP, dGTP, Cy3-dUTP) to check the fidelity of the polymerase; and a negative control with no polymerase present. Each SC-dCTP mix is introduced with an equal concentration of dATP, dGTP, Cy3-dUTP and each SC-dCTP. Once each reaction mixture had been placed within a chamber the slide was incubated for 30 minutes at 25°C. After the incubation was complete the slide was thoroughly washed to remove any Cy3-dUTP present and then imaged using a Genepix 4100A. As a quality control for the printing of the slide each reaction chamber contains a control line which has been pre-functionalised with an additional Cy3-modification. This should be seen in each chamber as a bright green vertical line during fluorescence visualisation.

In all arrayed primer extension reactions the template HM484 (figure 4.26) was used which incorporates a single dCTP analogue in the first position. Successful incorporation of the modified SC-dCTP analogues is confirmed by the subsequent incorporation of a C5-Cy3-labelled dUTP at the second position. Therefore if the modified dCTP is not

incorporated or if the template cannot be extended after the incorporation of the modified nucleotide no fluorescence will be seen (figure 4.28).

Figure 4.28: (a) Incorporation of SC-dCTP and subsequent elongation confirmed by the incorporation of a Cy3-labelled dUTP (b) Incorporation of SC-dCTP but no elongation and therefore no Cy3-labelled dUTP incorporation.

The arrayed primer extension experiment was initially performed on all but the oligonucleotide modified dCTP analogues (SC-dCTP-dendrimer4, SC-dCTP-dendrimer8, SC-dCTP-PEG4, SC-dCTP-PEG8, SC-dCTP-photocleavable) using Vent (exo-) as the DNA polymerase and incubating for 30 minutes at 25°C. A shorter incubation period was used due to the shorter template and in order to be comparable to the arrayed primer extension reactions previously performed by QuantuMDx. The slide was then imaged using fluorescence and the results shown in figure 4.29.

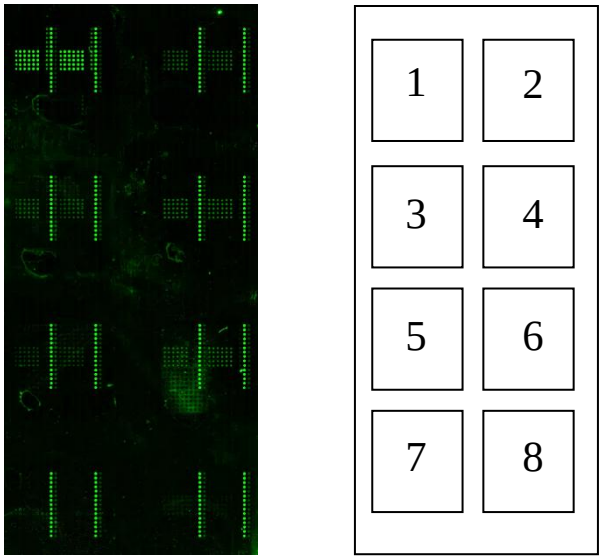


Figure 4.29: Fluorescence imaging of arrayed primer extension using GenePix 4100A. Chambers: (1) dCTP, dGTP, dATP, Cy3-dUTP (2) SC-dCTP-PEG8, dGTP, dATP, Cy3-dUTP (3) SC-dCTP-PEG4, dGTP, dATP, Cy3-dUTP (4) SC-dCTP-Dendrimer8, dCTP, dGTP, dATP, Cy3-dUTP (5) SC-dCTP-Dendrimer4, dGTP, dATP, Cy3-dUTP (6) SC-dCTP-Photocleavable, dGTP, dATP, Cy3-dUTP, (7) dCTP, dGTP, dATP, Cy3-dUTP, no polymerase (8) dGTP, dATP, Cy3-dUTP.

As expected from the data previously established from the primer extension reactions analysed by PAGE (figures 4.4, 4.12 and 4.22) all modified dCTP analogues showed a positive result for fluorescence, indicating that all SC-dCTP analogues were incorporated and the primer successfully extended. The fluorescence can then be quantified by removing background fluorescence present and averaging the intensity of each pixel containing fluorescence within the functionalised square chamber (figure 4.30). If the regular control is included in the table the fluorescent data for the SC-dCTP analogues is dwarfed due to the more efficient incorporation of natural dCTP to modified dCTP leading to significantly higher fluorescence. Therefore the regular control is omitted from figure 4.30 for clarity. SC-dCTP-dendrimer4 has a particularly low fluorescence given the higher fluorescence value of the larger dendrimer, SC-dCTP-dendrimer8, this may be an anomaly and further repeats would be needed in order to fully determine the efficiency of incorporation.

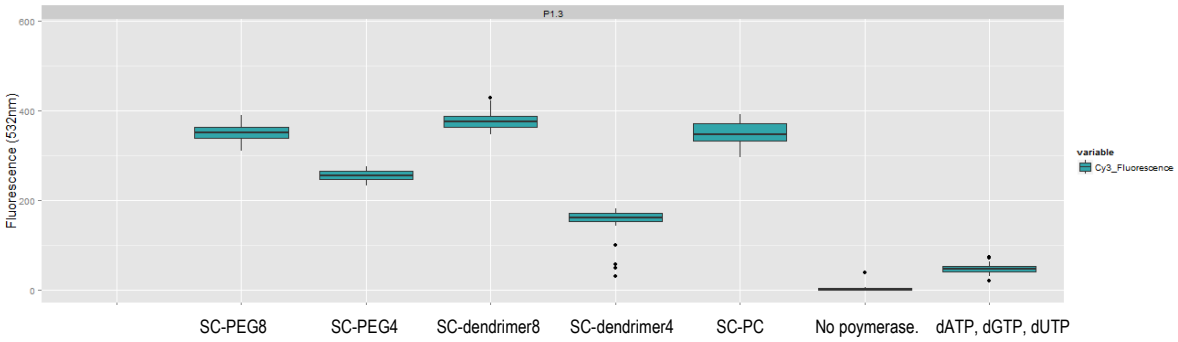


Figure 4.30: Quantified fluorescence data from GenePix 4100A analysis of figure 4.29 with the exclusion of chamber 1.

Given the results from the PAGE primer extension assays (figure 4.16) in which the SC-dCTP-ssdT24 analogue was not a substrate for Vent (exo-) polymerase after 1 h at 25°C the incubation time for the SC-dCTP-ssdT analogues was extended to 2 hours at 40°C. Therminator and Vent (exo-) DNA polymerases were used to test the incorporation of SC-dCTP-ssdT_{24} and SC-dCTP-ssdT_6. As previously described three controls were also performed and loaded into individual chambers. The two SC-dCTP primer extension reactions were simultaneously set up and placed in individual chambers before incubating the slide at 40°C for 2 hours.

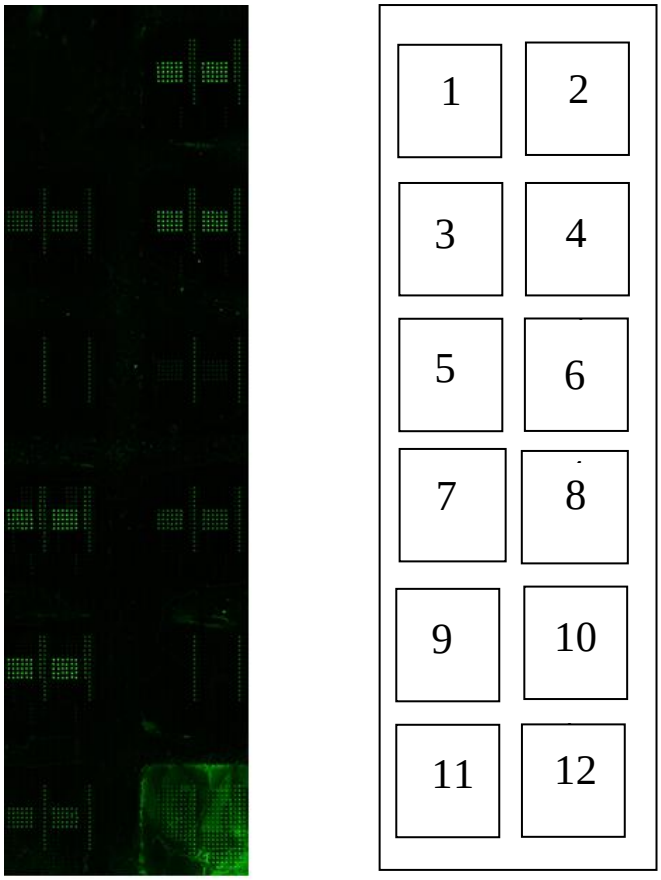
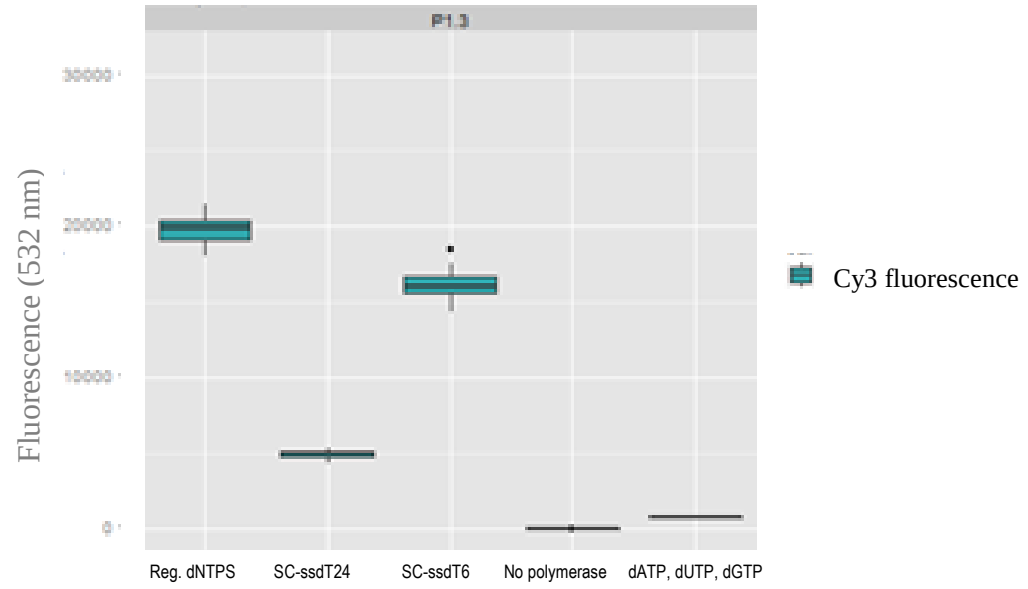


Figure 4.31: Fluorescence imaging of arrayed primer extension using GenePix 4100A. (1) Blank **Chambers 2 - 6 Vent (exo-)** : (2) dCTP, dGTP, dATP, Cy3-dUTP (3) SC-dCTP-ssdT24, dGTP, dATP, Cy3-dUTP (4) SC-dCTP-ssdT6, dGTP, dATP, Cy3-dUTP (5) dCTP, dGTP, dATP, Cy3-dUTP, no polymerase, (6) dGTP, dATP, Cy3-dUTP **Chambers 7 - 11 Therminator :** (7) dCTP, dGTP, dATP, Cy3-dUTP (8) SC-dCTP-ssdT24, dGTP, dATP, Cy3-dUTP (9) SC-dCTP-ssdT6, dGTP, dATP, Cy3-dUTP (10) dCTP, dGTP, dATP, Cy3-dUTP, no polymerase, (11) dGTP, dATP, Cy3-dUTP (12) Blank.

From visualisation of the slide it can be seen that both SC-dCTP-ssdT analogues are incorporated and extended using both DNA polymerases as determined by a green fluorescence visible in chambers 3, 4, 8 and 9 (figure 4.27). Therefore given the previous PAGE data (figure 4.17) this result implies that only a single SC-dCTP-ssdT24 can be incorporated into a template but that its incorporation does not hinder the elongation of the primer with natural dNTPS.

The fluorescence intensity from imaging the glass slide shown in figure 4.31 can then be quantified as shown in figure 4.32.

(a) Vent (exo-)



(b) Therminator

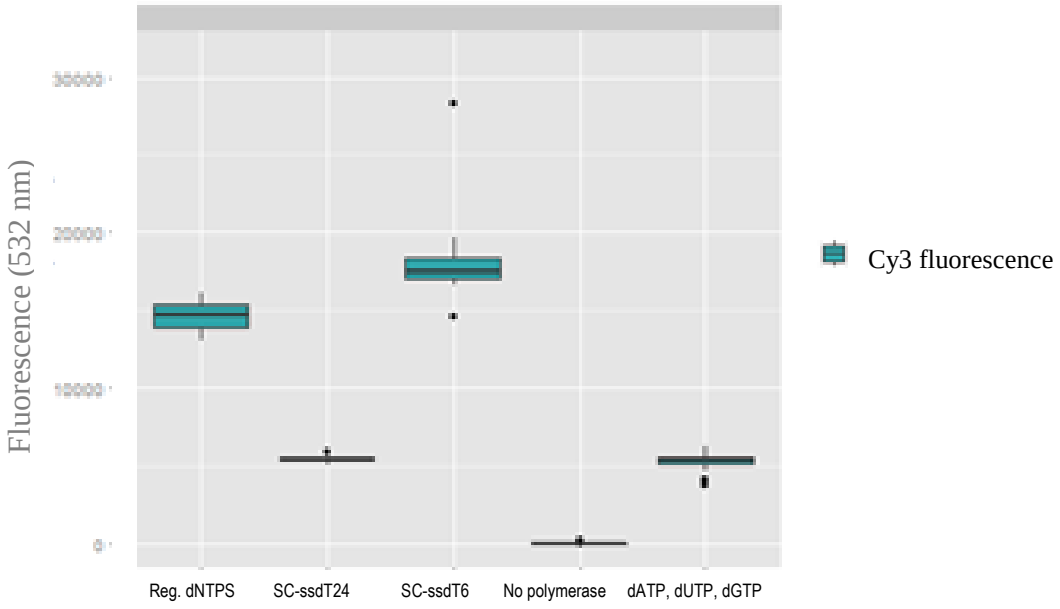


Figure 4.32: Quantified fluorescence data from GenePix 4100A analysis of figure 4.31 for (a) Vent (exo-) DNA polymerase and (b) Therminator DNA polymerase demonstrating the incorporation of SC-dCTP-ssdT24 and SC-dCTP-ssdT6.

Through the data presented in figure 4.31 and 4.32 it can be confirmed that both SC-dCTP-ssdT analogues are incorporated by Vent (exo-) and Therminator polymerases with limited misincorporation occurring. The smaller SC-dCTP-ssdT6 analogue appears

to have a higher efficiency of incorporation than the longer SC-dCTP-ssdT24 and is a good substrate for both polymerases. This is shown by a comparable efficiency of incorporation to regular dCTP under the conditions tested.

4.7 Conclusions

From our initial polymerase screen we found that SC-dCTP-PEG8 and SC-dCTP-PEG4 were substrates for a number of polymerases (table 4.2). Taking into account the speed of primer extension and the fidelity shown by each polymerase we focussed on two promising polymerases for future primer extension reactions: Therminator and Vent (exo-).

	HotStarTaq	*Phusion*	*VeraSeq 2.0*	*Vent (exo-)*	*Omni-Klentaq*	*Therminator*	*Klenow (exo-)*
SC-dCTP-PEG8	N	Y	Y	Y	Y	Y	Y
SC-dCTP-PEG4	N	Y	N	Y	Y	Y	Y

Table 4.2: Results from the initial polymerase screen, testing incorporation of a single SC-dCTP-PEG8 or SC-dCTP-PEG4 through primer extension reactions using HM284. Where (Y) indicates the SC-dCTP was a substrate and (N) indicates the SC-dCTP was not a substrate for the DNA polymerase.

Primer extension reactions were then performed for all 7 modified dCTP analogues synthesised. Although the conditions required for incorporation varied all analogues were found to be substrates for Therminator or Vent (exo-) DNA polymerase (table 4.3).

	Therminator	**Vent (exo-)**
SC-dCTP-PEG8	Y	Y
SC-dCTP-PEG4	Y	Y
SC-dCTP-Dendrimer8	Y	Y
SC-dCTP-Dendrimer4	Y	Y
SC-dCTP-ssdT24	Y	Y
SC-dCTP-ssdT6	Y	Y
SC-dCTP-PC	Y	Y

Table 4.3: Results from the primer extension reactions using Therminator and Vent (exo-) polymerases testing for single incorporation of a modified dCTP using the template HM284 or HM224.

Incorporation of multiple modified dCTP analogues into the DNA strand using the template HM245 (figure 4.13) was attempted. This was found to be particularly challenging for both Therminator and Vent (exo-) with most of the modified dCTP analogues. For example the inclusion of the dendrimer modified dCTPs appeared to stall the elongation once incorporated (figure 4.12) and as such both polymerases struggled to fully extend the primer where multiple incorporations were required (figure 4.14). This was also the case for the SC-dCTP-ssdT24 analogue, where only a single incorporation appears to be possible (figure 4.17), which was postulated as either the elongation stalling immediately after the SC-dCTP-ssdT24 incorporation or stalling prior to the incorporation of the next SC-dCTP-ssdT24.

By using the arrayed primer extension reactions it is possible to distinguish between the two reasons proposed for the polymerase stalling. It was found that in all cases where significant stalling had been observed (the incorporation of SC-dCTP-ssdT24 and both SC-dCTP-Dendrimer analogues) the polymerase was capable of elongating the primer after incorporation. This can be seen in the arrayed primer extension reactions by the successful incorporation of the Cy3-dUTP after the SC-dCTP incorporation (figure 4.29 and 4.31). This implies that the stall bands seen are due to the inability of the polymerase to incorporate two bulky modified dCTP analogues in close proximity. In the final device this will not be a concern as the reporter group, and therefore a large proportion of the modification, will need to be cleaved prior to the introduction of the next nucleotide.

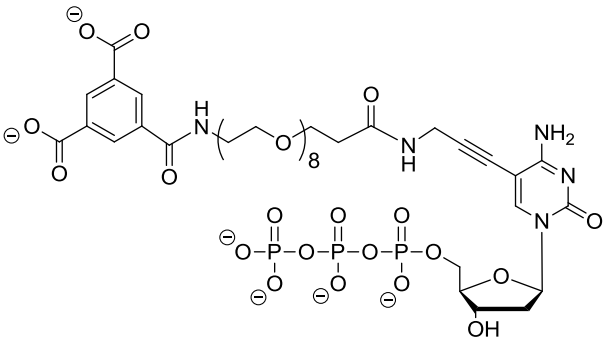
Chapter 5 –*Conclusions and future work*

5. Conclusions and Future work

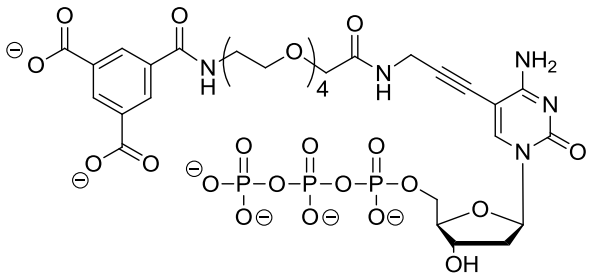
5.1 Conclusions

Synthesis

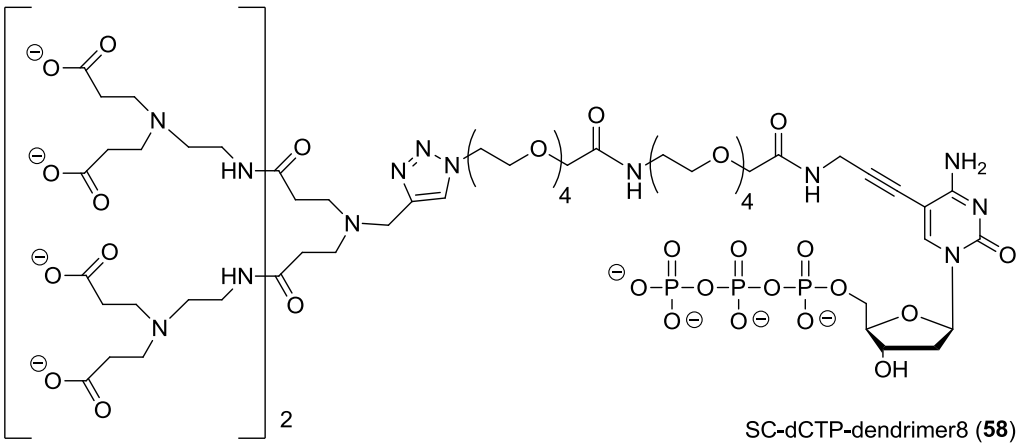
During this project seven novel C5-modified 2'-deoxycytidine-5'-triphosphates have been synthesised as proof-of-concept molecules for use in QuantuMDx's nanowire-based DNA sequencing device (figure 5.1).



