



**Synthesis of Nucleoside Analogues and Peptides
for Nanopore Analysis and Controlled
Bioactivity**

Mikiembo Nsimba Kukwikila

*A thesis submitted in partial fulfilment of the requirements for the
degree award of:*

Doctor of Philosophy

University College London

Department of Chemistry

2013

DECLARATION

I, Mikiembo Nsimba Kukwikila, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

This research describes the synthesis of nucleoside analogues and peptides, and their application to nanopore technology and cell biology. Nanopore technology is a technique in which analytes are detected by disturbances in ionic current caused as they traverse a pore, at a fixed applied potential. Here potential applications of the technology are explored.

A nanopore-based method to detect enzymatic activity was investigated. The technique involved the detection of peptide fragments proteolytically released from a solid-support. The released fragments gave rise to distinct electrical signals which facilitated their identification and characterisation. The frequency with which they traversed the pore was indicative of enzyme activity. The technique was successfully applied to the detection of the protease enzyme, renin, in the presence of human serum.

An approach to count nucleotide repeat sequences, which are the basis of forensic DNA fingerprinting, was also investigated. It was hypothesised that chemically modified copies of the repeat regions could be generated using primer extension and nucleoside analogues. The copying could be performed such that a single analogue was incorporated per repeat. Ensuing nanopore analysis would then give rise to electrical signals in which the analogues could be identified from unique signals. The number of repeat sequences could simply be determined by counting the number of signals. It was envisaged that this could be achieved using adamantane-modified nucleosides. However, these did not give rise to the expected result.

Inspired by the adamantane project, analogues of the drug, azacytidine were synthesised. Azacytidine is used to treat cancers that arise due to epigenetic modifications of DNA. It was hypothesised that its side-effects could be mitigated using photocaged-derivatives of the nucleoside, as these would enable greater temporal and spatial control of drug release. Analogues of azacytidine were successfully synthesised and shown to be uncaged back to the parent nucleoside.

CONTENTS

	<i>Page</i>
Aknowledgments.....	7
List of Figures and Tables.....	8
Abbreviations.....	15
Publications and Presentations.....	18
Chapter	
1 Introduction.....	19
1.1 Basics of Nanopores.....	20
1.1.1 Nanopores.....	21
1.2.1 Solid-State Nanopores.....	22
1.2 Sensing of Nucleic Acids.....	24
1.2.1 Current DNA Sequencing Techniques.....	24
1.2.2 Nanopore Sequencing: Third Generation Sequencing Technologies.....	27
1.2.3 Nanopore Sequencing: Seminal Investigations.....	27
1.2.4 Approaches to Nanopore Detection of Oligonucleotides.....	29
1.2.4.1 Employing Chemically Modified Oligonucleotides.....	29
1.2.4.2 Using an Engineered α HL Pore.....	32
1.2.4.3 Utilising Enzymes as Molecular Motors.....	33
1.3 Sensing Enzymatic Activity.....	36
1.3.1 Probing Enzyme Mechanism.....	36
1.3.2 Determining Enzyme Kinetics.....	37
1.4 Sensing Peptides.....	39
1.5 Aims.....	43
2 Chemical Tools for Nanopore Analysis of Enzymatic Activity in Blood Plasma.....	45
2.1 Summary.....	45
2.2 Introduction.....	45
2.2.1 Overview of the Renin-Angiotensin-Aldosterone System.....	45

2.2.2	Biochemistry of Renin-Angiotensin-Aldosterone System.....	47
2.2.3	Measurement of Plasma Renin Activity.....	49
2.3	Aim.....	51
2.4	Solid-Phase Peptide Synthesis.....	52
2.5	Results and Discussion.....	57
2.5.1	Proof-of-Principle Experiments using Trypsin.....	57
2.5.2	Preparation of Immobilised Substrate via Convergent Solid-Phase Peptide Synthesis.....	59
2.5.3	Enzymatic Digestion of Substrate Prepared by CSPPS.....	64
2.5.4	Preparation of Immobilised Substrate via Manual Solid-Phase Peptide Synthesis..	65
2.5.5	Enzymatic Digestion of Substrate Prepared by MSPPS.....	66
2.5.6	Nanopore Analysis.....	67
2.6	Sensing Strategy Applied to the Detection of Renin Activity.....	71
2.6.1	Developing the Sensing Approach: Control Reactions.....	73
2.7	Nanopore Analysis.....	78
2.8	Conclusions.....	80
2.9	Materials and Methods.....	81
3	Sizing Nucleotide Repeat Sequences using Nanopore Technology.....	90
3.1	Summary.....	90
3.2	Introduction.....	91
3.2.1	DNA Structure.....	91
3.2.2	DNA Fingerprinting.....	93
3.3	Aim.....	97
3.4	Resultsing and Discussion.....	99
3.4.1	Synthesis of Chemically Modified Nucleosides.....	99
3.4.2	Phosphorylation of Nucleosides.....	105
3.4.3	Enzymatic Incorporation of Modified Nucleosides.....	112
3.4.4	Oligonucelotide Synthesis.....	121
3.4.5	Nanopore Analysis.....	127

3.5	Conclusion.....	131
3.6	Materials and Methods.....	133
4	Photocaging of 5-Azacytidine for Drug Delivery Purposes.....	149
4.1	Summary.....	149
4.2	Introduction.....	149
4.2.1	DNA Methyltransferases and their Inhibition by Azanucleosides.....	149
4.2.2	Principles of Photocaging.....	153
4.2.3	Photodynamic Therapy.....	156
4.2.4	Design of Photocaged Azacytidine Derivatives.....	160
4.3	Aim.....	164
4.4	Results and Discussion.....	166
4.4.1	Synthesis of Caged Nucleosides.....	166
4.4.2	Uncaging of Caged Cytidine and Azacytidine.....	171
4.4.3	Preliminary Cellular Assays.....	174
4.5	Conclusion.....	177
4.6	Materials and Methods.....	178
5	Outlook.....	187
	References.....	192
	Appendix.....	201

ACKNOWLEDGEMENTS

Firstly, and most importantly, I would like to thank my supervisor, Dr. Stefan Howorka for giving me the opportunity to work on this project and for his excellent guidance throughout. I would also like to extend my gratitude to the past members of the Howorka group, in particular Dr. Nick Mitchell and Dr. Vinciane Borsenberger who provided much need guidance in the early years. I am also truly grateful to Dr. Abil Aliev and Dr. Lisa Haigh for their help with NMR and mass spectroscopy analysis, respectively. All of these people have supported and taught me so much over the years; without them, this project would not have proceeded as smoothly.

A massive thank you to Richard Kerr, Dr. Rachel Morgan, Dr. Jonathan Burns, Dr. Felix Schumacher, Bhavesh Premdjee, Paul Moody and all the users of lab 330 for all their science (and non-science) related advice, good humor, tea and most importantly, friendship. The last few months especially would have been impossible without their help and support; I am truly indebted. A special shout-out to my long suffering friends, especially Sanjayan Sathasivam and Daniza Mahenthiran. It has taken a long time to get to this stage, but they have stuck by me through it all (even though they had little or no clue about what I was doing). I would also like to mention David Kierce and all the ladies that attend his class at the Central School of Ballet. His class was a welcomed distraction from research and on several occasions helped me relax after a difficult day in the lab.

And finally, to my sisters, parents and extended family; my very own cheerleading squad. Everything begins and ends with them. Thank you.

LIST OF FIGURES AND TABLES

Figure 1: (A) Principle of nanopore analysis. A pore is placed within a membrane that separates two electrolyte-filled compartments. Applying a voltage across the membrane results in the flow of ions, this in turn generates a detectable ionic current. (B) Typical electrical trace. When the channel is open and unobstructed, the current level is termed 'open channel current'. As an analyte passes the inner constriction of the pore, the flow of ions is temporarily interrupted resulting in a reduction in the ionic current. The channel is said to be blocked.19

Figure 2: Cross-section of α -hemolysin protein pore in plasma membrane. Adapted from [2].21

Figure 3: Nanopore traces for poly[A], poly[U] and poly[C]. (A) Each polymer gave rise to distinct patterns of blockades. (B) Nanopore traces for co-polymer of poly[A] and poly[C]. Bi-level blockade were observed as the nanopore was able to detect the transition from adenine to cytosine (indicated by the arrows). Adapted from [33].28

Figure 4: Nanopore analysis of chemically modified oligonucleotides. (A) Peptide-modified nucleotides have been used to reduce the translocation speed of oligonucleotides. The modification increases the cross-sectional diameter of the DNA and gives rise to distinct electrical signals. (B) Peptide-modified nucleotides have been incorporated into oligonucleotide probes to enable nanopore-based detection of SNPs. (C) 8-oxo-7,8-dihydroguanine, an oxidised derivative of guanine was detected using α HL pore. The oxidised base was derivatised with a range of amines in order to increase the signal-to-noise ratio.30

Figure 5: DNA sequencing via the exonuclease approach. (A) Nucleotides are cleaved from ssDNA by an exonuclease enzyme (scissors) then detected by a pore modified with a cyclodextrin ring (turquoise). (B) Structure of cyclodextrin ring. For covalent attachment to the protein, a single amino group was coupled to the reactive bi-functional linker, succinimidyl 3-(2-pyridyldithio)propionate (SPDP), and the resultant construct coupled to a cysteine residue engineered into the β -barrel [15].32

Figure 6: Utilising Phi29 DNA polymerase as a molecular motor to sequence DNA. (A) Structure and dimensions of MspA pore. (B) 1, the DNA template (black) complementary base-pairs to the primer (blue). The 'blocking oligomer' (red) prevents extension of the primer in the bulk solution; 2, Applying a voltage causes the captured DNA-enzyme (green) complex to be drawn into the pore. This results in unzipping of the blocking oligomer; 3, The enzyme can now catalyse the incorporation of nucleotides into the

deprotected primer; 4, As DNA synthesis occurs, the DNA complex is rached into the *cis* chamber. When a nucleotide in the template is held within the key sensing region of the pore, a distinct electrical signal is obtained. (C) Representative electrical trace. Adapted from [52].35

Figure 7: Determining enzyme kinetics using nanopores. (A) Genetically engineered α HL pore carrying a single inhibitory peptide (shown in yellow and turquoise). A subunit of the enzyme (blue) is docked onto the inhibitory peptide.^{62,63} (B) Charged-based detection of enzymatic activity. Alkaline phosphatase catalyses the hydrolysis of phosphate groups attached to the pore.⁵² The resultant local change in conductivity is detected by the pore giving rise to distinct modulations in the ionic current. (C) In-situ enzymatic digestion. In the absence of the enzyme (a), current modulations are due solely to the starting substrate. On addition of the enzyme, new blockade translocation events (b) with distinct amplitudes and duration are observed.¹⁴38

Figure 8: Kinetics of polypeptide translocations in the presence of negatively-charged electrostatic traps. The frequency and duration of polypeptide-induced blockades were dependent upon the position of the traps.41

Figure 26: General structure of a DNA nucleoside and the chemical structures of pyrimidine and purine bases. The arrows indicate hydrogen-bonding sites. Each base has hydrogen-donating capabilities, represented by the arrows pointing away from the structures, and hydrogen acceptor capabilities represented by the arrows pointing towards the structure. Adapted from [135].91

Figure 27: Formation of phosphodiester bond. The hydroxyl group of the growing chain performs a nucleophilic attack on the α -phosphorous of the incoming nucleotide. The DNA duplex is maintained by specific hydrogen bonding pattern between the Watson-Crick base-pair; G•C and A•T. DNA strands are said to be anti-parallel as one strand runs in the 5'→3' and the other in the 3'→5'.92

Figure 28 Schematic representation of a microsatellite (STR) DNA marker. PCR primers are designed to target the flanking sequence regions. The number of tandem repeat units in the repeat region varies among individuals making them useful markers for identification. Adapted from [¹³⁹].94

Figure 30: Electrical traces for an oligonucleotide modified with two tyrosine peptide tags. As each tag is sequentially pulled through the nanopore, step-like alterations in the ionic current are observed. Step 1 (red) corresponds to the presence of leading non-modified region of the oligonucleotide strand. As the first bulky tag passes through the inner constriction, the ionic current is reduced (step 2). The intermediate non-modified region

of the oligonucleotide causes the ionic current to increase (step 3). Finally, as the second tag translocates the inner constriction, another reduction in the ionic current is observed (step 4). Adapted from [10].99

Figure 31: Synthesis of adamantane-modified nucleotides. (i) Adamantane methylamine, then Method A: EEDQ, DMF, 18 h or Method B: HBTU, Et₃N, cat. DMAP, DMF 18 h; (ii) 5-iodo-2'-deoxyuridine, CuI, Et₃N, Pd(PPh₃)₄, DMF, 16 h; (iii) POCl₃, proton sponge, 2 h; (iv) nBu₃NH⁺.H₃P₂O₇⁻, 0.1 M TEAB, 45 min..... 101

Figure 32: Mechanism of Sonogashira coupling. The Pd(0) undergoes oxidative addition to the nucleobase (the aryl halide) and in the process is oxidised to Pd(II). The addition of CuI as co-catalyst gives better results as CuI activates alkynes by forming a copper acetylide, which is more reactive and undergoes transmetalation with the palladium complex. The ligands attached to the Pd are converted from the *trans* to the *cis* isomers in an isomerism step. Finally, reductive elimination occurs to afford the product and regenerate the catalyst.103

Figure 33: Possible conformation of nucleosides and diagnostic NOE interactions.104

Figure 35: HPLC traces for the reaction of 5-iodo-2'-deoxyuridine with 1.5 equivalents of POCl₃ at 0°C. Peak A: 5',3'-diphosphate; Peak B: 5'-monophosphate; Peak C: 3'-phosphate; Peak S: Starting nucleoside. Y-axis: Absorbance at 260 nm. Adapted from [¹⁵³]. 108

Figure 36: (A) HPLC traces showing the formation of the phosphorodichloridate intermediate of 3c. Maximal formation occurs within 2 h. The peak between 17.5 and 19 min is due the proton sponge. Analytical HPLC analysis was performed using Varian Pursuit C18 column. See Materials and Methods for details. (B) HPLC trace showing the crude reaction mixture for the triphosphorylation of 3c. Peak 1: pyrophosphate salts; peak 2: a mixture of mono-, di-, tri and tetraphosphorylated derivatives of 2b; peak 3: proton sponge. Preparative HPLC was performed using a Polaris C18 column. See Materials and Methods for details. (C) Representative ion exchange trace. Peak 1, 2, 3 and 4 correspond to the mono-, di-, tri and tetra-phosphorylated nucleosides, respectively. Performed using Resource S anion exchange column packed with methyl sulfonate (strong cation exchanger). See Materials and Methods for details.109

Figure 37: Structure of adamantane-modified nucleotides 110

Figure 39: Proposed mechanism for triphosphorylation via a morpholidate chemistry. B: 5-iodouridine..... 112

Figure 44: Denaturing PAGE analysis of extension reactions with adamantane-modified nucleotides (4a-c) using Deep Vent exo^- . Lane 1: dNTP mix F; positive control in the presence of natural nucleotides. Lane 2: dNTP mix 0; negative control primer extension without dTTP or modified nucleotide. Lane 3: dNTP mix 1; extension of the primer by one modified nucleotide. Lane 4: dNTP mix 3; extension of the primer by 3 nucleotides. Lane 5: dNTP mix 6; extension of the primer by six nucleotides and concomitant incorporation of two modified nucleotides. Lane 6: dNTP mix 12; full extension of the primer and concomitant incorporation of three modified nucleotides. See Figure 40 for the sequence of the template and primer.	120
Table 6: Sequence of PentaD STR and synthesised oligonucleotides.....	122
Figure 46: Synthesis of 5'-DMT-5-adamantane modified nucleoside phosphoramidites. (i) DMT-Cl, Et ₃ N, cat. DMAP, pyridine; (ii) CuI, Et ₃ N, Pd(PPh ₃) ₄ , adamantane-acetylene linker, DMF, 16 h; (iii) CEP-Cl, DIPEA, DCM, 2 h.....	125
Figure 47: ³¹ P NMR analysis of residual 15b. NMR spectra recorded at 121.4 MHz in CDCl ₄	127
Figure 48: Translocation of tagged oligonucleotides and origin of step-like modifications in electrical traces. Adapted from [9].	128
Figure 49: Representative traces for oligonucleotides modified with short adamantane linkers, 4C. The electrical signals from oligonucleotides with one (O4) and two (O5) O4 modified nucleotides did not display any significant difference. Oligonucleotides with three modified nucleotides (O6) showed similar electrical signals as O4 and O6. However, several signals also possessed one or two 'shoulders'.	129
Table 7: Translocation Parameters of Adamantane-Modified Oligonucleosides.	130
Table 8: Chemical characterisation of adamantane-modified oligonucleotides	135
Figure 50: (A) Mechanism of DNA methylation by DNA methyltransferase (DNMT). (B) Mechanism of DNMT inhibition by azanucleosides. The additional nitrogen at C5 prevents β -elimination and reformation of the C5-C6 double bond. This results in a permanent covalent bond between the nucleoside and the enzyme. The DNA-DNMT adduct triggers apoptosis. Abbreviations: Ado, adenosyl; Met, methionine; SAH, S-adenosyl-L-homocysteine. Adapted from [191]......	151
Figure 51: Commonly used <i>o</i> -nitrobenzyl photolabile protecting groups.	155

Figure 52: Schematic representation of photodynamic therapy (PDT). The photosensitizer (PS, green) is a photoactivatable chemical which can act as both an imaging and chemotherapeutic agent. Upon administration (1), the PS enters the systemic circulation and is widely distributed (2). The PS preferentially accumulates within the tumour tissue (3). Alternatively, the PS can be formulated for topical administration to enable direct application to diseased tissue. Irradiation of the tumour or a specific area results in fluorescence emission which enables the tumour to be clearly located. Additionally, a singlet oxygen (PS^*) is generated which results in cytotoxicity. Adapted from [218].....157

Figure 53: Simplified Jablonski diagram showing the mechanism of action of photodynamic therapy (PDT). When the PS is illuminated by light of a particular wavelength, it is excited from the ground state to the singlet state (A). Energy can be released in the form of fluorescence emission (B) returning it to the ground state. Alternatively, it can enter the excited triplet state (C). From here, energy can be released in the form of phosphorescence (D) or transferred to other molecules (E). The energy can be transferred to oxygen resulting in singlet oxygen which is highly cytotoxic (F). Adapted from [219]..158

Figure 54: Molecular structure of some photosensitizers used in clinical photodynamic therapy. [225,226].....159

Figure 56: Intracellular metabolism of azanucleosides. The nucleosides enter the cell through protein pores (shown in orange). After cellular uptake, azacytidine is monophosphorylated by uridine-cytidine kinase, whilst 5-aza-2'-deoxycytidine is phosphorylated by deoxycytidine kinase. The monophosphates are subsequently phosphorylated to the active triphosphate, which is then incorporated into either DNA or RNA. Abbreviations: NMP, nucleoside monophosphate kinase; NDP, nucleoside diphosphate.163

Figure 57: The specificity of uridine-cytosine kinase for pyrimidine ribonucleosides. (A) Sugar recognition by UCK. The carbon atoms are coloured in grey, nitrogen atoms are coloured in blue and the oxygen atoms are coloured in red. (B) Hydrogen-bonding network of the substrate binding site of UCL and cytosine. Hydrogen-bonding interactions are shown as dashed lines. Adapted from [242].....164

Figure 58: Schematic representation of the objectives of this project. Objective 1: attach a photolabile protecting group at N4 of azacytidine, 2. It is hypothesised that the modified nucleoside will be able to enter the cell through a protein pore or passively diffuse across the membrane, 3. Objective 2: determine whether the nucleoside is triphosphorylated by intracellular kinase enzymes, 4 and subsequently incorporated into RNA or DNA, 5.

Irradiation of the cell with light of a particular wavelength should remove the protecting group, releasing azacytidine which can form a covalent bond to DNMT, 6.	165
Figure 59: Caged nucleosides synthesised in this project.	166
Figure 60: (A) Initial strategy towards the synthesis of caged nucleoside, 2, NOVC-aza. Regents and Conditions; Method A: TBSOTf, pyridine, 0°C 20 min then 6 h at rt; Method B: TBDMS-Cl, imidazole, DMF, 60-65 °C 8h. (B) Synthesis of photolabels. Regents and Conditions: Phosgene, THF, 0°C for 1 h, then rt overnight.	167
Figure 61: Alternative NOVC-based acylating agents.	169
Figure 62: One-point transient protection strategy. Photocaged derivatives of azacytidine (X = N) were synthesised via Method A. Regents and conditions: a) HMDS, DMF, 2 h, rt; b) chloroformate 7 or 9, pyridine, rt overnight; c) TAS-F, DMF, 2 h, rt. Photocaged derivative of cytidine (X = C) was synthesised via Method B. Regents and conditions: d) TMS-Cl, pyridine, 1 h, rt; e) NOVC-Cl, pyridine, overnight, rt; f) MeOH/H ₂ O (1:1, v/v), 30 min, rt.	170
Figure 63: Ring opening and hydrolysis of azacytidine. I is in dynamic equilibrium with its ring-opened formylated derivative, III. Irreversible deformylation results in the formation of a guanylurea derivative, IV.	170
Figure 65: Mechanism of uncaging nucleoside analogue 3.	172
Figure 64: (A) Absorbance spectra of di-NPE-Aza, 1. The absorbance spectra of azacytidine, 4, and NPE-OH, 8, are superimposed. (B) Absorbance spectra of NOVC-cytidine, 3. The absorbance spectra of cytidine, 13, and NOVC-OH, 6 are superimposed. All components at 100 µM, ACN:H ₂ O, 1:1, v/v, with 1% acetic acid.	172
Figure 66: Mechanism of uncaging nucleoside analogue 1.	173
Table 9: Chemical Properties of Caged Nucleosides.	174
Figure 67: (A) HPLC traces for the photolysis of analogue 1 (measured at 260 nm). (B) Time-course for the photolysis of analogues 1 (filled circles) and 2 (empty circles). Samples (100 µM, ACN/H ₂ O, 1:1, v/v, 1% acetic acid) were irradiated at 350 nm. Aliquots were taken at regular intervals then analysed by HPLC.	174
Figure 68: Growth curves for SaOS-2 cells under different conditions over 96 h period. Analogue 1 was dissolved into HPLC-grade acetonitrile. (A) Growth media alone. (B) Growth media with acetonitrile. (C) 0.5 µM 1. (D) 1.5 µM 1. (E) 0.5 µM 1 irradiated. (F) 1.5 µM 1 irradiated. For (E) and (F) a 1 mM stock solution of analogue 1 was irradiated for 24 h at 365 nm. The sample was diluted to the appropriate concentration then applied to the	

cells. All treatment conditions were performed in triplicate. The mean confluence has been plotted per treatment condition, with the error bars representing the standard deviation.176

ABBREVIATIONS

ABC	Abacavir, [4R-(2-Amino-6-cyclopropylamino-purin-9-yl)-cyclopent-2-en-1S-yl]-methanol
ACN	Acetonitrile
α HL	Alpha hemolysin
Amp.	Amplitude
Ang 0-His ₆	Angiotensinogen-His ₆
Ang I-His ₆	Angiotensin I-His ₆
ATP	Adenosine triphosphate
AZT	Zidovudine, 3'-azido-3'-deoxythymidine
Aza	Azacytidine
BSA	Bovine serum albumin
CEP-Cl	2-cyanoethylchloro- <i>N,N</i> -diisopropylphosphoramidite
CNB	α -carboxy-3-nitrobenzyl
CSPPS	Convergent solid-phase peptide synthesis
CuI	Copper iodide
Cyt	Cytidine
D4t	Stavudine, 2',3'-didehydro-3'-deoxythymidine
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCC	<i>N,N</i> -Dicyclohexylcarbodiimide
DCM	Dichloromethane
ddC	2',3'-Dideoxycytidine
ddH ₂ O	Double-distilled water
DIC	<i>N,N</i> -Diisopropylcarbodiimide
DMAP	4-(Dimethylamino)pyridine
DMF	<i>N,N</i> -Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT-Cl	4,4'-Dimethoxytrityl chloride
DNA	Deoxyribonucleic acid
DNA pol	DNA polymerase
DNMT	DNA methyltransferase
dNTP	Deoxynucleotide triphosphate
DIPEA	<i>N,N</i> -Diisopropylethylamine
dsDNA	Double stranded DNA
EDT	1,2-Ethanedithiol
EEDQ	2-Ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline
Et ₃ N	Triethylamine
Foo	Frequency of occurrence

h	hours
HBTU	O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate
His ₆	Hexahistidine
HIV	Human immunodeficiency virus
HMBA	4-Hydroxymethyl benzoic acid
HOBt	1-Hydroxybenzotriazole
HPLC	High-performance liquid chromatography (reverse phase)
IPA	Isopropanol
KCl	Potassium chloride
kDa	Kilodaltons
KCN	Potassium cyanide
LCMS	Liquid chromatography–mass spectrometry
MALDI	Matrix-assisted laser desorption/ionization
MeOH	Methanol
MgSO ₄	Magnesium sulphate
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectrometry
MSPPS	Manual solid-phase peptide synthesis
mV	Millivolts
NaOH	Sodium hydroxide
NB	<i>o</i> -Nitrobenzyl
NBOC	Nitrobenzyloxycarbonyl
NDP	Nucleoside diphosphate
NH ₄ OH	Ammonium hydroxide
Ni ²⁺ -NTA	Nickel-nitriloacetic
nm	Nanometre
NMP	Nucleoside monophosphate
NMR	Nuclear magnetic resonance
NOSEY	Nuclear Overhauser effect spectroscopy
NOVC	6-Nitroveratryloxycarbonyl
NPE	[2-(2-nitrophenyl)ethoxy]carbonyl
pA	Picoampres
PCR	Polymerase chain reaction
Pd(PPh ₃) ₄	Tetrakis(triphenylphosphine)palladium(0)
Phi29 DNAP	Phi29 DNA polymerase
PDT	Photodynamic Therapy
POCl ₃	Phosphorous oxychloride
poly[A]	Polyadenine
poly[C]	Polycytosine

poly[U]	Polyuridine
PRA	Plasma renin activity
PRC	Plasma renin concentration
PS	Photosensitizer
RAAS	Renin-angiotensin-aldosterone system
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
SAM	S-Adenosyl methionine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNPs	Single nucleotide polymorphisms
SPDP	Succinimidyl 3-(2-pyridyldithio)propionate
SPPS	Solid-phase peptide synthesis
ssDNA	Single stranded DNA
STR	Short tandem repeats
t	time
TAS-F	Tris(dimethylamino)sulfonium difluorotrimethylsilicate
TBE	Tris/Borate/EDTA
t-BuOH	<i>tert</i> -Butanol
TEAB	Triethylammonium bicarbonate
TEMED	Tetramethylethylenediamine
TES	Triethylsilane
TFA	Trifluoroacetic acid
TFE	2,2,2-Trifluoroethanol
TLC	Thin layer chromatography
t _R	Retention time
τ _{off}	Event duration
τ _{on}	Inter-event interval
UCK	Uridine cytidine kinase
VNTR	Variant number tandem repeat
μM	Micromolar

PUBLICATIONS AND PRESENTATIONS

This work has been published and discussed in the following forms:

Publications:

1. Synthesis and Enzymatic Incorporation of Modified Deoxyuridine Triphosphates. Vinciane Borsenberger, Mikiembo Kukwikila, Stefan Howorka., *Org. Biomol. Chem.*, 2009, **7**, 3826–3835
2. Electically Sensing Protease Activity with Nanopores. Mikiembo Kukwikila, Stefan Howorka., *J. Phys.: Condens. Matter*, 2010, **22**, 454103.