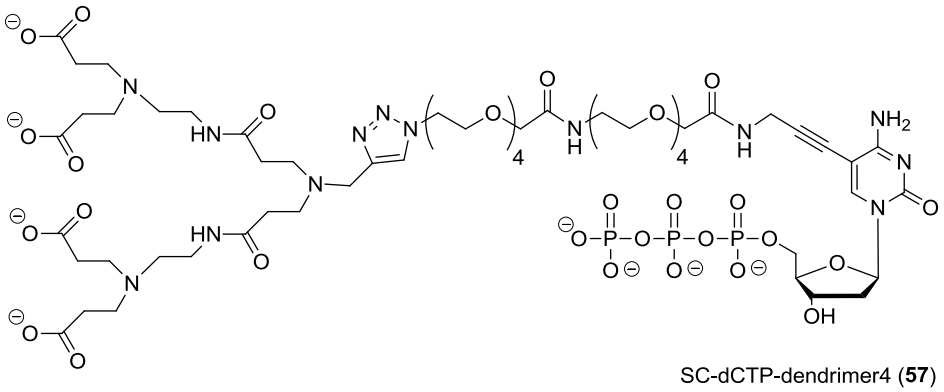
SC-dCTP-PEG8 (**34**)



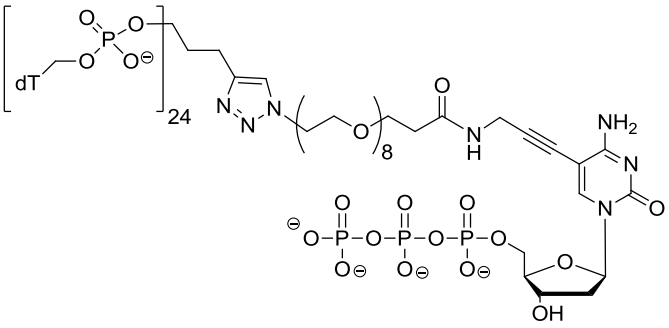
SC-dCTP-PEG4 (**41**)



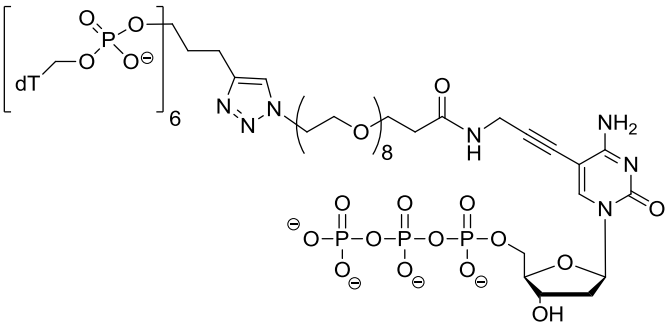
SC-dCTP-dendrimer8 (**58**)



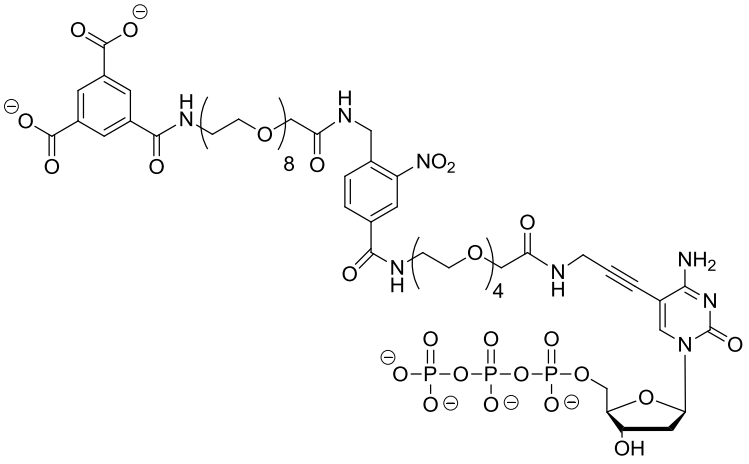
SC-dCTP-dendrimer4 (**57**)



SC-dCTP-ssdT24 (**62**)



SC-dCTP-ssdT6 (**64**)



SC-dCTP-photocleavable (**75**)

Figure 5.1: The chemical structures of the seven modified dCTP analogues synthesised.

The synthetic route to the seven dCTP analogues was through the functionalisation of the C5-anchor-modified dCTP (**26**). The anchor was introduced through a Sonogashira palladium catalysed cross coupling reaction with 5-iodo-2'-deoxycytidine. The subsequent phosphorylation was performed following P(V) phosphorylation techniques after attempts to synthesis 3'-hydroxyl protected 2'-deoxycytidine for P(III) phosphorylation were unsuccessful. With a suitable 'stock' dCTP (**26**) in hand, the effect of three types of modification on the polymerases substrate properties of the modified dCTPs were explored; Linker length, reporter group size and charge and the introduction of a photocleavable linker.

The first proof-of-concept compounds were synthesised with the exploration of linker length in mind. All linkers were functionalised with the trimesic acid reporter group for continuity. The commercially available $CA(PEG)_8$ linker (**29**) was functionalised with the PFP-ester of trimesic acid (**31**) and coupled to C5-propargylamino-dCTP (**26**) to give the SC-dCTP-PEG8 analogue (**34**). A shorter linker, $CA(PEG)_4$ (**38**), was synthesised from tetraethylene glycol, again functionalised with the trimesic acid reporter group and coupled to (**26**) to form the shorter SC-dCTP-PEG4 (**41**) in an analogous manner to SC-dCTP-PEG8.

It is assumed that as DNA sequencing occurs and the modified dNTPs are incorporated further and further away from the nanowire increasingly long linkers will be required for charge detection to continue to occur. The aim of synthesising a PEG_4 linker as a starting block to synthesising long linkers was demonstrated in the synthesis of the double length, azido-$(PEG_4)_2$-dCTP (**45**). This process could be repeated as needed to gain long linkers from a shorter 'monomer' and would allow a great deal of control over linker length.

Prior to this research only the trimesic acid reporter group had been explored as a suitable reporter group for use in the SBS device. Therefore new reporter groups with varying amounts of negative charge were synthesised to explore their influence on the substrate properties of the modified dCTPs. Two main types of reporter group were used; dendrimer reporter groups which are branched in shape or ssDNA which is linear. The azido-$(PEG_4)_2$–dCTP (**45**) was functionalised with one of two novel PAMAM dendrimers synthesised (PAMAM-1-D2 (**52**) or PAMAM-1-D3 (**53**)) through Cu(I) catalysed click chemistry to form the SC-dCTP-dendrimer4 and SC-dCTP-dendrimer8 analogues, (**57**) and (**58**) respectively. Although the initial yields for $CuSO_4$ catalysed

click reactions using sodium ascorbate were low it was found that the addition of TBTA, a Cu(I) stabilising compound, to the click reaction resulted in a two-fold increase in the yield.

To investigate ssDNA as a reporter groups a azido-PEG_8-NHS ester linker (**59**) was coupled to the anchor modified-dCTP (**26**) and functionalised with 5'-alkyne-modified ssDNA through click chemistry. Two ssDNA reporter groups were used for this purpose; poly-2'-deoxythymidine ODNs with a repeating sequence of either 6 or 24 bases. Both were successfully clicked onto the modified dCTP in the synthesis of modified dCTPs (**62**) and (**64**)**.** The suitability of ssDNA as a reporter group was of particular interest as its length can be easily varied to change the amount of charge present.

In the final SBS device the modified dNTPs will need to possess a cleavable moiety capable of removing the reporter group after incorporation into the growing DNA strand. The exploration of using a photocleavable linker positioned in-between the reporter group and DNA base has been described for this purpose. The synthesis of a photocleavable linker based upon a core 2-nitrobenzyl unit has been described (**67**) and its subsequent addition to a modified dCTP successfully demonstrated (**72**). The introduction of the anionic trimesic acid reporter group proved more challenging than expected but was achieved with the successful synthesis of SC-dCTP-Photocleavable analogue (**75**).

Primer extension reactions

The seven modified dCTP analogues synthesised possess five discrete reporter groups attached by a PEG_4, PEG_8, double length $(PEG_4)_2$ or cleavable linker. In order to investigate these features for the design of future modified dNTPs a series of primer extension reactions were performed to ensure the modified dCTPs were recognised as polymerase substrates. Initially a large variety of DNA polymerases were screened to test the polymerase recognition towards the SC-dCTP-PEG8 (**33**) and SC-dCTP-PEG4 (**41**) analogues. These included; HotStarTaq, Phusion, Veraseq 2.0, Vent (exo-), Omni-Klentaq, Therminator and Klenow (exo-). From this initial polymerase screen two polymerases, Vent (exo-) and Therminator, were chosen for future evaluation as they incorporated both (**33**) and (**41**) at low temperature in a short incubation time and demonstrated high fidelity.

The remaining modified dCTP analogues were tested as polymerase substrates using primer extension reactions and all were found to be substrates for both polymerases. It was also observed that dCTPs modified with either PAMAM dendrimer as the reporter group (SC-dCTP-dendrimer8 (**58**) or SC-dCTP-dendrimer4 (**57**)) partially stalled the elongation of the primer after incorporation. This was replicated with the use of the larger SC-dCTP-ssdT24 (**62**) although was not observed with the trimesic acid modified dCTP's SC-dCTP-PEG4 or SC-dCTP-PEG8. This is therefore likely due to the size of the reporter group interfering with the polymerase. However as in the final design the reporter group will be cleaved after incorporation this does not pose a serious problem. The paramount information is that they are polymerase substrates. It is interesting to note that linker length appeared to have little bearing on the substrate properties of the modified dCTPs, with the shorter SC-dCTP-PEG_4 being a substrate for five out of the seven polymerases screened.

All primer extension reactions were subsequently confirmed with the use of solid surface arrayed primer extension reactions. This successfully demonstrated the incorporation of the modified dCTPs synthesised by the extension of a primer upon a solid surface. It also demonstrated that the incorporation of the modified dNTPs does not terminate DNA synthesis.

Due to all of the seven modified dCTPs being substrates for one or more DNA polymerases no modification used has been excluded from the design of future modified dNTPs for use in QuantuMDx's SBS device.

5.2 Future work

Exploring longer linkers

The potential need for linkers of increasing length as the DNA sequencing reaction occurs and the reporter group is moved progressively further away from the nanowire has been previously discussed. Therefore the synthesis of longer linkers built from the coupling of PEG_4 monomers together would be an effective means (in terms of cost and synthetic ease) of extending the linker in PEG_4 increments. It is estimated that each PEG_4 linker is equivalent in length to the incorporation of 6 nucleotides.

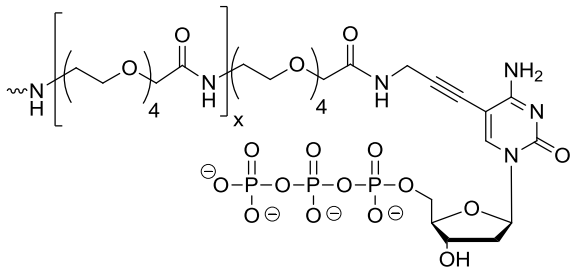


Figure 5.2: Schematic showing the potential synthesis of long linkers based upon the PEG_4 monomer.

Expanding the library of reporter groups

Incorporation of a number of novel reporter groups has been shown (chapter 3, section B) however, as more information as to the sensitivity of the nanowire becomes available more will be known about the required properties of the reporter group. The demonstrated incorporation of C5 modified dCTPs bearing ssDNA reporter groups (**62** and **64**) was of particular interest as the charge can be easily varied by synthesising different length ssDNA strands. The PAMAM dendrimer (**53**) can be extended to create larger, more polyanionic reporter groups if needed. Both of these reporter group types could be easily exploited in the future to provide the four discrete reporter groups needed for each dNTP.

Photolysis studies

Although the successful synthesis of a modified dNTP possessing a photocleavable linker has been described (**75**), further work is required to explore the subsequent photolysis. Cleavage of (**75**) could be confirmed through examining the difference in PAGE mobility between ODNs containing SC-dCTP-Photocleavable analogues which have been exposed to UV light (340 nm) and those which have not. As the cleavage would result in the loss of the PEG_8 linker and attached trimesic acid reporter group the ODN would contain a modified dCTP with a smaller C5 modification figure 5.3). This would travel further through the polyacrylamide gel than the ODN containing the larger, un-cleaved modified dCTP.

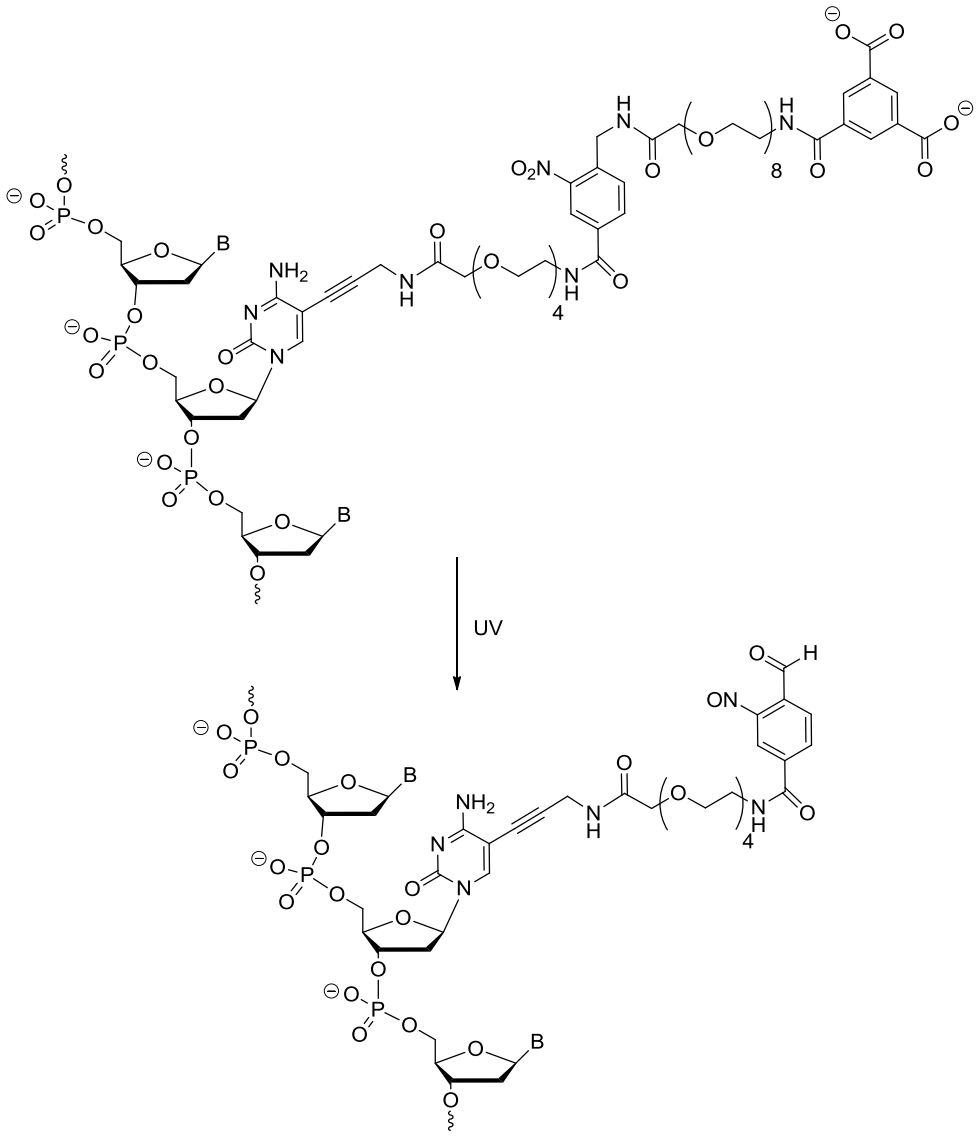


Figure 5.3: The *in situ* photolysis of an ODN containing the SC-dCTP-photocleavable dCTP (**75**). (where B represents one of the four DNA bases; A, G, C or T).

A 3'-blocking group

A key challenge in the future will be the design of a cleavable 3'-blocking group. As DNA polymerases have been shown to be more sensitive towards modifications made at the 3'-hydroxyl than those made at the C5 position of nucleotides, this may in turn require the engineering of a DNA polymerase. However if the photocleavable linker proves to be efficient at the cleavage of the reporter group a precedent for its use as a 3'-blocking group has already been established by Metzker and co-workers.[104]

Expanding biological testing

We have predominately focussed on the use of PAGE for examining the incorporation of modified dCTPs through primer extension reactions however QuantuMDx have recently started to develop solid-surface primer extension assays (such as the arrayed primer extension reactions in section 4.6) in progression towards mimicking the conditions of the final device. The next foreseen step is to introduce micro-fluidics to the arrayed primer extension reactions to investigate the sequential incorporation of dNTPs and move towards automation of sequencing.

Testing within the nanowire-based sequencing device

Ultimately more will be known about the suitability of the modified dCTP analogues synthesised for use in SBS once the nanowire-based sequencing device is operational. At this point further exploration into the required linker length for suitable DNA read-lengths will be required. Also once the nanowires sensitivity towards recognising variations in anionic charge between reporter groups has been determined appropriate reporter groups can be selected and tested. Although we have been focussed on the synthesis of modified dCTPs in due course all four dNTPs will require synthesis with the chosen modifications for use in the SBS device.

Chapter 6- *Experimental*

6. Experimental

All reagents were obtained from commercial suppliers and used without further purification unless stated otherwise. The 5'-alkyne-modified ODNs (**61** and **63**) were synthesised using a pentyne phosphoramidite (Biosearch) by DNA Technology, Denmark. All reactions performed in anhydrous solvents were also performed in flame dried glassware under an inert atmosphere of nitrogen unless stated otherwise. Dry solvents were obtained from the University of Sheffield Grubbs apparatus with the exception of anhydrous pyridine and 1,4-dioxane which were purchased from Sigma Aldrich.

Thin layer chromatography (TLC) was performed on pre-coated Merck silica gel 60 F_{254} aluminium backed plates. TLCs visualised under UV (254nm) and/or by staining with *p*-anisaldehyde for identifying nucleosides. Column chromatography was performed using silica gel for flash chromatography provided by Merck (30 – 70 μm).

Analytical RP-HPLC performed on Walters 2695 or 2690 instrumentation using a Phenomenex Gemini C18 5 μm 4.6 x 250 mm column, flow rate 1 mL/min, UV detection was recorded at 295 nm unless specified otherwise. Preparative RP- HPLC performed using a Phenomenex Gemini C18 5 μm 110Å 21.2 x 250 mm column at a flow rate of 21 mL / min. UV detection was recorded at 295 nm unless specified otherwise. All retention times are quoted from those found by analytical HPLC.

MPLC ion exchange chromatography was run on Gilson apparatus using DEAE sephadex A25 purchased from GE Healthcare. Sephadex media was loaded into a GE heathcare XK50/60 (600 x 50 mm) column for purifications greater than 100 μmol and an XK26/20 (200 x 26 mm) column for purifications less than 100 μmol.

NMR spectra were recorded on either a Bruker AC250 or AC400 spectrometer (individually stated for all data) and chemical shifts are reported in δ values relative to tetramethylsilane as an external standard. J values are given in Hz.

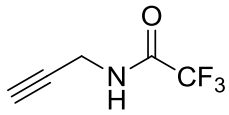
All mass spectrometry was kindly performed by the University of Sheffield Mass Spectroscopy service

DNA polymerases and ODN primer-template pairs were provided by QuantuMDx. Primer-template pairs had been designed by QuantuMDx and synthesised commercially.

Primer extension reactions were incubated in a Bio-Rad T100 gradient thermal cycler. PAGE gels were 17.5% polyacrylamide and were imaged using an Amersham Biosciences Typhoon Variable Mode Imager (excitation at 649 nm, emission at 670 nm for the Cy5 fluorophore). Aldehyde modified glass slides used for the arrayed primer extension reactions were purchased from Schott and imaged using a GenePix 4100A (Cy3 excitation at 532 nm).

UV measurements were performed using a NanoDrop 2000 spectrophotometer and triphosphate yields calculated using Beer Lambert law ($A = \Sigma\ c\ l$) where A = abs at 295 nm, Σ_{295} = 7700 $cm_2\mu M^{-1}$, l = path length (0.1 cm).[109] Extinction coefficients for the ODN modified dCTPs (**62** and **63**) were calculated using the online ATDbio oligonucleotide calculator using a dT extinction coefficient of Σ_{260} = 8700 $cm_2\mu M^{-1}$ and adjusted for base stacking. Fluorescent measurements were performed using a FluoroMax-4 spectrofluorometer.

2,2,2-Trifluoro-*N*-(prop-2-yn-1-yl)acetamide (14)



Prepared with modification to the literature procedure.[110]

Propargylamine (18.2 mL, 284.5 mmol) was added portionwise to a stirred solution of methyl trifluoroacetate (45 mL, 477.2 mmol, 1.6 eq.) in DCM (250 mL) at 0°C over 10 mins. After stirring for a further 30 mins the ice bath was removed and the reaction stirred at room temp. for 48 h. The reaction mixture was evaporated to dryness and purified by silica column chromatography (0 - 10% MeOH in DCM) to give a pale yellow oil (32.65 g, 216.1 mmol, 76%).

TLC: R_f 0.6 (10% MeOH/DCM)

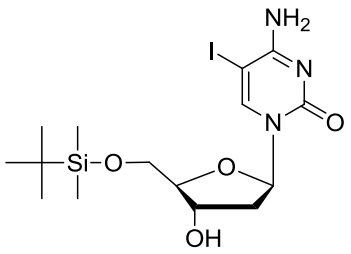
^{1}H NMR (250 MHz, $CDCl_3$) δ: 7.38 (s, 1H, N*H*), 4.13 (dd, *J* = 5.6, 2.5 Hz, 2H, CH_2), 2.32 (t, *J* = 2.6 Hz, 1H, C*H*)

^{13}C NMR (101 MHz, $CDCl_3$) δ 157.32 (q, *J* = 37.9 Hz), 115.59 (q, *J* = 287.2 Hz), 72.81, 29.57.

^{19}F NMR (377 MHz, $CDCl_3$) δ -76.09.

Mass Spec : ESI+ m/z 152.0 $[M+H]^+$

5'-*O*-(*tert*-Butyldimethylsilyl)-5-iodo-2'-deoxycytidine (16)[111]



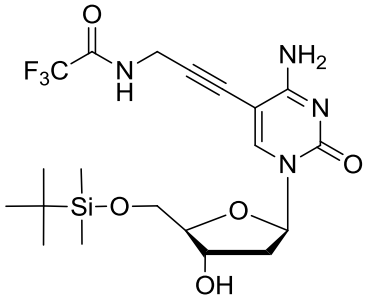
Imidazole (1.27 g, 18.70 mmol, 2.2 eq.) and *tert*-butyldimethylsilyl chloride (1.41 g, 9.35 mmol, 1.1 eq.) were added sequentially to a solution of 5-iodo-2'-deoxycytidine (3 g, 8.50 mmol) in anhydrous DMF (10 mL) and stirred overnight. After which time the solvent was removed under reduced pressure and the crude solid purified by silica column chromatography (5 - 20% MeOH in DCM) to give the title compound as a white solid (2.91g, 6.23 mmol, 74%).

TLC : R_f 0.60 (10% MeOH in DCM)

^{1}H NMR (400 MHz, DMSO) δ 7.95 (s, 1H, *H*6), 7.85 (s, 1H, N*H*), 6.65 (s, 1H, N*H*), 6.09 (dd, *J* = 7.8, 5.8 Hz, 1H, *H*1'), 5.26 (d, *J* = 4.0 Hz, 1H, 3'-O*H*), 4.20 – 4.12 (m, *J* = 2.2 Hz, 1H, *H*3'), 3.90 – 3.85 (m, 1H, *H*4'), 3.77 (ddd, *J* = 34.7, 11.5, 3.0 Hz, 2H, *H*5', *H*5''), 2.18 (ddd, *J* = 13.0, 5.7, 2.2 Hz, 1H, *H*2'), 1.90 (ddd, *J* = 13.4, 7.9, 5.9 Hz, 1H, *H*2'), 0.90 (s, 9H, ^{t}Bu), 0.12 (d, *J* = 5.7 Hz, 6H, Si(C*H*$_3$)).

Acc. Mass [ESI+]: Calculated for $C_{15}H_{27}IN_3O_4Si$ $[M+H]^+$: 468.0737 Observed:468.0826 $[M+H]^+$

5'-*O*-(*tert*-Butyldimethylsilyl)-5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (17)



Prepared with modification to the literature procedure. [54]

CuI (0.31 g, 1.61 mmol, 0.2 eq.), triethylamine (2.25 mL, 16.20 mmol, 2 eq.), and $Pd(PPh_3)_4$ (0.94 g, 0.81 mmol, 0.1 eq.) were added sequentially to a stirred solution of *O*-5'-(*tert*-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (**16**) (3.8 g, 8.10 mmol) in anhydrous DMF (40 mL) in the absence of light. The reaction flask was cooled to 0°C and 2,2,2-trifluoro-N-(prop-2-yn-1-yl)acetamide (**14**) (1.42mL, 12.50 mmol, 1.5 eq.) added portionwise over 10 mins. The reaction was stirred overnight at room temp. before being evaporated to dryness and purified by silica column chromatography (5% MeOH in DCM) to give a white solid (2.96 g, 6.00 mmol, 75%).

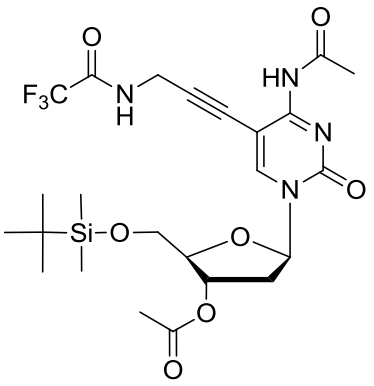
TLC: R_f 0.46 (5% MeOH in DCM)

^{1}H NMR (400 MHz, DMSO) δ 9.93 (t, *J* = 5.2 Hz, 1H, N*H*), 7.98 (s, 1H, *H*6), 7.88 (s, 1H, N*H*), 6.92 (s, 1H, N*H*), 6.12 (t, *J* = 6.6 Hz, 1H, *H*1'), 5.27 (d, *J* = 4.1 Hz, 1H, 3'-O*H*), 4.25 (d, *J* = 5.2 Hz, 2H, CH_2), 4.21 – 4.14 (m, 1H, *H*3'), 3.90 – 3.86 (m, 1H, *H*4'), 3.77 (ddd, *J* = 36.2, 11.5, 2.8 Hz, 2H, *H*5', *H*5''), 2.26 – 2.16 (m, 1H, *H*2'), 1.98 – 1.86 (m, 1H, *H*2'), 0.87 (s, 9H, ^{t}Bu), 0.08 (t, *J* = 2.8 Hz, 6H, Si(CH_3)).

Mass Spec : ESI+ m/z $[M+H]^+$ 491.2

N^4-Acetyl-3'-*O*-acetyl-5'-*O*-(tert-butyldimethylsilyl)-5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (18)

Attempted synthesis:

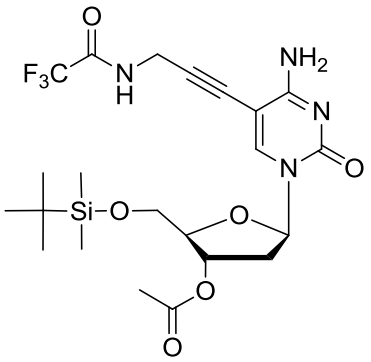


DMAP (0.77 g, 2.30 mmol, 0.25 eq.), triethylamine (2.5 mL, 18.06 mmol, 2 eq.) and acetic anhydride (6 mL, 63.00 mmol, 7 eq.) were added sequentially to a stirred solution of 5'-*O*-(*tert*-butyldimethylsilyl)-5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine **(17)** (4.43 g, 9.03 mmol) in anhydrous pyridine (25 mL). After 7 h the reaction was deemed incomplete upon TLC analysis, further acetic anhydride (1.7 mL, 2 eq.) was added and the reaction flask stirred overnight. A further amount of acetic anhydride (0.85 mL, 1 eq.) was added the following day and the reaction was warmed to 35°C for 4 h. Subsequently the reaction was quenched with the addition of MeOH (10 mL) and evaporated to dryness. The crude product was redissolved in EtOAc (30 mL) and washed with H_2O (3 x 10 mL). The organic layer was dried ($MgSO_4$) and concentrated to give a brown oil. Purification of this was attempted by silica column chromatography (5 - 20% MeOH in DCM) to give dark brown oil, although complete separation of products could not be achieved.

Mass spec (crude reaction mixture): ESI+ m/z $[M+H]^+$: 479.2, 521.2, 533.2, 575.2,

ESI+ m/z $[M+K]^+$: 635.2

3'-*O*-Acetyl-5'-*O*-(*tert*-butyldimethylsilyl)-5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (20)



Prepared with modification to the literature procedure. [112]

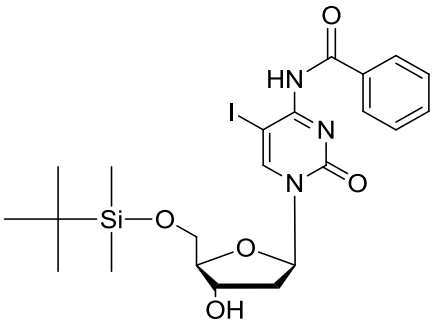
DMAP (61 mg, 0.5 mmol, 0.5 eq.), triethylamine (0.56 mL, 4 mmol, 2 eq.) and acetic anhydride (0.48 mL, 5.1 mmol, 2.5 eq.) were added sequentially to a stirred solution of 5'-*O*-(*tert*-butyldimethylsilyl)-5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (**17**) (1.0 g, 2.04 mmol) in anhydrous pyridine (4 mL). The reaction was stirred overnight before being quenched by the addition of MeOH (10 mL) and concentrated under reduced pressure. The crude product was redissolved in EtOAc (50 mL), the organic layer washed with H_2O (3 x 10 mL), dried ($MgSO_4$) and concentrated under reduced pressure. The crude product was purified by silica column chromatography (5 - 20% MeOH in DCM) to give a yellow crystalline solid (590 mg, 1.10 mmol, 54%).

TLC : R_f 0.57 (10% MeOH in $CHCl_3$)

^{1}H NMR (400 MHz, DMSO) δ 9.95 (t, *J* = 5.1 Hz, 1H, N*H*), 8.00 – 7.86 (m, 2H, N*H*, H6), 6.99 (s, 1H, N*H*), 6.15 (dd, *J* = 8.3, 5.6 Hz, 1H, *H*1'), 5.15 (br-d, *J* = 6.0 Hz, 1H, *H*3'), 4.25 (d, *J* = 5.2 Hz, 2H, CH_2), 4.12 (d, *J* = 1.7 Hz, 1H, *H*4'), 3.82 (ddd, *J* = 22.8, 11.5, 2.8 Hz, 2H, *H5′*, *H*5''), 2.42 – 2.31 (m, 1H, *H*2'), 2.08 – 2.04 (m, 4H, H2', CH_3), 0.86 (s, 9H, ^{t}Bu), 0.09 (br-s, 6H, Si(CH_3)$_2$).

Mass Spec : ESI+ m/z $[M+H]^+$ 533.2

N^4-Benzoyl-5'-*O*-(tert-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (21) [79]



O-5'-(*tert*-butyldimethylsilyl)-5-iodo-2'-deoxycytidine **(16)** (1 g, 2.12 mmol) was dissolved in anhydrous pyridine (2 mL) and evaporated to dryness (x 3) prior to use.

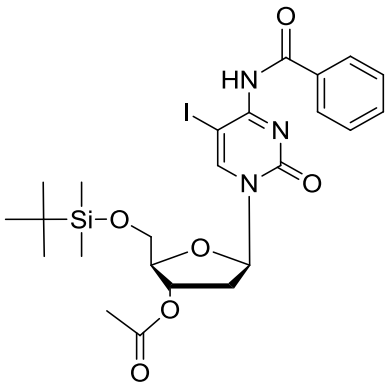
TMSCl (1.1 mL, 8.5 mmol, 4eq.) was added to a stirred solution of *O*-5'-(*tert*-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (**16**) in anhydrous pyridine (10 mL). After 3 h benzoyl chloride (0.32 mL, 2.75 mmol, 1.3 eq.) was added and the reaction stirred for a further 90 mins. After which time the reaction was cooled using an ice bath, distilled H_2O added (10 mL) and stirring continued for a further 5 h before evaporating the mixture to dryness (co-evaporated with toluene, 3 x 10 mL). The crude product was redissolved in DCM (50 mL) and washed with sat. sodium bicarbonate solution (3 x 20 mL), the organic layer dried ($MgSO_4$) and concentrated under reduced pressure. The resulting crude product was purified by silica column chromatography (1 - 5% MeOH in DCM) to give a yellow crystalline solid (1.11 g, 1.94 mmol, 92%).

TLC: R_f 0.64 (5% MeOH in DCM)

^{1}H NMR (400 MHz, d^6- DMSO) δ 12.92 (s, 1H, N*H*), 8.40 – 8.07 (m, 3H, *H*6, 2 x Ar-C*H*), 7.63 (t, *J* = 7.3 Hz, 1H, Ar-C*H*), 7.53 (t, *J* = 7.5 Hz, 2H, 2 x Ar-C*H*), 6.11 (t, *J* = 6.7 Hz, 1H, *H*1'), 5.35 (d, *J* = 4.1 Hz, 1H, 3'-O*H*), 4.23 – 4.18 (m, 1H, *H*3'), 3.94 (d, *J* = 2.5 Hz, 1H, *H*4'), 3.81 (ddd, *J* = 15.1, 11.6, 3.1 Hz, 2H, *H*5', *H*5''), 2.30 – 2.20 (m, 1H, *H*2'), 2.18 – 2.07 (m, 1H, *H*2'), 0.92 (s, *J* = 12.7 Hz, 9H, ^{t}Bu), 0.15 (d, *J* = 4.4 Hz, 6H, 2 x Si(CH_3)).

Acc. Mass [ESI+]: Calculated for $C_{22}H_{31}IN_3O_5Si$ $[M+H]^+$: 572.1033 Observed: 572.1083 $[M+H]^+$

N^4-Benzoyl-3'-*O*-acetyl-5'-*O*-(*tert*-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (22)



Acetic anhydride (1.5 mL, 5% v:v) was added to a stirred solution of N^4-benzoyl-5'-*O*-(*tert*-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (**21**) (600 mg, 1.05 mmol) in anhydrous pyridine (30 mL) and stirred overnight. The reaction was then quenched by the addition of MeOH (10 mL) and evaporated to dryness. The crude reaction mixture was washed with sat. sodium bicarbonate and extracted with DCM (30 mL x 3). The organic layers were combined, dried ($MgSO_4$) and evaporated to dryness. The crude product was then purified by silica column chromatography (5% EtOAc in DCM) to give a pale yellow solid (490 mg, 0.80 mmol, 76%).

TLC: R_f 0.30 (5% EtOAc in DCM)

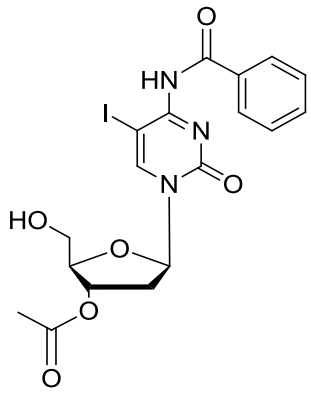
^{1}H NMR (250 MHz, $CDCl_3$) δ 8.44 – 8.35 (m, 2H, 2 x Ar-*CH*), 8.29 (s, 1H, *H*6), 7.60 – 7.43 (m, 3H, 3 x Ar-C*H*), 6.32 (dd, *J* = 9.0, 5.3 Hz, 1H, *H*1'), 5.27 (d, *J* = 5.9 Hz, 1H, *H*3'), 4.20 (d, *J* = 0.9 Hz, 1H, *H*4'), 4.02 – 3.90 (m, 2H, *H*5', *H*5''), 2.58 (dd, *J* = 13.9, 5.4 Hz, 1H, *H*2'), 2.21 – 2.13 (m, 1H, *H*2'), 2.12 (s, *J* = 3.2 Hz, 3H, CH_3), 0.97 (s, *J* = 2.8 Hz, 9H, ^{t}Bu), 0.22 (s, J = 3.0 Hz, 3H, Si(CH_3)), 0.20 (s, 3H, Si(CH_3)).

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.55, 145.12, 132.91, 130.28, 128.27, 86.99, 86.19, 75.61, 63.53), 38.90, 26.18, 21.00, -5.04, -5.28.

Mass Spec : ESI+ m/z $[M+H]^+$ 614.1

Acc. Mass [ESI+]: Calculated for $C_{24}H_{33}IN_3O_6Si$ $[M+H]^+$: 614.1105 Observed: 614.1200 $[M+H]^+$

N^4-Benzoyl -3'-*O*-acetyl-5-iodo-2'-deoxycytidine (23)



Prepared with modification to the literature procedure.[113]

1 M tetrabutylammonium fluoride in THF (1.5 mL, 1.5 mmol, 1.2 eq.) was added portionwise to a solution of N^4-benzoyl-3'-*O*-acetyl-5'-*O*-(tert-butyldimethylsilyl)-5-iodo-2'-deoxycytidine (**22**) (760 mg, 1.24 mmol) in anhydrous THF (10 mL). The reaction was stirred overnight before being evaporated to dryness. The crude solid was purified by silica column chromatography (10% EtOAc in DCM) to give a pale yellow solid (530 mg, 1.06 mmol, 86%).

TLC : R_f 0.37 (10% EtOAc in DCM)

^{1}H NMR (250 MHz, $CDCl_3$) δ 8.48 (s, 1H, *H*6), 8.38 (d, *J* = 7.0 Hz, 2H, 2 x Ar-C*H*), 7.61 – 7.53 (m, 1H, Ar-C*H*), 7.52 – 7.43 (m, 2H, 2 x Ar-C*H*), 6.31 (dd, *J* = 7.9, 5.9 Hz, 1H, *H*1'), 5.38 (dt, *J* = 6.1, 2.2 Hz, 1H, *H*3'), 4.19 (dd, *J* = 4.2, 2.0 Hz, 1H, *H*4'), 4.08 – 3.91 (m, 2H, *H*5', *H*5''), 2.55 (ddd, *J* = 14.1, 5.9, 2.3 Hz, 1H, *H*2'), 2.46 – 2.33 (m, 1H, *H*2'), 2.13 (s, *J* = 4.8 Hz, 3H, CH_3).

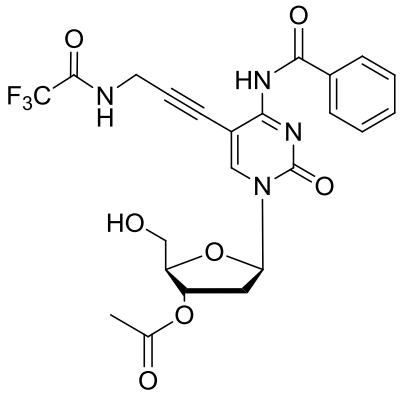
^{13}C NMR (63 MHz, $CDCl_3$) δ 170.62, 146.21, 132.91, 130.28, 128.26, 86.59, 85.61, 74.65, 62.48, 38.31, 20.93.

Mass Spec : ESI+ m/z $[M+H]^+$ 500.0

Acc. Mass [ESI+]: Calculated for $C_{18}H_{18}IN_3O_6$ $[M+H]^+$: 500.0274 Observed: 500.0309 $[M+H]^+$

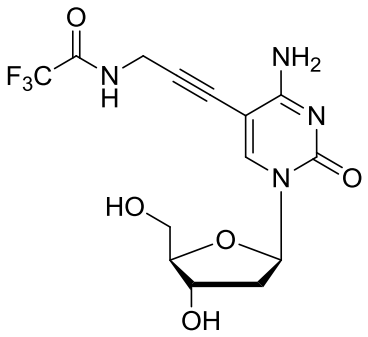
N^4-Benzoyl-3'-*O*-acetyl-5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (24) [114]

Attempted synthesis:



CuI (7.6 mg, 0.04 mmol, 0.2 eq.), triethylamine (56 µL, 0.4 mmol, 2 eq.) and $Pd(PPh_3)_4$ (23 mg, 0.02 mmol, 0.1 eq.) were added sequentially to a stirred solution of N^4-benzoyl-3'-*O*-acetyl-5-iodo-2'-deoxycytidine **(23)** (100 mg, 0.20 mmol) in anhydrous DMF (2 mL) in the absence of light. The reaction flask was cooled to 0°C followed by the addition of 2,2,2-trifluoro-*N*-(prop-2-yn-1-yl)acetamide (**14**) (35 µL, 0.3 mmol, 1.5 eq.) and stirred overnight. The reaction was then evaporated to dryness (co-evaporation with toluene 3 x 30 mL) and the crude product redissolved in EtOAc (10 mL) to which a 5% aq. EDTA (3 mL) solution was added. The two layers were mixed vigorously until the aqueous layer had a green/blue colour. The organic layer was separated, dried ($NaSO_4$) and evaporated to dryness. The crude material was purified by silica column chromatography (20 % DCM in EtOAC – 1% MeOH in EtOAC) eluting the title compound as an impure mixture of two compounds**.** The crude material (60 mg) was dissolved in EtOAc and purification re-attempted by centrifugal TLC (eluent system: EtOAc). Resolution of the title product as a pure compound could not be achieved.

5-(3-Trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (25)



Prepared with modification to the literature procedure.[54]

CuI (0.114 g, 0.6 mmol, 0.1 eq.), triethylamine (1.7 mL, 12 mmol, 2 eq.), $Pd(PPh_3)_4$ (0.347 g, 0.3 mmol, 0.05 eq.) and 2,2,2-trifluoro-*N*-(prop-2-yn-1-yl)acetamide (**14**) (2.72 g, 18 mmol, 3 eq.) were added sequentially to a stirring solution of 5-iodo-2'-deoxycytidine (2.12g, 6 mmol) in anhydrous DMF (30 mL) in the absence of light. The reaction was deemed complete by TLC analysis after 12 h at which point the solution was evaporated to dryness and purified by silica column chromatography (5 – 20 % MeOH in DCM). The title compound was isolated as a yellow crystalline solid (2.04 g, 5.95 mmol, 90%).

TLC: R_f 0.3 (10% MeOH in DCM)

^{1}H NMR (400 MHz, DMSO) δ 9.97 (t, *J* = 4.8 Hz, 1H, N*H*), 8.16 (s, 1H, *H*6), 7.85 (s, 1H, N*H*), 6.88 (s, 1H, N*H*), 6.11 (t, *J* = 6.5 Hz, 1H, *H*1'), 5.22 (d, *J* = 4.2 Hz, 1H, 3'-O*H*), 5.08 (t, *J* = 5.1 Hz, 1H, 5'-O*H*), 4.29 (d, *J* = 5.1 Hz, 2H, CH_2), 4.19 (td, *J* = 7.0, 3.5 Hz, 1H, *H*3'), 3.79 (q, *J* = 3.3 Hz, 1H, *H*4'), 3.64 – 3.50 (m, 2H, *H*5', *H*5''), 2.15 (m, 1H, *H*2'), 1.98 (m, 1H, *H*2').

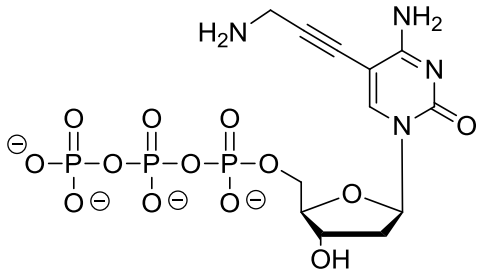
^{13}C NMR (101 MHz, DMSO) δ 164.82, 156.27, 153.85, 144.80, 90.38, 87.96, 85.88, 75.81, 70.63, 61.48, 46.17, 41.26, 30.42.

^{19}F NMR (377 MHz, DMSO) δ -74.43.