Presentations:

1. Electically Sensing Protease Activity with Nanopores. Poster presented at the American Chemical Society National Meeting, Denver, 2011.
2. Electically Sensing Protease Activity with Nanopores. Oral presentation at the Institute of Structural and Molecular Biology Retreat, Cambridge, 2011.
3. Chemical Tools for Nanopore Analysis of Enzymatic Activity in Blood Plasma. Poster presented at the Nanopore Workshop, Imperial College London, 2011.

1 INTRODUCTION

1.1. Basics of Nanopore Analysis

In recent times there has been growing interest in technologies which enable the detection and analysis of individual biomolecules without the need for prior labelling. At present much of the time and cost involved in analysing biomolecules is expended on their purification, amplification and labelling so that they provide sufficient signal-to-noise ratios for optics-based techniques, such as fluorescence spectroscopy and high performance liquid chromatography (HPLC), or for use in biological assays.¹ As a result of labelling, these techniques often require costly reagents and sophisticated instrumentation. It is believed that by analysing individual biomolecules directly, sample preparation would be simplified and the cost of analysis reduced. Additionally, direct analysis would be more sensitive as smaller sample sizes would be required to yield results, and sub-populations which ordinarily may be undetected by conventional ensemble techniques, may be detected.² Nanopore technology is one single-molecule technique which is currently being pursued as a tool to analyse biomolecules of interest.

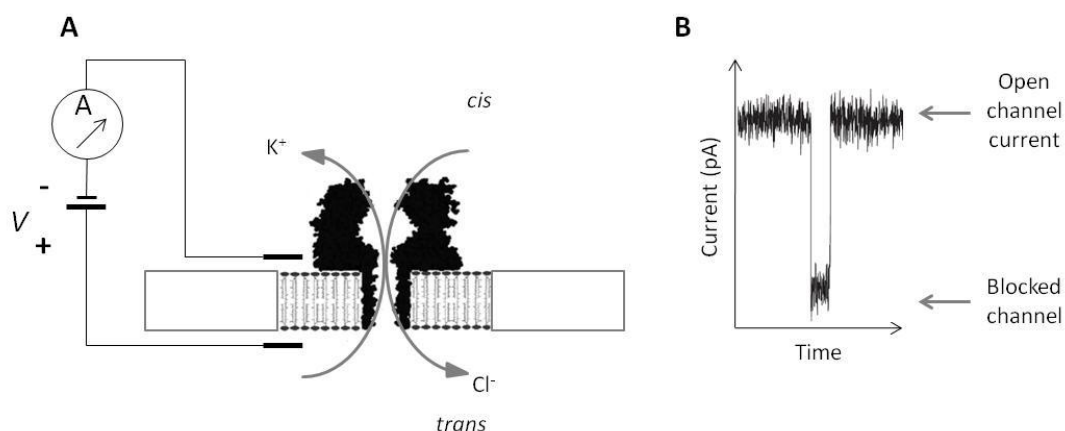


Figure 1: (A) Principle of nanopore analysis. A pore is placed within a membrane that separates two electrolyte-filled compartments. Applying a voltage across the membrane results in the flow of ions, this in turn generates a detectable ionic current. (B) Typical electrical trace. When the channel is open and unobstructed, the current level is termed 'open channel current'. As an analyte passes the inner constriction of the pore, the flow of ions is temporarily interrupted resulting in a reduction in the ionic current. The channel is said to be blocked.

At the heart of nanopore technology is a pore with nanometer-scale dimensions which is placed within an insulating membrane that separates two electrolyte-filled chambers (Figure 1A). When a voltage is applied across the membrane, a flow of ions is induced which in turn generates a current. The current is readily detected using standard electrophysiological techniques. When the pore is open and unobstructed, the ions are able to flow freely across the membrane resulting in a constant current, termed the open-channel current. However, when a charged analyte is added to one chamber, the analyte is electrophoretically driven through the pore. As it does so, the flow of ions is temporarily blocked resulting in a reduction in the ionic current (Figure 1B). Once the translocating analyte clears the pore, the flow of ions is restored and the current returns to the open-channel level. By monitoring the magnitude (i.e. amplitude), duration and frequency with which the blockades occur, it is possible to characterise the translocating molecule with respect to size and concentration.

1.1.1. Nanopores

1.1.2. Biological Nanopores

The α -hemolysin (α HL) nanopore is the most widely used pore in research.^{3,4} It is a protein toxin secreted by the *Staphylococcus aureus* pathogen. The toxin attacks target cells by spontaneously inserting itself into the plasma membrane thus disrupting the cell's ability to regulate its osmotic and electrochemical gradient.^{5,6} The protein is composed of seven identical subunits which self-assemble in the lipid-bilayer to give a heptameric mushroom-shaped pore. The structure and the approximate dimensions of the pore are shown in Figure 2. The pore consists of a large cap domain which sits on the extracellular side of the membrane (in nanopore experiments, this side is denoted *cis*) and has dimensions of 10 x 10 nm.⁶ The opening of the pore measures 2.6 nm in diameter and widens to 4.6 nm at the inner vestibule.¹ The inner constriction measures 1.4 nm and widens to 2 nm at the intracellular side of the membrane (in nanopore experiments, this side is denoted *trans*). The stem domain is composed of a water filled, slightly anion selective β -barrel which is approximately 5.2 nm in length and completely spans the bilayer.

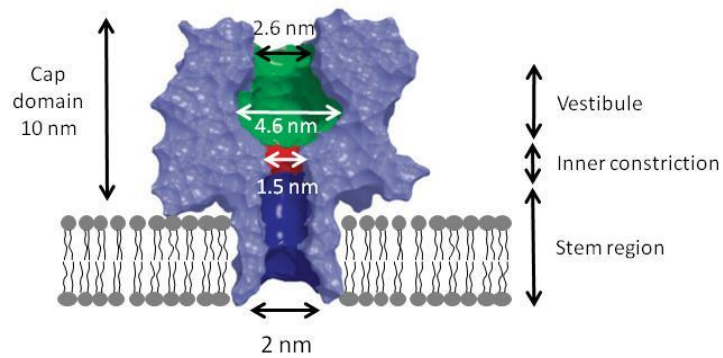


Figure 2: Cross-section of α -hemolysin protein pore in plasma membrane. Adapted from [2].

Several features help explain the popularity of α HL in nanopore experiments. Firstly, the electrical properties of the pore have been well characterised. The pore displays Ohmic characteristics with conductivity proportional to electrolyte concentration, and current directly proportional to voltage.^{1,6} For instance, in a 1 M KCl solution, a typical α HL pore carries a current of approximately 120 pA when 120 mV is applied from the *cis* to *trans* side. Furthermore, as the pore lacks any moving parts which could potentially obstruct the flow of ions, it is open indefinitely.^{2,7} These factors facilitate the detection of analyte-induced blockades against an unchanging current level.² Secondly, the dimensions of the inner constriction are comparable to the dimensions of many biomolecules of interest, such as single stranded DNA (ssDNA) and peptides.^{8,9} Consequently, these biomolecules are able to effectively occlude the ionic current as they traverse the pore, giving rise to distinct reductions in conductivity which facilitates their detection. This feature has fuelled the research and development of α HL-based technologies for next-generation bioanalytical and diagnostic tools.^{3,10} Finally, as the protein is able to withstand mutagenic changes and harsh purification procedures, it can be readily engineered or ‘tuned’ to detect specific analytes.^{2,11,12} As the crystal structure of the protein has been solved, such modifications can be performed with atomic-scale precision.⁶ This is typically achieved by substitution or replacement of amino acids at the inner constriction or β -barrel. For instance, introducing a ring of aromatic residues at the inner constriction of the pore, will impart specificity for hydrophobic analytes.^{13,14} Alternatively, the pore can be tuned to detect smaller analytes which would not ordinarily perturb the ionic current and hence be detected. This can be achieved by lodging circular adapters into the vestibule, thereby

reducing the inner diameter of the pore.¹⁵ As a result of such modifications, the range of molecules that can be detected by nanopores is not limited to biopolymers, but also encompasses smaller organic molecules.^{16,13}

Although α HL pores are the mainstay of nanopore technology, they are associated with some inherent drawbacks which have limited their use.³ The most significant is the instability of the experimental setup. The protein itself has been shown to be remarkably stable, tolerating temperatures nearing 100 °C.¹⁷ The instability arises from the lipid bilayer which supports the nanopore. In a typical nanopore experiment, the lipid bilayer is formed across a 30-100 μ M orifice in a Teflon cassette. Ideally, the formed membrane should be stable enough to allow several hours of nanopore analysis to be performed. However, in reality, as a result of the non-covalent and fluid nature of the lipids, the bilayer often ruptures within 2 hours.² The stability of the bilayer can be improved by employing chemical reagents to change its composition¹⁸ or by tethering it to a support.¹⁹ Nonetheless, destabilisation by diffusion of the lipid molecules out of the membrane and into the electrolyte solution can still occur.² An additional limitation is the lack of control over the number of pores which insert into the bilayer. Pore insertion is typically achieved by adding a protein solution to one of the electrolyte-filled chambers. Once the pore has spontaneously formed and been inserted into the bilayer, excess non-inserted protein is flushed out by changing the electrolyte solution. The process of insertion is completely random and can often result in the insertion of several pores. In this instance, the membrane must be broken, a new bilayer formed and insertion repeated. As α HL can only function when supported by a lipid bilayer,² its use and commercialisation has been restricted. Consequently, alternative nanopores and sensing modalities have been sought.

1.1.3. Solid-State Nanopores

Solid-state nanopores are synthetic pores typically formed from inorganic materials or organic polymers.¹⁰ The most common solid-state systems are silicon-based.⁴ The pores are typically fabricated by using an ion beam to either drill or etch an opening into the material.^{2,4} Solid-state nanopores offer a more robust and durable alternative to biological

pores, enabling nanopore analysis to be performed over a wider range of experimental parameters such as temperature, pH and ionic strength.² As a result of their superior stability, they are more readily integrated into sensing devices and are thus more amenable to commercialisation.¹⁰

Fabrication of nanopores from solid-state material presents another advantage over their biological counterparts; the ability to control the pore diameter and length. Whilst biological nanopores have fixed geometries, the diameter of solid-state nanopores varies depending on the fabrication techniques employed.⁴ Nanopores have been fabricated with diameters ranging from 50 nm (or larger) to as small as 1 nm; smaller than the inner constriction of α HL.^{4,10} This offers a mechanism through which the nanopore can be tuned according to the size analyte under investigation. Akin to biological pores, the specificity and sensitivity of solid-state pores can also be altered through chemical modification. For instance, it has been observed that silicon-based pores coated in aluminium oxide display enhanced sensitivity and electrical properties compared to untreated pores.^{20,21} The translocation speed of negatively charged biopolymers analysed using the treated pores, was reduced compared translocations through untreated pores. This was attributed to favourable electrostatic interactions between the biopolymers and the positively charged surface. The net effect was an improvement in the sensitivity of the pore and a reduction in the background noise. Similar affects can be achieved through chemical modification using organic reagents.²²

Although diameters comparable to biological pores have been achieved, attaining sensitivities comparable to biological pores remains a challenge. The sensitivity of nanopores is determined by two factors; the thickness of the pore (i.e. volume of the pore) and the level of background noise.⁴ Pore thickness determines how many analyte molecules can occupy the pore and hence whether single-molecule detection can be achieved. To date the thicknesses achieved using solid-state materials have been greater than the length of biological pores (> 10 nm). As a result, during nanopore experiments, perturbation of the electric current is generally attributed to the presence of more than one analyte molecule rather than a single molecule. Additionally, solid-state nanopores

exhibit much higher background noise than biological pores. A further limitation is the structural variability between fabricated pores. No two fabricated pores are exactly the same even if they undergo the same process with the same specifications.⁴ Consequently it is not possible to assure reproducibility between experiments performed using nanopores fabricated in different laboratories under identical conditions. This is in stark contrast to α HL which is produced by cells thus possesses a well-defined geometry and heterogeneity.

1.2. Sensing of Nucleic Acids

By far the most widely reported application of nanopore technology is as a potentially ultra-fast and cheap alternative to current DNA sequencing techniques. Since the successful completion of the Human Genome Project in 2003, the main goal of which was to elucidate the nucleotide sequence of a complete genome, a great deal of information has been gained about the genetic causes of disease and the effect of mutations.^{23,24} Completion of the project also spurred the development of a new idea, 'personalised genomic medicine'. The core of this idea is that a patients' health can be managed by tailoring treatment to particular genomic attributes in addition to the traditional biological aspects.²⁵ For instance, genomic data may enable earlier diagnosis, more effective prevention and treatment of disease and avoidance of drug side effects.^{4,25,26} However, such a strategy is highly dependent upon the ability to sequence an individuals' genome cheaply and quickly.

1.2.1. Current DNA sequencing Techniques

Current commercially available sequencing technologies are all loosely based upon the Sanger method pioneered in the 1970s.^{27,28} For each of these technologies, a significant amount of sample preparation is required prior to sequencing. The DNA to be sequenced must first be isolated then split into smaller more manageable fragments. As the technologies do not possess the sensitivities required to directly detect and analyze the fragments, sample amplification is required. Depending on the nature of the sample, this

is achieved using bacterial host organisms or amplification enzymes.²⁸ Sequencing is accomplished with the aid of short oligonucleotide fragments, termed primers, which are complementary to the DNA flanking the region of interest. A cyclic, multi-step reaction is performed in which the primer is extended. However, the extension of the primer is terminated prematurely due to the incorporation of one of four fluorescent nucleoside analogues. As termination occurs at every possible nucleotide position in the DNA, a mixture of extension products are obtained. The sequence is elucidated by high-resolution gel electrophoresis in combination with capillary electrophoresis. As the DNA fragments exit the capillary, the fluorescence emission of the fragments is measured and bioinformatics software used to re-assemble the sequence of the original DNA sample. While this approach has been the mainstay of DNA sequencing for many years, and was successfully used to complete the Human Genome Project, it is associated with some drawbacks. Namely, the amount of DNA that can be processed per unit time, that is, the throughput, is low. Indeed, the Human Genome Project took ten years to complete.^{23,29}

Second generation technologies emerged in response to the low throughput associated with the Sanger-based technologies (also known as first-generation methods).³⁰ Second generation technologies are conceptually similar to the first-generation methods in that sample amplification is required and fluorescent nucleotides are employed. However, sequencing is achieved in an array-based manner. Multiple copies of the DNA fragments of interest are immobilised onto a solid-support then sequentially flooded with fluorescent nucleotides. The support is then washed to remove excess reagents and subsequently scanned to determine which nucleotide has been incorporated. This can involve exciting the array with a laser and monitoring the fluorescence emission of the incorporated nucleoside. The cycle is repeated until the reaction is no longer viable.³⁰ This approach has been successfully developed into commercial sequencing instrument by Solexa (now part of Illumina, San Diego). The data from each immobilised fragment is used to reassemble the sequence of the original DNA sample. As the tens of thousands of strands can be immobilised on a single support, the throughput is significantly increased compared to first-generation methods. Furthermore,

this approach also allows for a greater degree of parallelism.²⁸ However, due to large number of washing and scanning cycles, second generation technologies can take a greater amount of time to garner sequencing results (typically several days). Second generation technologies are also limited by the read length, that is, the number of DNA bases that can be identified sequentially.²⁸ As the DNA fragments utilised in second generation technologies are immobilised onto a solid-support, they are required to be significantly shorter in order to avoid any unfavourable steric interactions. However, the increased number of fragments is a potential source of sequencing errors as data from more fragments must be combined in order to determine the sequence of the original DNA sample. First generation technologies are able to achieve read lengths approaching 1000 bases, whilst second generation can only achieve half that and have been shown to be ten-fold less accurate.^{28,30}

A drawback common to both first and second generation technologies is the cost. In 2011, it was possible to sequence an entire human genome in eight days for approximately \$10,000.²⁶ This estimation merely covers the cost of analysis and not the instrumentation which can vary from hundreds of thousands to millions of dollars.²⁸ Although these estimations are still significantly lower than the 3 billion dollars required to initially complete the Human Genome Project,²³ in order for personalised genomic medicine to become a reality, complete sequencing would need to be performed within 24 hours for \$1000 or less. This was a challenge set by the US National Institutes of Health in 2004.³¹ Current technologies are also limited by the need to amplify the sample prior to sequencing. Not only does this increase the complexity, time and cost of sample preparation, but can also be an additional source of sequencing errors. Specifically as, these amplification steps are typically performed by enzymes, the fidelity of which will dictate the accuracy with which the DNA is copied.²⁸ It is believed that these limitations could be mitigated or completely eliminated if a nanopore-based approach to sequencing was employed.

1.2.2. Nanopore Sequencing: Third Generation Sequencing Technology

The use of nanopores as a means of sequencing DNA was initially proposed by Dekker *et al.* in the 1990s.¹⁰ It was believed that as DNA is a polyanionic molecule, in the presence of an applied potential, it would be electrophoretically driven through the pore in a sequential linear fashion. Each of the four DNA bases (A, T, C and G) was expected to generate distinct electrical signals reflecting differences in their chemical structure. Hence by monitoring changes in the ionic current that occurred as the DNA strand translocated the pore, the identity of the bases could be elucidated and the sequence of the DNA determined. Due to the sequence complementarity of double-stranded DNA (dsDNA), only one strand from the DNA duplex would need to be sequenced. This approach raised the prospect of faster sequencing as it would require very minimal sample preparation. Furthermore, the elimination of expensive reagents such as fluorescent nucleosides or polymerase enzymes was expected to lower the analysis costs. As the approach had very few infrastructure requirements, that is, was solely dependent upon the nanopore and electrophysiological equipment, the initial start-up costs were also expected to be significantly lower than optics-based approaches. This would enable DNA sequencing to be decentralised from large institutions to smaller laboratories. Nanopore devices could also be more readily miniaturised to produce hand-held or bench-top devices. Finally, by directly sequencing the DNA, fragments as long as 50,000 base pairs could be analysed sequentially which would greatly simplify the genome assembly process.³¹

1.2.3. Nanopore Sequencing: Seminal Investigations

Seminal experiments demonstrating the voltage driven transport of ssDNA and ssRNA were performed by Kasianowicz *et al.* using the α HL pore.³² α HL was chosen as the inner diameter could only accommodate ssDNA, and not dsDNA which with a diameter of 2.2 nm, is too wide.¹⁰ Using polyuridine polymers (poly[U]) of varying lengths, it was demonstrated that the passage of individual DNA molecules could be detected as transient reductions in ionic conductivity. Furthermore, it was observed that the duration and magnitude of blockades were proportional to the lengths of the polymers, indicative

of the polymers translocating the pore in linear extended states. The realisation that nanopore recordings could be used to determine the lengths of individual oligonucleotides suggested that other properties of oligonucleotides could be measured as they translocated the pore. In theory, purine and pyrimidine bases passing through an ion channel should produce different electrical signals which reflected their size and chemical properties. This notion was substantiated in experiments performed using homopolymers of adenine and cytosine, which were of equal length.^{33,34} The nanopore recordings revealed that the patterns of blockades caused by poly[C] were easily distinguished from the patterns caused by poly[A] whether the polymers were analysed separately or in a mixture (Figure 3A).³³ In further experiments which utilised co-polymers of poly [A] and poly [C], bi-level blockades were observed as the nanopore was able to detect the transition from an adenine to a cytosine nucleotide or more specifically from a purine to a pyrimidine nucleotide (Figure 3B).³⁴

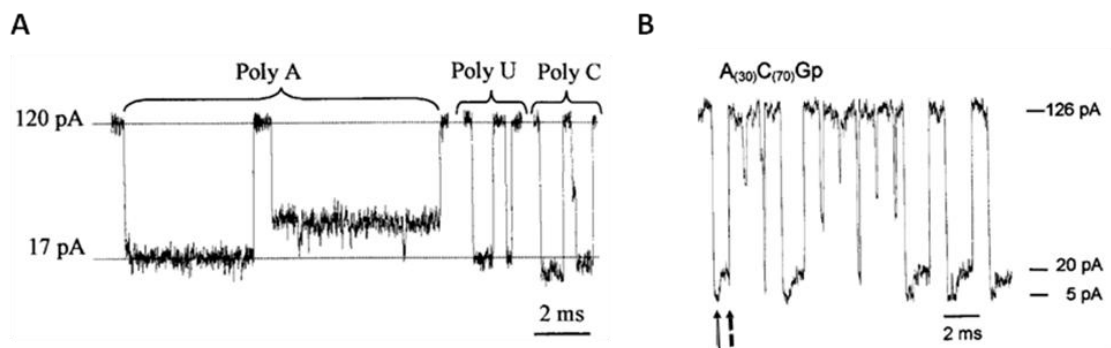


Figure 3: Nanopore traces for poly[A], poly[U] and poly[C]. (A) Each polymer gave rise to distinct patterns of blockades. (B) Nanopore traces for co-polymer of poly[A] and poly[C]. Bi-level blockades were observed as the nanopore was able to detect the transition from adenine to cytosine (indicated by the arrows). Adapted from [33].

This result was particularly appealing given the desire to sequence DNA by directly analysing a translocating DNA strand. However, it was subsequently realised that under standard experimental conditions, single-base resolution could not be achieved.^{33,34} The DNA strands were translocating the pore at average rates of approximately 1-20 nucleosides per microsecond when driven by voltages of 120-150 mV;³¹ much too fast to enable individual bases to be resolved. It was hypothesised that by reducing the translocation speed so that each base occupied the detector for approximately 1

millisecond or longer, it could be possible to resolve the bases.³¹ As a result, to date, nanopore research has been focused on reducing the translocation speed of ssDNA rather than obtaining sequence data.

1.2.4. Approaches to Nanopore Detection of Oligonucleotides

1.2.4.1. Employing Chemically Modified Oligonucleotides

One strategy to facilitate the detection of oligonucleotides is to modify nucleotides with bulky chemical tags which act as ‘molecular brakes’³ during translocation. Whilst developing a nanopore-based method to detect DNA mutations, Mitchell *et al.* observed that chemical tags attached to individual bases increased the cross-sectional diameter of the translocating oligonucleotides and hence reduced the translocation speed as they passed the narrow inner constriction of the α HL pore.⁹ Using synthetic oligonucleotides carrying peptide tags of varying length, charge and hydrophobicity, it was demonstrated that characteristic electrical signatures specific to the composition of the tag could be obtained. The chemical composition of a representative tag is shown in Figure 4A. The presence of a single tag in the oligonucleotide reduced the translocation speed by a factor of 10. Further reductions in the translocation speed were observed when the oligonucleotides were modified with two peptide tags. Interestingly, each of the tags in the di-modified oligonucleotides acted independently to give characteristic electrical signatures. Hence the α HL pore was able to detect multiple separate bases. This observation raised the prospect of modifying each of the four DNA nucleotides with a different tag to facilitate base resolution during nanopore sequencing. While such an approach would enable real-time sequencing, modifying all the bases of a genomic DNA sample would be challenging due to the small distances between the bases. The length of the linker and the size of the tags would need to be significantly reduced to avoid unfavourable steric interactions. However, such tagging approaches have shown greater utility in the nanopore-based detection of single-nucleotide polymorphisms (SNPs) where fewer bases in the oligonucleotide require modification.