Acc. Mass [ESI+]: Calculated for $C_{14}H_{15}F_3N_4O_5$ $[M+H]^+$: 377.1070 Observed: 377.1073 $[M+H]^+$

P(V) phosphorylation of (25):

5-(3-Aminoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (26)



Prepared with modification to the literature procedure.[115]

To stirred a solution of 5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (**25**) (375 mg, 1 mmol) in anhydrous trimethyl phosphate (3.4 mL) at 0°C, freshly distilled phosphorus oxychloride (105 µL, 1.1 mmol, 1.1 eq.) was added and the reaction stirred on ice for 3 h. Separately a solution of 0.5 M tributylammonium pyrophosphate in anhydrous DMF (8 mL, 4 mmol, 4eq.) and dry tributylamine (2 mL, 8 mmol, 8eq.) was mixed vigorously. After 3 h the pyrophosphate solution was added to the main reaction flask. After 20 mins TEAB buffer (20 mL, 0.1 M) was added and the reaction allowed to return to room temp. After a further 3 h stirring, 35% ammonium hydroxide (20 mL) was added and the reaction stirred overnight. After this time the ammonia was removed under reduced pressure and the resulting solution purified by MPLC ion-exchange chromatography (0.05 M – 0.9 M TEAB, 8 mL/min, 12 h). The collected triphosphate (eluting between approx. 0.7 and 0.8 M TEAB) was further purified by prep-RP-HPLC and the resulting product fractions pooled, evaporated to dryness and the triphosphate isolated in a 20% yield (198 µmol).

HPLC gradient: 5 – 65% B over 30 min where A = 0.1 M TEAB and B =30% MeCN / 0.1 M TEAB. Retention time of product = 10 min.

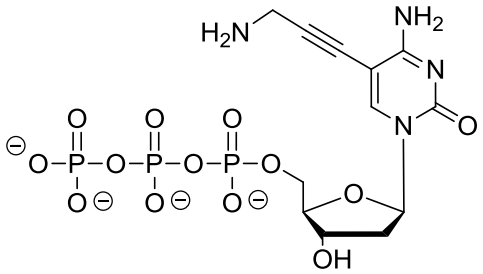
^{31}P NMR (101 MHz, D_2O) δ -8.95 (d, *J* = 16.6 Hz), -11.59 (d, *J* = 19.1 Hz), -22.78 (br).

^{1}H NMR (400 MHz, D_2O) δ 8.31 (s, 1H, *H*6), 6.14 (t, *J* = 5.6 Hz, 1H, *H*1'), 4.54 – 4.47 (m, 1H, *H*3'), 4.24 – 4.16 (m, 1H, *H*4'), 4.15 – 4.07 (m, 2H, C*H*$_2$), 3.90 (bs, 2H, *H*5', *H*5''), 2.43 – 2.32 (m, 1H, *H*2'), 2.30 – 2.19 (m, 1H, *H*2')

Acc. Mass [ESI-]: Calculated for $C_{12}H_{19}N4O_{13}P_3$ $[M-H]^-$: 519.0108 Observed: 519.0083 $[M-H]^-$

P(III) phosphorylation of (25):

5-(3-Aminoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (26)



Prepared with modification to the literature procedure. [84, 85]

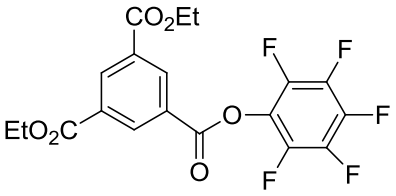
A solution of salicyl chlorophosphite (243 mg, 1.2 mmol, 1.2 eq.) in anhydrous DMF (7.5 mL) was added to a stirred solution of 0.5 M tributylammonium pyrophosphate (4 mL, 2 mmol, 2 eq.) and dry tributylamine (3.25 mL, 14 mmol, 7 eq.) in anhydrous DMF and the resulting solution stirred for 1 h at room temp. The contents of the flask were then transferred via a syringe into a stirred solution of 5-(3-trifluoroacetamidoprop-1-ynyl)-2'-deoxycytidine (**25**) (375 mg, 1 mmol) in anhydrous DMF (5 mL) at -13°C. After 3 hrs iodine (3% in 9:1 pyridine/water) was added dropwise until a permanent brown colour was observed and the solution stirred for 30 min at which point the reaction was hydrolysed with the addition of TEAB buffer (30 mL, 0.1 M). After 3 h 35% ammonia hydroxide (20 mL) was added at stirred at room temp for a further 3 h before removing the ammonia under reduced pressure. The resulting crude solution was purified by MPLC ion-exchange chromatography (0.05 M – 0.9 M TEAB, 8 mL/min, 12 h) and the collected triphosphate further purified by prep-RP-HPLC. The product fractions were pooled, evaporated and the title triphosphate isolated in an 18% yield (182 µmol).

HPLC gradient: 5 – 65% B over 30 min where A= 0.1 M TEAB and B= 30% MeCN / 0.1 M TEAB. Retention time of product = 10 min.

^{31}P NMR (101 MHz, H_2O+D_2O) δ -8.25 (d, *J* = 19.8 Hz), -11.52 (d, *J* = 19.3 Hz), -22.57 (t, *J* = 19.4 Hz).

Acc. Mass [ESI+]: Calculated for $C_{12}H_{19}N_4O_{13}P_3$ $[M+Na]^+$: 543.0059 Observed: 543.0050 $[M+Na]^+$

1,3-Diethyl-5-(pentafluorophenyl)benzene-1,3,5-tricarboxylate (31)



Diethyl-1,3,5-benzenetricarboxylate (**30**) (1.5 g, 5.63 mmol) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride were dissolved in anhydrous DMF (20 mL) and stirred for 30 mins. After this time a solution of pentafluorophenol (1.14 g, 6.19 mmol, 1.1 eq.) in anhydrous DMF (5 mL) was added portionwise to the reaction. After 90 mins the reaction was deemed to be complete by TLC and the solution evaporated under reduced pressure to give a white oil. The crude oil was redissolved in acetone (100 mL) and sodium phosphate buffer (50 mL, 0.1 M, pH 6.2) and extracted with ethyl acetate (3 x 100 mL). The organic layers were combined, dried ($MgSO_4$) and evaporated under reduced pressure. The crude solid isolated was purified by silica column chromatography (1-6 % EtOAc in Hexanes) and the title compound isolated as a white solid (1.48 g, 3.42 mmol, 61%).

TLC: R_f 0.3 (5% EtOAc in Hexanes)

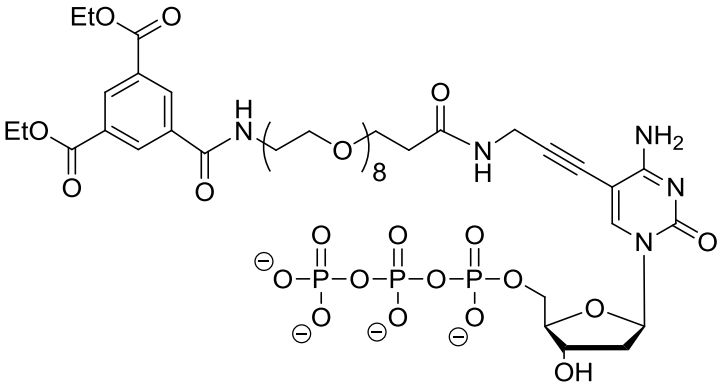
^{1}H NMR (250 MHz, $CDCl_3$) δ 9.01 (m, 3H, 3 x Ar-C*H*), 4.49 (q, *J* = 7.1 Hz, 4H, 2 x C*H*$_2$), 1.47 (t, *J* = 7.1 Hz, 6H, 2 x C*H*$_3$).

^{19}F NMR (235 MHz, $CDCl_3$) δ -152.21 (m, 2F), -157.09 (m, F), -161.81 (m, 2F).

Acc. Mass [ESI+]: Calc. for $C_{19}H_{13}O_6{}^{19}F_5$ $[M+H]^+$: 434.0739 found: 434.0739 $[M+H]^+$

Calc. for $C_{19}H_{13}O_6{}^{18}F_5$ $[M+H]^+$: 433.0705 found: 433.0708 $[M+H]^+$

5-(3,5-Diethoxycarbonylbenzamido-PEG$_8$-ethylcarboxamidoprop-1-ynyl)- 2'-deoxycytidine-5'-triphosphate (33)



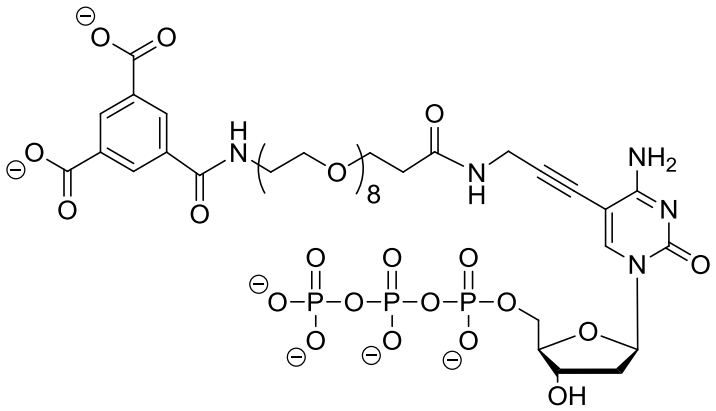
TSTU (90 mg, 300 µmol, 2.1 eq.) and DIPEA (105 µL, 600 µmol, 4.2 eq.) were added successively to a solution of 3,5-diethoxycarbonylbenzamido-PEG$_8$-acid (**23**) (207 mg, 300 µmol, 2.1 eq.) in anhydrous DMF (2.25 mL). After 2 h the content of the flask was syringed into a solution of 5-(4-amino-5-(3-aminoprop-1-yn-1-yl)-2'-deoxycytidine-5'-triphosphate (**26**) (143 µmol) stirring in sodium borate buffer (0.1 M, pH 8.5, 2.25 mL) at 0°C and stirred for 24 h. Once complete the reaction was evaporated to dryness and purified by prep-RP-HPLC to isolate the title triphosphate in a 57% yield (82 µmol).

HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 22.2 min.

^{31}P NMR (101 MHz, D_2O) δ -7.70 (d, *J* = 20.3 Hz), -11.48 (d, *J* = 19.7 Hz), -22.78 (t, *J* = 19.9 Hz).

Acc. Mass [ESI-]: Calculated for $C_{44}H_{69}N_5O_{27}P_3$ $[M-H]^-$: 1192.3393 Observed: 1192.3375 $[M-H]^-$

5-(3,5-Dicarboxybenzamido-PEG$_8$-ethylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (34)



A solution of sodium hydroxide (2 M, 1.5 mL) was added to stirred solution of 5-(3,5-diethoxycarbonylbenzamido-PEG$_8$-ethylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**33**) (82 µmol) in distilled water (1.5 mL) at 0°C. After 2 h the reaction was neutralised by the dropwise addition of aqueous acetic acid (1 M) before evaporating to dryness. The triphosphate was isolated by prep-RP-HPLC and analysed by mass spectrometry and ^{31}P NMR before subsequently being converted to the Na^+ salt using Dowex ion exchange resin (Na^+ form). The triphosphate was isolated as the sodium salt in an 82% yield (67 µmol).

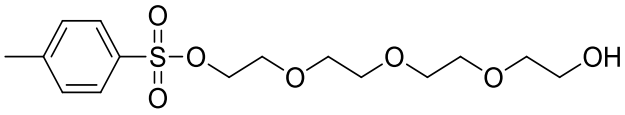
HPLC gradient: 5 – 95% B over 45 min where A= 0.1 M TEAB and B= 50%MeCN/0.1 M. Retention time of product = 13 min.

^{31}P NMR (101 MHz, D_2O) δ -10.64 (d, *J* = 20.0 Hz), -11.55 (d, *J* = 20.3 Hz), -23.34 (t, *J* = 19.9 Hz).

Acc. Mass [ESI-]:Calculated for $C_{40}H_{61}N_5O_{27}P_3$ $[M-H]^-$:1136.2767 Observed: 1136.2736 $[M-H]^-$

2-(2-(2-(2-hydroxyethoxy)ethoxy)ethoxy)ethyl 4-methylbenzenesulfonate (35)

[PEG_4 monotosylate]



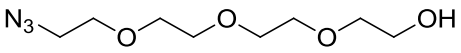
Triethylamine (54 mL, 0.39 M, 1.5 eq.) was added to a stirred solution of tetraethylene glycol (50 g, 0.26 M) in anhydrous DCM (20 mL). The reaction was cooled to 0°C and freshly recrystalised 4-tolenesulfonyl chloride (4.9 g, 25.7 mmol, 0.1 eq.) added and the reaction stirred overnight before being concentrated *in vacuo*. The crude oil was redissolved in DCM (100 mL) and washed with water (3 x 20 mL). The aquous layers were combined and extracted with DCM (3 x 20 mL). The organic layers were then combined, washed with acetic acid (0.1 M, 30 mL), water (30 mL), dried ($MgSO_4$) and evapourated to dryness to give a clear oil (8.25 g, 23.7 mmol, 92%).

^{1}H NMR (250 MHz, $CDCl_3$) δ 7.82 – 7.72 (m, 2H, 2 x Ar-*H*), 7.32 (dd, *J* = 8.6, 0.6 Hz, 2H, 2 x Ar-*H*), 4.17 – 4.09 (m, 2H, C*H*$_2$), 3.72 – 3.54 (m, 14H, 7 x C*H*$_2$), 2.89 (bs, 1H, O*H*), 2.42 (s, 3H, C*H*$_3$).

Acc. Mass [ESI+]: Calculated for $C_{15}H_{24}O_7S$ $[M+H]^+$: 349.1316 Observed: 349.1317 $[M+H]^+$

1-Azido-3,6,9-trioxaundecane-11-ol (36)

[PEG_4 azide]



Sodium azide (3.85 g, 59.2 mmol, 2.5 eq.) was added portionwise to a stirred solution of PEG_4 monotosylate (**35**) (8.25 g, 23.68 mmol) in ethanol (100 mL) over a period of 15 mins. Once homogeneous the reaction was heated to 70°C. After 20 h the reaction was cooled to room temp. and the mixture evaporated to dryness to give a crude orange oil. The crude oil was dissolved in DCM (100 mL) and washed with water (3 x 30 mL), the aquous layers subsequently combined and extracted with DCM (3 x 30 mL). The organic layers were then combined, dried ($MgSO_4$) and evaporated to dryness to give a pale yellow oil (4.1 g, 18.7 mmol, 79%).

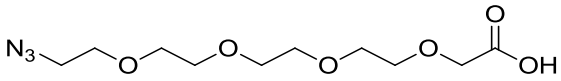
^{1}H NMR (250 MHz, $CDCl_3$) δ 3.74 – 3.55 (m, 14H, 7 x C*H*$_2$), 3.38 (t, *J* = 5.0 Hz, 2H, C*H*$_2$), 2.78 (bs, 1H, O*H*)

^{13}C NMR (63 MHz, $CDCl_3$) δ 72.49, 70.66, 70.62, 70.55, 70.31, 70.02, 61.65, 50.63.

Acc. Mass [ESI+]: Calculated for $C_8H_{17}N_3O_4$ $[M+H]^+$: 220.1292 Observed: 220.1289 $[M+H]^+$

14-Azido-3,6,9,12-tetraoxatetradecan-1-oic acid (37)[116]

[CAz(PEG)$_4$]



Sodium hydride (2.46 g, 61.5 mmol, 3 eq.) was added portion-wise to a stirring solution of 1-azido-3,6,9-trioxaundecane-11-ol (**36**) in anhydrous THF (60 mL) at 0°C over a period of 30 mins. Once in solution, bromoacetic acid (3.56 g, 25.6 mmol, 1.25 eq.) was added and the reaction stirred overnight at room temp. The reaction was then quenched with the slow addition of water (5 mL) and stirred for 15 mins. The solvent was removed under reduced pressure and the residue dissolved in DCM (150 mL), washed with HCl (2 M, 50 mL), sat. brine (50 mL) and dried ($MgSO_4$) before being concentrated to give the title compound as a light orange oil (5.52 g, 18.1 mmol, 88%).

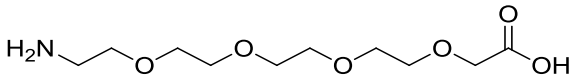
^{1}H NMR (250 MHz, $CDCl_3$) δ 4.15 (s, 2H, CH_2), 3.77 – 3.70 (m, 2H, CH_2), 3.69 – 3.62 (m, 12H, 6 x CH_2), 3.37 (t, *J* = 5.0 Hz, 2H, CH_2).

^{13}C NMR (63 MHz, $CDCl_3$) δ 173.18, 71.16, 70.62, 70.56, 70.49, 70.29, 69.96, 68.62, 50.59, 29.65.

Acc. Mass [ESI+]: Calculated for $C_{10}H_{19}N_3O_6$ $[M+H]^+$: 278.1347 Observed: 278.1350 $[M+H]^+$

14-Amino-3,6,9,12-tetraoxatetradecan-1-oic acid (38)

[CA(PEG)$_4$]



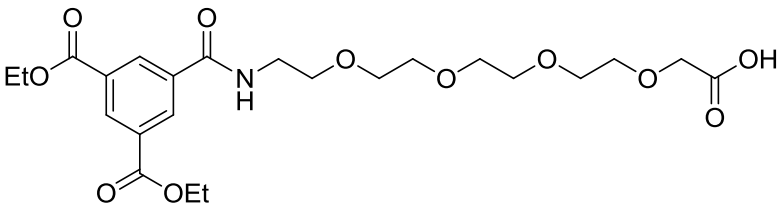
14-azido-3,6,9,12-tetraoxatetradecan-1-oic acid (**37**) (1.87 g, 6.74 mmol) in MeOH (50 mL, 0.14 M) was placed under a H_2 atmosphere (30 bar) at 50°C using a ThalesNano H-Cube. The solution was then passed over a 10% Pd/C catcart at 1 mL/min. The solvent was removed under reduced pressure to give the title compound as pale yellow oil (1.53 g, 6.08 mmol, 90%).

^{1}H NMR (250 MHz, DMSO) δ 3.98 (s, 2H, C*H*$_2$), 3.58 (t, *J* = 5.3 Hz, 2H, C*H*$_2$), 3.54 – 3.41 (m, 12H, 6 x C*H*$_2$), 2.90 (m, 2H, C*H*$_2$).

^{13}C NMR (101 MHz, D_2O) δ 174.18, 70.07, 69.58, 69.51, 69.45, 67.58, 66.31, 39.15.

Acc. Mass [ESI+]: Calculated for $C_{10}H_{21}NO_6$ $[M+H]^+$: 253.1442 Observed: 253.1445 $[M+H]^+$

14-(3,5-Diethoxycarboxamido)-3,6,9,12-tetraoxatetradecan-1-oic acid (39)



DIPEA (170 µL, 1 mmol, 1 eq.) and 1,3-diethyl-5-(pentafluorophenyl)benzene-1,3,5-tricarboxylate (**22**) (0.52 g, 1.2 mmol, 1.2 eq.) were added sequentially to a stirred solution of 14-amino-3,6,9,12-tetraoxatetradecan-1-oic acid (**38**) (0.25 g, 1 mmol) in anhydrous DMF (10 mL) and the reaction stirred overnight. The solvent was removed under reduced pressure and the crude product purified by prep-RP-HPLC. The product fractions were pooled and evaporated to dryness to give the title product in a 22 % yield (0.11 g, 0.22 mmol).

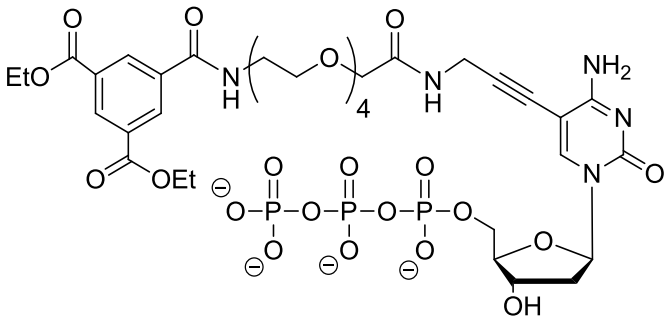
HPLC gradient: 20 – 70% B over 20 min where A= H_2O + 0.1% trifluoroacetic acid and B= MeCN + 0.1 % trifluoroacetic acid. Retention time of product = 10.5 min

^{1}H NMR (250 MHz, $CDCl_3$) δ 8.80 (t, *J* = 1.6 Hz, 1H, Ar-C*H*), 8.70 (d, *J* = 1.6 Hz, 2H, 2 x Ar-C*H*), 7.57 (bs, 1H, N*H*), 4.44 (q, *J* = 7.1 Hz, 4H, 2 x CH_2), 4.16 (s, 2H, CH_2), 3.78 – 3.59 (m, 16H, 8 x CH_2), 1.44 (t, *J* = 7.1 Hz, 6H, 2 x CH_3).

^{13}C NMR (101 MHz, $CDCl_3$) δ 173.05, 165.40, 134.60, 133.39, 132.48, 131.40, 70.79, 70.27, 70.06, 69.67, 68.59, 61.94, 40.37, 14.20.

Acc. Mass [ESI+]: Calculated for $C_{23}H_{33}NO_{11}$ $[M+H]^+$: 500.2116 Observed: 500.2132 $[M+H]^+$

5-(3,5-Diethoxycarbonylbenzamido-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (40)



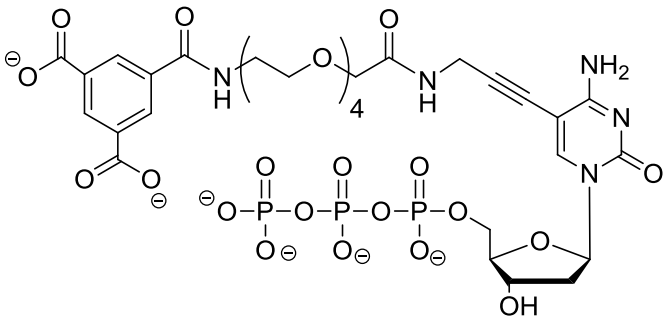
TSTU (82 mg, 272 µmol, 2.1 eq.) and DIPEA (95 µL, 544 µmol, 4.2 eq.) were added successively to a stirred solution of 14-(3,5-diethoxycarboxamido)-3,6,9,12-tetraoxatetradecan-1-oic acid (**30**) (136 mg, 272 µmol, 2.1 eq) in anhydrous DMF (2.25 mL) under N_2. After 2 h the contents of the flask were transferred via syringe into a solution of 5-(4-amino-5-(3-aminoprop-1-yn-1-yl)-2'-deoxycytidine-5'-triphosphate (**26**) (129 µmol) stirring in sodium borate buffer (0.1 M, pH 8.5, 2.25 mL) at 0°C. The reaction was stirred for 24 h before removing the solvent under reduced pressure. The resulting crude triphosphate was purified by prep-RP-HPLC (96 µmol, 74%).

HPLC gradient: 5 – 65% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 26.7 min.

^{31}P NMR (101 MHz, D_2O) δ -6.93 (d, *J* = 20.5 Hz), -11.45 (d, *J* = 19.8 Hz), -22.68 (t, *J* = 20.1 Hz).

Mass Spec : ESI+ m/z 1002 $[M+H]^+$

5-(3,5-Dicarboxylate-benzamido-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (41)



A solution of sodium hydroxide (2 M, 1.5 mL) was added to a stirring solution of 5-(3,5-diethoxycarbonylbenzamido-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**40**) (96 µmol) in water (1.5 mL) at 0°C. After 2 h the reaction was neutralised through the dropwise addition of aqueous acetic acid (1 M) before evaporating to dryness. The crude product was purified by prep-RP-HPLC to give the title triphosphate in a 82% yield (79 µmol).

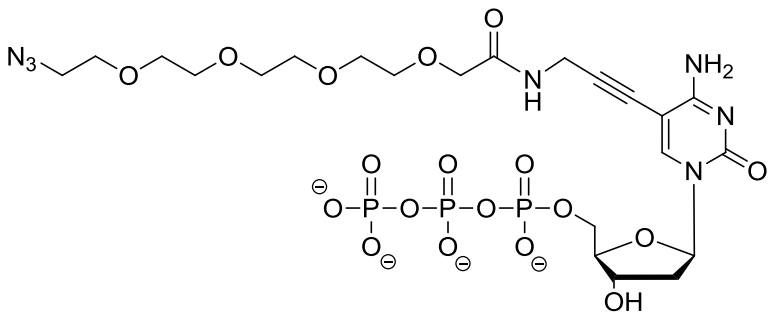
HPLC gradient: 5 – 65% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 10.3 min.

^{1}H NMR (250 MHz, D_2O) δ 8.29 (m, 1H, Ar-*H*), 8.15 (d, *J* = 1.4 Hz, 2H, 2 x Ar-*H*), 7.94 (s, 1H, *H*6), 6.06 (t, *J* = 6.4 Hz, 1H, *H*1'), 4.13 – 3.98 (m, 5H, *H*4', *H*5', *H*5" and C*H*$_2$), 3.90 (s, 2H, C*H*$_2$), 3.65 – 3.34 (m, 16H, 8 x C*H*$_2$), 2.38 – 2.02 (m, 1H, *H*2').

^{31}P NMR (101 MHz, D_2O) δ -6.58 (d, *J* = 21.1 Hz), -11.42 (d, *J* = 19.9 Hz), -22.72 (t, *J* = 20.6 Hz).

Acc. Mass [ESI-]: Calculated for $C_{31}H_{41}N_5O_{23}P_3$ $[M-H]^-$: 944.1442 Observed: 944.1405 $[M-H]^-$

5-(Azido-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (43)



TSTU (55 mg, 184 µmol, 2.1 eq.) and DIPEA (64 µL, 368 µmol, 4.2 eq.) were added sequentially to a solution of 14-azido-3,6,9,12-tetraoxatetradecan-1-oic acid (**28**) (60 mg, 184 µmol, 2.1 eq.) in anhydrous DMF (2.25 mL). After 3 h the contents of the flask were transferred via syringe into a stirred solution of 5-(3-aminoprop-1-yl)-2'-deoxycytidine-5'-triphosphate (**26**) (86 µmol) in aqueous sodium borate buffer (0.1 M, pH 8.5, 2.25 mL) at 0°C. After 24 h the reaction was evaporated to dryness and purified by prep-RP-HPLC (67 µmol, 78%).

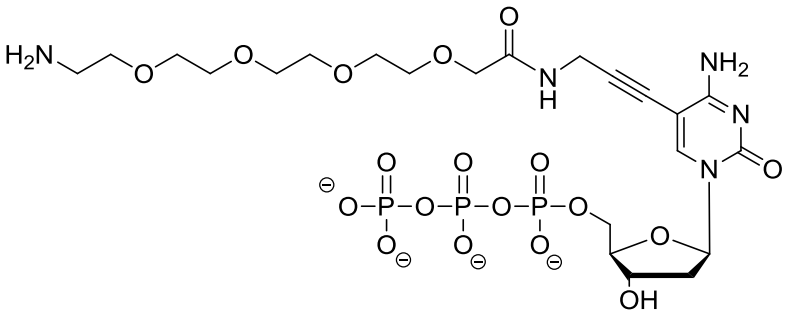
HPLC gradient: 5 – 65% B over 30 min where A= 0.1 M TEAB and B= 30% MeCN/0.1 M TEAB. Retention time of product = 23.3 min.

^{31}P NMR (101 MHz, D_2O) δ -9.80 (d, *J* = 19.8 Hz), -11.50 (d, *J* = 20.2 Hz), -23.18 (t, *J* = 20.2 Hz).

^{1}H NMR (250 MHz, D_2O) δ 8.02 (s, 1H, *H*6), 6.10 (t, *J* = 6.6 Hz, 1H. *H*1'), 4.48 – 4.40 (m, 1H, *H*3'), 4.13 (s, 2H, C*H*$_2$), 4.10 – 4.02 (m, 3H, *H*4' and *H*5', *H*5''), 3.99 (s, 2H, C*H*$_2$), 3.62 – 3.50 (m, 14H, 7 x C*H*$_2$), 3.36 – 3.30 (m, 2H, C*H*$_2$), 2.35 – 2.24 (m, 1H, *H*2'), 2.21 – 2.08 (m, 1H, *H*2').

Acc. Mass [ESI-]: Calculated for $C_{22}H_{35}N_7O_{18}P_3$ $[M-H]^-$: 778.1222 Observed: 778.1251 $[M-H]^-$

5-(Amino-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (44)



Prepared with modification to the literature procedure.[117]

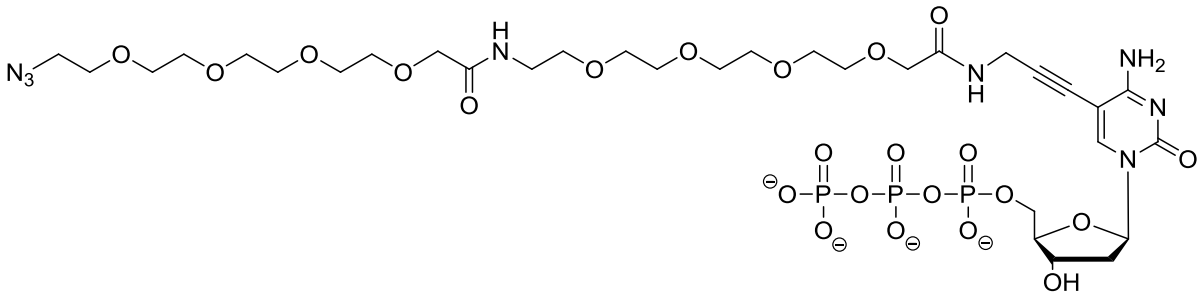
Triphenylphosphine (27 mg, 102 µmol, 2.1 eq.) was added to a stirring solution of 5-(azido-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**43**) (49 µmol) in anhydrous pyridine (0.75 mL). After 7 h a 5% NH_3 solution (4 mL) was added and the reaction stirred overnight. The reaction was then evaporated to dryness and purified by MPLC ion exchange chromatography using a TEAB gradient (0.05 M–0.8 M, 4 mL/min, 6 h) on DEAE A-25 sephadex. The triphosphate fractions (eluting between approx. 0.5 and 0.6 M TEAB) were pooled and evaporated to dryness to give the title compound in a 59% yield (29 µmol).

^{1}H NMR (250 MHz, D_2O) δ 8.05 (s, 1H, *H*6), 6.10 (t, *J* = 6.3 Hz, 1H, *H*1'), 4.50 – 4.39 (m, 1H, *H*3'), 4.15 – 4.02 (m, 5H, *H*4', *H*5', *H*5" and C*H*$_2$), 3.99 (s, 2H, C*H*$_2$), 3.68 – 3.47 (m, 7 x C*H*$_2$), 3.14 – 3.09 (m, 2H, C*H*$_{2,\ \text{partially obsc.}}$), 2.35 – 2.07 (m, 2H, *H*2' and *H*2").

^{31}P NMR (162 MHz, D_2O) δ -6.38 (d, *J* = 21.1 Hz), -11.34 (d, *J* = 20.0 Hz), -22.49 (t, *J* = 20.5 Hz).

Acc. Mass [ESI-]: Calculated for $C_{22}H_{38}N_5O_{18}P_3$ $[M-H]^-$: 752.1352 Observed: 752.1351 $[M-H]^-$

5-(Azido-(PEG$_4$)$_2$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (45)



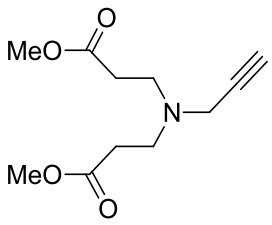
TSTU (57 mg, 188 µmol, 2.1 eq.) and DIPEA (65 µL, 376 µmol, 4.2 eq.) were added sequentially to a stirred solution of 14-azido-3,6,9,12-tetraoxatetradecan-1-oic acid (**37**) (52 mg, 188 µmol, 2.1 eq.) in anhydrous DMF (2.25 mL). After 3 h the contents of the flask were transferred via syringe into a stirred solution of 5-(amino-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**44**) (89 µmol) in sodium borate buffer (0.1 M pH 8.5, 2.25 mL) at 0°C. After 24 h the reaction was evaporated to dryness and purified by prep-RP-HPLC (58 µmol, 65%).

HPLC gradient: 5 – 100% B over 45 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 24.2 mins.

^{31}P NMR (162 MHz, D_2O) δ -6.40 (d, *J* = 21.2 Hz), -11.45 (t, *J* = 28.4 Hz), -22.63 (t, *J* = 20.6 Hz).

Acc. Mass [ESI-]: Calculated for $C_{32}H_{55}N_8O_{23}P_3$ $[M-H]^-$:1011.2520 Observed: 1011.2519 $[M-H]^-$

PAMAM 1-D1 (50)[92]



Methyl acrylate (32.8 g, 381.5 mmol, 3.5 eq.) was added dropwise to a solution of propargylamine (6 g, 109 mmol) in anhydrous methanol (200 mL) at 0°C over 30 min. After the addition was complete the reaction was allowed to return to room temp. before being gently heated at 30°C for 48 h. The solution was then evaporated to dryness and purified by silica column chromatography (30% EtOAc in Et_2O) to give a pale yellow oil (23.9 g, 105.3 mmol, 97%).

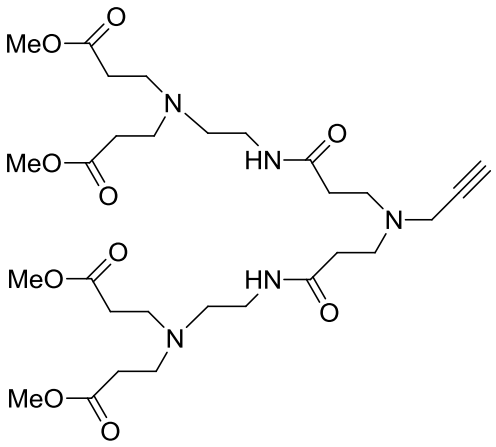
TLC: R_f 0.5 (30% EtOAc in Et_2O) visualised with $KMnO_4$

^{1}H NMR (250 MHz, $CDCl_3$) δ 3.60 (s, 6H, 2 x C*H*$_3$), 3.36 (d, *J* = 2.3 Hz, 2H, C*H*$_2$), 2.77 (t, *J* = 7.1 Hz, 4H, 2 x C*H*$_2$), 2.40 (t, *J* = 7.0 Hz, 4H, 2 x C*H*$_2$), 2.17 (t, *J* = 2.4 Hz, 1H, C*H*)

^{13}C NMR (63 MHz, $CDCl_3$) δ 172.56, 77.96, 73.31, 51.50, 48.86, 41.80, 32.81.

Acc. Mass [ESI+]: Calculated for $C_{11}H_{18}NO_4$ $[M+H]^+$: 228.1236 Observed : 228.1226 $[M+H]^+$

PAMAM 1-D2 (52)[118]



Ethylenediamine (134 mL, 2 M, 20eq.) was added portionwise to a stirred solution of PAMAM 1-D1 (**50**) (23.4 g, 102.9 mmol) in anhydrous methanol (50 mL) and gently warmed to 30°C under argon. After 48 h the solution was evaporated to dryness, redissolved in methanol (20 mL) and diethylether (~ 50 mL) added until a white, oil-like top layer formed. The top layer was then separated and evaporated to dryness to give a pale yellow oil (**51**). This was used without characterisation.

Methyl acrylate (65 mL, 720 mmol, 7 eq.) was then added to a stirred solution of the crude reaction mixture in MeOH (50 mL). After 72 h the reaction was evaporated to dryness and a crude mass of 54.85 g obtained. A portion of the crude (12.49 g) was purified by silica column chromatography (10 % MeOH in EtOAc) to give a yellow oil (7.73 g, 12.3 mmol, 52% over 2 steps).

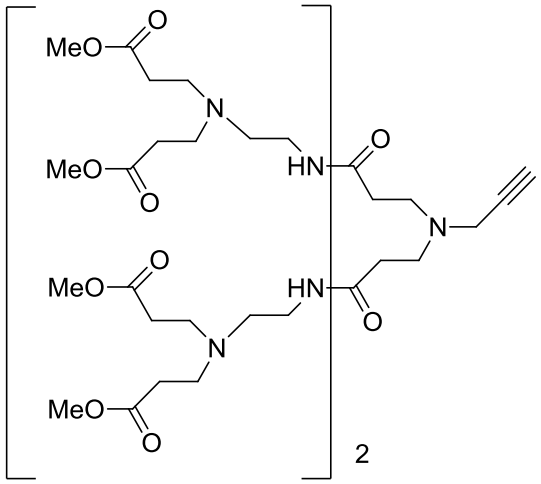
TLC: R_f 0.2 (10% MeOH in EtOAc) visualised with $KMnO_4$

^{1}H NMR (250 MHz, $CDCl_3$) δ 7.10 (bs, 2H, 2 x N*H*), 3.64 (s, 12H, 3 x C*H*$_3$), 3.44 (d, *J* = 2.3 Hz, 2H, C*H*$_2$), 3.27 (dd, *J* = 11.5, 5.6 Hz, 4H, 2 x C*H*$_2$), 2.82 (t, *J* = 6.6 Hz, 4H, 2 x C*H*$_2$), 2.74 (t, *J* = 6.7 Hz, 8H, 4 x C*H*$_2$), 2.52 (t, *J* = 5.9 Hz, 4H, 2 x C*H*$_2$), 2.39 (dt, *J* = 13.5, 6.6 Hz, 12H, 4 x C*H*$_2$), 2.18 (t, *J* = 2.2 Hz, 1H, C*H*).

^{13}C NMR (63 MHz, $CDCl_3$) δ 172.93, 77.97, 73.36, 52.99, 51.56, 49.40, 49.28, 41.16, 37.04, 33.84, 32.68.

Acc. Mass [ESI+]: Calculated for $C_{29}H_{49}N_5O_{10}$ $[M+H]^+$: 628.3552 Observed: 628.3553 $[M+H]^+$

PAMAM 1-D3 (53)[118]



Ethylenediamine (15.2 mL, 227 mmol, 20eq.) was added portionwise to a stirred solution of PAMAM 1-D2 (**52**) (7.13g, 11.35 mmol) in anhydrous methanol (20 mL) under argon. After 48 h the solution was evaporated to dryness, redissolved in methanol (20 mL) and diethyl ether (50 mL) added until a white oil-like top layer formed. The top layer was then separated and evaporated to dryness to give a clear white oil (6.85 g, 9.26 mmol, 82 %). This was used without characterisation.

Methyl acrylate (12 mL, 129.64 mmol, 14 eq.) was then added to a stirred solution of the crude reaction mixture (6.85 g, 9.26 mmol) in methanol (50 mL). After 72 h the reaction was evaporated to dryness and purified by silica column chromatography (30 - 65% MeOH in EtOAc) to give a light orange oil (8.45 g, 5.91 mmol, 52% over 2 steps).

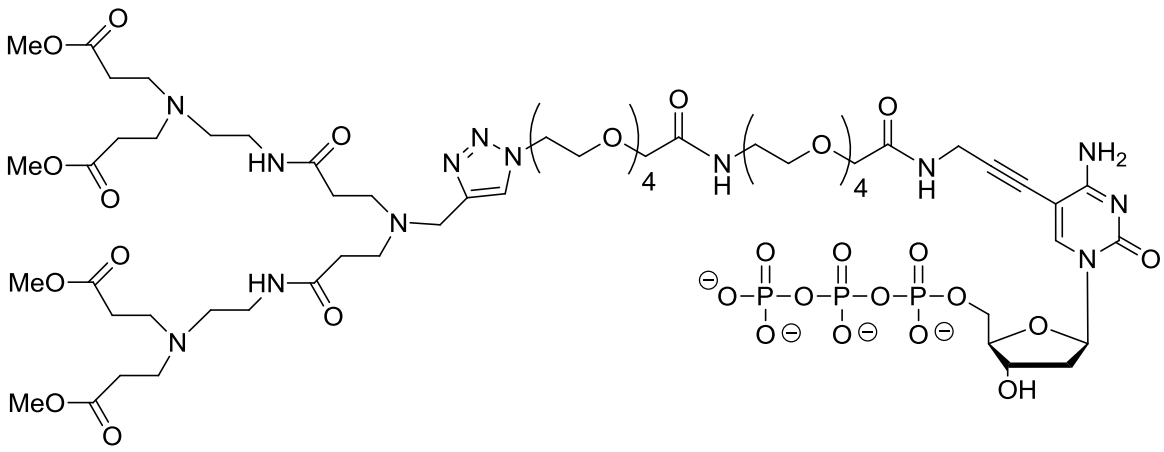
TLC: R_f 0.1 (30% MeOH in EtOAc) visualised with KMnO4

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75 (bs, 2H, 2 x N*H*), 7.10 (bt, *J* = 5.2 Hz, 4H, 4 x N*H*), 3.65 (s, 24H, 8 x CH_3), 3.44 (s, 2H, CH_2), 3.31 – 3.23 (m, 12H, 6 x CH_2), 2.82 – 2.70 (m, 28H, 14 x CH_2), 2.60 – 2.49 (m, 12H, 6 x CH_2), 2.45 – 2.32 (m, 28H, 14 x CH_2), 2.20 (t, *J* = 2.2 Hz, 1H, C*H*).

^{13}C NMR (101 MHz, $CDCl_3$) δ 173.06, 172.35, 172.12, 78.01, 73.48, 52.90, 52.50, 51.64, 50.52, 49.85, 49.38, 49.21, 40.98, 37.37, 37.16, 33.79, 32.66.

Acc. Mass [ESI+]: Calculated for $C_{65}H_{114}N_{13}O_{22}$ $[M+H]^+$: 1428.8201 Observed: 1428.8235 $[M+H]^+$

5-(PAMAM 1-D2-(PEG$_4$)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (55)



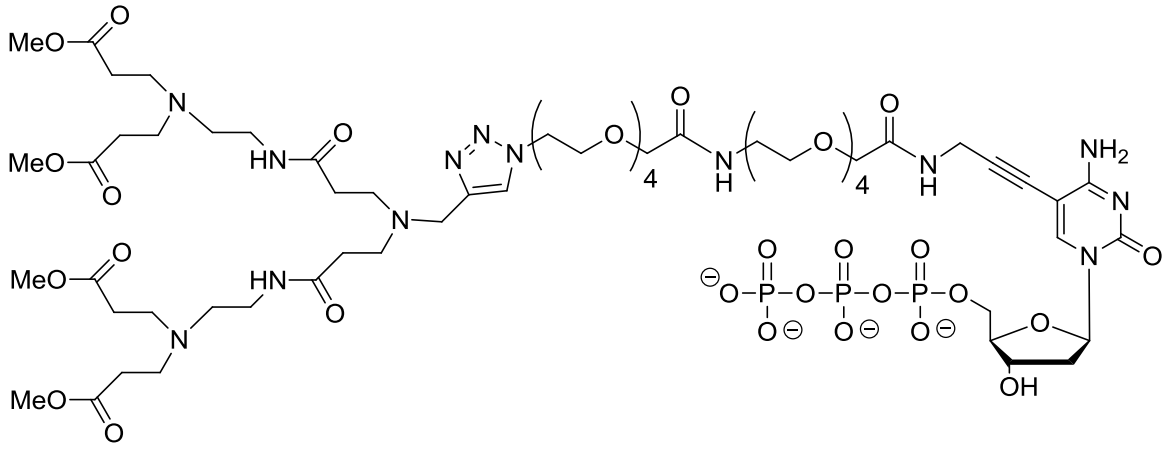
A solution of PAMAM-D2 (**52**) (22 mg, 33 µmol) in anhydrous THF (1.4 mL) was added to a stirred solution of 5-(azido-(PEG$_4$)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**45**) (39 µmol, 1.2 eq.) in distilled H_2O (1 mL). Once in solution $CuSO_4$ (83 µL, 10mg/1mL aq. solution, 3.3 µmol, 0.1 eq.) and sodium ascorbate (264 µL, 100 mg/1mL aq. solution, 132 µmol, 4 eq.) were added sequentially. After 2 h the solvent was removed under reduced pressure and the crude product purified by prep-RP-HPLC (11 µmol, 35 %).

HPLC gradient: 5- 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 19.5 min.

^{31}P NMR (162 MHz, D_2O) δ -8.29 (d, *J* = 20.1 Hz), -11.40 (d, *J* = 20.0 Hz), -22.77 (t, *J* = 20.1 Hz).