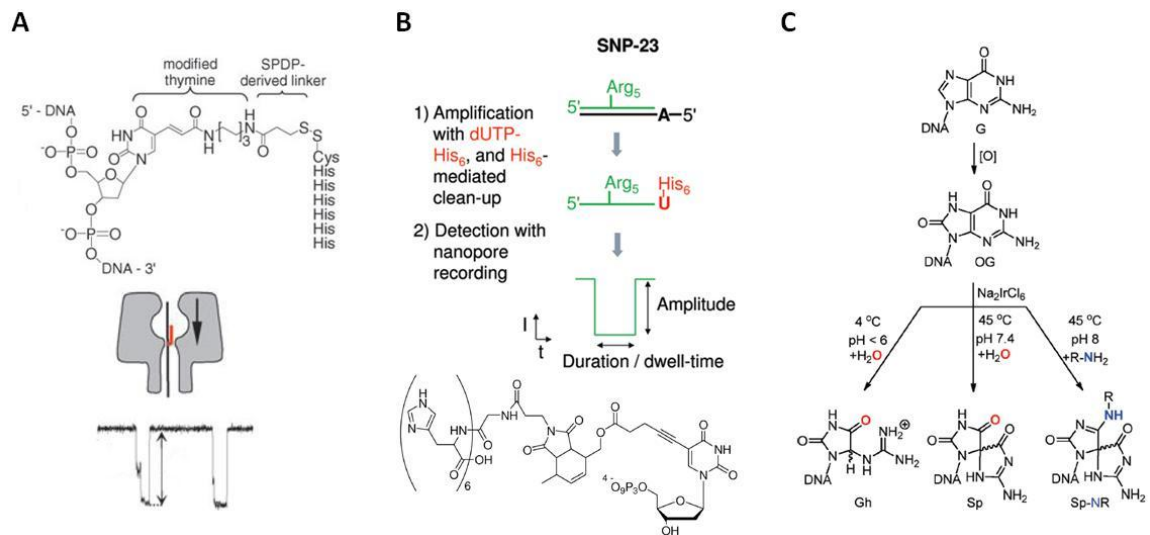


Figure 4: Nanopore analysis of chemically modified oligonucleotides. (A) Peptide-modified nucleotides have been used to reduce the translocation speed of oligonucleotides. The modification increases the cross-sectional diameter of the DNA and gives rise to distinct electrical signals. (B) Peptide-modified nucleotides have been incorporated into oligonucleotide probes to enable nanopore-based detection of SNPs. (C) 8-oxo-7,8-dihydroguanine, an oxidised derivative of guanine was detected using α HL pore. The oxidised base was derivatised with a range of amines in order to increase the signal-to-noise ratio.

SNPs are single mutations within DNA that result in the substitution of one nucleotide for another.³⁵ Such mutations are the single greatest cause of genetic variability between individuals and as such generally have little or no effect on health. However, mutations in key regions of the genome have been shown to trigger the onset of certain diseases including cancer, confer resistance to drugs and render individuals susceptible to external factors such as toxins.³⁶ Accordingly, SNPs can act as biological markers, enabling researchers to locate genes associated with diseases and conditions.³⁵ Borsenberger *et al.*, utilised peptide-tagged oligonucleotides to propose a nanopore-based method to detect SNPs.³⁷ The approach is outlined in Figure 4B. Short peptide-tagged oligonucleotides were used as probes to bind sequence-specifically to longer SNP containing oligonucleotides. The probes bound to the regions flanking the suspected SNP. The presence of the SNP was determined by enzymatically incorporating a hexahistidine (His₆)-tagged nucleotide analogue. Specifically, if the SNP was present, then the modified nucleotide would be incorporated. However, if the SNP was absent a natural nucleotide would be incorporated. Incorporation of the His₆-tagged nucleoside enabled the SNP derived probes

to be isolated and subsequently characterised by nanopore analysis in which characteristic electrical signals were observed.

In a similar fashion, modified nucleotides have also been used to detect 8-oxo-7,8-dihydroguanine, an oxidised derivative of guanine that forms as a result of oxidative stress on cells. Oxidative stress, which is the underlying cause of many age-related disorders and diseased states, arises when reactive oxygen species oxidises cellular components.³⁸ Schibel *et al.*³⁹ utilised a previously developed immobilisation technique⁴⁰ to achieve single-base resolution of a number of guanine analogues, and in doing so avoided the issues related to the rapid translocation speed described above. In brief, a biotin molecule was directly attached to the 3'-end of the oligonucleotides and the resultant biotin-DNA (Btn-DNA) conjugates bound to streptavidin (Strep) protein. On application of voltage, the Strep-Btn-DNA complex was driven into the α HL pore. As Strep was too large to traverse the pore, the tethered DNA was held within the pore for 1-2 s to collect current blockage data, then released by reversing the applied potential. The capture and release cycle was repeated several times. The longer residence time enabled a single oxidised base to be distinguished when placed within a homopolymer sequence or a more biologically relevant heteropolymer. Akin to the studies described above, the oxidised guanine was derivatised with amines to form bulky spirocyclic adducts in order to facilitate discrimination within the nanopore (Figure 4C). In doing so, the amplitude of blockade increased by almost 10% in homopolymers and 2% in heteropolymers compared to the underderivatised, unoxidised control. These preliminary findings highlight the discriminatory abilities of the α HL pore as such small modifications were readily detected. With optimisation of the derivatising agents, it may be possible to detect the oxidized nucleoside in freely translocating DNA. This would provide a rapid means of quantifying the number of oxidised residues within a single DNA strand. Furthermore, information regarding the flanking sequences could also be obtained. Neither of these characteristics can be determined using enzymatic assays or HPLC which are typically used to analyse oxidised guanines. The authors also demonstrated the broader applicability of nucleic sensing for purposes other than sequencing.

1.2.4.2. Using an Engineered α HL Pore

By engineering the α HL pore, its sensitivity can be greatly increased and varied to enable other sequencing approaches to be explored. One such approach is exonuclease sequencing. Sequencing is achieved through the detection of individual nucleotide monophosphates, cleaved from the end of ssDNA by the action of an exonuclease enzyme. The monomers are released and detected in sequential order reflecting the sequence of the original DNA strand (Figure 5A).

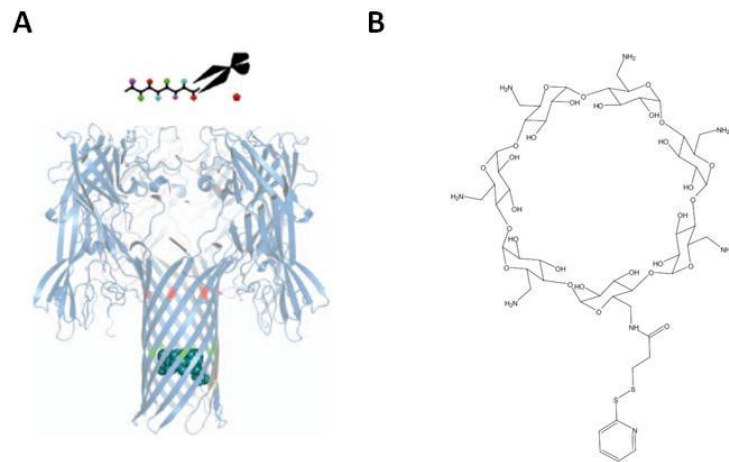


Figure 5: DNA sequencing via the exonuclease approach. (A) Nucleotides are cleaved from ssDNA by an exonuclease enzyme (scissors) then detected by a pore modified with a cyclodextrin ring (turquoise). (B) Structure of cyclodextrin ring. For covalent attachment to the protein, a single amino group was coupled to the reactive bi-functional linker, succinimidyl 3-(2-pyridyldithio)propionate (SPDP), and the resultant construct coupled to a cysteine residue engineered into the β -barrel [15].

This approach was pioneered by researchers in the Bayley group at the University of Oxford collaboration with Oxford Nanopore Technologies.^{15,41,42} In the wild-type α HL pore, the nucleotide monophosphates translocate the pore with minimal effect on the ionic current due to the incomparable dimensions of the inner constriction.⁴¹ To bring the dimensions of the pore closer to the size of the nucleotide monophosphates, an aminocyclodextrin adapter was covalently attached to the β -barrel (Figure 5B).¹⁵ Covalent attachment was achieved by engineering a single cysteine residue into the barrel and using a bifunctional linker, succinimidyl 3-(2-pyridyldithio)propionate (SPDP), to link the cyclodextrin ring via a disulphide bond.⁴² In doing so, the small nucleotides were able to modulate the ionic current by temporarily binding to the adapter as they translocated the

pore. On the basis of amplitude of blockade and duration the binding events, the natural nucleotide monophosphates and 5-methylcytosine, a nucleotide analogue which is essential in epigenetic studies,⁴³⁻⁴⁵ could be identified with 99.9% accuracy.¹⁵ This approach is one of two being developed for commercial purposes by Oxford Nanopore Technologies.⁴⁶ However, it is obvious that there a number of problems that must be solved before the method can be established as a viable sequencing technique. Namely, a mechanism to ensure that each nucleotide is captured and traverses the pore in sequential order is required. As the exonuclease enzyme is added to the bulk-phase *cis* chamber,⁴ the probability of the nucleotides traversing in a random order is high. It has been proposed that this could be overcome through chemical attachment or genetic fusion of the enzyme to pore.^{15,31} This would enable each released nucleotide to be captured and funnelled directly into the pore. Similarly, a mechanism to ensure that the released nucleotide exits into the *trans* chamber and it not recaptured, is required. It has been demonstrated that the probability of a nucleotide exiting into the *trans* chamber is a function of applied voltage, with some nucleotides appearing to have a greater probability than others.¹⁵ For accurate sequencing, it is necessary that each nucleotide is captured and exits into the *trans* chamber with equal probability. Finally, to ensure that only a single nucleotide is present in the pore any one time, the turnover of the enzyme must be slower than the dwell time of the nucleotide. This will go towards ensuring the accuracy of the sequencing data. It may be necessary to alter exonuclease turnover through direct mutation or changes to physical conditions such as temperature and pH.^{15,41}

1.2.4.3. Utilising Enzymes as Molecular Motors

Recently there has been a shift towards using enzymes to reduce and control the translocation speed of otherwise rapid ssDNA through pores. Such techniques take advantage of the highly processive nature of polymerase or exonuclease enzymes and their ability to move along ssDNA one base at a time. A schematic representation of the molecular motors approach is shown in Figure 6. In brief, an enzyme is added to the *cis* compartment of the nanopore resulting in an enzyme-DNA complex in bulk solution.

When a voltage is applied across the membrane, the free ssDNA is pulled into the lumen of the pore and the large enzyme-DNA complex resides on top of the pore. As the enzyme processes the DNA, the strand is pulled through the pore in a slow and controlled manner. Importantly, rather than a continuous smooth motion, the sequential nature of the enzymes results in a ratcheting effect as each base is paused within the pore as the enzyme completes its catalytic cycle.^{4,47} In order to function as a molecular motor for the purposes of DNA sequencing, enzymes must meet two key requirements.^{4,48} Firstly, the ability to form enzyme-DNA complexes which can withstand the forces generated as the complex is electrophoretically driven into the pore. Secondly, during DNA processing, the enzymes must generate a force greater than the force required to hold the complex on top of the nanopore. These requirements ensure that in the presence of an applied potential, the enzyme is able to move along the DNA strand without concomitant dissociation of the enzyme-DNA complex. Lieberman *et al.* recently advanced the field by demonstrating that the phi29 DNA polymerase (phi29 DNAP) was a suitable enzyme as it was able to meet both criteria (Figure 6).⁴⁸ Phi29 DNAP, an enzyme responsible for viral replication, is able to catalyse the extension of DNA in the 5'→3', in addition to possessing exonuclease activity in the 3'→5' direction.^{49,50} Both the extension and exonuclease activity move the DNA in single steps which can be used to control DNA translocation through the pore. Furthermore, the enzyme is able to catalyse the incorporation of at least 70 kilobases of DNA after a single binding event, without the aid of any accessory proteins.⁴⁸⁻⁵⁰ Manrao *et al.* were able to build on the work of Lieberman *et al.* to associate observed modulations of electrical ionic current to sequence information.⁵¹ To achieve this, Phi29 DNAP was bound to ssDNA template that was hybridised to a primer (Figure 6B, 1). The primer had a terminal hairpin loop on the 5'-end to protect it from the actions of the enzyme. The technical hurdle of activating a single polymerase in the nanopore, whilst inhibiting the enzyme in the bulk solution prior to sequencing, was overcome by employing a 'blocking oligomer' (Figure 6B, 2).^{47,51} This short oligonucleotide fragment bound downstream of the primer, resulting in stable complexes in the bulk solution that readily dissociated once the complex had been captured by the nanopore. The dissociation process termed

unzipping, liberated the 3'-end of the primer to enable DNA synthesis (Figure 6B, 3). When the nucleotide triphosphates and divalent ions necessary for primer extension were added to the *cis* chamber, the enzyme catalysed the extension of the primer. In doing so, it acted as a motor ratcheting the DNA through the *cis* side of the pore a single-nucleotide at a time (Figure 6B, 4). Once DNA synthesis of the template strand was complete, both the DNA and phi29 DNAP exited to the *cis* side.

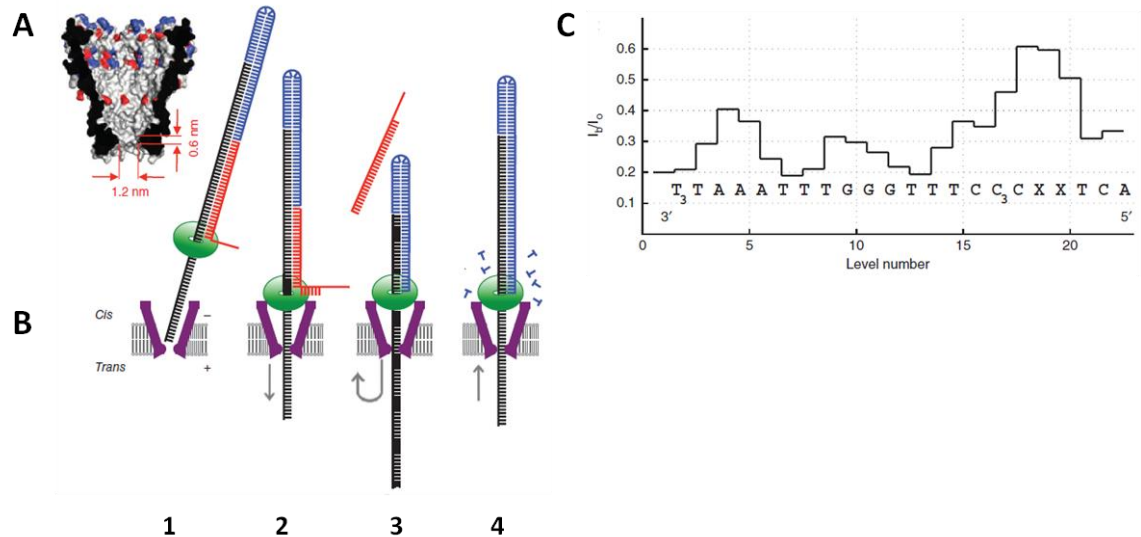


Figure 6: Utilising Phi29 DNA polymerase as a molecular motor to sequence DNA. (A) Structure and dimensions of MspA pore. (B) 1, the DNA template (black) complementary base-pairs to the primer (blue). The 'blocking oligomer' (red) prevents extension of the primer in the bulk solution; 2, Applying a voltage causes the captured DNA-enzyme (green) complex to be drawn into the pore. This results in unzipping of the blocking oligomer; 3, The enzyme can now catalyse the incorporation of nucleotides into the deprotected primer; 4, As DNA synthesis occurs, the DNA complex is ratcheted into the *cis* chamber. When a nucleotide in the template is held within the key sensing region of the pore, a distinct electrical signal is obtained. (C) Representative electrical trace. Adapted from [52].

Rather than using α HL, a mutated version of the protein pore MspA, which has a slightly different geometry, was used (Figure 6A). In essence, the key sensing region in MspA is shorter than that of α HL so that as few as four nucleotides contribute to measured changes in electric current compared to 10-12 nucleotides in α HL.^{53,54} Due to these structural properties, Manrao *et al.* were unable to resolve individual nucleotides as they passed the inner constriction. However, they were able to translocate the DNA in single-nucleotide steps at a rate appropriate for sequencing. With the aid of computer

algorithms it may be possible to decipher the electrical traces to gain single-nucleotide resolution.

1.3. Sensing of Enzymatic Activity

In addition to providing information on the structural properties of substrates, nanopore analysis can also be used to directly probe the mechanics and kinetics of enzymes themselves. To date, much of the information regarding enzyme activity and kinetics is garnered from large ensemble techniques which monitor changes in substrate or product concentration over time.^{55,56} These techniques do not possess the sensitivities required to examine individual enzyme molecules, but rather provide information on the average behaviour of a collection of molecules.⁵⁷ Single-molecule techniques, such as nanopore analysis, provide a means of studying individual enzymatic reactions. As a result, they can give insight into individual reaction steps and intermediates which may be masked in ensemble techniques.^{44,55,57} A nanopore-based method for detecting enzymatic reactions would offer similar benefits as those discussed for DNA nanopore sequencing. Namely, it would obviate the need for sample amplification thus could potentially reduce the cost of analysis and simplify sample preparation. Furthermore, ultra low levels of enzyme activity could be detected, which would be beneficial for clinical diagnosis, where enzyme activity can be indicative of normal or abnormal cellular function.⁵⁶ Finally, it would enable enzymatic reactions to be followed in real-time. A variety of different strategies have been described to assess enzymatic activity using nanopores.

1.3.1. Probing Enzyme Mechanism

The α HL nanopore has emerged as a tool for studying the catalytic cycle of polymerase enzymes at a single-molecule level. Using DNA-enzyme complexes, akin to those used to probe the enzymatic molecular motors described above, polymerase-catalysed primer extensions have been resolved at a single-nucleotide level.^{47,58} Furthermore, primer extension has been measured in real-time and the extent controlled by varying the applied potential.⁵⁹ Nanopores have also been used to study the structural changes that occur as the enzyme selects nucleotides to pair against the DNA template.⁶⁰ During primer

extension, DNA polymerase selects the correct nucleotide substrate from a pool that contains both complementary and non-complementary nucleotides.^{60,61} This selection process involves the formation of a series of enzyme-DNA complexes in which the nucleotides are scrutinised prior to the formation of the phosphodiester bond.⁶¹ It was observed that enzyme-DNA complexes formed with complementary and non-complementary nucleotides could be distinguished based on their properties when captured on top of the α HL pore. Specifically, complexes formed with complementary nucleotides were more resilient to voltage-induced dissociation than those formed with non-complementary nucleotides.

1.3.2. Determining Enzyme Kinetics

Nanopore technology can also be used to estimate the kinetics of enzymes. Depending on the nature of the enzyme and its substrate, two different experimental platforms are possible. The enzyme can be characterised by monitoring its action on a substrate which is attached to the pore or alternatively, by monitoring its action on an unbound substrate free in solution. The former has been used on occasions when the substrate has been too large to enter the pore. For instance, Cheley *et al.* developed an approach to detect kinase enzymes and their inhibitors.^{62,63} A peptide inhibitor was covalently attached to the entrance of the β -barrel at the *trans* side of the membrane (Figure 7A). Binding of the enzyme to the peptide resulted in a partial block of the channel and subsequent disturbance in the ionic current. The peptide was placed in close enough proximity that the binding events that occurred in the bulk *trans* phase were transmitted to the interior of the pore.⁶³ The relationship between applied potential and event duration and frequency, gave an estimation of the dissociation and association rate constants, and hence binding affinities. Given that nanopore technology is amenable to miniaturization and parallelisation, this approach could be used to develop large array-based devices for screening peptide inhibitors for the enzyme.

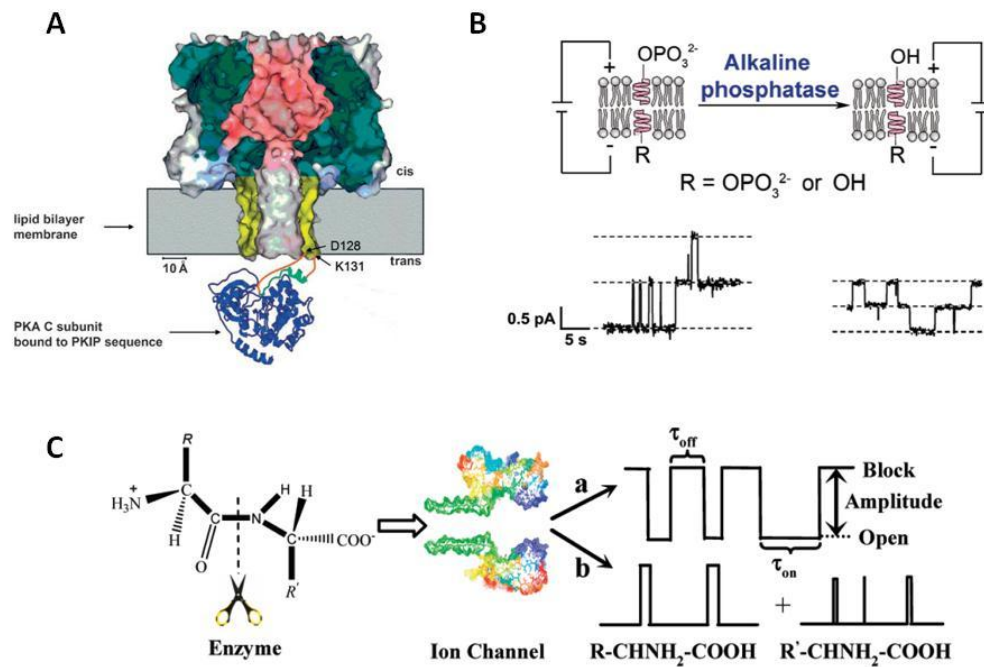


Figure 7: Determining enzyme kinetics using nanopores. (A) Genetically engineered α HL pore carrying a single inhibitory peptide (shown in yellow and turquoise). A subunit of the enzyme (blue) is docked onto the inhibitory peptide.^{62,63} (B) Charged-based detection of enzymatic activity. Alkaline phosphatase catalyses the hydrolysis of phosphate groups attached to the pore.⁵² The resultant local change in conductivity is detected by the pore giving rise to distinct modulations in the ionic current. (C) In-situ enzymatic digestion. In the absence of the enzyme (a), current modulations are due solely to the starting substrate. On addition of the enzyme, new blockade translocation events (b) with distinct amplitudes and duration are observed.¹⁴

Macrae *et al.* utilised a similar approach to develop a strategy to detect and characterise the activity of alkaline phosphatase.⁵² Using a semi-synthetic pore, the ability of the enzyme to catalyse the hydrolysis of phosphate groups placed at the entrance of the pore, were detected. The approach is schematically represented in Figure 7B. Modulations in ionic current were not due to obstruction of the pore, but rather alterations in the ionic charge on the functional groups around the opening of the pore. When the negatively charged phosphate groups were present, a greater conductance was observed. This was attributed to an electrostatically induced local increase in cations near the negatively charged phosphate groups. Upon hydrolysis to neutral hydroxyl groups, a reduction in conductivity was observed. The Michaelis-Menton kinetics and catalytic efficiency were determined by following the conversion of substrate to product over time, in a range of enzyme concentrations. Using this approach, enzymatic activity could be detected with

only picomolar concentrations of substrate and pico- to nanomolar concentrations of enzymes.

Zhao *et al.* took a different approach to the study the kinetics of the trypsin protease.¹⁴ Enzymatic digestion of a peptide substrate was performed in the solution; directly in the *cis* chamber (Figure 7C). The kinetics of the reaction was determined by measuring translocation events in the absence and presence of the enzyme. In the absence of the enzyme, only a single type of translocation events, corresponding to the intact substrate, was observed. However, upon addition of the enzyme to the *cis* chamber, two new types of translocation events were detected. These were attributed to the enzymatically released peptide fragments. By monitoring changes in the frequency of these blockades, the kinetics of the enzyme could be determined in real-time. However, this sensing approach was limited by the requirement for an engineered protein pore. In particular, the pores were modified to enable greater discrimination between the peptides. Such a requirement limits the broad applicability of this approach, as different substrates would necessitate different pore modifications. Furthermore, as the pore blockades originated from multiple peptides (i.e. the starting material and the released fragments) the data processing was somewhat lengthy and complicated.

1.4. Sensing of Peptides

Inspired by nanopore studies involving nucleic acids, researchers are increasingly utilising the knowledge gained to sense polypeptides directly. As this field of research has not been pursued as aggressively as nucleic acid research, comparatively little is known about the mechanics of peptide translocations. However, in general, it has been observed that single-molecule analysis of polypeptides is slightly more difficult to achieve. Whilst ssDNA is linear, possesses a dense net negative charge and can be driven through a nanopore with little effort, the same cannot be said of polypeptides.⁶⁴ Polypeptides can be extremely heterogeneously charged containing positive, negative, polar and hydrophobic side chains, which can complicate nanopore experiments. If a peptide analyte does not possess enough, or the correct net charge, very few translocation events will be

detected.⁸ However, the lower charge densities of peptides can prolong the translocation time by 1-2 orders of magnitude, giving rise to increased resolution compared to nucleic acids.⁸ Furthermore, polypeptides can form a number of secondary structures in solution which can reduce the probability of entering the pore.⁶⁵ Despite these experimental challenges, a nanopore-based approach to analyse peptides could be beneficial given the important role they play in many cellular process. Many methods are currently available to detect and characterise peptides including mass spectrometry, nuclear magnetic resonance (NMR) and HPLC.⁶⁶⁻⁶⁸ In addition to the advantages previously described, it is hoped that nanopore sensors will enable peptides to be detected in physiological conditions without the need for any pretreatment.⁶⁵ Furthermore, studying the translocation of peptides through a nanopore may give some insight into the protein translocations that occurs in cells *in vivo*.⁶⁹

Polypeptides must overcome two large energy barriers in order to translocate a pore. The first energy barrier exists as the polypeptide enters the nanopore lumen from the aqueous phase and begins to translocate the pore. This is associated with a concomitant loss of conformational entropy. A second energy barrier is encountered as the peptide exits the other side of the pore back into the aqueous phase.⁷⁰ Translocating nucleic acids also experience similar energy barriers.⁷¹ These energy barriers can be lowered by either structurally altering the nanopore or the polypeptide itself. Such modifications will in turn alter the kinetics of translocation.

Engineering electrostatic 'traps' into the interior cavity of the pore has been shown to reduce the energy barrier and facilitate translocation.⁶⁹ These traps consisted of a ring of negatively charged aspartic acid residues at the entrance and exit of the α HL β -barrel (Figure 8). The ability of positively charged peptides to translocate pores which contained one, two or no traps, was assessed. In this study, the positively charged peptide was added to the *trans* chamber to avoid the entropic penalty associated with entering the vestibule on the *cis* side. It was observed that electrostatic traps placed at the *trans* entrance significantly increased the association rate constant, which is proportional to the

frequency of the single-channel current blockades (Figure 8B).⁶⁵ Conversely, a trap placed at the *cis* entrance of the β -barrel had a greater affect on the dissociation rate constant, which is proportional to the inverse of the duration of the translocation events (Figure 8C).⁶⁵ When traps were placed at both ends of the β -barrel, significant increases were seen in both constants (Figure 8D). Furthermore, the net flow of polypeptides in the *trans* to *cis* direction was significantly increased compared to the unmodified and singly-modified pores.

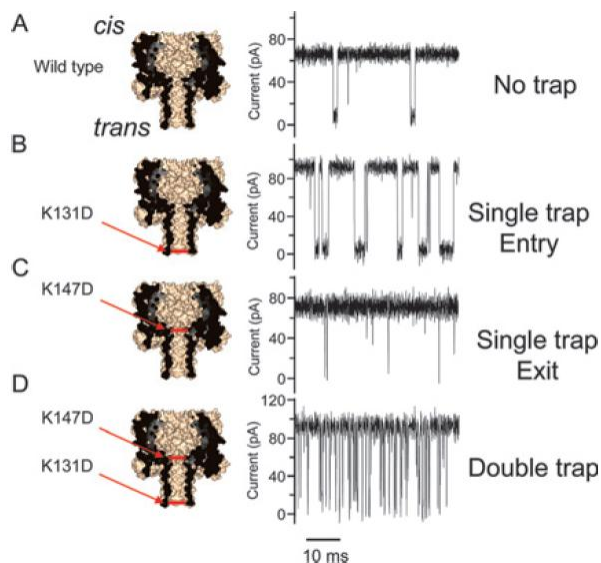


Figure 8: Kinetics of polypeptide translocations in the presence of negatively-charged electrostatic traps. The frequency and duration of polypeptide-induced blockades were dependent upon the position of the traps.

The results suggested that the traps behaved as binding sites for the polypeptides pulling them through the pore and thus promoting translocation events. Interestingly, it was also observed that the peptide-pore interaction was not solely dependent upon the traps, but also the nature of the peptides. Specifically, peptides which possessed more hydrophobic character exhibited lower association rate constants, that is, fewer translocation events. This suggested that energy barrier experienced by hydrophobic polypeptides was greater than their hydrophilic counterparts due to the hydrophilic nature of the pore interior. Thus the binding interaction between the pore and the peptide, and hence the nature of the pore, played a significant role in either facilitating or retarding translocation.

Similar results were obtained in a study in which aromatic residues were engineered into the inner constriction of the α HL pore.^{14,72} Analysis of peptides composed primarily of aromatic residues revealed that the duration and the amplitudes of the blockades were significantly increased, compared to results obtained using unmodified pores. This was attributed to the hydrophobic interactions between the aromatic residues. Essentially, the interactions retarded the translocation of the peptides, enabling them to occupy the inner constriction for a greater amount of time. A direct correlation was also observed between the length of the peptides and the translocation parameters. Intuitively, the findings were not unreasonable as longer peptides do diffuse more slowly than shorter peptides.⁷² Furthermore, as described above, longer peptides are likely to experience a greater entropic penalty prior to translocation, which may also explain this observation. The effect of peptide structure on translocation was also investigated. Peptides that contained more aromatic residues possessed greater binding affinities with the pore. Interestingly, peptides which were of the same length but differed in sequence by only a single aromatic residue displayed a six-fold difference in the event duration and a two-fold difference in the blockade amplitude. Hence modifying the pore with aromatic residues had significantly increased the sensitivity and resolution such that a single amino acid difference could be discerned.

1.5. Aims

The overarching aim of this project was to develop novel sensing strategies which made use of native α HL pores and thus could be readily extended to solid-state nanopores. Solid-state nanopores are more robust and amenable to commercialisation, but cannot be modified to the same atomic-scale precision as biological pores thereby limiting their scope.² Attempts were made to overcome this inherent discrepancy by developing sensing approaches where selectivity is not achieved via pore modification but through modification of the analyte. Two new sensing strategies are described in this thesis. They were both tested with native α HL pores to achieve consistency with existing studies. However, it is expected that they are also compatible with solid-state nanopores. Chapter

2 describes efforts to develop a novel nanopore-based strategy to detect the activity of a protease enzyme. Previously reported strategies¹⁴ have been based upon the detection of multiple peptide fragments generated as the enzyme acted upon a substrate in bulk solution, or required engineering of the pore to improve sensitivity. Using renin as a model protease enzyme, a simpler approach in which a single peptide fragment is enzymatically released and driven through the pore, was envisaged. It was hoped that such an approach would greatly simplify data analysis. Furthermore, by including a hexahistidine tag (His₆-tag) at the terminus of the substrate and performing the enzymatic reaction in the presence of nickel-nitriloacetic (Ni²⁺-NTA) beads, the enzymatically released peptide could be captured and purified from the crude reaction mixture. This approach could provide a means of detecting an analyte directly within a complex biological matrix. Chapter 3 continues in the same vein, describing efforts to develop a nanopore-based strategy to characterise forensically important nucleotide repeat sequences, which are the basis of DNA fingerprinting. Nanopore analysis of ssDNA has been hampered by rapid translocation speeds and subsequent poor resolution of individual bases. It was imagined that a nanopore-based sensing strategy could be realised by using DNA strands modified with nucleotides bearing bulky adamantane tags. It was hypothesised that the modifications would increase the steric bulk of the DNA reducing the translocation speed and hence improving base resolution. Thus, the specific aims of this project were to synthesise nucleoside analogues, incorporate them into oligonucleotides and assess their transport through the nanopore. The motivation of the work presented in Chapter 4 deviates from that of the previous chapters, however it builds on the notion of employing nucleoside analogues to overcome challenges encountered in arena of chemical biology. The specific aim of this project was to synthesise photocaged derivatives of the nucleoside drug azacytidine, which is used to treat cancers that arise due to aberrant methylation of particular regions of the genome. Hypermethylation of these regions can abnormally enhance the function of some genes, such as the oncogenes, or cause other, such as tumour-suppressor genes, to be silenced. Whilst the drug is a potent anti-cancer agent, it is associated with some severe side-

effects which limit its use. It was hypothesised that the side-effects could be mitigated with greater spatial and temporal control over drug release.

2 CHEMICAL TOOLS FOR NANOPORE ANALYSIS OF ENZYMATIC ACTIVITY IN BLOOD PLASMA

2.1. Summary

In this chapter, a label-free strategy to detect enzymatic activity via nanopore recordings is presented. In short, multiple copies of peptide substrates were tethered via a flexible linker to microscale beads. Upon enzymatic cleavage of the immobilised substrate by a protease, soluble fragments were released. These were subsequently electrically detected as they translocated through the nanopore causing reversible blockades of the ionic pore current. Initial proof-of-principle experiments were performed with the trypsin protease and a simple peptide substrate. The developed technique was applied to the detection of renin, a protease enzyme involved in blood pressure regulation, using His-tag analogues of its natural substrate, angiotensinogen. The strategy simplifies the currently used protocols for monitoring renin activity which are based on radioimmunoassays. Furthermore it enables the detection of an analyte within a biological sample; a feat which has been traditionally difficult to achieve using nanopores due to instability of the membrane by sample components such as plasma proteins, and low signal-to-noise ratios. The initial proof-of-concept using trypsin has been published in the *Journal of Physics: Condensed Matter* (*J. Phys.: Condens. Matter*, 2010, **22**, 454103). The work describing the application to renin has been submitted to *Analytical Chemistry*.

2.2. Introduction

2.2.1. Overview of the Renin-Angiotensin-Aldosterone System

The renin-angiotensin-aldosterone system (RAAS) is a physiological system which plays a major role in maintaining and regulating blood pressure, salt and water homeostasis.⁷³ Renin is a proteolytic enzyme which is released from the kidneys directly into the blood stream. Once in circulation, the enzyme acts upon its substrate, angiotensinogen to produce angiotensin I, a small decapeptide. Angiotensin I has no biological action itself, but is converted to angiotensin II by the action of angiotensin-converting enzyme (ACE),

an enzyme present primarily in vascular cells supplying the pulmonary system. Angiotensin II is considered the main effector peptide of the RAAS. It acts upon a number of different cells to ultimately induce an increase in blood pressure (Figure 9).⁷⁴

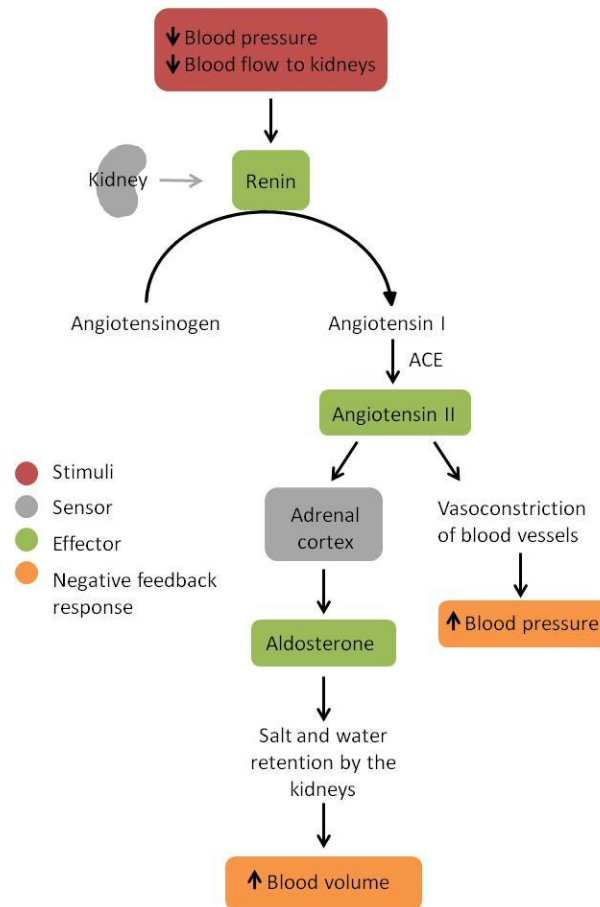


Figure 9: The renin-angiotensin-aldosterone system (ACE: angiotensin-converting enzyme). Adapted from [75].

The main stimuli for renin secretion are conditions of low blood volume and pressure (Figure 9). A fall in blood pressure immediately results in reduced blood flow through the kidneys. This change in pressure is detected mechanically by specialised cells termed baroreceptors, and chemically by macula densa cells which detect the subsequent reduction in sodium levels. Both these changes trigger a feedback mechanism which results in increased renin secretion from a group of cells in the kidney called juxtaglomerular apparatus, and ultimately, an increase in angiotensin II.⁷⁵ Angiotensin II is one of the most potent vasoconstrictor agents known.⁷⁶ It exerts its major effect by binding to receptors located on smooth muscles cells which line arteries and arterioles.

This causes the vessels to contract and consequently increases blood pressure. The aforementioned mechanism is a result of the systemic action of RAAS and the affects are exerted within a matter of minutes.⁷⁷

Components of the RAAS system may also be expressed locally within tissue, enabling angiotensin II to augment blood pressure indirectly.^{74,77} The main mechanisms through which this occurs are stimulation of the thirst centre of the hypothalamus which is located in the brain, and promotion aldosterone secretion from the adrenal cortex which is located on top of the kidney.⁷⁵ Activation of these endocrine glands promotes a rise in blood volume and thus an increase in blood pressure. Specifically, when the thirst centre is stimulated, more water is absorbed from the intestines into the blood. Stimulation of the adrenal cortex, promotes release of aldosterone which subsequently causes the kidneys to retain more salt and consequently more water. The net effect is that an individual will drink more, retain more water, urinate less and thereby increase blood volume and restore renal perfusion. Such local affects are typically mediated over the course of several hours or days.⁷⁷

The RAAS system can also work in reverse. High salt intake can lead to high blood volume and pressure which in turn inhibits renin secretion. With less angiotensin II formation, less salt is retained by the kidneys and more water is excreted as urine. Unfortunately, one in three adults worldwide suffers from chronically high blood pressure, which may be attributed to elevated levels of renin secretion.⁷⁸ In this case, the intake of salt must be lowered to match the impaired ability to excrete salt in urine.

2.2.2. Biochemistry of the Renin-Angiotensin-Aldosterone System

Renin is classified as an aspartic protease enzyme.⁷⁹ This class of enzyme is characterised by inhibition by pepstatin, a hexapeptide from *Streptomyces*, and pH optima in the acid range due to the involvement of two conserved aspartic acids within the active site.⁷⁹ However, the designation of acidic protease is not properly applicable to renin, since its pH optimum is close to neutral when acting upon its natural substrates. The renin substrate, angiotensinogen, is a α_2 -globulin protein produced and released into circulation

by the liver.⁸⁰ The substrate is a glycoprotein with 14% carbohydrate and a molecular weight of approximately 70 kDa.^{80,81} Seminal investigations by Skeggs *et al.*,⁸² showed that treating a substrate protein with trypsin releases a tetradecapeptide from the *N*-terminal of the protein. Furthermore, it was observed that this tetradecapeptide was in fact susceptible to the actions of renin and could mediate the effects of RAAS.

Table 1: Comparison of the amino acid sequence of human angiotensinogen and angiotensin I.

Material	Sequence																									
	P3	P2	P1	P1'	P2'	P3'																				
Human Angiotensinogen ^a	H-Asp-	Arg-	Val-	Tyr-	Ile-	His-	Pro-	Phe-	His-	Leu-	Val-	Ile-	His-	X-	Glu-	Ser-	Thr-	Ser-	Glu-	Gln-	Leu-	Ala-	Lys-	Ala-	Asp-	
Tetradecapeptide ^b	H-Asp-	Arg-	Val-	Tyr-	Ile-	His-	Pro-	Phe-	His-	Leu-	Val-	Ile-	His-	X-	OH											
Angiotensin I ^c	H-Asp-	Arg-	Val-	Tyr-	Ile-	His-	Pro-	Phe-	His-	Leu-	OH															

^a The sequence of the first 25 amino acids of human angiotensinogen determined by Tewksbury *et al.* X has been proposed to be the site of carbohydrate attachment.

^b Tetradecapeptide released by the action of trypsin. In experiments requiring human angiotensinogen, the tetradecapeptide often serves as a synthetic substrate.

^c Angiotensin I is released by the action of renin. The scissile bond lies between Leu (P1) and Val (P1'). Angiotensin I is enzymatically released from both the human angiotensinogen and the tetradecapeptide

The amino acid sequence of the first 25 residues of human angiotensinogen were determined by Tewksbury *et al.*, by Edman degradation of the protein (Table 1).⁸³ Interestingly, the identity of the fourteenth amino acid residue could not be determined suggesting that it was a site of carbohydrate attachment. It was observed that the first ten amino acids of human angiotensinogen corresponded to a previously reported sequence of human angiotensin I.⁸⁴ Hence the scissile bond in human angiotensinogen was leu-val.

Renin is known to be a species specific enzyme.⁸⁵ Human angiotensinogen is the only naturally occurring substrate for human renin and is resistant to cleavage by non-primate renin. Conversely, angiotensinogen isolated from most other species can be cleaved by human renin. This species specificity has been attributed to the differing amino acid sequences in positions P₁'-P₃' as the residues at the P₅-P₁ sites are identical among all angiotensinogens.^{86,87} [The notation of Schechter and Berger is used here,⁸⁸ in which the residues of the substrate are numbered away from the scissile bond P₁---P_n on the *N*-

terminal side and P₁'---P'_n on the C-terminal side. Corresponding enzyme subsites are labeled S_n].

2.2.3. Measurement of Plasma Renin Activity

Abnormal renin secretion and irregularities in the RAAS have been implicated in a number of diseases and conditions.^{89,90} For this reason, measurement of circulating renin as an indicator of RAAS activity can be a useful clinical marker and diagnostic tool. As angiotensin II is the main effector of the RAAS system (Figure 9), one would expect renin evaluation to be based upon its production. However, due to its minute quantities and rapid degradation in the blood, measurement can be challenging. As a result, circulating renin is measured either directly using antibodies which recognise its active site, or indirectly in kinetic assays which measure the capacity of the enzyme to produce angiotensin I. These two approaches result in the determination of plasma renin concentration (PRC) and plasma renin activity (PRA), respectively.⁹¹

PRA is usually determined using radioimmunoassay of angiotensin I generated by endogenous enzyme and substrate.⁹² Typically, a blood plasma sample containing angiotensin I is placed in an antibody-coated tube and antibody-antigen complex formation ensues. ¹²⁵I-labeled angiotensin I is added to the reaction vessel, and competes with unlabelled angiotensin I for a binding site. After incubation at 37 °C,⁹² the amount of radioactivity remaining on the reaction vessel is measured, from which the amount of unlabelled angiotensin I in the plasma sample can be estimated.^{93,92} The use of PRA is based upon the assumption that measuring the generation of angiotensin I *in vitro*, assesses the action of renin *in vivo*.⁹⁴ In order for this assumption to hold true, the plasma angiotensinogen levels and renin catalytic activity should be preserved as present *in vivo*. This can be achieved by carefully controlling the temperature whilst collecting plasma samples.⁹⁵ Plasma samples must be collected and immediately stored at -20 °C to prevent cryoactivation of prorenin, the inactive precursor of renin, which can be activated when plasma is cooled to between -5 °C and 4 °C. The accuracy of the PRA can be further improved by subtracting background enzymatic activity. At ambient temperature, renin

will consume angiotensinogen and produce angiotensin I. Thus the amount of angiotensin I generated in this instance will need to be subtracted from that generated in the radioimmunoassay.⁹² Taking the aforementioned precautions prevents uncontrolled angiotensin I production and subsequently overestimation of renin activity.

PRC is also determined using a radioimmuno-based assay. However, the assay uses antibodies to measure the number of active renin molecules independently of their enzymatic activity.⁹⁶ The assay makes use of two different antibodies: a less selective antibody that is immobilized on a solid support which is used to 'capture' angiotensin I from plasma; and a second, radioactive antibody reacts with the 'captured' angiotensin I to enable its quantification.^{95,97} This form of analysis can also be described as a sandwich assay as the analyte being measured is 'sandwiched' between two different antibodies.⁹⁵

Recently DiaSorin S.P.A. (Saluggia, Vercelli, Italy), introduced an alternative sandwich assay, in which the second antibody possesses chemiluminescent properties.^{92,98,99} This assay permits the direct measurement of renin using antibodies which are modified with acridinium esters. Once these antibodies have reacted with the 'captured' renin, the resultant complexes are treated with an alkaline peroxide solution which initiates a chemiluminescent reaction. The amount of bound labelled antibody, is determined by relative light units, which are directly proportional to the concentration of the active renin in the sample.⁹² Chemiluminescent labels enable signal amplification and therefore offer increased sensitivity over standard PRC techniques. Furthermore, the process has been fully automated which enables higher throughput of samples. This is in stark contrast to radioimmuno-based techniques which are completely manual.⁹⁷

Chemiluminescent assays offer several other advantages over standard PRC assays. The most significant is elimination of the use of radioactive material. Use of isotopes requires that special measures be taken for their procurement and disposal. Furthermore, due to the hazards involved in using such materials, preparation of the starting materials can be time-consuming and require a high level of expertise.⁹⁷ Additionally, whereas non-isotopic labels have indefinite stability, radiolabels are inherently unstable and may only have a shelf-life of a few weeks or months.⁹⁷

HPLC-based methods for the quantification of renin activity have also been developed. These typically make use of a fluorogenic peptide substrates which are hydrolyzed by renin to generate fluorogenic and non-flourogenic products.¹⁰⁰ The amount of hydrolysis product is directly measured by reverse-phase HPLC, and the peak area compared to that of a standard in order to quantify the amount of peptide present. Assays based on fluorescence resonance energy transfer within peptide substrate, have been developed.^{101,102} The assays make use of peptide substrates that possess fluorescence doner and acceptor moieties at either the *N*- or *C*-terminal of a substrate analogue. In the intact peptide, the fluorescence is quenched due to favourable energetic overlap of the fluorophores excited state and the absorbance of the quencher. Cleavage of the substrate by renin liberates a fluorogenic peptide fragment from the proximity of the quencher. This results in a gradual increase in fluorescence, the rate of which is proportional to rate of peptide hydrolysis and thus the activity of the enzyme.

While many techniques for measuring renin activity are currently available, they all require at least one component of the assay to be labelled. This represents a major limitation as the use of labels can increase the cost of an assay and complicate the preparation of components. It is envisioned that employing a nanopore-based approach will obviate the need for any labels and thus provide a simplified assay. Furthermore, it may permit the development of portable devices, which in turn could facilitate point-of-care analysis.

2.3. Aim

In light of the current techniques available for the detection of renin activity, the ultimate aim of this project was to develop a nanopore-based technique which did not require engineering of the protein pore or labelling of any components. Furthermore, it was endeavoured that enzymatic digestion would liberate a single peptide fragment and thus vastly simplify data analysis. The proposed strategy is schematically summarised in Figure 10. Multiple copies of a peptide substrate are tethered via a flexible linker to microscale beads (Figure 10A). Upon enzymatic cleavage of the substrate by the enzyme (Figure 10B),

soluble fragments are released and subsequently electrically detected as they translocate through a nanopore (Figure 10C).

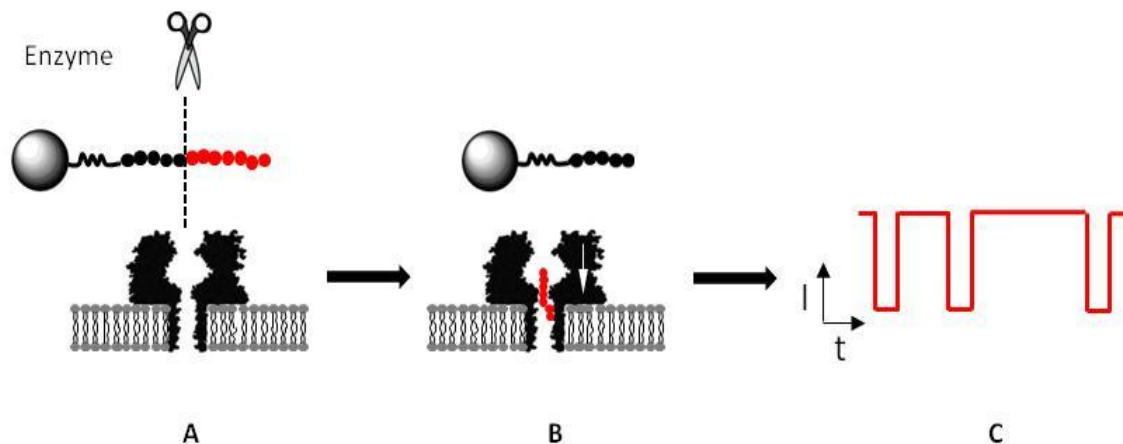


Figure 10: Proposed strategy for nanopore-based technique to detect enzymatic activity. (1) Multiple copies of a peptide substrate, tethered via a flexible linker onto a microscale solid support. The substrate is cleaved by a proteolytic enzyme. For reasons of visual clarity, only one peptide substrate is shown. (2) The released peptide fragments (shown in red) are electrophoretically driven through the nanopore. (3) This gives rise to reversible blockades of an ionic pore current.

It was hypothesised that by including a His₆-tag at the terminus of the substrate and performing the digestion in the presence of nitrilotriacetic acid (Ni²⁺-NTA), the enzymatically released peptide could be captured and purified from the crude reaction mixture. It was hoped that this two-step approach would enable the analysis of peptides from a complex biological matrix such as human serum. This has traditionally been difficult to achieve using nanopores due to instability of the lipid bilayer and increases in background noise triggered by other molecules which can also block the pore.