Acc. Mass [ESI-]: Calculated for $C_{61}H_{103}N_{13}O_{33}P_3$ $[M-H]^-$: 1638.6000 Observed: 1638.6009 $[M-H]^-$.

5-((PAMAM-1-D2)-(PEG$_4$)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (55)



The Cu:TBTA complex was preformed through mixing a solution of TBTA (3.3 mg, 6 µmol, 0.2 eq) in 40 µL H_2O:DMSO:*t*BuOH (4:3:1) with an aq. $CuSO_4$ solution (100 µL, 10 mg/1 mL, 6 µmol, 0.2 eq.) The solution was stored in the fridge until use.

To a stirred solution of 5-(Azido-(PEG$_4$)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**45)** (30 µmol) in H_2O (125 µL) an aq. solution of sodium ascorbate (120 µL, 200 mg/mL, 120 µmol, 4 eq.), TBTA:Cu(II) mix (140 µL) and a solution of PAMAM-D2 (**52**) (30 mg, 48 µmol, 1.6 eq.) in anhydrous THF (125 µL) were added sequentially. After 2 h the solvent was removed under reduced pressure and the crude triphosphate purified by prep-RP-HPLC. The resulting product fractions were pooled and evaporated to give the title triphosphate in a 70% yield (21 µmol).

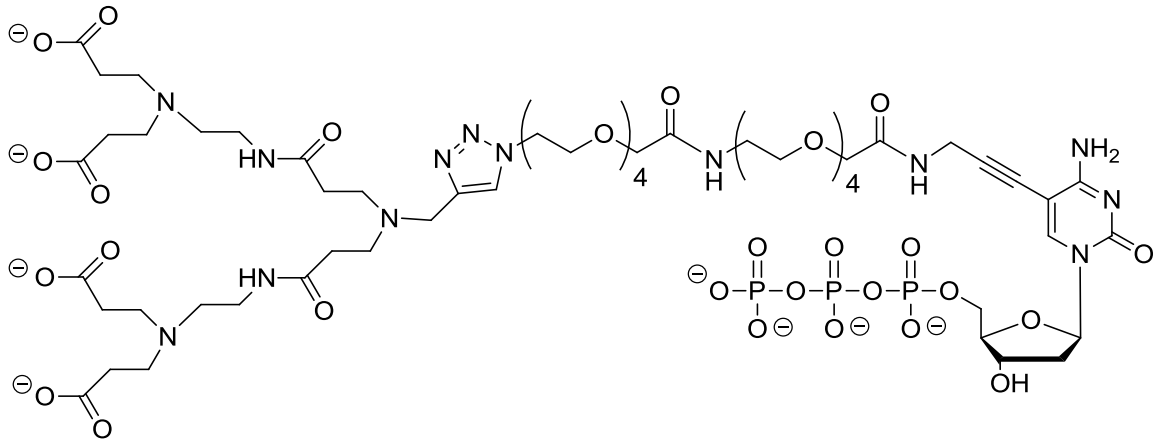
HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 19.5 min.

^{1}H NMR (400 MHz, D_2O) δ 8.09 (s, 1H, *H*6), 7.93 (s, 1H, Triazole-C*H*), 6.14 (br-s, 1H, *H*1'), 4.53 (s, 1H, *H*3'), 4.25 – 4.07 (m, 5H, *H*4', *H*5', *H*5'' and CH_2), 4.04 (s, 2H, CH_2), 3.97 (s, 2H, CH_2), 3.88 (s, 2H, CH_2), 3.74 (s, 2H, CH_2), 3.69 – 3.48 (m, 38H, 4 x CH_3, 13 x CH_2), 3.34 (br-s, 2H, CH_2), 3.17 (br-s, 4H, 2 x CH_2), 2.76 – 2.65 (m, 12H, 6 x CH_2), 2.46 (m, 12H, 4 x CH_2) (2*H*' and 2*H*'' obscured).

^{31}P NMR (162 MHz, D_2O) δ -6.39 (d, *J* = 21.2 Hz), -11.36 (d, *J* = 19.8 Hz), -22.58 (t, *J* = 20.7 Hz).

Acc. Mass [ESI-]: Calculated for $C_{61}H_{103}N_{13}O_{33}P_3$ $[M-H]^-$: 1638.6000 Observed: 1638.5975 $[M-H]^-$

5-((Polycarboxylate-(PAMAM-1-D2)-$(PEG_4)_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (57)



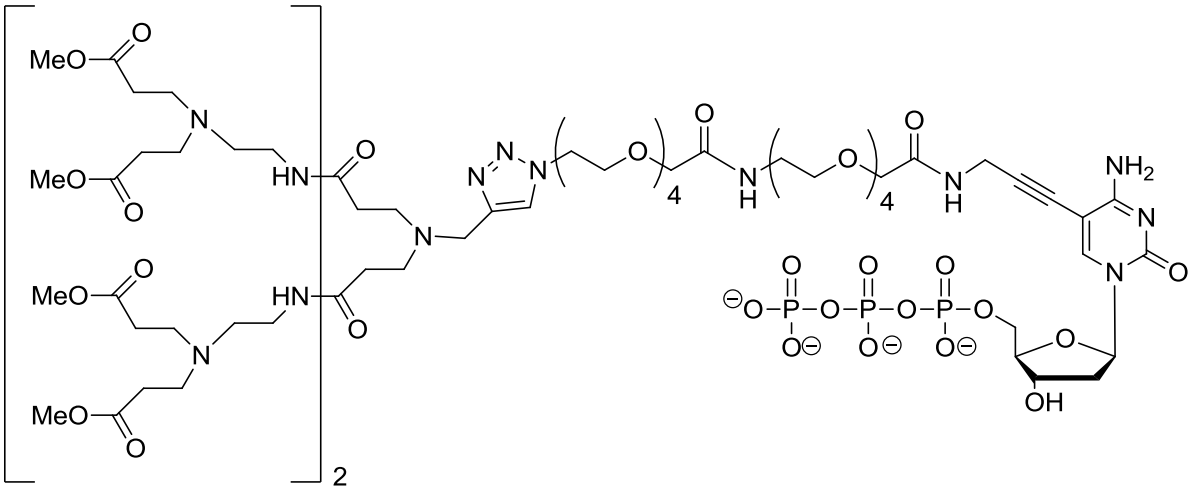
A solution of sodium hydroxide (2 M, 0.25 mL) was added to a stirred solution of 5-((PAMAM-1-D2)-$(PEG_4)_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**55**) (21 µmol) in water (0.75 mL) at 0°C. After 2 h the reaction was neutralised by the dropwise addition of aqueous acetic acid (1 M) before evaporating to dryness. The resulting crude product was purified by prep-RP-HPLC (14 µmol, 68%).

HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 10.7 min

^{31}P NMR (162 MHz, D_2O) δ -10.36 (d, *J* = 17.7 Hz), -11.52 (d, *J* = 21.6 Hz), -23.22 (t, *J* = 18.1 Hz).

Acc. Mass [ESI-]: Calculated for $C_{57}H_{96}N_{13}O_{33}P_3$ $[M-H]^-$: 1583.5403 observed: 1583.5398 $[M-H]^-$

5-((PAMAM-1-D3)-(PEG_4)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (56)

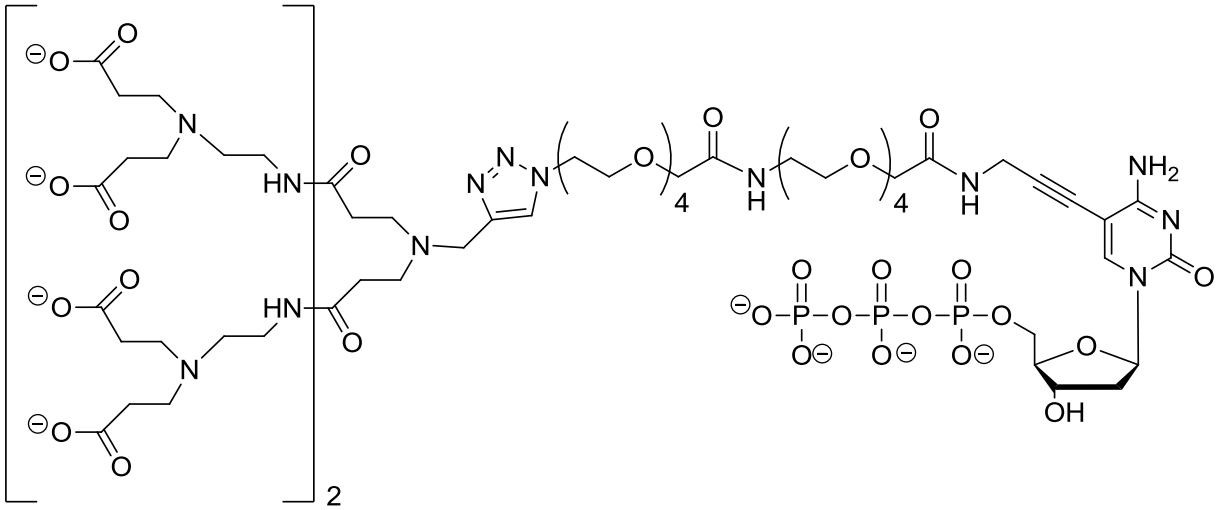


A solution of PAMAM 1-D3 (**53**) (47 mg, 33 µmol, 1.6 eq.) in THF (200 µL) was added to a stirring solution of 5-(azido-(PEG_4)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**45**) (21 µmol) in H_2O (50 µL). To this an aq. sodium ascorbate solution (85 µL, 200 mg/mL, 85 µmol, 4 eq.) followed by an aq. $CuSO_4$ solution (70 µL, 10 mg/mL, 4 µmol, 0.2 eq.) were added in quick succession. After 2 h the reaction was evaporated to dryness and the crude triphosphate purified by prep-RP-HPLC to give the title compound in a 37% yield (7.8 µmol).

HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB +. Retention time of product = 23.5 min

^{31}P NMR (162 MHz, D_2O) δ -6.31 (d, *J* = 21.1 Hz), -11.31 (d, *J* = 20.0 Hz), -22.42 (t, *J* = 20.8 Hz).

5-((Polycarboxylate-PAMAM-1-D3)-(PEG_4)$_2$-methylcarboxamido-prop-1-ynyl)- 2'-deoxycytidine-5'-triphosphate (58)



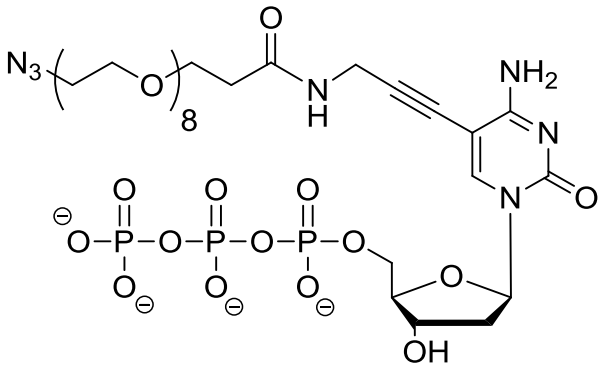
A solution of sodium hydroxide (2 M, 0.25 mL) was added to a stirring solution of 5-((PAMAM-1-D3)-(PEG_4)$_2$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**56**) (5 µmol) in water (0.75 mL) at 0°C. After 2 h the reaction was neutralised by the dropwise addition of aqueous acetic acid (1 M) before evaporating to dryness and purification by prep-RP-HPLC purification (3.1 µmol, 63%).

HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 6.6 min

^{31}P NMR (162 MHz, D_2O) δ -6.32 (d, *J* = 20.8 Hz), -11.27 (d, *J* = 19.8 Hz), -22.47 (t, *J* = 20.4 Hz).

Acc. Mass [ESI-]: Calculated for $C_{89}H_{151}N_{21}O_{45}P_3$ $[M-H]^-$: 2328.1869 Observed: 2328.9458 $[M-H]^-$

5-(Azido-PEG$_8$-ethylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (60)



Prepared with modification to the literature procedure.[119]

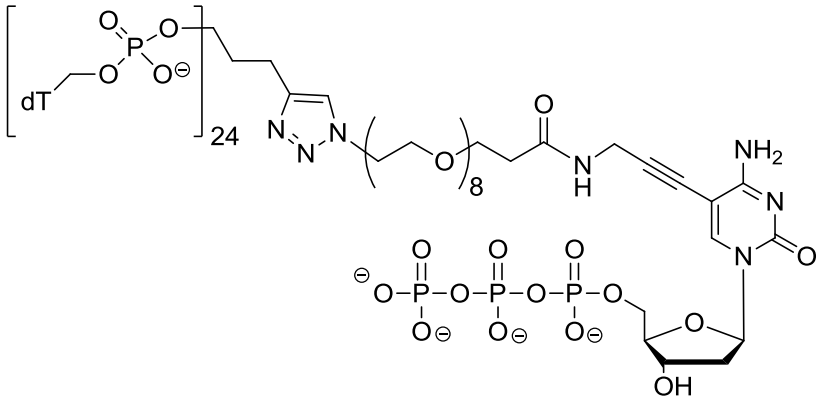
A solution of azido-PEG$_8$-NHS ester (**59**) (56 mg, 100 µmol, 5 eq.) in acetonitrile (0.5 mL) was added to a stirring solution of 5-(3-aminoprop-1-yl)-2'-deoxycytidine-5'-triphosphate (**26**) (19 µmol) in HEPES buffer (0.5 mL, 800 mM, pH8) and the reaction mixture stirred overnight at room temp. After this time, the reaction was evaporated to dryness and the crude mixture purified by prep-RP-HPLC to give the title compound in a 70% yield (13 µmol).

HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 15.6 min.

^{31}P NMR (162 MHz, D_2O) δ -10.33 (d, *J* = 20.1 Hz), -11.61 (d, *J* = 20.3 Hz), -23.31 (t, *J* = 20.2 Hz).

Acc. Mass [ESI-]: Calculated for $C_{31}H_{53}N_7O_{22}P_3$ $[M-H]^-$: 968.2462 Observed: 968.2489 $[M-H]^-$

5-[4-(dT$_{24}$)-1,2,3-triazole-PEG$_8$-ethylcarboxamido-prop-1-ynyl]-2'-deoxycytidine-5'-triphosphate (62)



Prepared with modification to the literature procedure.[96]

The Cu:TBTA complex was preformed through mixing 100 µL (2 µmol, 1.5 eq) of a of 20 mM TBTA solution in H_2O/DMSO/*t*BuOH (4:3:1) with 67.5 µL (1.35 µmol) of a 20 mM $Cu(OAc)_2$ solution in H_2O/DMSO/*t*BuOH (4:3:1). The solution was stored in the fridge until use.

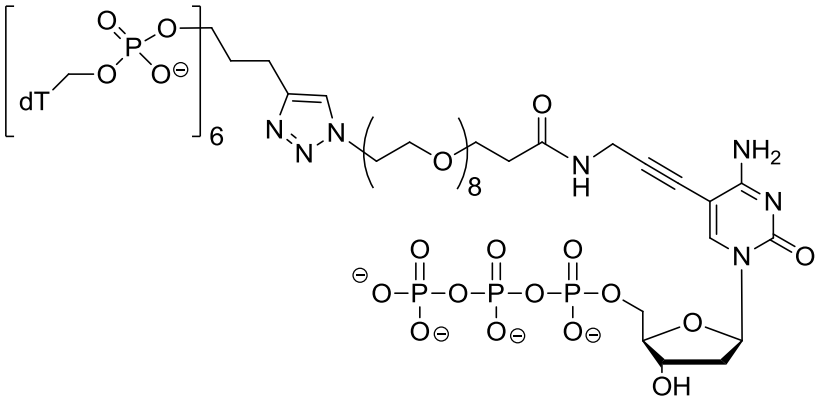
To a stirred degassed solution of 5-(azido-PEG$_8$-ethylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**45**) (400 nmol, 4 eq.) in H_2O (12 µL) under argon an aq. solution of sodium ascorbate (20 mM) (27 µL, 0.54 µmol), the preformed TBTA:Cu(II) mix (167 µL) and an aq. solution of the 5'-alkyne modified ODN (**61**) (100 µL, 100 nmol) were added sequentially followed by aq. sodium bicarbonate (67.5 µL, 0.2 M) and DMSO (37.5 µL). After 8 h the solvent was removed under reduced pressure and the crude triphosphate purified by RP-HPLC. The resulting product fractions were pooled and evaporated to give the title triphosphate in an 62 % yield (62 nmol).

HPLC gradient: 5 – 100 % B over 30 minutes where A= 0.1 M TEAB and B= 50% MeCN/ 0.1 M TEAB. UV detection = 270 nm. Retention time of product = 12.3 minutes.

Mass Spec : Calculated ESI- m/z 8354.00 $[M-H]^-$

Observed ESI- m/z 8354.61 $[M-H]^-$

5-[4-(dT$_6$)-1,2,3-triazole-PEG$_8$-ethylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (64)



Prepared with modification to the literature procedure.[96]

The Cu:TBTA complex was preformed through mixing 100 µL (2 µmol, 1.5 eq) of a of 20 mM TBTA solution in H_2O/DMSO/*t*BuOH (4:3:1) with 67.5 µL (1.35 µmol) of a 20 mM $Cu(OAc)_2$ solution in H_2O/DMSO/*t*BuOH (4:3:1). The solution was stored in the fridge until use.

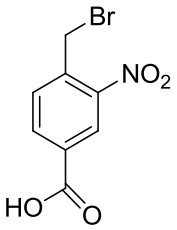
To a stirred degassed solution of 5-(azido-PEG$_8$-ethylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**45**) (400 nmol, 4 eq.) in H_2O (12 µL) under argon an aq. solution of sodium ascorbate (20 mM) (27 µL, 0.54 µmol), the preformed TBTA:Cu(II) mix (167 µL) and an aq. solution of the 5'-alkyne modified ODN (**61**) (100 µL, 100 nmol) were added sequentially followed by aq. sodium bicarbonate (67.5 µL, 0.2 M) and DMSO (37.5 µL). After 8 h the solvent was removed under reduced pressure and the crude triphosphate purified by RP-HPLC. The resulting product fractions were pooled and evaporated to give the title triphosphate in an 68 % yield (68 nmol).

HPLC gradient: 5 – 100 % B over 30 minutes where A= 0.1 M TEAB and B= 50% MeCN/ 0.1 M TEAB. UV detection = 270 nm. Retention time of product = 12.4 minutes.

Mass Spec : Calculated ESI- m/z 2878.52 $[M-H]^-$

Observed ESI- m/z 2878.54 $[M-H]^-$

4-(Bromomethyl)-3-nitrobenzoic acid (66)[120]



4-Bromomethylbenzoic acid (**65**) (1 g, 4.65 mmol) was added portionwise to fuming nitric acid (6 mL) at -10°C over 30 mins. After 1 h of stirring maintaining a temperature of -10°C the reaction was quenched by pouring the contents onto ice water. The resulting precipitate was filtered and washed with H_2O (2 x 50 mL). The collected white solid was dissolved in the minimum volume of hot DCM and precipitated with the addition of pentane followed by cooling. The precipitated was filtered and dried under vacuum to give the title compound as white solid (0.76 g, 63%).

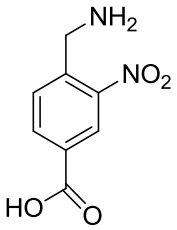
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.78 (d, *J* = 1.7 Hz, 1H, Ar-C*H*), 8.35 (dd, *J* = 8.0, 1.7 Hz, 1H, Ar-C*H*), 7.76 (d, *J* = 8.1 Hz, 1H, Ar-C*H*), 4.90 (s, 2H, CH_2).

^{13}C NMR (101 MHz, $CDCl_3$) δ 169.51, 148.10, 138.20, 134.68, 133.11, 130.66, 127.23, 119.93, 27.80.

Acc. Mass [ESI+]: Calc. for $C_8H_6NO_4{}^{79}Br$ $[M+H]^+$: 259.9552 found: 259.9553 $[M+H]^+$

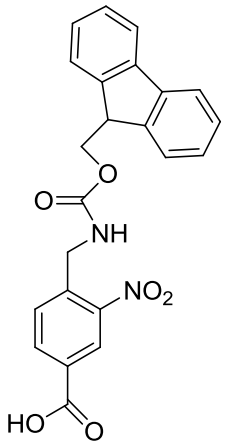
Calc. for $C_8H_6NO_4{}^{81}Br$ $[M+H]^+$: 261.9528 found: 261.9533 $[M+H]^+$

4-(Aminomethyl)-3-nitrobenzoic acid (67) [121]



4-(Bromomethyl)-3-nitrobenzoic acid (**66**) (1 g, 4.65 mmol) was dissolved in 8% NH_3 in ethanol (100 mL). After 20 h the reaction mixture was filtered and the collected solid washed with H_2O (50 mL), MeOH (50 mL) and Et_2O (50 mL). The light brown solid (0.31 g, 1.58 mmol, 34 %) was collected, dried under vacuum and used without characterisation.

4-(Fluorenylmethyloxycarbonyl-aminomethyl)-3-nitrobenzoic acid (68)[121]



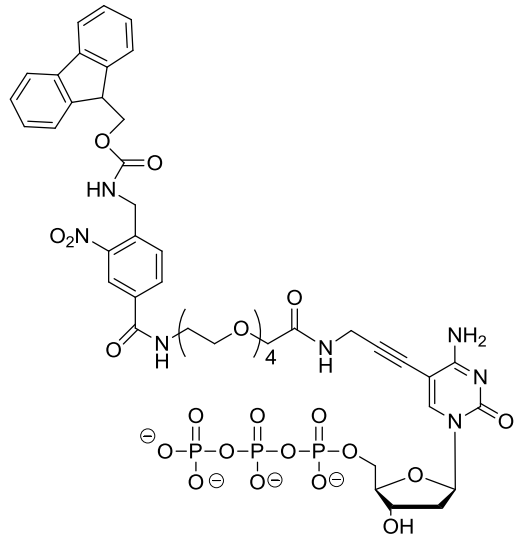
A solution of fluorenylmethoxycarbonyl chloride (410 mg, 1.6 mmol, 1 eq.) in 1,4-dioxane (1.5 mL) was added dropwise to a solution of 4-(aminomethyl)-3-nitrobenzoic acid (**67**) (310 mg, 1.6 mmol) in 10% sodium carbonate (10 mL) and 1,4-dioxane (5 mL) at 0°C. The reaction was returned to room temp and left to stir overnight. 1 M HCl was added dropwise to the reaction flask until pH 5 was reached and no more precipitate was seen to form. The precipitate was then filtered, washed with water (20 mL) and dried under vacuum to give the title compound as a white solid (540 mg, 1.3 mmol, 28%).