As this project involved a significant amount of peptide synthesis, a brief overview of the techniques, reagents and reaction mechanisms will be provided before presenting the results.

2.4. Solid-Phase-Peptide Synthesis

The peptides used during this investigation were synthesized via solid phase peptide synthesis (SPPS), using Fmoc-chemistry.¹⁰³⁻¹⁰⁵ SPPS involves the construction of a peptide chain in the C→N-terminal direction on an insoluble solid support.¹⁰⁶ The principles of SPPS are illustrated in Figure 11. The C-terminal amino acid of the target peptide is

attached to the support through its carboxyl group. The protecting group masking the α -amino group is removed (deprotected) and the soluble by-products removed by washing and filtration. An excess of the second amino acid is added, the carboxyl terminal of which is activated towards amide bond formation. Activation is typically achieved through the use of a coupling agent (*in situ*). The N- α -amino protecting group of the resultant dipeptide is removed and the coupling process repeated until the desired peptide is synthesised. After each coupling step, excess amino acids and soluble by-products are removed. In the final step, the bond between the peptide and the resin is cleaved. The peptide is liberated into solution where it can be isolated and purified.

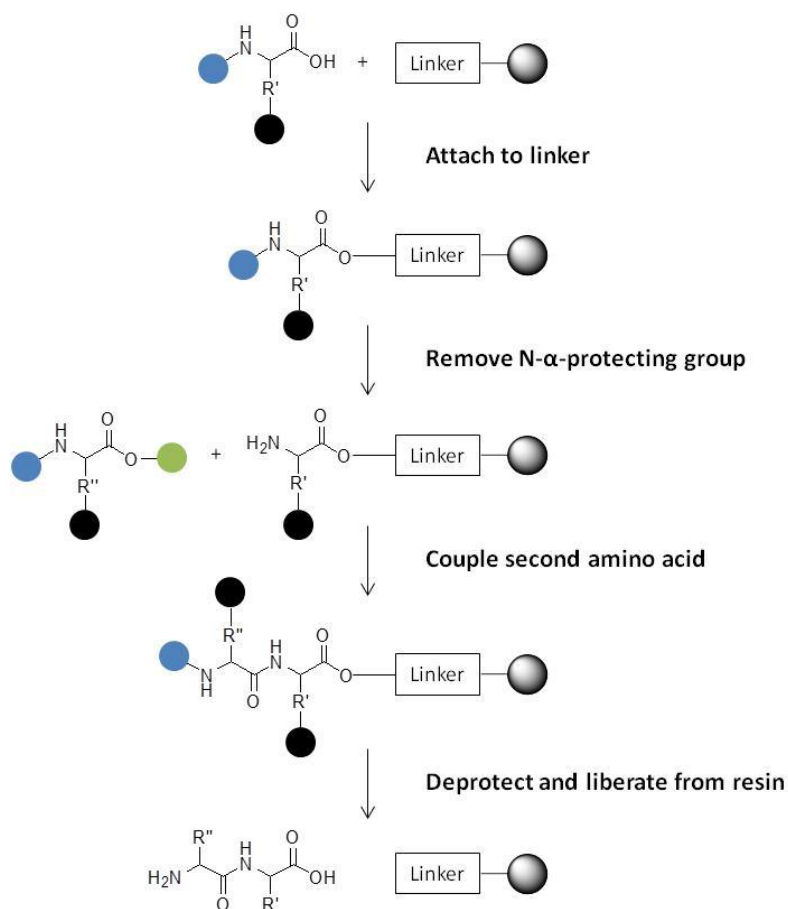
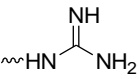
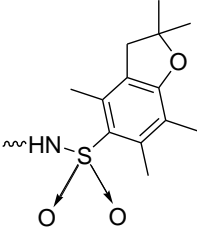
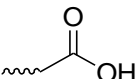
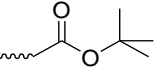
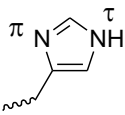
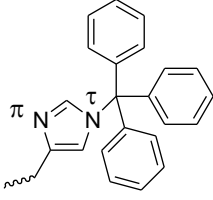

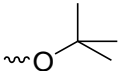

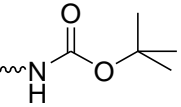


Figure 11: Principle of solid phase peptide (SPPS) synthesis. All non-reacting functional groups are protected during the synthesis. The N- α -protecting group (blue circle) is typically protected with Fmoc. The side-chain protecting groups (black circle) are typically orthogonal with the bond between the peptide and the solid support. Hence the peptide can be liberated from the solid-support and side-chain deprotected simultaneously. Green circle is a C-terminal activating group.

Generally, the bonds between the peptide, the solid-support and side-chain protecting

groups, are such that liberation and deprotection occur simultaneously. The *N*- α -amino protecting groups can be removed independently and selectively in the presence of the side-chain protecting groups. Hence these groups are said to be orthogonal. In Fmoc SPPS, orthogonality is achieved by use of the base-labile Fmoc group for protection of the α -amino group, and acid-labile side-chain protecting groups. The bond to the resin is usually also acid labile. Fmoc cleavage is traditionally achieved with 20-50% piperidine in DMF, whereas global side-chain deprotection and release of the peptide with 95% TFA. The protecting groups most commonly used in this project have been summarised in Table 2.

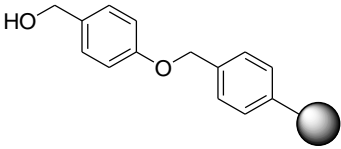
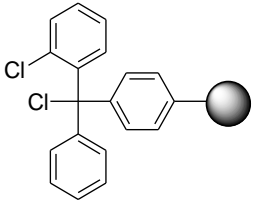
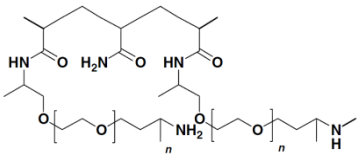
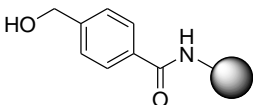
Table 2 Common side-chain protecting groups used in Fmoc-SPPS.¹⁰⁶

Entry	Side-chain functionality	Protecting group	Abbreviation
1	Arg 		Pbf
2	Asp 		OtBu
3	His 		Trt
4	Ser/Tyr 		tBu
5	Lys 		Boc

The solid supports or resins employed in SPPS are typically polystyrene-based.¹⁰⁴ The resins are able to swell in solvents to allow reagents to access reactive sites and enable the synthesis and release of the peptide. In addition to this, the resin must be chemically inert and resistant to degradation during synthesis. The most important feature of resins is that they can be readily functionalised with linkers to which the *C*-terminus of a peptide is attached. The linker serves as a reversible attachment between the peptide chain and the solid support, and also protects the *C*-terminal α -carboxyl terminal during synthesis.¹⁰⁶ The choice of linker dictates the ease with which the peptide is liberated from the resin and also the functional group at the *C*-terminus. Most linkers are designed to release the peptide with either a carboxylic acid or an amine upon treatment with TFA.¹⁰⁶ The loading of the resin, that is, the concentration of the linker functional groups, determines the theoretical yield of the final peptide. The linkers used throughout this investigation are summarised in Table 3.

The concentration of TFA employed dictates whether the peptide is released in a completely deprotected state, or with the side-chain protecting groups intact. The strength of the acid required to release the peptide is dependent on the nature of the linker and the leaving group. In general, peptides are more easily liberated from linkers which are electron-rich. Conversely, electron-poor linkers may require the use of a stronger acid such as HF. Regardless of the acid which is used, peptide deprotection and liberation results in a reaction medium which is rich in electrophilic alkylating agents. These species' can alkylate electron-rich functional groups on the peptide such as hydroxyls and thiols. These unwanted reactions can be minimised by performing the deprotection and liberation steps in the presence of scavengers. The most commonly used are thiols such as 1,2 ethanedithiol (EDT), silanes such as triethylsilane (TES) and water.

Table 3 Linker resins used for the synthesis of peptides.

Entry	Linker	Cleavage conditions	Loading
Wang resin			
1		90-95% TFA v/v	0.65 mmolg ⁻¹ ^[a]
2-Chlorotrityl chloride resin			
2		1. 90-95% TFA v/v ^[c] 2. 20% TFE:DCM v/v ^[d]	0.60 mmolg ⁻¹ ^[b]
PEGA			
3		Enzymatic digestion ^[e]	0.20 mmolg ⁻¹ ^[f]
HMBA			
4		0.1 M NaOH (aq)	0.089 mmolg ⁻¹ ^[g]

^[a] This loading refers to the pre-loaded glycine derivative which was utilised in this investigation. ^[b] This loading refers to the unsubstituted resin. The resin was loaded with the desired amino acid prior to peptide synthesis. Refer to the Materials and Methods section for details. ^[c]Applying these conditions simultaneously liberates and side-chain deprotects the peptide. ^[d] Applying these conditions liberates a fully protected peptide. ^[e] The peptides were directly attached to PEGA via a permanent amide bond. Peptide fragments were liberated by enzymatic digestion. Peptides attached via HMBA linker could be released by treatment with 0.1 M NaOH. ^[f] Refers to the loading of the un-substituted resin. ^[g] This is an approximation. The loading was determined by monitoring Fmoc deprotection after peptide synthesis.

Amide bond formation requires activation of the carboxyl terminal of the incoming amino acid. This is typically achieved with the use of coupling agents which render the carbonyl of the carboxylic acid more electrophilic and thus more susceptible to nucleophilic attack by the amine of the growing peptide chain. The most commonly used coupling agents are uronium salts which are preferred for their solubility, stability and comparatively lower side reactions.^{107,108} The most popular reagent of this class is *N,N,N',N'*-tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU). Its mechanism of action is shown in Figure 12. The deprotonated acid reacts with the uronium salt to generate an activated acyl-uronium species and 1-hydroxybenzotriazole (HOBt). HOBt readily reacts with the activated acid to generate a more reactive benzotriazolyl (OBt) ester which undergoes nucleophilic attack by the amine to generate an amide bond.

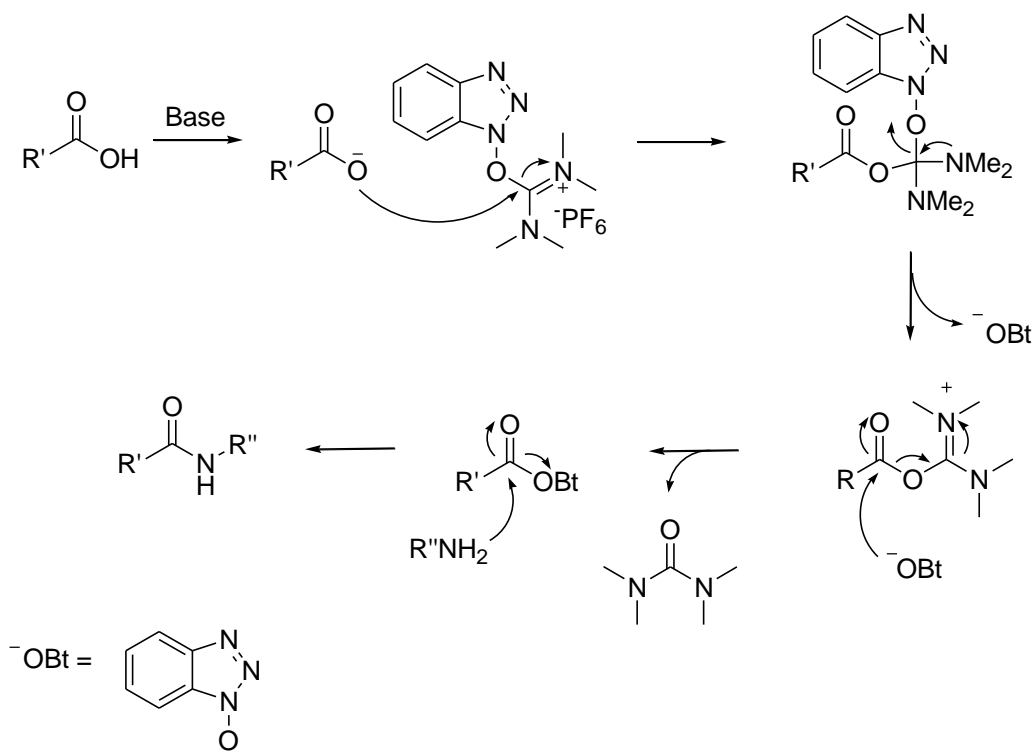


Figure 12: Amide bond formation mediated by HBTU. A tertiary base is employed to deprotonate the carboxyl terminal of the peptide. The reaction is performed in one-pot.

2.5. Results and Discussion

2.5.1. Proof-of-Principle Experiments using Trypsin

Initial proof-of-principle experiments were performed using a simple peptide, Y₃KIG₄, and

the trypsin protease.¹⁰⁹ The peptide sequence was designed by considering the specificity of trypsin as well as the requirements for nanopore analysis. Trypsin is a pancreatic serine protease which cleaves peptides selectively at the carboxyl-side of lysine and arginine amino acid residues. Its specificity arises due to an electrostatic interaction between these residues, which are positively charged at physiological pH, and a negatively charged aspartate residue within the active site (which located in the S_1 catalytic pocket).¹¹⁰ The rate of enzymatic hydrolysis is also determined by the amino acid at the carboxyl side of the scissile bond, that is, position P1'.^{110,111} In particular, the enzyme favours large hydrophobic residues, such as isoleucine, in this position.¹¹² Taking these factors into consideration, it was expected that the enzyme would cleave the peptide between the isoleucine and lysine residues to liberate an Y_3K fragment (Figure 13A).

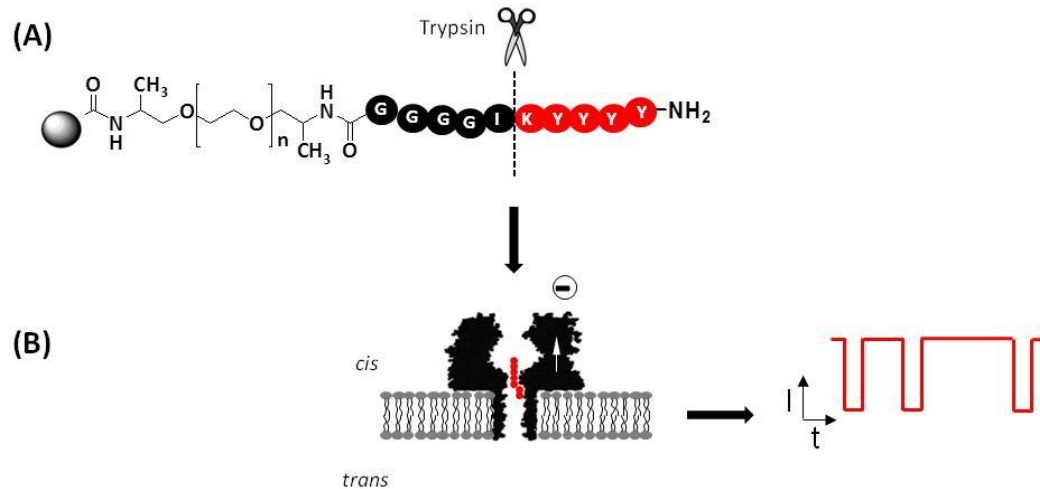


Figure 13: The approach to electrically detect trypsin activity via nanopore recordings. (A) Multiple copies of a peptide substrate, tethered via a flexible poly(ethylene) glycol linker onto a solid support, are cleaved by trypsin enzyme. The sequence of the substrate and the cleavage site are shown. For visual clarity, only a single substrate molecule is shown. (B) The released peptide fragments are electrophoretically driven through the α HL pore giving rise to reversible blockades of the ionic pore current. Drawing not to scale.

The detection of the peptide by nanopore analysis was considered by placing three tyrosine residues at the N -terminus. It was anticipated that these bulky residues would increase the extent of pore blockade and thereby facilitate identification of the intact peptide and the released fragment. Additionally, as a result of the lysine residue, the peptide was expected to carry a net positive charge which would facilitate the electrophoretic passage through the pore when a negative transmembrane potential was

applied (Figure 13B). Finally, four small glycine residues were placed at the C-terminus. It had previously been reported that these residues acted as a 'spacer' to aid the synthetic incorporation of the bulky tyrosine and isoleucine residues, and increase the steric accessibility of the protease.¹¹³

2.5.2. Preparation of Immobilised Substrate via Convergent Solid-Phase Peptide Synthesis

In order to determine whether the model peptide was indeed a suitable substrate for trypsin, digest reactions were performed using a soluble form of the peptide. The peptide was synthesised using standard Fmoc-based SPPS on a Wang resin (Table 3, entry 1) pre-loaded with glycine, and HBTU coupling chemistry. The peptide was isolated in a crude yield of 77% which corresponded to a mass of 59 mg. A small subset of the peptide was purified by preparative HPLC, characterised by LC-MS analysis then subjected to the action of trypsin. A ratio of 1:50 enzyme:peptide was employed and the reaction mixture incubated at 37 °C for 18 h. An aliquot of the resultant solution was analysed by HPLC which showed a decrease in the peak area of the Y₃KIG₄ peptide and the emergence of a new peak (Figure 14B). MS analysis confirmed that the latter was indeed the Y₃K fragment (Figure 14B). The release of the fragment was certainly due to the action of the enzyme as blank experiments in which the peptide was incubated in buffer in the absence of trypsin did not produce any peptide fragments. This notion was supported in a supplementary experiment in which Y₃KG₅ was incubated with trypsin. Substitution of the isoleucine for a glycine residue inhibited the release of the Y₃K fragment. This experiment also confirmed the importance of having a hydrophobic amino acid residue in the S1' binding site of the enzyme.

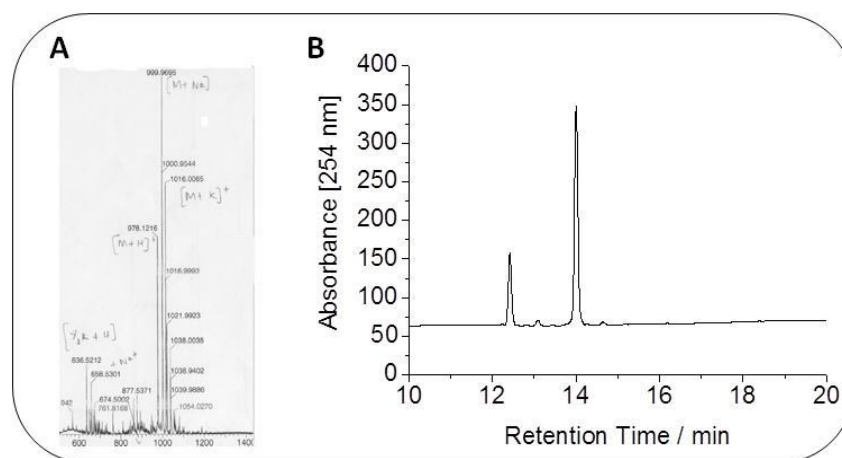


Figure 14: (A) MALDI trace of solution phase digestion of Y_3KIG_4 . M corresponds to the undigested substrate. (B) HPLC trace of digestion mixture. The peak at 12.4 min corresponds to the Y_3K fragment. The peak at 14.0 min corresponds to intact, undigested substrate. HPLC analysis was performed on a Varian Pursuit XRS C18 column and UV absorbance measured at 254 nm. See Materials and Methods for details.

Having shown that Y_3KIG_4 was a substrate for trypsin, efforts were made to synthesise the substrate on a PEGA (bis 2-acrylamidoprop-1-yl polyethyleneglycol cross-linked with dimethyl acrylamide) resin (Table 3, entry 3). PEGA is one of the most widely used resins for heterogeneous (i.e. on-bead) enzymatic assays.¹¹⁴ It possesses 2-aminopropyl groups at the end of polyethylene glycol chains, which are attached to a polyacrylamide core (Table 3, entry 3).¹¹³ This feature enables the resin to be readily functionalised using standard coupling chemistry. Furthermore, the presence of PEG chains inside the resin, furnishes it with hydrophilic properties which enable it to swell in a range of organic solvents and aqueous media.¹¹⁵ Accordingly, the swelling property of the resin determines the ability of an enzyme to diffuse through the resin and access the immobilised substrate. PEGA₁₉₀₀ which was used in this investigation, is freely permeable to enzymes with a molecular weight of up to 35-45 kD.¹¹⁶ Thus the resin was ideally suited for assays involving trypsin and renin which have molecular weights of 23.3 kD¹¹⁷ and 42.0 kD⁸⁰, respectively.

The initial attempts to synthesise Y_3KIG_4 on PEGA resin were based on the convergent solid-phase peptide synthesis (CSPPS). CSPPS involves the synthesis of small fully protected peptide fragments using standard SPPS described above (Figure 11). After

purification, the fragments are coupled together on a solid-support to construct a larger peptide. The target peptide is then deprotected and can be cleaved from the resin. Thus the aim was to synthesise the fully protected Y₃KIG₄, couple it to PEGA then remove the protecting groups to give the immobilised substrate.

Synthesis of the fully protected peptide was achieved using standard SPPS with polystyrene resin bearing a 2-chlorotrityl linker (Table 3, entry 2). In a slight alteration to conventional SPPS, the final Fmoc deprotection was not performed. Furthermore, cleavage of the peptide from the resin was achieved in far milder conditions using 20% TFE/DCM (v/v) which enabled the fully protected peptide to be isolated in 30% crude yield. Dissolution of a small amount of the crude material into neat DMSO enabled analytical HPLC and LCMS analysis to be performed. However purification by preparative HPLC was ruled out as addition of water to the DMSO solution resulted in precipitation of the peptide. Using analytical HPLC, the purity of the protected peptide was determined as 69% (

Figure 15B). Indeed, peaks with retention time of 18.27 and 24.92 min were observed. LCMS analysis revealed that these were the Fmoc-deprotected product (1246 m/z [M+2Na]²⁺) and the desired fully-protected peptide (1468 m/z [M-H]⁻), respectively (

Figure 15A). A partially protected adduct with mass 1353 m/z was also observed. This corresponded to the fully protected peptide minus a *tert*-butyl protecting group (Table 2, entry 4).

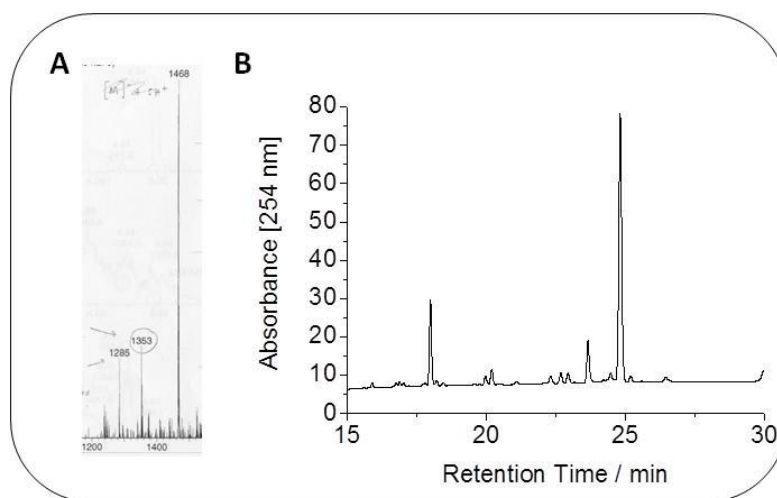


Figure 15: (A) Mass spectrometric analysis of fully protected Y_3KIG_4 , 1468 m/z [M-H]⁻; Fmoc-deprotected product, 1246 m/z [M+2Na]²⁺; partially protected adduct (minus *tert*-butyl protecting group), 1353 m/z. (B) HPLC analysis of the crude fully protected Y_3KIG_4 . The peak at 25 min corresponds to the desired peptide. HPLC analysis was performed on a Varian Pursuit XRS C18 column and UV absorbance measured at 254 nm. See Materials and Methods for details.

The crude peptide was subsequently coupled to the PEGA resin via an amide bond. The peptide was dissolved in the minimum amount of DMSO producing a highly concentrated solution to which HOBt and *N,N'*-diisopropylcarbodiimide (DIC) were added. The activated peptide solution was added to the PEGA resin and the resultant suspension shaken at room temperature. Completeness of the coupling step was determined by testing the resin for the presence of free amines using the Kaiser test. The test is based on the reaction of ninhydrin with primary amines, which gives a characteristic dark blue colour.¹¹⁸ The resin gave a negative result, indicating that no free amines were present and thus coupling was complete. The bound peptide was subsequently Fmoc-deprotected by incubation with 20% piperidine/DMF (v/v), and side-chain deprotected using standard cleavage cocktail comprised of TFA:TES:water, 95:2.5:2.5. Fmoc deprotection was monitored by analytical HPLC which revealed a peak corresponding to the dibenzofluvene-piperidine cleavage product (Figure 16, **2**), confirming that deprotection had occurred.