^{1}H NMR (400 MHz, DMSO) δ 8.46 (d, *J* = 1.5 Hz, 1H, Ar-C*H*), 8.23 (dd, *J* = 8.1, 1.5 Hz, 1H, Ar-C*H*), 8.04 (t, *J* = 5.9 Hz, 1H, N*H*), 7.91 (d, *J* = 7.5 Hz, 2H, 2 x Ar-C*H*), 7.71 (d, *J* = 7.4 Hz, 2H, 2 x Ar-C*H*), 7.56 (d, *J* = 8.1 Hz, 1H, Ar-C*H*), 7.43 (t, *J* = 7.4 Hz, 2H, Ar-C*H*), 7.34 (t, *J* = 7.2 Hz, 2H, 2 x Ar-CH), 4.55 (d, *J* = 5.8 Hz, 2H, CH_2), 4.40 (d, *J* = 6.7 Hz, 2H, CH_2), 4.25 (t, *J* = 6.6 Hz, 1H, C*H*).

^{13}C NMR (101 MHz, DMSO) δ 165.88, 156.89, 148.16, 144.26, 141.26, 139.64, 134.27, 131.31, 130.01, 128.10, 127.54, 125.73, 125.62, 120.61, 65.95, 47.25, 41.66.

Acc. Mass [ESI+]: Calculated for $C_{23}H_{18}N_2O_6$ $[M+H]^+$: 419.1238 found: 419.1236 $[M+H]^+$

5-(Fluorenylmethyloxycarbonyl-aminomethyl)-3-nitrobenzamido)-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (71)



TSTU (36 mg, 120 µmol, 2.1 eq.) and DIPEA (42 µL, 240 µmol, 4.2 eq.) were added sequentially to a stirring solution of 4-(fluorenylmethyloxycarbonyl-aminomethyl)-3-nitrobenzoic acid (**68**) (50 mg, 120 µmol, 2.1 eq.) in anhydrous DMF (2 mL). After 3 h the contents of the flask were syringed into a solution of 5-amino-(PEG)$_4$-methylcarboxamido-prop-1-ynyl-2'-deoxycytidine-5'-triphosphate (**43**) (57 µmol) in sodium borate buffer (pH 8.5, 0.1 M, 1 mL) at 0°C. After 24 h the reaction was evaporated to dryness and purified by prep-RP-HPLC.

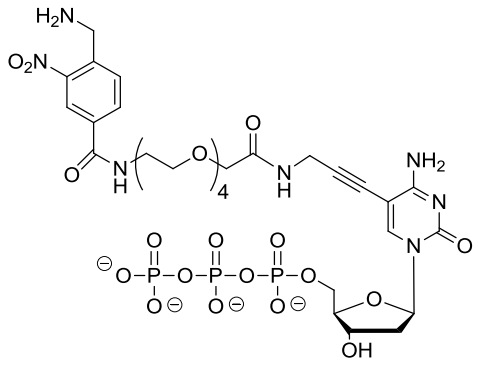
HPLC gradient: 50 – 100% B over 30 min where A= 0.1 M TEAB and B= 0.1 M TEAB + 50% MeCN. Retention time of product = 16.1 min.

^{1}H NMR (400 MHz, MeOD) δ 8.52 (s, 1H, Ar-C*H*), 8.13 (s, 2H, H6, Ar-C*H*), 7.82 (d, *J* = 7.5 Hz, 2H, 2 x Ar-CH_2), 7.68 (d, *J* = 7.3 Hz, 2H, 2 x Ar-C*H*), 7.56 (d, *J* = 8.1 Hz, 1H, Ar-C*H*), 7.45 – 7.30 (m, 4H, 4 x Ar-C*H*), 6.20 (s, 1H, *H*1'), 4.56 (bs, 1H, *H*3'), 4.47 (d, *J* = 6.2 Hz, 1H, *H*4'), 4.36 – 4.18 (m, 5H, C*H*, 2 x CH_2), 4.09 – 4.00 (m, 4H, *H*5', *H*5", CH_2), 3.70 – 3.62 (m, 18H, 9 x CH_2), 2.40 – 2.32 (m, 1H, *H*2'), 2.22 – 2.14 (m, 1H, *H*2").

^{31}P NMR (162 MHz, MeOD) δ -10.12 (d, *J* = 21.0 Hz), -11.22 (d, *J* = 21.1 Hz), -23.45 (t, *J* = 20.1 Hz).

Acc. Mass [ESI-]: Calculated for $C_{45}H_{54}N_7O_{23}P_3$ $[M-H]^-$: 1152.2411 Observed: 1152.2376 $[M-H]^-$

5-(Aminomethyl)-3-nitrobenzamido)-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (72)



5-(4-(fluorenylmethyloxycarbonyl-aminomethyl)-3-nitrobenzamido)-PEG$_4$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**71**) was dissolved in a 1:1 solution of H_2O:NH_4OH (35%) (2 mL total volume) and stirred overnight at room temp. The reaction was then evaporated to dryness and purified by HPLC to give a yellow solid (20 µmol, 35% over two steps).

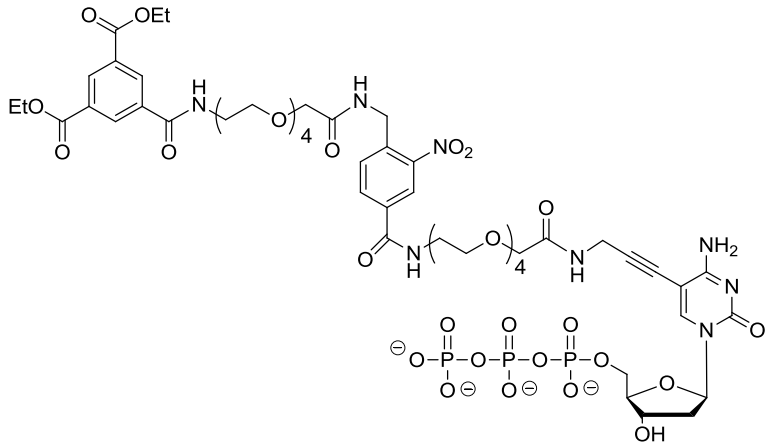
HPLC gradient: 50 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 13.6 min.

^{1}H NMR (400 MHz, D_2O) δ 8.40 (s, 1H, Ar-C*H*), 8.00 (d, *J* = 7.3 Hz, 1H, Ar-C*H*), 7.92 (s, 1H, *H*6), 7.66 (d, *J* = 7.8 Hz, 1H, Ar-C*H*), 6.03 (t, *J* = 6.0 Hz, 1H, *H*1'), 4.48 (s, 1H, *H*3'), 4.15 – 4.07 (m, 5H, *H*4', *H*5', *H*5'', CH_2), 4.02 (s, 2H, CH_2), 3.69 – 3.52 (m, 18H, 9 x CH_2), 2.34 – 2.25 (m, 1H, *H*2'), 2.19 – 2.09 (m, 1H, *H*2'').

^{31}P NMR (162 MHz, D_2O) δ -6.35 (d, *J* = 19.8 Hz), -11.32 (d, *J* = 19.7 Hz), -22.48 (t, *J* = 19.7 Hz).

Acc. Mass [ESI-]: Calculated for $C_{30}H_{44}N_7O_{21}P_3$ $[M-H]^-$: 930.1730 Observed: 930.1753 $[M-H]^-$

5-[(3,5-diethoxycarbonylbenzamido-PEG$_4$-acetamido]methyl)-3-nitrobenzamido))-PEG$_4$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (73)



TSTU (24 mg, 84 µmol, 4.1 eq.) and DIPEA (28 µL, 160 µmol, 8 eq.) were added sequentially to a solution of 3,5-diethoxycarbonylbenzamido-PEG$_4$-acid (**39**) (40 mg, 84 µmol, 4.2 eq.) in anhydrous DMF (1 mL). After 3 h the contents of the flask were syringed into a stirring solution of 5-((4-(aminomethyl)-3-nitrobenzamido)-PEG$_4$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**62**) (20 µmol) in sodium borate buffer (pH 8.5, 0.1 M, 1 mL) at 0°C. After 24 h the reaction was evaporated to dryness and purified by HPLC. The product fractions were subsequently pooled and evaporated to dryness to give the title triphosphate in an 81% yield (16.2 µmol).

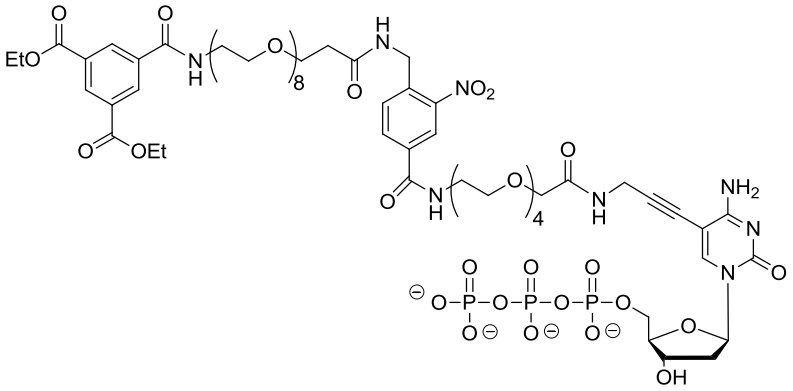
HPLC gradient: 5-100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 23.3 min.

^{1}H NMR (400 MHz, D_2O) δ 8.40 (s, 1H, Ar-C*H*), 8.32 (s, 2H, 2 x Ar-C*H*), 7.99 – 7.92 (m, 2H, Ar-C*H*), 7.93 (s, 1H. H6), 7.85 (d, *J* = 7.7 Hz, 1H, Ar-C*H*), 7.41 (d, *J* = 7.9 Hz, 1H, Ar-C*H*), 6.25 (t, *J* = 6.8 Hz, 2H, CH_2), 6.00 (t, *J* = 5.4 Hz, 1H, *H*1'), 4.59 – 4.47 (m, 1H, *H*3'), 4.32 (m, 4H, 2 x CH_2), 4.20 – 3.97 (m, 9H, *H*4', *H*5', *H*5'', 3 x CH_2), 3.77 – 3.41 (m, 36H, 18 x CH_2), 2.33 – 2.25 (m, 1H, *H*2'), 2.22 – 2.04 (m, 1H, *H*2''), 1.33 (t, *J* = 7.0 Hz, 6H, 2 x CH_3).

^{31}P NMR (162 MHz, D_2O) δ -6.37 (d, *J* = 21.1 Hz), -11.39 (d, *J* = 19.9 Hz), -22.56 (t, *J* = 20.7 Hz).

Acc. Mass [ESI-]: Calculated for $C_{53}H_{75}N_8O_{31}P_3$ $[M-H]^-$: 1411.3678 Observed: 1411.3663 $[M-H]^-$

5-[(3,5-diethoxycarbonylbenzamido-PEG$_8$-acetamido)ethyl)]-3-nitrobenzamido))-PEG$_4$-methylcarboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (74)



TSTU (6 mg, 20 µmol, 2 eq.) and DIPEA (7 µL, 40 µmol, 4 eq.) were added sequentially to a solution of 3,5-diethoxycarbonylbenzamido-PEG$_8$-acid (**32**) (14 mg, 20 µmol, 2 eq.) in anhydrous DMF (1 mL). After 3 h the contents of the flask were transferred via a syringe into a stirred solution of 5-((4-(Aminomethyl)-3-nitrobenzamido)-PEG$_4$-carboxamido-prop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**72**) (10 µmol) in sodium borate buffer (pH 8.5, 0.1 M, 1 mL) at 0°C. After 24 h the reaction was evaporated to dryness and purified by RP-HPLC. The product fractions pooled and evaporated to give the title product in a 4% yield (380 nmol).

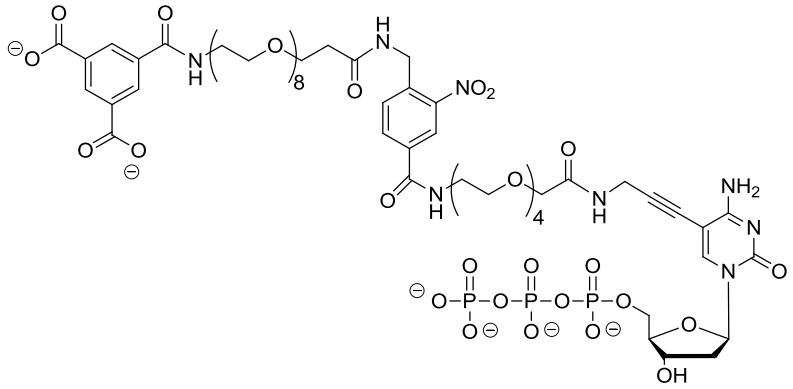
HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN/0.1 M TEAB. Retention time of product = 25.4 min.

^{1}H NMR (400 MHz, D_2O) δ 8.54 (d, *J* = 1.8 Hz, 1H, Ar-C*H*), 8.44 (m, 2H, 2 x Ar-C*H*), 8.29 (s, 1H, Ar-C*H*), 7.93 (m, 2H, Ar-C*H*, *H*6), 7.51 (d, *J* = 8.2 Hz, 1H, Ar-C*H*), 6.02 (t, *J* = 6.3 Hz, 1H, *H*1'), 4.35 (m, 4H, 2x CH_2), 4.23 – 4.05 (m, 5H, *H*4', *H*5', *H*5", CH_2), 4.00 (s, 2H, CH_2), 3.76 – 3.47 (m, 52H, 26 x CH_2), 2.30 (dd, *J* = 13.0, 7.3 Hz, 1H, *H*2'), 2.18 – 2.10 (m, 1H, *H*2"), 1.36 – 1.30 (m, 6H, 2 x CH_3).

^{31}P NMR (162 MHz, D_2O) δ -6.13, -11.18, -22.02

Acc. Mass [ESI-]: Calculated for $C_{62}H_{93}N_8O_{35}P_3$ $[M-H]^-$: 1601.4883 Observed: 1601.4822 $[M-H]^-$

5-[(3,5-dicarboxylbenzamido-PEG$_8$-acetamido)ethyl)-3-nitrobenzamido)]-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (75)



A solution of sodium hydroxide (2 M, 12.5 µL) was added to a stirring solution of 5-[(3,5-diethoxycarbonylbenzamido-PEG$_8$-acetamido)methyl)-3-nitrobenzamido)]-PEG$_4$-methylcarboxamidoprop-1-ynyl)-2'-deoxycytidine-5'-triphosphate (**74**) (380 nm) in water (87.5 µL) at 0°C. After 1 h the reaction was neutralised with the dropwise addition of aqueous 1 M acetic acid before evaporating to dryness. The crude product was isolated by RP-HPLC purification (340 nmol, 89% yield).

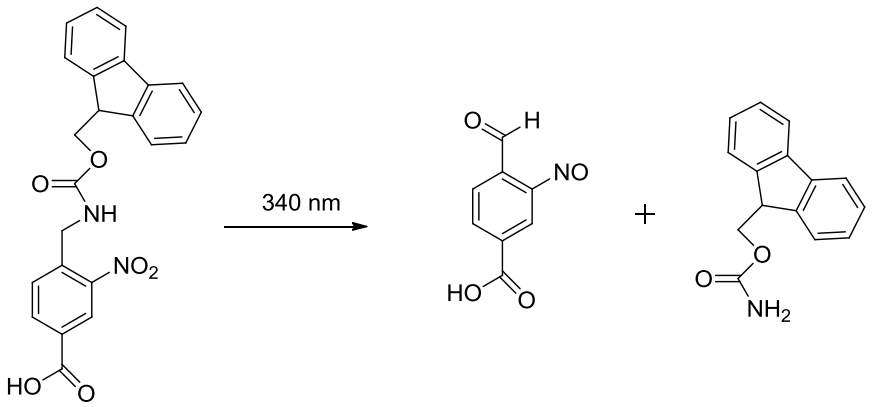
HPLC gradient: 5 – 100% B over 30 min where A= 0.1 M TEAB and B= 50% MeCN / 0.1 M TEAB. Retention time of product = 15.1 min.

^{1}H NMR (400 MHz, D_2O) δ 8.37 (d, *J* = 15.7 Hz, 2H, 2 x Ar-*H*), 8.21 (s, 2H, 2 x Ar-*H*), 7.96 (m, 2H, Ar-*H*, *H*6), 7.57 (d, *J* = 7.9 Hz, 1H, Ar-*H*), 6.07 (t, *J* = 6.2 Hz, 1H, *H*1'), 4.17 – 4.11 (m, 5H, *H*4', *H*5', *H*5", CH_2), 4.00 (s, 2H, CH_2), 3.71 – 3.46 (m, 56H, 28 x CH_2), 2.37 – 2.30 (m, 1H, *H*2'), 2.21 – 2.13 (m, 1H, *H*2').

^{31}P NMR (162 MHz, D_2O) δ -7.59, -10.95, -20.44.

Acc. Mass [ESI-]: Calculated for $C_{58}H_{85}N_8O_{35}P_3$ $[M-H]^-$: 1545.4257 Observed: 1545.4235 $[M-H]^-$

General procedure for the photolysis of (68)



A solution of 4-(Fluorenylmethyloxycarbonyl-aminomethyl)-3-nitrobenzoic acid (**68**) (2 mg, 5 μmol) was dissolved 1:1 H_2O/DMSO solvent mix (1 mL) and placed in a 0.5 mL silica (quartz) cuvette. The photolysis reaction was performed using a FluoroMax-4 spectrofluorometer; $\lambda_{excitation}$ 340 nm, slit width 5 nm.

The emission was monitored every 150 seconds over the range λ 350 – 600 nm with the maxima at 400 nm collected. Measurements were continued in this manner for 500 mins.

General procedure for primer extension reactions

Using either HotStarTaq*, Veraseq 2.0, Phusion, Vent (exo-), Klenow (exo-), Therminator or Omni-KlenTaq DNA polymerase, primer extension reactions were performed as 20 µL reactions comprising of: 1 U polymerase in 1x polymerase specific buffer, 20 nM Cy5-5'-labelled primer and 15 nM template. To investigate the incorporation of a modified dCTP a SC-dCTP mix (SC-dCTP, dATP, dGTP, dTTP) was included alongside the three controls described below. All dNTP mixes were used as a concentration of 625 nM. In all cases the polymerase was added lastly to the reaction mixes before the desired incubation period.

After incubation, each reaction was stopped with the addition of 5 µL complementary sequence (100 µM) in a solution of EDTA (250 mM, pH8), formamide and Orange G gel loading dye. The reaction mixtures were denatured by heating at 95°C for 3 minutes and analysed by electrophoresis on a 17.5% polyacrylamide gel (PAGE). The gel was imaged by fluorescence visualisation facilitated by the use of the Cy5-labelled primer.

Controls used;

Positive control: dATP, dGTP, dTTP, dCTP

Polymerase fidelity control: dATP, dGTP, dTTP

Negative control: No polymerase.

*HotStarTaq polymerase was heated to 95°C for 3 mins prior to use.

General procedure for arrayed primer extension reactions

The aldehyde pre-functionalised glass slides were purchased from Schott and the immobilisation of the 5'-amine terminated primer performed by Heather Murton at QuantuMDx.

Using Therminator or Vent (exo-) DNA polymerases arrayed primer extension reactions were performed as 40 µL reactions comprising of 8 U polymerase in a 1x polymerase specific buffer and 80 nM template. To investigate the incorporation of a modified dCTP a SC-dCTP mix (SC-dCTP, dATP, dGTP, Cy3-dUTP) was included alongside the three controls described below.

Positive control: dATP, dGTP, Cy3-dUTP, dCTP

Polymerase fidelity control: dATP, dGTP, Cy3-dUTP

Negative control: No polymerase

Each SC-dCTP mix is introduced with an equal concentration of dATP, dGTP, Cy3-dUTP and SC-dCTP with a total concentration of 2.5 µM dNTPs. Each control also has a 2.5 µM total dNTP concentration. Once each reaction mixture had been placed within a chamber the slide was incubated for 30 minutes at 25°C or 2 hours at 40°C. After the incubation was complete the chambers were emptied using a pipette, the adhesive chamber barriers removed and the slide thoroughly washed to remove any Cy3-dUTP present. The slide was imaged using a GenePix 4100A by fluorescence visualisation facilitated by the use of the Cy3-labelled dUTP.

Chapter 7 –*References*

7. References

1. Crick,F.H.C. and Watson,J.D. (1954) The Complementary Structure of Deoxyribonucleic Acid. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, **223**, 80–96.

2. Brown,W.H., Foote,C.S., Iverson,B.L. and Anslyn,E. V. (2010) Organic Chemistry Cengage Learning.

3. Corfù,N.A. and Sigel,H. (1991) Acid-base properties of nucleosides and nucleotides as a function of concentration. Comparison of the proton affinity of the nucleic base residues in the monomeric and self-associated, oligomeric 5'-triphosphates of inosine (ITP), guanosine (GTP), and ade. *Eur. J. Biochem.*, **199**, 659–69.

4. Watson, J.D. and Crick,F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–8.

5. Franklin,R.E. and Gosling,R.G. (1953) Molecular Configuration in Sodium Thymonucleate. *Nature*, **171**, 740–41.

6. Rich,A. and Zhang,S. (2003) Timeline: Z-DNA: the long road to biological function. *Nat. Rev. Genet.*, **4**, 566–72.

7. Rich,A., Nordheim,A. and Wang,A.H. (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.*, **53**, 791–846.

8. Meselson,M. and Stahl,F.W. (1958) The replication of DNA in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.*, **44**, 671–82.