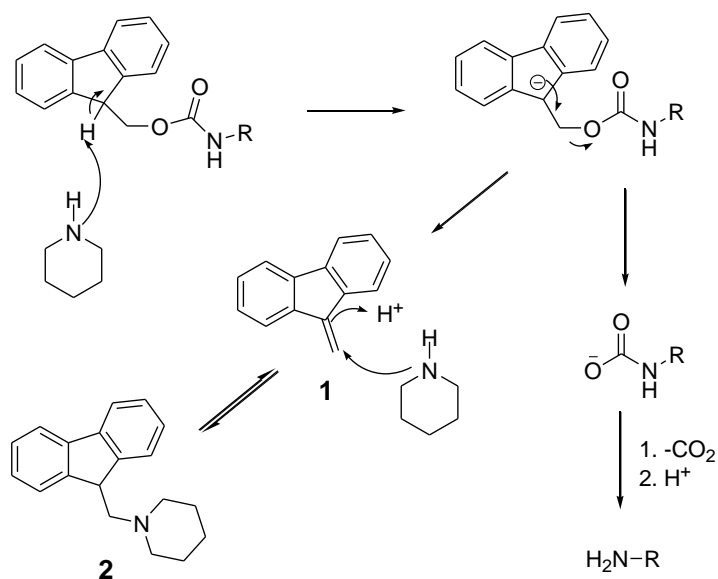


Figure 16: Removal of *N*-Fmoc group with piperidine. The reaction proceeds via initial abstraction to give a stabilized dicyclopentadiene anion, which on elimination gives rise to dibenzofluorene, **1**. **1** is scavenged by piperidine to give adduct **2**. The products of the deprotection reaction absorb UV strongly, thus provide a method of monitoring this reaction. Adapted from [119].

The Fmoc-deprotection solution was used to estimate the loading of the resin, which in turn, was used to estimate the yield of the coupling. The UV absorbance of the deprotection solution was measured in triplicate using equation (1)¹²⁰ⁱ:

$$\text{Resin loading (mmol g}^{-1}\text{)} = \frac{\text{Abs}_{\text{sample}}}{\text{mg of sample} \times 1.75} \quad (1)$$

The loading was estimated to be 0.03 mmol g⁻¹. As the unsubstituted PEGA resin was purchased with a loading of 0.2 mmol g⁻¹, the coupling yield was determined as 15%. The relatively low yield was unexpected considering that the Kaiser test suggested that no free amines were present on the PEGA resin. The loading may have been improved by performing a second coupling using a fresh solution of the activated peptide however this was not possible as the supply of peptide was limited. Synthesis of the protected peptide was performed using 120 mg of chlorotriyl resin pre-loaded with glycine (resin loading: 0.8 mmol g⁻¹) and typically yielded only 17-60 mg of material (12-42% yield). One plausible

ⁱ This is only a rough estimation the equation calls for the use of the dry weight of the resin. However, the

explanation of the low yields is the hydrophobic nature of the protected peptide. It was hypothesized that the cleaved peptide was interacting with the resin rather than being released into solution. Attempts were made to extract the peptide from the resin by suspending the resin in DMSO. However, subsequent HPLC analysis revealed that the very little additional peptide was released. The low yield could also be attributed to the poor swelling properties of PEGA in DMSO. The protected peptide was soluble in DMSO but completely insoluble in aqueous solutions or DMF. Conversely, PEGA displays optimum swelling in water and DMF but does not swell in DMSO.¹²¹ The swelling properties of PEGA are summarised in Figure 17. It is likely that the coupling reactions only took place on the outer surface of the resin, which was accessible to the peptide.

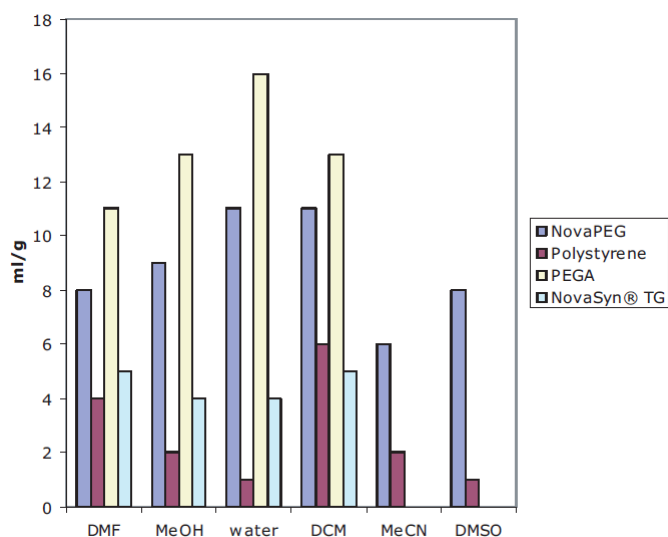


Figure 17: Swelling properties of PEGA and other resins [121].

2.5.3. Enzymatic Digestion of Substrate Prepared by CSPPS

Having successfully produced PEGA-bound substrate, albeit in a low yield, a heterogeneous enzymatic assay was performed using trypsin. Incubation of the PEGA derivative with the enzyme did indeed result in the release of a single fragment as determined by HPLC analysis (Figure 18). MS analysis identified the peptide to be of the sequence Y₃K. This confirmed that the cleavage site correlated to that of the soluble peptide and indicated that the resin did not interfere with the activity of the enzyme. The purity of the released peptide was surprisingly high considering that the protected peptide had not been purified.

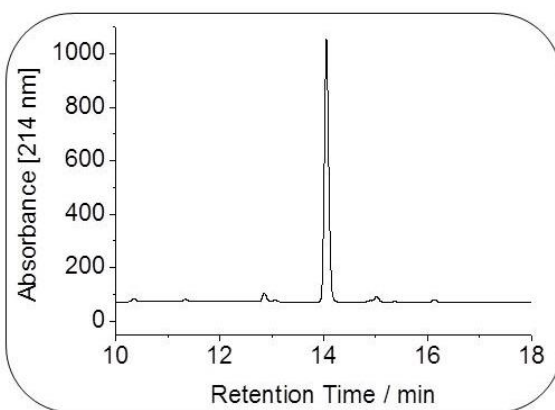


Figure 18: HPLC trace of Y₃K enzymatically released from PEGA-Y₃KIG₄. HPLC analysis was performed on a Varian Pursuit XRS C18 column and UV absorbance measured at 214 nm.

2.5.4. Preparation of Immobilised Substrate via Manual Solid-Phase Peptide Synthesis

Although the PEGA-bound substrate was successfully prepared using the CSPPS-based approach, an alternative method through which larger quantities of PEGA-substrate could be prepared, was required. Manual solid-phase peptide synthesis (MSPPS) directly onto the PEGA resin was seen as a better alternative.

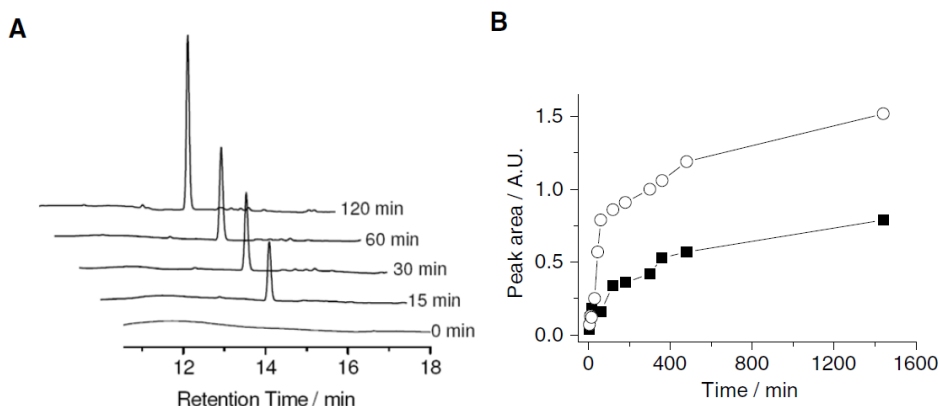
The principle behind MSPPS is the same as automated SPPS. A peptide is assembled through repetitive coupling and deprotection of *N*- α -protected amino acids on a solid support; however, there are some significant differences. Most notably, during MSPPS, the coupling efficiency of each step is carefully monitored and can be improved upon. Monitoring is typically achieved using a qualitative technique in which a reagent reacts with free amino groups to produce a chromophoric compound (e.g. Kaiser test).¹⁰⁵ Any observed reduction in the coupling efficiency can be improved by performing an additional coupling until tests show that the reaction is complete. A further distinction from automated SPPS is that unreacted amino groups are chemically blocked (capped) at each step. Regardless of the efficiency of the coupling agent used, each step of the synthesis will contain peptides to which couplings have failed, giving rise to deletion peptides.¹⁰⁵ The number of deletion sequences can be reduced by treating the resin with acetic anhydride, in the presence of an organic base. As a result of these practices, peptides synthesised using MSPPS can benefit from improved yield and homogeneity.

Furthermore, MSPPS offers greater control over synthesis and can often be performed on a larger scale compared to automated SPPS.

Using MSPPS, PEGA-Y₄KIG₄ was synthesised starting with 2.33 g of PEGA. With the exception of the first step, only a single coupling reaction was needed to drive each step to completion (as determined by the Kaiser test). Amide bond formation was mediated by HOBt/DIC rather than HBTU/DIPEA, as the former has greater solubility in DMF compared to the latter. At each step, capping was performed and Fmoc-deprotection monitored by HPLC. The loading of the resin was estimated as 0.16 mmolg⁻¹ using equation 1 and spectroscopic analysis at 290 nm of the dibenzofluvene-piperidine adduct formed upon deprotection of the terminal tyrosine residue. The coupling yield was determined as 80% which is almost 5-times higher than yield achieved using CSPPS-based approach to peptide synthesis.

2.5.5. Enzymatic Digestion of Substrate Prepared by MSPPS

The PEGA derivative was incubated with trypsin and the supernatant subsequently analysed by HPLC and MS analysis. Analytical HPLC showed that a single peptide fragment had been released; the purity of which was slightly greater than that released from the CSPPS-based substrate. MS analysis identified the peptide to be of sequence Y₄K.ⁱⁱ



ⁱⁱAn additional tyrosine residue was coupled.

Figure 19: Kinetics of substrate cleavage by trypsin. (A) HPLC traces for the time-dependent enzymatic release of the YK₄ peptide fragment from PEGA. Y-axis: absorbance at 214 nm (B) Kinetics of enzymatic cleavage of PEGA-bound Y₄KIG₄ (solid squares) and in solution (empty circles). The data points are the averages of the peak areas from two independent experiments. Adapted from [109].

The rate of Y₄K generation in the homogenous and heterogeneous assays was assessed by HPLC (Figure 19). In both instances, the amount of peptide increased with incubation time, as would be expected for an enzymatic digestion (Figure 19A). The resulting kinetics yielded an apparent initial rate constant of $2.6 \times 10^{-3} \text{ min}^{-1}$ for the cleavage from PEGA (Figure 19B, solid squares). By comparison, cleavage of soluble peptide substrate under otherwise comparable conditions proceeded with a higher rate constant of $4.0 \times 10^{-3} \text{ min}^{-1}$ with a 1.9-times higher amount of the peptide at the endpoint (Figure 19B, empty circles). The difference in kinetics likely reflects the longer diffusion time of the enzyme to the bead surface. Porous supports generally have a high loading of functional groups, which are distributed over the surfaces inside the support so that enzymes need to penetrate the support for successful catalysis.¹²² These accessibility problems do not occur in homogenous assays.

2.5.6. Nanopore Analysis

The solution of Y₄K peptide fragments released from the PEGA resin was examined using the α HL nanopore. As discussed earlier, nanopore analysis is an electrical approach to sense individual molecules that temporarily reduce the ionic current flowing through a nanoscale pore. The passage of the analyte through the pore is termed a ‘translocation event’ and the analyte is said to have caused a ‘blockade’. In practical terms, modulations of the ionic current are measured using electrophysiology data acquisition software called pClamp. The software analyses the details of the electrical traces and deposits the data into a spreadsheet for further numeric-graphical analysis using software called Clampfit. To ensure a statistically valid analysis, the blockade data are fitted to the appropriate distributions.¹²³ Three translocation parameters are obtained: the amplitude (or size of the blockade, A), the dwell time (or duration of translocation events, τ_{off}) and the duration between events (τ_{on}) (Figure 20).

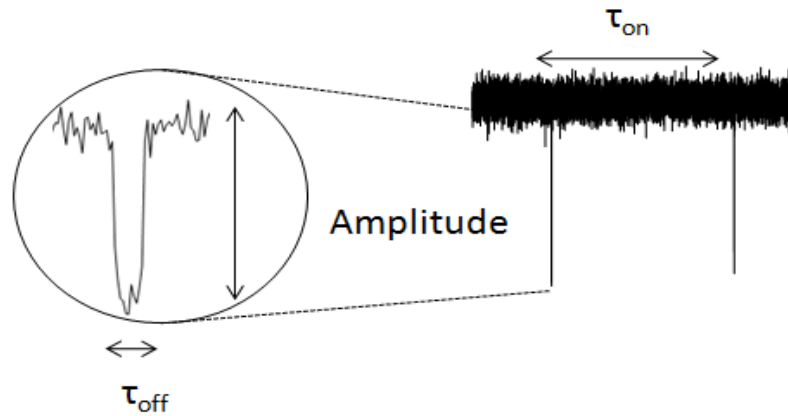


Figure 20: Definition of the amplitude, dwell time τ_{off} and inter-event interval τ_{on} .

The amplitude of blockade, A , is expressed as a percentage or a fraction of the open channel current with units of picoamperes (pA). pClamp measures the amplitude of individual events and the data plotted and fitted to a Gaussian distribution. In an experiment involving the analysis of a single analyte, the majority of translocation events will have similar amplitudes. Specifically, there will be very few events with uncharacteristically large (complete blockade of the pore) or small (incomplete translocations) amplitudes.¹²³ The peak of the Gaussian curve is taken as the statistical average blockade for that set of results. The average dwell time, τ_{off} , is derived by plotting the individual τ_{off} values against the number of translocation events within the data set. The data are fitted to a Poisson distribution which describes an exponential decay for which the exponential decay constant is the characteristic value. A logarithmic scale is applied to the y -axis and the average τ_{off} taken from the decay constant. The time between events, τ_{on} , is determined in the same way as τ_{off} . τ_{on} is dependent upon the concentration of the analyte being analyzed; the two variables are directly proportional. Consequently, in order to compare τ_{on} values determined from independent nanopore experiments, experiments are often performed with samples that have similar concentration. In situations where this is not possible, the τ_{on} value can be normalised for concentration after it has been determined. By taking the inverse of τ_{on} , the frequency of occurrence (Foo) of translocation events can be determined. Thus Foo corresponds to the frequency of translocation events per unit time. The nanopore data reported throughout this thesis was generated from at least three independent nanopore recordings, and at

least a thousand translocation events analysed. This was to ensure the statistical validity of the results.

Nanopore analysis of Y₄K peptide solution was conducted in 2 M KCl solution buffered with 50 mM Tris at pH 8.0, with a potential of +100 mV at the *trans* side of the α HL pore with the *cis* side grounded. This configuration amounted to an effective negative potential at the *cis* side relative to the *trans* side. Under these conditions, in the absence of any peptide, the α HL pore exhibited an open channel current of 191 ± 8 pA. Furthermore, the lysine residue, which has a pKa of 10.5,¹²⁴ was expected to carry a net positive charge and be driven through the pore. Indeed, addition of the peptide fragment solution to the *trans* side, with a final concentration of 1.03 μ M, caused short current blockades. The translocation parameters were derived from the analysis of 3700 translocation events of four independent recordings.

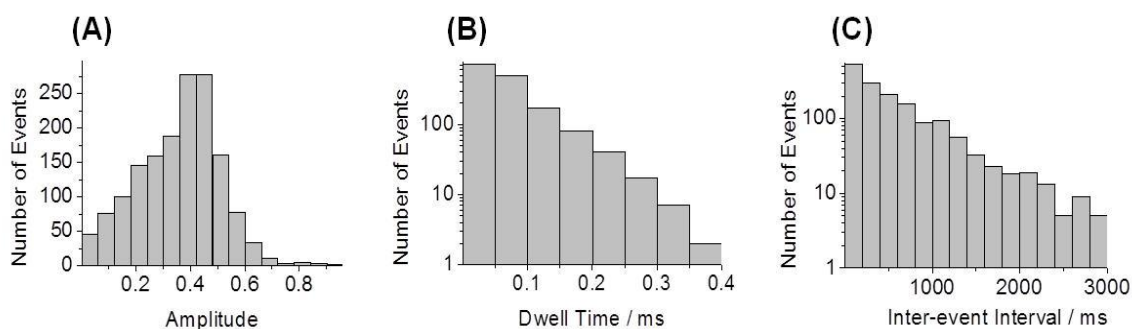


Figure 21: Characteristics of Y₄K-induced pore blockades. (A) Histogram of the amplitude normalized to the open-channel current. (B) Dwell time (τ_{off}). (C) Inter-event interval (τ_{on}). Data from one recording with 1565 events was used to generate this figure [109].

The amplitude was normalised to the open channel current to yield a histogram distribution which was fitted to a Gaussian curve (Figure 21A). The average amplitude was determined as $38.8 \pm 1.9\%$, which is similar to the pore blockades found for peptides in other studies.^{8,125} Thus the result was consistent with the translocation of the Y₄K fragment through the inner constriction of α HL pore. However, the possibility that the peptide temporarily blocked the pore without fully translocating it could not be ruled out. The τ_{off} data was fitted to a single exponential decay and yielded an average of 87 ± 15 μ s (Figure 21B). This value is in line with the timescales for the translocation of similar small peptides through non-modified α HL pores.^{8,125} The final translocation parameter τ_{on} , was

determined as 430 ± 70 ms, which corresponded to a frequency of occurrence of 2.3 ± 0.2 s⁻¹. After normalising for the peptide concentration, the translocation frequency was also found to be within the ranges observed in other studies.^{8,125} The observed translocation events were certainly due to the action of the Y₄K fragment, as control experiments in which the chemically synthesised fragment was analysed, yielded similar results. The results are summarised in Table 4. In particular, the translocation parameters of the enzymatically released fragment fall within experimental error of the chemically synthesised fragment. Furthermore, translocation of the intact, undigested peptide could also be ruled out, as the chemically synthesised peptide yielded characteristically different amplitude of blockade; the blockade of the intact peptide was almost twice that of the released fragment.

Table 4: Translocation parameters for enzymatically released Y₄K peptide fragments and synthesised controls.

Translocation parameter	Source of Y ₄ K fragment		
	Enzymatically released	Chemically synthesised ^[a]	Y ₄ KIG ₄ ^[a]
τ_{off}	87 ± 15 μ s	100 ± 20 μ s	110 ± 11 μ s
τ_{on}	430 ± 70 ms	442 ± 63 ms	512 ± 105 ms
Foo	2.3 ± 0.2 s ⁻¹	2.3 ± 0.3 s ⁻¹	3.2 ± 1.5 s ⁻¹
Amplitude	38.6%	32.8%	58.6%

^[a] Synthesised using standard SPPS.

2.6. Sensing Strategy Applied to the Detection of Renin Activity

The newly developed sensing strategy was applied to the detection of renin activity in the presence of plasma proteins. Akin to the proof-of-principle studies, the PEGA-bound substrate was treated with renin to release a soluble peptide fragment. The sequence of the bound substrate was derived from the first 14 amino acid residues on the *N*-terminus of human angiotensinogen (Table 1). Due to difficulties in isolating and purifying human angiotensinogen, experiments requiring a natural substrate are usually performed with

the tetradecapeptide (Ang0, Table 1).⁸⁶ Investigations have shown it to be a suitable substrate for renin and resistant to cleavage by proteases commonly found in plasma.⁸⁶ A hexahistidine (His₆) tag was placed at the *N*-terminus of the substrate to enable enzymatic digestion and downstream purification to be performed in a single reaction vessel. The sensing approach is summarised in Figure 22. In brief, angiotensin I-His₆ (AngI-His₆) was released as a result of enzymatic activity and concomitantly bound to nickel-nitriloacetic acid beads (Ni²⁺-NTA, N, Figure 22A). After washing the beads to remove residual plasma proteins (Figure 22B), the bound AngI-His₆ was eluted by addition of imidazole (Figure 22C). Finally, the purified peptide solution was subjected to nanopore analysis (Figure 22D).

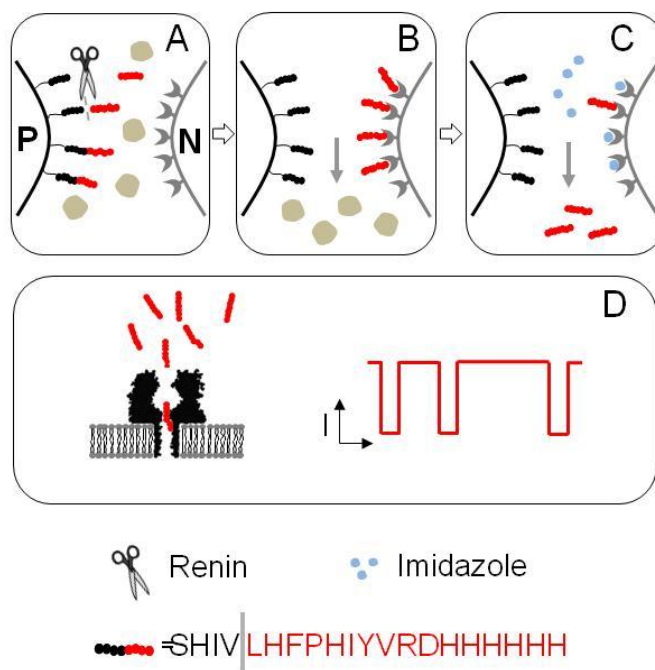


Figure 22: Sensing of renin activity in the presence of plasma proteins. (A) The peptide substrate is immobilized onto PEGA, P, via a flexible linker. Only a single substrate molecule is shown for clarity. AngI-His₆ (shown in red) is released by the action of renin and concomitantly bound by Ni²⁺-NTA, N, beads. (B) The unbound plasma proteins (brown circles) are removed by washing the beads. (C) Bound AngI-His₆ is eluted from Ni²⁺-NTA by the action of imidazole (blue circle). (D) The purified peptide solution is subjected to nanopore analysis. The analyte required for nanopore analysis is generated using two different types of beads within a one-column format.

The PEGA-bound substrate was prepared by MSPPS as described above. However, rather than coupling the peptide directly onto the resin, a cleavable 4-hydroxymethylbenzoic acid (HMBA, Table 3, entry 4) linker was employed. Although preparation of the trypsin

substrate PEGA-Y₃KIG₄ had proceeded smoothly, the lack of a cleavable linker meant that the progress of the synthesis could not be followed. Kaiser test analysis gave an indication of the coupling efficiency; yet, the purity of the peptide could not be determined until enzymatic digestion. The cleavable linker enabled the growing peptide to be cleaved from the resin at periodic intervals during the synthesis, and the purity to be assessed by HPLC or MS analysis. This was particularly important after the coupling of proline as removal of the *N*- α -Fmoc group, resulted in a secondary amine which gave rise to a brown colouration of the beads rather than the characteristic blue in the Kaiser test.¹¹⁸ The peptides were liberated from the resin using a solution of aqueous sodium hydroxide which nucleophilically attacked the ester bond between the peptide and the HMBA linker (Figure 23A). Once synthesis was complete, the loading of the resin was determined as 0.089 mmolg⁻¹ which corresponded to a yield of 48% based on the initial loading of PEGA.

2.6.1. Developing the Sensing Approach: Control Reactions.

The PEGA-bound substrate was initially subjected to renin digestion in the absence of plasma proteins. The reaction was monitored by analytical HPLC over the course of several hours. The resins were washed and equilibrated in the MOPS-NaOH-BSA buffer (pH 8.0) prior to the addition of renin, then incubated at 37 °C. The results are shown in Figure 23. Within 5 min of incubation, a new peak with retention time of 15.36 min emerged (Figure 23B). Using MALDI analysis, this was identified as AngI-His₆. The peak increased in magnitude with longer incubation times, confirming that release was due to the action of the enzyme (Figure 23C). Furthermore, HPLC analysis showed that the enzymatically released peptide co-migrated with the chemically synthesised peptide, further corroborating the identity of the peptide (Appendix 1).

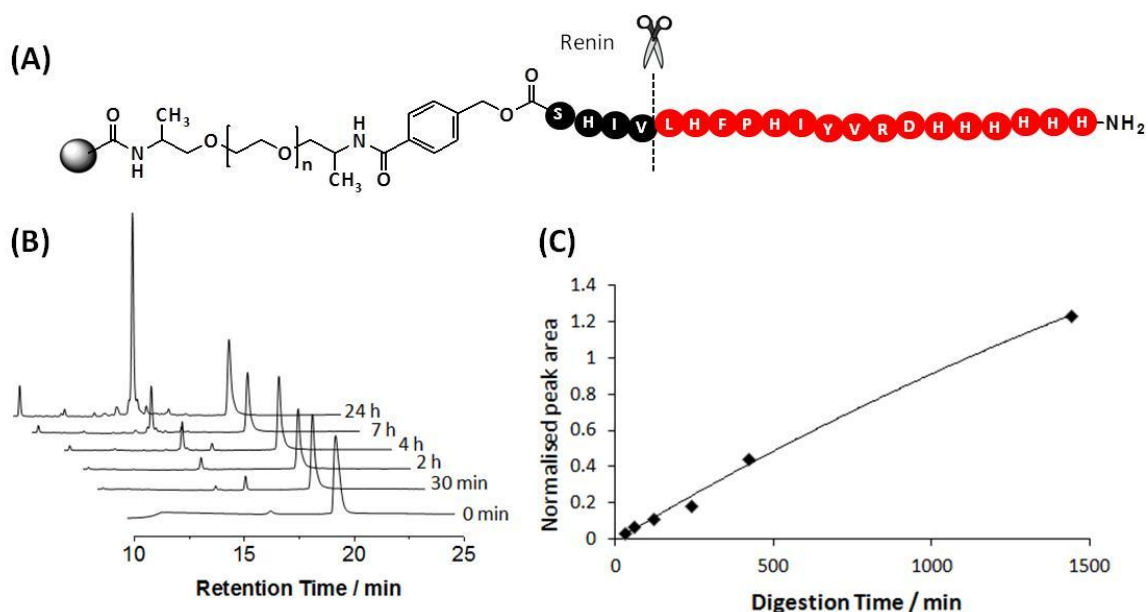


Figure 23: Digestion of PEGA-bound Ang0. (A) Structure of PEGA-HMBA-Ang0-His₆ and renin cleavage site. (B) HPLC trace for the time-dependant enzymatic release of AngI-His₆ peptide from PEGA $t_R = 15.36$ min corresponds to AngI-His₆. $t_R = 20$ min corresponds to BSA present in the buffer. (c) The peak area of the released AngI-His₆ was normalized to that of BSA. HPLC analysis was performed on a Varian Pursuit XRS C18 column and UV absorbance measured at 214 nm.

Control experiments, in which the PEGA-bound substrate was incubated in the absence of renin, were also performed. As expected, the AngI-His₆ was not detected however the reaction mixture did show traces of Ang0-His₆, the undigested intact substrate. This was attributed to hydrolysis of the ester bond between the peptide and the linker by the hydroxyl ions, which were used as a buffering agent. Hydrolytic cleavage was also observed in the presence of other buffers (25 mM sodium acetate and 50 mM Tris-base), but not in plain un-buffered water. These results were surprising, as the use of the HMBA linker in biotransformations has been widely reported,^{126,127} but to our knowledge, hydrolytic cleavage of the ester bond has not been observed. The rate of hydrolytic release from the resin was monitored in a time-course experiment. Using the same experimental conditions as those employed in digestion experiments, aliquots of buffer were removed at regular intervals then subjected to HPLC analysis. With the aid of a calibration curve, the amount of intact substrate released at each time-point was estimated. At 24 h, the amount of hydrolytically released peptide accounted for only 5%

of the amount released enzymatically. Hence for these investigations, the amount of hydrolytically released peptide was deemed to be negligible.

The ability of AngI-His₆ to bind to and be eluted from Ni²⁺-NTA beads was determined using a soluble form of the peptide. A buffered solution of the peptide was added to a suspension of Ni²⁺-NTA beads, which had previously been thoroughly washed and equilibrated. As MOPS contains a tertiary amine which could potentially reduce the Ni²⁺ to elemental nickel,¹²⁸ the experiment was performed in sodium phosphate buffer. After incubation for 24 h, the suspension was filtered, the beads thoroughly washed, and elution achieved by the addition of 250 mM imidazole buffer. The amount of peptide recovered was estimated as 48%. This was extrapolated from the area of the HPLC peak and a calibration curve. The experiment was also performed in the presence of BSA (0.18 mgml⁻¹), the protein required to stabilise renin during digest experiments. BSA was not detected in the eluted solution confirming that the protein did not bind to Ni²⁺-NTA. In terms of the sensing approach, this observation was seen as beneficial as the removal of BSA from the eluted AngI-His₆ solution would result in less background noise in the nanopore traces.

The one-pot reaction using the PEGA-Ang0-His₆ (Figure 22, A-C) was performed in a similar manner to the control experiments. To the PEGA-bound substrate, Ni²⁺-NTA beads were added followed by 5 U of renin. After incubation at 37 °C overnight, the suspension was transferred to a spin column and the supernatant removed. The mixture of beads were washed with water then treated with the imidazole solution. The HPLC trace of the supernatant featured a peak with a retention time corresponding to that of AngI-His₆. The identity of the peptide was confirmed by MALDI-MS. Hence it was shown that AngI-His₆ could be enzymatically released by renin, concomitantly captured by Ni²⁺-NTA and subsequently eluted in a clean fraction. The release of AngI-His₆ was certainly due to the action of renin, as blank experiments in which renin was not included did not yield any peptide fragments, but trace amounts of the hydrolytically released substrate was detected.

In order to simulate *in vivo* conditions, the digestion of PEGA-bound AngI-His₆ was performed in presence of human serum proteins. Plasma is the straw-coloured fluid portion of the blood which is responsible for keeping the cellular components in suspension. It is composed primarily of water and dissolved solutes, but also contains organic molecules such as hormones, enzymes, antibodies and a variety of proteins.⁷⁵ Blood serum has exactly the same composition as blood plasma, but lacks fibrinogen, an anticoagulant. Two problems were anticipated from the use of human serum in our sensing approach. First was the contamination of the eluted AngI-His₆ with serum proteins. As the protein concentration of serum is high, approximately 1 mM,¹²⁹ the beads would need to be thoroughly washed to prevent carry-over and subsequent complications in nanopore analysis. Secondly, human plasma is known to contain histidine-rich proteins which bind to Ni²⁺-NTA.¹³⁰ It was possible that these proteins could be eluted with AngI-His₆ and also cause contamination. In order to achieve good resolution, and hence selective elution, a gradient of imidazole would be needed.

Optimisation experiments were performed to determine the ideal concentration of imidazole to elute AngI-His₆ and the plasma proteins separately. Suspensions of Ni²⁺-NTA were incubated with buffered solutions of AngI-His₆ and plasma proteins. The suspensions were subsequently filtered using a spin column and the beads washed ten times with plain water then treated with increasing concentrations of imidazole. The resultant aqueous washes and elution filtrates from the experiments involving serum proteins were analysed by SDS-PAGE. The wash and elution solutions from experiments involving AngI-His₆ were analysed by and analytical HPLC. The results are shown in Figure 24.

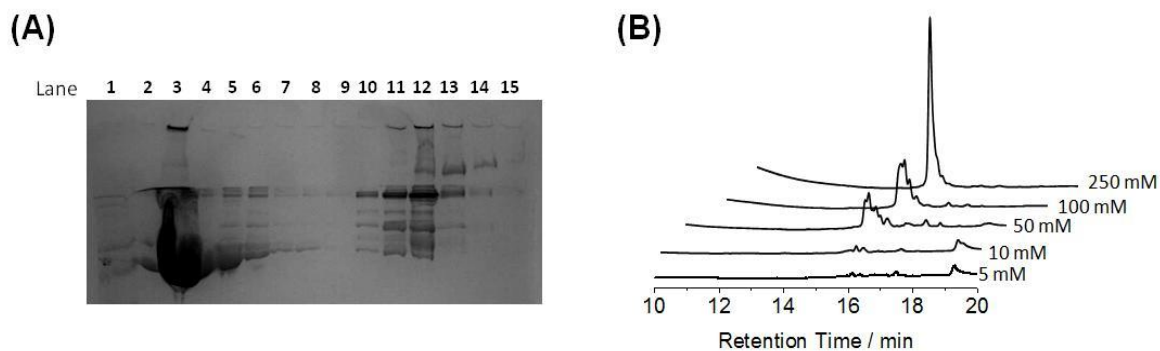


Figure 24: SDS-PAGE analysis of the wash solutions and HPLC analysis of the imidazole elution solutions. After 3.5 h incubation, the supernatant was removed by centrifugation. The Ni²⁺-NTA beads were washed with plain water (10 x 500 μ L) then increasing concentration of buffered solutions of imidazole (500 μ L, containing 50 mM sodium phosphate, 300 mM NaCl, 1-500 mM imidazole, pH 8.1). (A) SDS-PAGE analysis of the wash solutions. Lane 1: Ladder; Lane 2: Blank; Lane 3: Reaction filtrate. Contains the unbound AngI-His₆ and plasma; Lane 4: Blank; Lane 5: Wash 1; Lane 6: Wash 2; Lane 7: Wash 5; Lane 8: Wash 6; Lane 9: 1 mM imidazole; Lane 10: 5 mM imidazole; Lane 11: 10 mM imidazole; Lane 12: 50 mM imidazole, Lane 13: 100 mM imidazole; Lane 14: 250 mM imidazole; Lane 15: 500 mM imidazole. The gel was stained with Coomassie blue. (B) HPLC analysis of the imidazole elution solutions. HPLC analysis was performed on a Varian Pursuit XRS C18 column and UV absorbance measured at 214 nm.

In Figure 24A, Lane 3 corresponds to the reaction filtrate thus contains unbound plasma proteins. The high protein concentration of human serum is indicated by the intensity and the size of the band. The two flanking lanes (Lane 2 and 4) were left empty to accommodate any overspill from Lane 3. A significant amount of protein can be seen in the well at the top of Lane 3. These are most likely proteins which are too large to migrate through the gel. Lanes 5-8 contain wash solutions 1, 2, 5 and 6, respectively. After the beads had been washed six times with plain water, (Figure 24A, Lane 8), almost no protein was visible. Lanes 9-15 contain the imidazole elution solutions. At 1 mM imidazole (Lane 9), almost no His-rich serum proteins were eluted. Elution of the His-rich proteins began to occur at 5 mM imidazole (Lane 10) and the majority occurred at 50 mM imidazole (Lane 12). Between 100-500 mM imidazole (Lanes 13-15), no new His-rich proteins were eluted. The elution solutions of AngI-His₆, were analysed by analytical HPLC (Figure 24B). Akin to the experiment with the serum proteins, the Ni²⁺-NTA beads were washed with plain water then treated with increasing concentrations of imidazole. Between 1-10 mM imidazole, no peptide was detected. Elution began to occur at 50 mM imidazole and the

majority at 250 mM imidazole. No peptide was eluted at 500 mM imidazole.

These results suggested that optimum resolution would be achieved if the beads were first washed with plain water seven times to remove the majority or all of the unbound serum proteins (Figure 24A). A significant amount of the His-rich plasma proteins could be removed by treating the beads with 10 mM imidazole (Figure 24A, Lane 11). Finally, elution of AngI-His₆ could be achieved using 250 mM imidazole. Using this protocol, the eluted AngI-His₆ would most likely contain a small amount of His-rich serum proteins (those eluted at 50 mM imidazole, Figure 24A, Lane 12) however all of the eluted peptide would be recovered. Using the Bradford assay for protein concentration, it was determined that after the seventh aqueous wash, the protein concentration had been reduced by 300-fold. Furthermore, performing four washes with 10 mM imidazole reduced the protein concentration by further 250-times. Hence, the concentration of plasma proteins in the final elution solution was almost 10,000-fold lower than in human serum.

The final control experiment was to determine whether the endogenous enzymes within the human serum were still active. Human serum contains a range of enzymes including renin and angiotensinogenases which could act on the PEGA-bound substrate and skew the ensuing nanopore results. To determine whether this was the case, the PEGA-bound substrate was incubated with Ni²⁺-NTA and human serum without any additional enzyme. The beads were washed and eluted as described above, then subjected to HPLC and MS analysis. In this instance, no AngI-His₆ was detected, suggesting that the endogenous enzymes were not active. As the human serum utilised in this investigation was purchased rather than taken directly from a donor, the most plausible explanation for these observations were the age of the serum and the manner in which it was processed. Specifically, the serum could have been collected then stored for several weeks prior to purchase, during which the enzymatic activity could have waned. This is in contrast to current clinical practises, in which a blood sample is taken and analysed immediately after collection to ensure the activity of the enzyme.⁹² Additionally, purchased serum is sterile

filtered prior to vending to remove any bacteria or fungus which may have contaminated the sample. While, the filter is large enough to enable the passage of proteins including enzymes, if the proteins precipitated or aggregated they would not be able pass through the filter. Purchased serum is also treated with EDTA, which could have sequestered ions essential to the activity of the enzymes, thus rendered them inactive. Once the sensing technique is fully developed, it will be used to measure the activity of endogenous renin. However, so as to avoid any complications, in these early investigations, it was important that any AngI-His₆ generated and subsequent translocations in nanopore analysis were due solely to the added renin.

2.7. Nanopore Analysis

The enzymatically released AngI-His₆ generated using the optimised conditions and a synthesised standard of the peptide, were analysed using the α HL nanopore. The recordings were performed in 1 M KCl solution buffered with 50 mM Tris-HCl at pH 8.5. A potential of +100 mV was applied to the *trans* side of the α HL pore with the *cis* side grounded. Under these conditions, the α HL pore displayed an open channel current of 104.13 ± 8.8 pA ($n = 3$) and the peptide possessed a slight net positive charge.

A synthesised standard of AngI-His₆ was added to the *trans* chamber at a final concentration of 24 μ M. The translocation parameters were derived from the analysis of 7800 events over 4 independent recordings. The peptide was characterised by a τ_{off} value of 0.96 ± 0.18 ms and average amplitude of $83 \pm 1.6\%$. The average τ_{on} was determined as 158.2 ± 12 s which corresponded to a frequency of occurrence of 6.34 ± 0.49 ms⁻¹.

Having fully characterised the synthesised standard, nanopore analysis was used to monitor the enzymatic release of AngI-His₆ from PEGA, over time. As the frequency occurrence gives an indication of the relative concentration of an analyte, the frequency was expected to increase with the enzymatic digestion time. A time-course experiment was performed in which the PEGA-bound substrate was incubated with renin in the absence of Ni²⁺-NTA and plasma proteins. Aliquots of the supernatant were periodically removed, subjected to nanopore analysis and the frequency of occurrence determined.

The results are summarised in Figure 25.

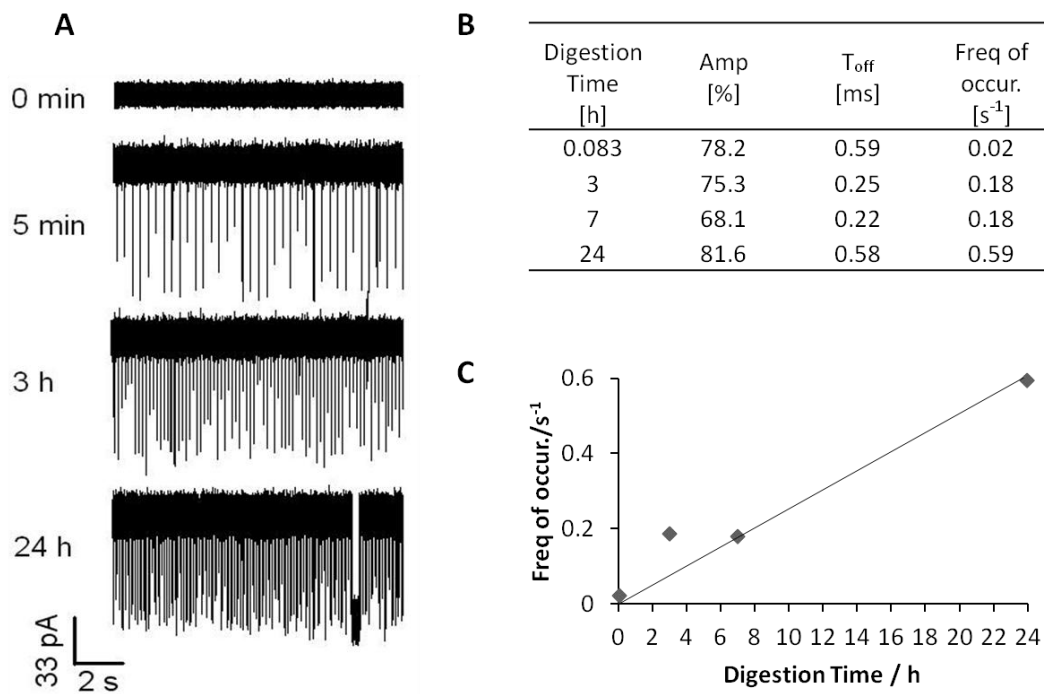


Figure 25: Monitoring enzymatic release of AngI-His₆. (A) Visual inspection of the electrical traces. The frequency of the translocation events increases with time. (B) Summary of the translocation parameters. (C) Plot of frequency of occurrence with digestion time.

In line with the expected results, the frequency of the translocation events increased with time (Figure 25A). A 30-fold increase in the frequency of the translocation events was observed over 24 h (Figure 25B). The amplitudes of blockade and the τ_{off} values were on a par with those of the synthesised standard, providing further evidence that the observed translocations were due to enzymatic release of the peptide. In Figure 25A, the sample at $t=0$ min was the supernatant prior to the addition of the enzyme, thus contained solely the MOPS-BSA buffer. As expected, no translocation events were observed as no peptide fragments were released. The result also showed that the BSA present in the buffer did not interfere with the nanopore recordings. Thus the serum protein contaminants discussed in section above are unlikely affect nanopore analysis. As the change in F_{oo} was readily detected (Figure 25C), this experimental approach represented a viable enzymatic assay.

2.8. Conclusion

In this chapter a new method of electrically detecting enzymatic activity has been presented. The approach involves the enzymatic release of peptides from a solid support followed by nanopore analysis. In previously reported nanopore-based techniques, the electrical signals originated from multiple peptide fragments which complicated data analysis.¹⁴ The presented approach is far simpler as the pore blockades originate from a single peptide. Furthermore, in the absence of any enzyme, no electrical signals are observed which confers enhanced sensitivity. As a non-engineered version of the α HL pore was utilised, the approach can be applied to a wide range of enzymes and substrates. Engineering of the pore would limit the broader applicability of the approach as the modifications would need to be tailored for different substrates. Furthermore, it would be more difficult to extend the approach to solid-state nanopores. Solid-state nanopores are more robust and amenable to commercialisation, however, cannot be engineered to atomic-scale precision. Hence, nanopore-based techniques developed using non-engineered α HL pores can be more readily applied to solid-state pores.

Using the described approach, the kinetics of peptide release mediated by trypsin was determined. A disparity was observed between the rates of peptide release in homogenous (solution-phase) and heterogeneous (on-bead) reactions, with heterogeneous reactions occurring 35% slower than homogenous. This was attributed to the enzyme having difficulty accessing the immobilised substrates. In the future, this could be overcome by employing solid-supports that possess higher surface-to-volume ratios, such as controlled-pore glass, or employing longer linkers.¹¹³ Alternatively, introducing positive charges into the PEGA resin has been shown to improve its swelling properties and thus enhance biotransformations.¹²⁰

The activity of renin enzyme was successfully detected in the presence of human serum. Employing a His-tag analogue of the naturally occurring substrate facilitated downstream purification from plasma proteins, which are known to disrupt the plasma

membrane during nanopore analysis (Stefan Howorka, unpublished data). Purification of the released Ang I-His₆ was optimised so that a minimal amount of plasma protein was applied to the nanopore chamber. The increase in the frequency of peptide release was readily detected with no interference or cross-selectivity from the plasma proteins. Hence the developed approach processes the attributes required for enzymatic testing.

2.9. Materials and methods

All protected amino acids, chlorotrityl and Wang resins used for solid phase peptide synthesis were purchased from Novabiochem Merck Biosciences. *O*-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) and 4-hydroxymethylbenzoic acid (HMBA) were also purchased from NovaBiochem Merck Biosciences. PEGA₁₉₀₀ (poly(ethylene glycol) amide with a molecular weight of 1900 Da) was purchased from Varian Inc. in a slurry of 24% ethanol with a free loading of 0.2 mmol g⁻¹ NH₂ dry resin. All organic solvents and water were purchased from Fischer Scientific. The acetonitrile and water were of HPLC grade quality. 1,3-diisopropyl carbodiimide (DIC), acetic anhydride, 1,8-diazabicycloundec-7-ene (DBU), 2,2,2-Trifluoroethanol (TFE) and *N,N*-diisopropylethylamine (DIPEA) were purchased from Sigma. Piperidine, 1-hydroxybenzotriazole (HOBT), triethylsilane (TES) and trifluoroacetic acid (TFA) were purchased from Alfa Aesar. All buffer components were purchased from Sigma, were of reagent grade or higher and were filtered through a 0.2 micron filter prior to use. QIAshredder microcentrifuge spin columns and Ni²⁺-NTA agarose beads were purchased from Qiagen. Modified trypsin (treated with L-(tosylamido-2-phenyl) ethyl chloromethyl ketone, TPCK) and native human kidney renin was purchased from New England Biolabs and Calbiochem, respectively. Both enzymes were used without any further purification. The purity of the peptides was assessed using analytical HPLC and MS analysis. HPLC analysis was performed using reverse phase Varian ProStar system with a Model 210 solvent delivery mode and a Model 320 dual wave-length detector. The mobile phase A consisted of 0.1% (v/v) TFA in water and mobile phase B consisted of 0.1% (v/v) TFA in acetonitrile. The UV absorbance was monitored at 214 and 254 nm. Analytical HPLC

analysis was performed using a Varian Pursuit XRS C18 column (250 x 4.6 mm). The gradient started with 2% eluent B and rose linearly to 60% B over 20 min with a flow rate of 1 mL min⁻¹. Semi-preparative purification was performed using a Phenomenex Luna C18 column (250 x 10 mm). The gradient started with 2% eluent B and rose linearly to 60% B over 30 min with a flow rate of 5 mL min⁻¹. Fractions containing the product were pooled, the solvent was removed under reduced pressure to approximately 1 mL and the solution lyophilised overnight. ESI-MS analysis was performed using a Waters Aquity Ultra performance LC-MS system equipped with an Aquity UPLC BEH C18 column (50 x 2.1 mm, 1.7 mm beads). MALDI-MS analysis was performed using Waters Micro MX machine. For MALDI analysis, the peptides were concentrated and desalted via aspiration through C18 Ziptip tips (Millipore) and eluted using a 50% ACN solution. The eluted peptides were mixed with α -cyano-4-hydroxycinnamic acid matrix then spotted onto a MALDI plate. The spots were allowed to dry prior to analysis. SDS-PAGE analysis was carried out in a tris-glycine buffer system. Ten microlitre aliquots of each sample were mixed with an equal volume of loading dye. The loading dye was composed of SDS (1 g), Coomassie Blue G-250 (2 mg), and glycerol (3 mL) with 0.5 M Tris-base pH 6.8 (6 mL) then made to a total volume of 10 mL with H₂O. The gels were run in native conditions (4% stacking gel, 12% separating gel) at 200 mV for 1 h. The gels were visualized by Coomassie Blue G-250 staining according to standard protocols.¹³¹ The protein concentration was determined using a Bradford Protein Assay kit from Bio-Rad.

2.9.1. Peptide synthesis

Y₄KIG₄, Y₄K, angiotensinogen, angiotensin I, angiotensinogen-His₆ and angiotensin I-His₆ controls were synthesised using standard Fmoc solid-phase peptide synthesis with HBTU/DIPEA coupling chemistry on a Syro automated system. Y₄KIG₄ and Y₄K were synthesised using Wang pre-loaded with glycine. The angiotensin derivatives were synthesised using pre-loaded chlorotriptyl resins. Angiotensinogen and angiotensinogen-

His₆ were synthesised using chlorotrityl resin preloaded with serine, whilst angiotensin and angiotensin-His₆ were synthesised using chlorotrityl resin preloaded with leucine.

2.9.1.1. Loading of chlorotrityl resin

Fmoc-protected amino acid (1.2 eq) was dissolved in anhydrous DCM (5 mL) and DIPEA (4 eq) added. To the solution, chlorotrityl resin (1.20 mmol g⁻¹ loading) was added and the resultant suspension was agitated at room temperature for 2 h using a Merrifield bubbler. After this time, the suspension was filtered and washed with the following: DCM/MeOH/DIPEA (17:2:1, 3 X 10 mL), DCM (3 X 6 mL), DMF (2 X 6 mL), and finally DCM (2 X 6 mL). The resin was then dried in a desiccator overnight. The loading of the resin was estimated using DBU method.¹³² Briefly, 2 mg of the resin (approx 10 μmol with respect to the Fmoc group) was suspended in 2% DBU/DMF (2 mL). The suspension was shaken at room temperature for 1 h, after which it was diluted to 10 mL with ACN. An aliquot of this solution (2 mL) was diluted to 25 mL with ACN and its absorbance at 304 nm measured. The loading of the Fmoc-Ser(tBu)-Cltr resin was determined to be 0.67 mmol g⁻¹ corresponding to a coupling yield of 56%. The loading of the Fmoc-Leu-Cltr resin was determined to be 0.64 mmol/g corresponding to a coupling yield of 53%.

2.9.1.2. Automated peptide synthesis to produce a fully deprotected peptide

Four equivalents of coupling agent and amino acid were used in the coupling step of 40 min duration. Fmoc deprotection was carried out by shaking the resin in 40% piperidine for 3 min. After completion of the synthesis, the peptide was liberated from resin and simultaneously using a cleavage cocktail composed of TFA/TES/H₂O (95:2.5:2.5) for 3 h. The peptide solution was collected and the peptide precipitated out by the addition of ice-cold diethyl ether. After centrifuging the solution to a pellet, the supernatant was decanted, replaced with fresh diethyl ether and the peptide re-suspended in the diethyl ether. This centrifugation, re-suspension purification process was repeated two additional times. The pellet was then dissolved in the minimum amount of deionised water, frozen on an acetone/dry ice bath and lyophilised overnight. The identity and purity of the peptides was determined by HPLC and MS analysis.