9. Nakamura,T., Zhao,Y., Yamagata,Y., Hua,Y. and Yang,W. (2012) Watching DNA polymerase η make a phosphodiester bond. *Nature*, **487**, 196–201.

10. Erlich,H., Gelfand,D. and Sninsky,J. (1991) Recent advances in the polymerase chain reaction. *Science*, **252**, 1643–51.

11. Hadidi,A., Flores,R., Randles,J. and Semancik,J. (2003) Viroids: Properties, Detection, Diseases and their Control (Google eBook) Csiro Publishing.

12. Kunkel,T.A. (2004) DNA replication fidelity. *J. Biol. Chem.*, **279**, 16895–8.

13. Cline,J. (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.*, **24**, 3546–51.

14. Bertina,R.M., Koeleman,B.P., Koster,T., Rosendaal,F.R., Dirven,R.J., de Ronde,H., van der Velden,P.A. and Reitsma,P.H. (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*, **369**, 64–67.

15. The Finished Human Genome - Wellcome To The Genomic Age - Wellcome Trust Sanger Institute.

16. Perkel,J. (2013) Finding the true $1000 genome. *Biotechiques*, **54**, 71–4.

17. Sanger,F. and Coulson,A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.*, **94**, 441–8.

18. Maxam, A. M., & Gilbert,W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci.*, **74(2)**, 560–4.

19. Mattes,W.B., Hartley,J.A. and Kohn,K.W. (1986) Mechanism of DNA strand breakage by piperidine at sites of N7-alkylguanines. *Biochim. Biophys. Acta - Gene Struct. Expr.*, **868**, 71–6.

20. Maniatis,T., Jeffrey,A. and Van deSande,H. (1975) Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry*, **14**, 3787–94.

21. Sanger,F. (1977) DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci.*, **74**, 5463–67.

22. Prober,J., Trainor,G., Dam,R., Hobbs,F., Robertson,C., Zagursky,R., Cocuzza,A., Jensen,M. and Baumeister,K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science*, **238**, 336–41.

23. Smith,L.M., Sanders,J.Z., Kaiser,R.J., Hughes,P., Dodd,C., Connell,C.R., Heiner,C., Kent,S.B. and Hood,L.E. Fluorescence detection in automated DNA sequence analysis. *Nature*, **321**, 674–9.

24. Hood,L.E., Hunkapiller,M.W. and Smith,L.M. (1987) Automated DNA sequencing and analysis of the human genome. *Genomics*, **1**, 201–12.

25. Ansorge,W., Sproat,B., Stegemann,J., Schwager,C. and Zenke,M. (1987) Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.*, **15**, 4593–4602.

26. Venter,J.C., Adams,M.D. and Myers,E.W. (2001) The sequence of the human genome. *Science*, **291**, 1304–51.

27. Mardis,E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, **9**, 387–402.

28. Wheeler,D.A., Srinivasan,M., Egholm,M., Shen,Y., Chen,L., McGuire,A., He,W., Chen,Y.-J., Makhijani,V., Roth,G.T., *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–6.

29. Levy,S., Sutton,G., Ng,P.C., Feuk,L., Halpern,A.L., Walenz,B.P., Axelrod,N., Huang,J., Kirkness,E.F., Denisov,G., *et al.* (2007) The diploid genome sequence of an individual human. *PLoS Biol.*, **5**, 2113–44.

30. Nyrén,P. and Lundin,A. (1985) Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal. Biochem.*, **151**, 504–9.

31. Nyrén,P. (2007) The history of pyrosequencing. *Methods Mol. Biol.*, **373**, 1–14.

32. Ruparel,H., Bi,L., Li,Z., Bai,X., Kim,D.H., Turro,N.J. and Ju,J. (2005) Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 5932–7.

33. Ronaghi,M. (1998) DNA sequencing:A Sequencing Method Based on Real-Time Pyrophosphate. *Science*, **281**, 363–365.

34. England,R. and Pettersson,M. (2005) Pyro Q-CpG™: quantitative analysis of methylation in multiple CpG sites by Pyrosequencing®. *Nat. Methods*, **2**.

35. Hutter,D., Kim,M.-J., Karalkar,N., Leal,N.A., Chen,F., Guggenheim,E., Visalakshi,V., Olejnik,J., Gordon,S. and Benner,S.A. (2010) Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups. *Nucleosides. Nucleotides Nucleic Acids*, **29**, 879–95.

36. Ranasinghe,R.T. and Brown,T. (2011) Ultrasensitive fluorescence-based methods for nucleic acid detection: towards amplification-free genetic analysis. *Chem. Commun. (Camb).*, **47**, 3717–35.

37. Schadt,E.E., Turner,S. and Kasarskis,A. (2010) A window into third-generation sequencing. *Hum. Mol. Genet.*, **19**, 227–40.

38. Eid,J., Fehr,A. and Gray,J. (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133–8.

39. Schloss,P.D., Westcott,S.L., Jenior,M.L. and Highlander,S.K. (2015) Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system PeerJ Inc.

40. Quail,M.A., Smith,M., Coupland,P., Otto,T.D., Harris,S.R., Connor,T.R., Bertoni,A., Swerdlow,H.P. and Gu,Y. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

41. Rothberg,J.M., Hinz,W., Rearick,T.M., Schultz,J., Mileski,W., Davey,M., Leamon,J.H., Johnson,K., Milgrew,M.J., Edwards,M., *et al.* (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, **475**, 348–52.

42. Merriman,B. and Rothberg,J.M. (2012) Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, **33**, 3397–417.

43. Perkel,J. (2011) Making Contact with Sequencing's Fourth Generation. *Biotechiques*, **50**, 93–5.

44. Eisenstein,M. (2012) Oxford Nanopore announcement sets sequencing sector abuzz. *Nat. Biotechnol.*, **30**, 295–6.

45. Meller,A. and Branton,D. (2002) Single molecule measurements of DNA transport through a nanopore. *Electrophoresis*, **23**, 2583–91.

46. Maitra,R.D., Kim,J. and Dunbar,W.B. (2012) Recent advances in nanopore sequencing. *Electrophoresis*, **33**, 3418–28.

47. Kasianowicz,J.J., Brandin,E., Branton,D. and Deamer,D.W. (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 13770–3.

48. DNA: An introduction to nanopore sequencing - Analytes and Applications: DNA, RNA, proteins - Technology - Oxford Nanopore Technologies.

49. Znosko,B.M., Barnes,T.W., Krugh,T.R. and Turner,D.H. (2003) NMR studies of DNA single strands and DNA:RNA hybrids with and without 1-propynylation at C5 of oligopyrimidines. *J. Am. Chem. Soc.*, **125**, 6090–7.

50. Obeid,S., Baccaro,A., Welte,W., Diederichs,K. and Marx,A. (2010) Structural basis for the synthesis of nucleobase modified DNA by Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 21327–31.

51. Laos,R., Thomson,J.M. and Benner,S.A. (2014) DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides. *Front. Microbiol.*, **5**, 565.

52. Bentley,D.R., Balasubramanian,S., Swerdlow,H.P., Smith,G.P., Milton,J., Brown,C.G., Hall,K.P., Evers,D.J., Barnes,C.L., Bignell,H.R., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53–9.

53. Li,Y., Mitaxov,V. and Waksman,G. (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 9491–6.

54. D.Williams, J.Brazier, J.Grasby, S.Lee,. (2001) Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Res.*, **29**, 1565–73.

55. Thoresen,L.H., Jiao,G.-S., Haaland,W.C., Metzker,M.L. and Burgess,K. (2003) Rigid, conjugated, fluoresceinated thymidine triphosphates: syntheses and polymerase mediated incorporation into DNA analogues. *Chemistry*, **9**, 4603–10.

56. Kuwahara,M., Nagashima,J., Hasegawa,M., Tamura,T., Kitagata,R., Hanawa,K., Hososhima,S., Kasamatsu,T., Ozaki,H. and Sawai,H. (2006) Systematic characterization of 2’-deoxynucleoside- 5'-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. *Nucleic Acids Res.*, **34**, 5383–94.

57. Jäger,S., Rasched,G., Kornreich-Leshem,H., Engeser,M., Thum,O. and Famulok,M. (2005) A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.*, **127**, 15071–82.

58. Obeid,S., Baccaro,A., Welte,W., Diederichs,K. and Marx,A. (2010) Structural basis for the synthesis of nucleobase modified DNA by Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 21327–31.

59. Hollenstein,M. (2013) Deoxynucleoside triphosphates bearing histamine, carboxylic acid, and hydroxyl residues--synthesis and biochemical characterization. *Org. Biomol. Chem.*, **11**, 5162–72.

60. Finn,P.J., Bull,M.G., Xiao,H., Phillips,P.D., Nelson,J.R., Grossmann,G., Nampalli,S., McArdle,B.F., Mamone,J.A., Flick,P.K., *et al.* (2003) Efficient incorporation of positively charged 2', 3'-dideoxynucleoside-5'-triphosphates by DNA polymerases and their application in 'direct-load' DNA sequencing. *Nucleic Acids Res.*, **31**, 4769–78.

61. Finn,P.J., Sun,L., Nampalli,S., Xiao,H., Nelson,J.R., Mamone,J.A., Grossmann,G., Flick,P.K., Fuller,C.W. and Kumar,S. (2002) Synthesis and application of charge-modified dye-labeled dideoxynucleoside-5'-triphosphates to 'direct-load' DNA sequencing. *Nucleic Acids Res.*, **30**, 2877–85.

62. Cui,Y., Wei,Q., Park,H. and Lieber,C.M. (2001) Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science*, **293**, 1289–92.

63. Zhang,G.-J., Zhang,G., Chua,J.H., Chee,R.-E., Wong,E.H., Agarwal,A., Buddharaju,K.D., Singh,N., Gao,Z. and Balasubramanian,N. (2008) DNA sensing by silicon nanowire: charge layer distance dependence. *Nano Lett.*, **8**, 1066–70.

64. Patolsky,F., Zheng,G. and Lieber,C.M. (2006) Nanowire sensors for medicine and the life sciences. *Nanomedicine (Lond).*, **1**, 51–65.

65. Baddiley,J., Michelson,A.M. and Todd,A.R. (1949) 124. Nucleotides. Part II. A synthesis of adenosine triphosphate. *J. Chem. Soc.*, **124**, 582–586.

66. Smith,M. and Khorana,H.G. (1958) Nucleoside Polyphosphates. VI. 1 An Improved and General Method for the Synthesis of Ribo- and Deoxyribonucleoside 5'-Triphosphates. *J. Am. Chem. Soc.*, **80**, 1141–1145.

67. Moffatt,J.G. (1964) A general synthesis of nucleoside-5'-triphosphates. *Can. J. Chem.*, **42**, 599–604.

68. Moffatt,J.G. and Khorana,H.G. (1961) Nucleoside Polyphosphates. X. 1 The Synthesis and Some Reactions of Nucleoside-5' Phosphoromorpholidates and Related Compounds. Improved Methods for the Preparation of Nucleoside-5' Polyphosphates 1. *J. Am. Chem. Soc.*, **83**, 649–658.

69. Yoshikawa,M., Kato,T. and Takenishi,T. (1967) A novel method for phosphorylation of nucleosides to 5′-nucleotides. *Tetrahedron Lett.*, **8**, 5065–5068.

70. Ludwig,J. (1981) A new route to nucleoside 5’-triphosphates. *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131–3.

71. Yoshikawa,M., Kato,T. and Takenishi,T. (1969) Studies of Phosphorylation. III. Selective Phosphorylation of Unprotected Nucleosides. *Bull. Chem. Soc. Jpn.*, **42**, 3505–3508.

72. Kovács,T. and Ötvös,L. (1988) Simple synthesis of 5-vinyl- and5-ethynyl-2′-deoxyuridine-5′-triphosphates. *Tetrahedron Lett.*, **29**, 4525–4528.

73. Alder,R.W., Bowman,P.S., Steele,W.R.S. and Winterman,D.R. (1968) The remarkable basicity of 1,8-bis(dimethylamino)naphthalene. *Chem. Commun.*, **13**, 723–724.

74. Ludwig,J. and Eckstein,F. (1989) Rapid and efficient synthesis of nucleoside 5’-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, **54**, 631–635.

75. Knochel,P. and Molander,G. (2014) Comprehensive Organic Synthesis second. Newnes.

76. Chemistry of Nucleosides and Nucleotides, Volume 1 (2013) Springer Science & Business Media.

77. Sonogashira,K., Tohda,Y. and Hagihara,N. (1975) A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Lett.*, **16**, 4467–4470.

78. Ogilvie,K. (1973) The tert-butyldimethylsilyl group as a protecting group in deoxynucleosides. *Can. J. Chem.*, **51**, 3799–3807.

79. Ti,G.S., Gaffney,B.L. and Jones,R.A. (1982) Transient protection: efficient one-flask syntheses of protected deoxynucleosides. *J. Am. Chem. Soc.*, **104**, 1316–1319.

80. Kocienski,P.J. (2005) Protecting Groups Thieme.

81. Wang,Y., Tkachenko,B.A., Schreiner,P.R. and Marx,A. (2011) Diamondoid-modified DNA. *Org. Biomol. Chem.*, **9**, 7482–90.

82. Schoetzau,T., Langner,J., Moyroud,E., Roehl,I., Vonhoff,S. and Klussmann,S. (2003) Aminomodified nucleobases: functionalized nucleoside triphosphates applicable for SELEX. *Bioconjug. Chem.*, **14**, 919–26.

83. Matulic-Adamic,J., Daniher,A.T., Karpeisky,A., Haeberli,P., Sweedler,D. and Beigelman,L. (2000) Functionalized nucleoside 5′-triphosphates for in vitro selection of new catalytic ribonucleic acids. *Bioorg. Med. Chem. Lett.*, **10**, 1299–1302.

84. Caton-Williams,J., Smith,M., Carrasco,N. and Huang,Z. (2011) Protection-free one-pot synthesis of 2'-deoxynucleoside 5'-triphosphates and DNA polymerization. *Org. Lett.*, **13**, 4156–9.

85. Caton-Williams,J., Lin,L., Smith,M. and Huang,Z. (2011) Convenient synthesis of nucleoside 5'-triphosphates for RNA transcription. *Chem. Commun. (Camb).*, **47**, 8142–4.

86. Caton-Williams,J., Lin,L., Smith,M. and Huang,Z. (2011) Convenient synthesis of nucleoside 5'-triphosphates for RNA transcription. *Chem. Commun. (Camb).*, **47**, 8142–4.

87. Zhu,Z., Chao,J., Yu,H. and Waggoner,A.S. (1994) Directly labeled DNA probes using fluorescent nucleotides with different length linkers. *Nucleic Acids Res.*, **22**, 3418–22.

88. Lacenere,C., Garg,M.K., Stoltz,B.M. and Quake,S.R. (2006) Effects of a modified dye-labeled nucleotide spacer arm on incorporation by thermophilic DNA polymerases. *Nucleosides. Nucleotides Nucleic Acids*, **25**, 9–15.

89. Bond Lengths and Energies (2015) *http://www.science.uwaterloo.ca/~cchieh/cact/c120/bondel.html*.

90. Alcantar,N.A., Aydil,E.S. and Israelachvili,J.N. (2000) Polyethylene glycol-coated biocompatible surfaces. *J. Biomed. Mater. Res.*, **51**, 343–51.

91. Lee,J.W., Kim,B.-K., Kim,H.J., Han,S.C., Shin,W.S. and Jin,S.-H. (2006) Convergent Synthesis of Symmetrical and Unsymmetrical PAMAM Dendrimers. *Macromolecules*, **39**, 2418–2422.

92. Yu,T., Liu,X., Bolcato-Bellemin,A.-L., Wang,Y., Liu,C., Erbacher,P., Qu,F., Rocchi,P., Behr,J.-P. and Peng,L. (2012) An amphiphilic dendrimer for effective delivery of small interfering RNA and gene silencing in vitro and in vivo. *Angew. Chem. Int. Ed. Engl.*, **51**, 8478–84.

93. Kolb,H.C., Finn,M.G. and Sharpless,K.B. (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed. Engl.*, **40**, 2004–2021.

94. Lee,J.W., Han,S.C., Yun,S.-H. and Jin,S.-H. (2013) Convergent Synthesis of Carbazole Core PAMAM Dendrimer via Click Chemistry. *Bull. Korean Chem. Soc.*, **34**, 971–974.

95. Rostovtsev,V. V., Green,L.G., Fokin,V. V. and Sharpless,K.B. (2002) A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective 'Ligation' of Azides and Terminal Alkynes. *Angew. Chemie*, **114**, 2708–2711.

96. Ingale,S.A. and Seela,F. (2013) Stepwise click functionalization of DNA through a bifunctional azide with a chelating and a nonchelating azido group. *J. Org. Chem.*, **78**, 3394–9.

97. Chan,T.R., Hilgraf,R., Sharpless,K.B. and Fokin,V. V (2004) Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.*, **6**, 2853–5.

98. Copper(II)-TBTA complex (2015) *http://www.lumiprobe.com/p/click-chemistry-catalys*.

99. Baccaro,A., Steck,A.-L. and Marx,A. (2012) Barcoded nucleotides. *Angew. Chem. Int. Ed. Engl.*, **51**, 254–7.

100. Li,Z., Bai,X., Ruparel,H., Kim,S., Turro,N.J. and Ju,J. (2003) A photocleavable fluorescent nucleotide for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 414–9.

101. Ramos,R., Manning,B., Aviñó,A., Gargallo,R. and Eritja,R. (2009) Photocleavage of Peptides and Oligodeoxynucleotides Carrying 2-Nitrobenzyl Groups. *Helv. Chim. Acta*, **92**, 613–622.

102. Li,Z., Bai,X., Ruparel,H., Kim,S., Turro,N.J. and Ju,J. (2003) A photocleavable fluorescent nucleotide for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 414–9.

103. Abramova,T. V. and Silnikov,V.N. (2006) 4-Aminomethyl-3-Nitrobenzoic Acid—A Photocleavable Linker for Oligonucleotides Containing Combinatorial Libraries. *Nucleosides, Nucleotides, and Nucleic Acids*, **24**, 1333–1343.

104. Metzker,M.L., Raghavachari,R., Richards,S., Jacutin,S.E., Civitello,A., Burgess,K. and Gibbs,R.A. (1994) Termination of DNA synthesis by novel 3'-modifieddeoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.*, **22**, 4259–4267.

105. Dutson,C. (2015) Syntheis of C5-modified dUTPs for use in SBS sequencing.

106. Nič,M., Jirát,J., Košata,B., Jenkins,A. and McNaught,A. eds. (2009) Norrish Type II photoreaction. In *IUPAC Compendium of Chemical Terminology*. IUPAC, Research Triagle Park, NC.

107. Pelliccioli,A.P. and Wirz,J. (2002) Photoremovable protecting groups: reaction mechanisms and applications. *Photochem. Photobiol. Sci.*, **1**, 441–458.

108. Therminator™ DNA Polymerase | NEB (2015) *https://www.neb.com/products/m0261-therminator-dna*.

109. Powell,M., Somers,R. and Froehler,B. (2015) Glen Research Extinction Coeffiecients. *http://www.glenresearch.com/Technical/Extinctions.html*.

110. Curphey,T.J. (1979) Trifluoroacetylation of amino acids and peptides by ethyl trifluoroacetate. *J. Org. Chem.*, **44**, 2805–2807.

111. Corey,E.J. and Venkateswarlu,A. (1972) Protection of hydroxyl groups as tert-butyldimethylsilyl derivatives. *J. Am. Chem. Soc.*, **94**, 6190–6191.

112. Xu,S., Held,I., Kempf,B., Mayr,H., Steglich,W. and Zipse,H. (2005) The DMAP-catalyzed acetylation of alcohols--a mechanistic study. *Chemistry*, **11**, 4751–7.

113. Zhang,W. and Robins,M.J. (1992) Removal of silyl protecting groups from hydroxyl functions with ammonium fluoride in methanol. *Tetrahedron Lett.*, **33**, 1177–1180.

114. Hobbs,F.W. (1989) Palladium-catalyzed synthesis of alkynylamino nucleosides. A universal linker for nucleic acids. *J. Org. Chem.*, **54**, 3420–3422.

115. Giller,G. (2003) Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. I. Chemical synthesis of various reporter group-labeled 2'-deoxyribonucleoside-5'-triphosphates. *Nucleic Acids Res.*, **31**, 2630–2635.

116. Spiegeld,D. and Parker,C. (2012) Bifunctional molecules with antibody-recruiting and entry inhibitory activity against the human immunodefiency virus, WO2012068366.

117. Wolfe,J.L. (2002) Synthesis and polymerase incorporation of 5'-amino-2',5'-dideoxy-5'-N-triphosphate nucleotides. *Nucleic Acids Res.*, **30**, 3739–3747.

118. Lee,J.W., Kim,J.H., Kim,B.-K., Kim,J.H., Shin,W.S. and Jin,S.-H. (2006) Convergent synthesis of PAMAM dendrimers using click chemistry of azide-functionalized PAMAM dendrons. *Tetrahedron*, **62**, 9193–9200.

119. Sørensen,R.S., Okholm,A.H., Schaffert,D., Kodal,A.L.B., Gothelf,K. V and Kjems,J. (2013) Enzymatic ligation of large biomolecules to DNA. *ACS Nano*, **7**, 8098–104.

120. Barany,G. and Albericio,F. (1985) Three-dimensional orthogonal protection scheme for solid-phase peptide synthesis under mild conditions. *J. Am. Chem. Soc.*, **107**, 4936–4942.

121. Lattuada,L., Cappelletti,E., Linder,K. and Nunn,A. (2011) Gastrin releasing peptide compounds. US 20110250133.

Chapter 8 –Appendix