2.9.1.3. Automated peptide synthesis to produce a fully protected peptide

Four equivalents of coupling agent and amino acid were used in the coupling step of 40 min duration. Upon completion of the synthesis, the protected peptide was liberated from the resin using a cleavage cocktail of 20% TFE in DCM for 30 min. The suspension was filtered then resuspended in fresh cleavage cocktail for an additional 30 min. This was repeated once more. The combined filtrates were concentrated *in vacuo* to a volume of approximately 2 mL. The peptide was precipitated from solution with the addition of ice-cold diethyl ether. After centrifuging the solution to a pellet, the supernatant was decanted, replaced with fresh diethyl ether and the peptide re-suspended in fresh diethyl ether. This centrifugation, re-suspension purification process was repeated two additional times. The pellet was dried under vacuum overnight. For analytical HPLC and MS analysis, approximately 1 mg of material was dissolved into DMSO (3 mL).

2.9.1.4. Coupling of protected peptide to PEGA resin

The protected peptide (5 eq relative to the free amines on the resin) was dissolved in the minimum amount of DMSO. To the resultant solution, HOBt (5 eq) which had been dissolved in anhydrous DMF, was added. The resultant mixture was added to the resin followed by DIC (5 eq) to initiate amide bond formation. After 2 h, Kaiser test (see section 2.9.1.6) on a small subset of beads indicated that coupling was complete. The resin was filtered and washed with anhydrous DMF (3 x 3 mL). Fmoc deprotection was achieved by shaking the resin in 20% piperidine/DMF for 1 h. The loading of the resin was determined by spectroscopic analysis of the Fmoc deprotection solution.

2.9.1.5. Manual peptide synthesis

Coupling of 4-hydroxymethylbenzoic acid to PEGA resin

HMBA (3 eq) and HOBt (6 eq) were dissolved into anhydrous DMF (4 mL) and the solution added to the PEGA resin (3.0 g, wet weight). The resultant solution was gently shaken before the addition of DIC (4 eq). The suspension was shaken at room temperature for 2 h after which a small subset of beads were removed, washed with ethanol, then tested for

free amines using the Kaiser test. If the beads tested positive for free amines (a blue colouration of the beads), the reaction was allowed to proceed overnight. If the coupling was not complete after the overnight reaction, the beads were filtered, washed with anhydrous DMF and the reagents replenished. Once the reaction had gone to completion, the beads were filtered then subjected to a series of wash cycles: 1 M NaOH:DMF (1:1 v/v, 10 mL) for 15 min, 3 x DMF:H₂O (1:1 v/v, 10 mL) followed by anhydrous 3 x DMF (10 mL). The filtered resin was stored at 4°C.

Synthesis of immobilised renin substrate (PEGA-AngiotensinogenH₆)

HMBA-PEGA was weighed into a solid phase extraction tube (SPE) and washed with DMF (3 x 10 mL) then re-suspended in anhydrous DMF (10 mL). The first amino acid (3 eq. relative to amount of NH₂ initially on the PEGA resin), DIC (4 eq) and DMAP (0.1 eq) were added to the suspension. After shaking at room temperature for 2 h, the resin was filtered then re-suspended in fresh DMF (10 mL) and the reagents replenished. Completion of the reaction was confirmed by Kaiser test. Once free amines were detected, the resin was washed with DMF (3 x 10 mL), and then acetic anhydride (6 eq) in anhydrous DMF (10 mL) was added to acetylate hydroxyl groups on the resin. The suspension was shaken at room temperature for 1 h after which the resin was filtered and washed with anhydrous DMF (3 x 10 mL). The Fmoc group was removed from the coupled amino acid by incubating with piperidine/DMF (20% piperidine, 10 mL). After shaking for 1 h, the suspension was filtered and the resin washed with more Fmoc cleavage solution (10 mL) followed by anhydrous DMF (3 x 10 mL). A small subset of beads was subjected to the Kaiser test to ensure that the terminal amine was released. The subsequent amino acids were coupled in the same manner, however, the coupling steps were performed only once for 1 h. Furthermore, after the incorporation of proline, Kaiser test for free amines gave rise to a brown colouration of the beads rather than the characteristic blue (due to the secondary amine on proline).¹⁰⁵ Further confirmation for the incorporation of proline was provided by cleaving the peptide from a subset of resin (see below for details) and analysing by MS.

Once synthesis of the peptide was complete, the side chain protecting groups were removed by suspending the resin in a TFA cleavage cocktail (TFA/TES/H₂O, 95:2.5:2.5, 10 mL) initially for 10 min then for 2 h in fresh cleavage solution. The resin was then filtered and washed with H₂O, DMF and DCM (3 X 5 mL each), and dried in a dessicator in preparation for the enzymatic reactions.

In order to determine the purity of the coupled peptide, a small subset of beads were released from the resin by treating with 0.1 M NaOH for 2 h at room. The filtrate was collected, neutralised with HCl and analysed by HPLC and LCMS. The released peptide was found to be >95% pure, and the coupling yield determined using the DBU method described above.

2.9.1.6. Kaiser test

Approximately 2 mg of the resin-bound peptide was washed with ethanol then transferred to a 1 mL tube from Eppendorf. Two drops of each of the Kaiser Test solutions was added and the tube then placed in a heat block at 100 °C for 15 min. The test solutions were composed of: **I** (0.001 M KCN, 2 % v/v pyridine), **II** (5 % w/v Ninhydrin in *t*-BuOH), and **III** (80 % v/v phenol in *t*-BuOH). A blue colour indicated a free amine, whereas yellow indicated a protected amine.

Table 5: Summary of peptides synthesised.

Peptide	Mass [m/z]		HPLC retention Time [min]
	Expected	Observed	
Angiotensinogen	1733.0	1733.8	16.90

Angiotensin I	1296.5	1296.5	16.40
Angiotensinogen-Histag	2555.9	2556.1	15.30
Angiotensin I-Histag	2119.3	2120.3	15.36

2.10. Enzymatic reactions

2.10.1. Tryptic digestion of Y₃KIG₄

For the enzymatic cleavage of peptide tethered on PEGA-beads, 2 mg of dried resin (containing 0.43 mg of peptide substrate assuming a load of 0.16 mmol g⁻¹) was suspended in water (1320 µL). The resin was allowed to equilibrate and swell for 30 min. The trypsin enzyme pellet (20 µg) was dissolved in 20 µL of HPLC grade water to a final concentration of 1 µg µl⁻¹, and 8.6 µL of the solution was added to the resin suspension (enzyme : peptide ratio 1 : 50 (w/w)). The suspension was shaken on a heat block at 37 °C overnight, after which the solution was placed on ice and subsequently stored at -20 °C. The peptide content in the supernatant was analysed via HPLC. The control experiment on the enzymatic cleavage of soluble peptide was conducted by employing a ratio of 1:20 enzyme:peptide. To 1970 µL of deionised water, 20 µL of peptide solution (1 µg µl⁻¹) followed by 10 µl trypsin solution (0.1 µg µl⁻¹). The solution was incubated at 37 °C before being analysed by HPLC and mass spectrometry.

2.10.2. Solution-phase digestion using renin

For the time-course experiments, 4 mg of dried resin was suspended in 50 mM MOPS buffer (1400 µL, pH 6.8) containing 0.18 mg mL⁻¹ BSA. The suspension was equilibrated at 37 °C for 1.5 h with gentle shaking to ensure that the beads had swollen and dispersed. The reaction was started by the addition of renin solution (5U, 0.2U µl⁻¹). At regular intervals, 50 µL aliquots were withdrawn and immediately placed on ice then stored at 4 °C until HPLC, MS or nanopore analysis. Analytical HPLC was used to determine the degree of proteolysis. The peaks area of the released peptide were normalised to the peak area of BSA in the buffer, and a calibration curve used to determine the amount of peptide present.

2.10.3. Solution-phase digestion in the presence of human serum and Ni²⁺-NTA

PEGA-angiotensinogenH₆ (4 mg) was suspended in 1 mL ddH₂O then equilibrated at 37°C for 1 h with gentle shaking. Ni²⁺-NTA slurry (200 µL) were centrifuged (10 min, 0.8 rpm) to remove the storage solution then re-suspended in ddH₂O (1 mL), washed and re-centrifuged. This equilibration and washing process was repeated two additional times. The beads were washed and equilibrated a third time with sodium phosphate buffer (100 mM, pH 7.4, 0.19 mg mL⁻¹ BSA, 1 mL). The washings were removed and replaced with fresh phosphate-BSA buffer (400 µL). The PEGA-angiotensinogenH₆ suspension was centrifuged (10 min, 0.8 rpm), the water removed and replaced with the Ni²⁺-NTA-buffer suspension. To the resultant mixture, human serum (125 µL) was added followed renin solution (5U, 1.25 U µL⁻¹). The resultant suspension was incubated at 37 °C, with gentle shaking at 950 rpm for 18 h. The suspension was transferred to a QiaShredder spin column (the inner homogenising frit removed), and centrifuged (30 s, 0.8 rpm) to separate the supernatant. The beads were washed with ddH₂O (7 x 500 µL), then wash solution (10 mM imidazole, 300 mM NaCl, 50 mM sodium phosphate, pH 7.9, 4 x 500 µL), then treated with elution solution (250 mM imidazole, 200 mM NaCl, 50 mM sodium phosphate, pH 7.9, 4 x 500 µL). The beads were incubated with each wash and elution solution for 1 min, with gentle shaking, before centrifugation to collect each filtrate in a fresh tube. The eluted peptides analysed by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE), HPLC and MALDI MS.

2.11. Nanopore analysis

Single-channel current recordings were performed by using a planar lipid bilayer apparatus as described in [133]. Briefly, a bilayer of 1,2- diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) was formed on an aperture (80 µm in diameter) in a Teflon septum (Goodfellow Corporation, Malvern, PA) separating the *cis* and *trans* chambers of the apparatus. Each compartment contained 2 M KCl, 50 mM Tris-HCl, pH 8.0, unless otherwise stated. Gel-purified heptameric αHL protein (final concentration

0.01–0.1 ng mL⁻¹) was added to the *cis* compartment to achieve insertion of a single channel into the bilayer. Subsequently, peptide-containing samples were added to the *trans* side. Transmembrane currents were recorded at a holding potential of +100 mV (with the *cis* side grounded) by using a patch-clamp amplifier (Axopatch 200B, Axon Instruments, Union City, CA). For analysis, currents were low-pass filtered at 20 kHz and sampled at 50 kHz using a Digidata 1200 A/D converter (Axon Instruments), as described in [134].

3 SIZING NUCLEOTIDE REPEAT SEQUENCES USING NANOPORE TECHNOLOGY

3.1. Summary

The aim of this project was to establish a nanopore-based strategy to size forensically important nucleotide repeat sequences, which are the basis for DNA fingerprinting. Short nucleotide repeat sequences occur throughout the genome. The number and the length of a repeat sequence, within a given region of the genome is characteristic for an individual. To date, the rapid speed with which non-modified single stranded DNA translocates the pore has prevented resolution of individual bases and hence the development of a nanopore-based sensing strategy. It was envisaged that such a strategy could be realized by employing DNA strands modified with nucleotides bearing adamantane tags. The modification could be performed in such a way that a single adamantane tag is incorporated per repeat sequence. It was hoped that chemical modification would increase the steric bulk of the DNA strands to reduce the translocation speed in ensuing nanopore analysis. The resultant unique electrical signatures could be analysed to determine the length and number of repeat sequences and thus discriminate between individuals. A successful nanopore-based sensing strategy would enable the rapid pre-screening of forensic samples in a point-of-care setting, and development of portable devices for use at the scene of a crime. The modified nucleotides were successfully synthesised and incorporated into oligonucleotides using a protocol published in the *Journal of Organic and Biomolecular Chemistry* (*Org. Biomol. Chem.*, 2009, **7**, 3826–3835). However, nanopore analysis did not yield the expected results. Unique electrical signatures were not obtained although a correlation between the number of tags and the translocation speed was observed.

3.2. Introduction

3.2.1. DNA structure

DNA (2'-deoxyribonucleic) is the molecular store of genetic information in nearly all living organisms.¹²⁴ It is a linear polymer composed of phosphorylated monomer units. Each monomer is composed of a pentose sugar (2'-deoxy-D-ribofuranose), a phosphate group,

and a nitrogenous heteroaromatic base.¹²⁴ Four different bases are commonly found in DNA; the two purines adenine (A) and guanine (G) and the two pyrimidines cytosine (C) and thymine (T). The structure of these heterocyclic compounds and the numbering system for the carbon and nitrogen atoms is shown in Figure 26. The pyrimidines have a single ring containing four carbon atoms and two nitrogen atoms. Conversely, purines have a fused pyrimidine-imidazole ring system. Bases are typically coupled to the 1'-carbon of the 2'-deoxyribose sugar via a glycosidic bond giving rise to a nucleoside monomer. Nucleotides are phosphorylated derivatives of nucleosides. The phosphate groups are bound via a phosphodiester linkage to either the 5' or 3' position of the sugar (Figure 26).

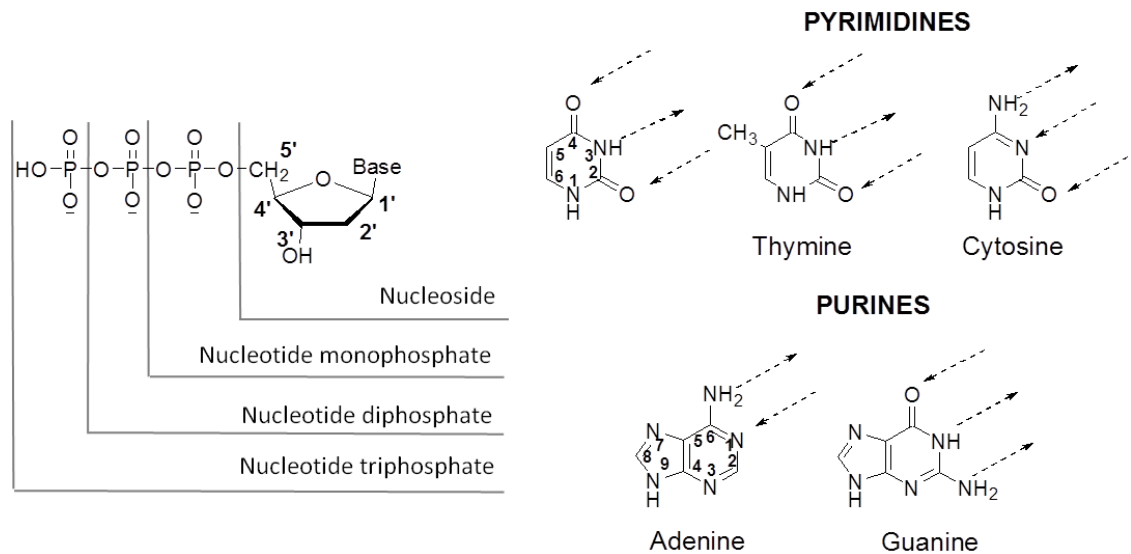


Figure 26: General structure of a DNA nucleoside and the chemical structures of pyrimidine and purine bases. The arrows indicate hydrogen-bonding sites. Each base has hydrogen-donating capabilities, represented by the arrows pointing away from the structures, and hydrogen acceptor capabilities represented by the arrows pointing towards the structure. Adapted from [135].

Individual monomer units are linked together through phosphodiester bonds. Specifically, 5'-phosphate group of one nucleotide is covalently attached to the 3'-hydroxyl group of another. When several nucleotides are linked in this way, the resulting polymer is termed an oligonucleotide. As all the monomers in an oligonucleotide have the same orientation, oligonucleotide chains are said to have directionality (Figure 27). One end of the chain is said to be the 5'-end as the 5'-hydroxyl group is free. The other end of the chain is said to

be the 3'-end as the 3'-hydroxyl is free. Each phosphate group which participates in a phosphodiester linkage has a pK_a of approximately 2. As a result, at near neutral pH, DNA will exist as a polyanion.^{124,136} By convention, the sequence of an oligonucleotide is defined by reading the sequence of the nucleotide bases in the 5'→3' direction.

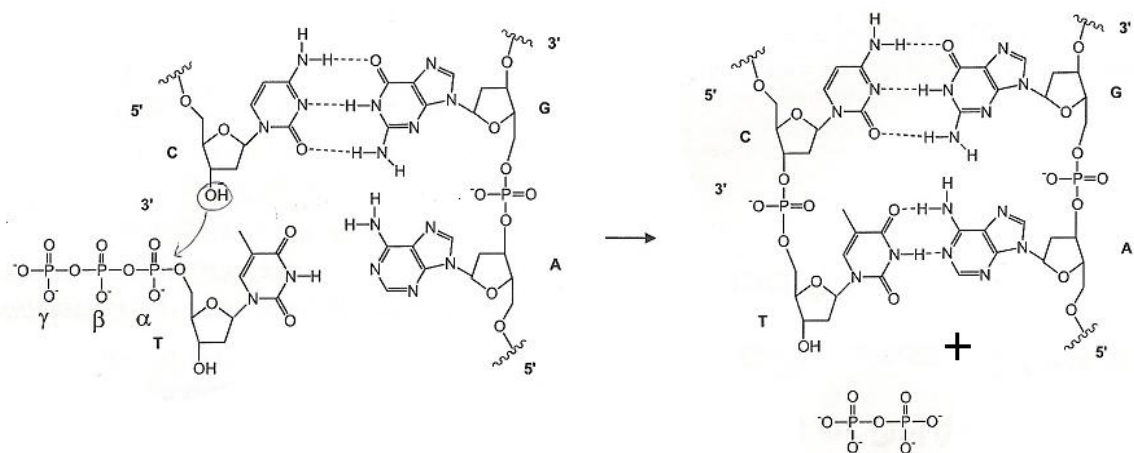


Figure 27: Formation of phosphodiester bond. The hydroxyl group of the growing chain performs a nucleophilic attack on the α -phosphorous of the incoming nucleotide. The DNA duplex is maintained by specific hydrogen bonding pattern between the Watson-Crick base-pair; G•C and A•T. DNA strands are said to be anti-parallel as one strand runs in the 5'→3' and the other in the 3'→5'.

In living organisms, DNA rarely exists as a single oligonucleotide chain, but rather as a pair of oligonucleotides which are tightly associated in a helix, each coiled around a central axis.¹³⁷ The two DNA strands are held together by hydrogen bonding between the base pairs (Figure 27). To maintain the geometry of the double helical structure, a larger purine monomer base-pairs with a smaller pyrimidine. Accordingly, G and C always pair via three hydrogen bonds and A and T via two hydrogen bonds (Figure 27). Hence, the polynucleotide chains are complementary and can serve as templates for each other. The complementary strands are said to be antiparallel as one runs in the 5'→3' direction and the other in the 3'→5' direction. The helical structure of DNA is impart due to the anti-parallel relationship between the strands, but is mainly attributed to the interactions between the base-pairs.^{124,137} In the helix, the base-pairs are stacked one above another and are almost perpendicular to the long axis of the molecule. The non-covalent interaction that exists between the bases is termed the stacking interaction. This

interaction brings the base-pairs closer together, resulting in a hydrophobic interior that causes the sugar phosphate backbone to twist into a helix.^{124,137} DNA typically exists as a B-DNA; a right-handed helix. Due to the manner in which the bases stack and the sugar-phosphate backbone twists, the double helix has two grooves of unequal width. These grooves are called the major and minor groove.

3.2.2. DNA Fingerprinting

DNA fingerprinting describes the technique employed by forensic scientists to identify individuals based on differences within their genetic makeup. Within human cells, DNA is found within the nucleus and organized into threadlike structures called chromosomes. The DNA material within the chromosomes is composed of 'coding' and 'non-coding' regions. The coding regions are known as genes and contain all the information necessary for a cell to make vital proteins. The genes in turn are composed of exons, the protein-coding regions, and introns, the intervening sequences. Identifying and mapping all the genes, was the focus of the Human Genome Project, which announced a final reference sequence for the human genome in 2003.¹³⁸ One of the most surprising findings of the project was that humans have less than 30,000 protein-coding genes rather than the 50,000-100,000 previously thought.^{138,139} Furthermore, it has since been observed that the genetic diversity between individuals is very small. Only a fraction of the DNA sequence (0.3%, approximately 10 million nucleotides) varies between individuals to make them unique.¹³⁹ It is these variable regions of the genome which are exploited for human identity purposes. Methods have been developed to locate and characterize the genetic variation at specific sites within the genome.¹³⁹

The most widely used method for genetic fingerprinting is based upon the detection of nucleotide repeat sequences that occur within introns. Nucleotide repeat sequences are sets of nucleotides that are repeated in tandem. These are found throughout the genome. The number of repeat units and the length of the repeat vary significantly from one individual to the next. These regions are often referred to as satellite DNA.¹⁴⁰ A schematic representation is shown in Figure 28. Thousands of these satellite regions have been characterized and catalogued, as a result, some are now

commonly used as markers for forensic DNA fingerprinting. This typically involves comparing DNA isolated from biological evidence (e.g. blood or hair) found at the scene of a crime, to DNA isolated from a suspect or a victim. The DNA is analyzed for the presence of a specific set of nucleotide repeat sequences (markers) and their lengths determined in order to identify an individual. When the core repeat unit is of medium length (typically between 10-100 nucleotides), this is referred to as a minisatellite or a VNTR (variant number tandem repeat).¹⁴¹

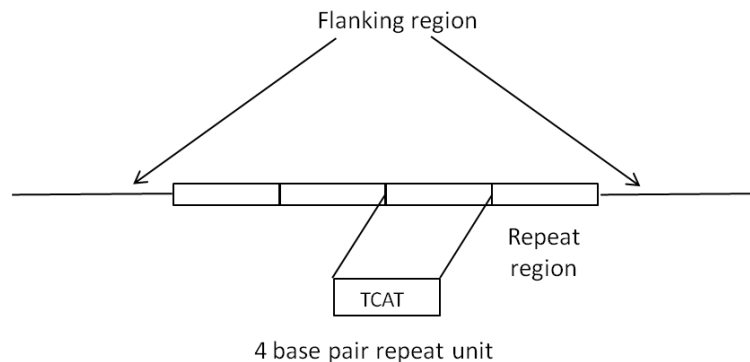


Figure 28 Schematic representation of a microsatellite (STR) DNA marker. PCR primers are designed to target the flanking sequence regions. The number of tandem repeat units in the repeat region varies among individuals making them useful markers for identification. Adapted from [¹³⁹].

These were first exploited as a molecular tool in the early 1980s to develop the first example of genetic fingerprinting called restriction fragment length polymorphism (RFLP).^{142,143} The technique was based upon the ability of restriction enzymes, to recognize and cleave the DNA at sites flanking specific minisatellites. Digestion of genomic DNA with these enzymes generated DNA fragments of varying lengths which were then resolved by gel electrophoresis. The separated fragments were then denatured, that is, rendered single stranded, in the gel by treatment with a strong alkaline solution. A nylon membrane was placed over the gel, and the DNA transferred by capillary action. A DNA fingerprint was obtained after the hybridization of a radiolabeled oligonucleotide probe that was able to bind specific minisatellites by complementary base pairing. The DNA fingerprint was subsequently visualized by autoradiography and compared to that of either the victim or the suspects found at the scene of a crime. By comparing the repeat sequences at multiple positions of the genome, or loci, the power of discrimination

between individuals was greatly increased. However, this approach was extremely labor intensive and interpretation of the data required a great deal of expertise. Furthermore, the process did not easily lend itself to automation. The entire process was time consuming; requiring almost one month to generate each fingerprint.¹⁴⁴ An additional limitation was that its success was highly dependent upon the quality of the DNA sample. The technique could only be applied to whole DNA samples in which the restriction sites were intact. As a result, the technique was often incompatible with DNA samples from criminal cases which are often degraded or contaminated. The demand for rapid turnaround times and high volumes of sample processing soon saw RFLP superseded by techniques based on the polymerase chain reaction (PCR).

PCR is an enzyme catalysed cyclic-process in which a specific region of DNA is replicated repeatedly to yield many copies of a particular sequence.¹⁴⁵ During each cycle, a copy of the target DNA sequence is generated for every molecule of the starting DNA. The boundary of amplification is determined by short oligonucleotides called primers. These are complementary to the DNA sequence flanking the region to be copied. They bind to the target DNA sequence by complementary base pairing, then are extended at their 3'-end by enzymes, in a manner analogous to the biological process of DNA replication. As PCR progresses, the DNA generated is itself used as a template for replication so that the amount of DNA is increased exponentially.^{139,144} PCR-based strategies for amplifying nucleotide repeat sequences have several advantages over RFLP analysis. Firstly, PCR requires only trace quantities of DNA. Typically, 25 ng of DNA is sufficient for PCR compared to 300-500 ng required for RFLP analysis.¹⁴⁴ Thus PCR is ideally suited to forensic DNA analysis where DNA samples are only available in small quantities. Secondly, PCR is able to generate a large quantity of a relatively pure product within a shorter time frame. Finally, degradation of the DNA sample is less of concern as PCR targets smaller segments of the DNA.

The method of DNA fingerprinting used today is based upon PCR amplification of short tandem repeat (STR) regions within introns. STRs or microsatellites, are regions within DNA in which the repeat units are 2-6 base pairs in length (Figure 28). STR repeat

sequences are named by the length of the repeat unit. For example, dinucleotide repeats have two nucleotides repeated in tandem. Trinucleotides have three nucleotides in the repeat unit, tetranucleotides have four, and so on and so forth.¹³⁹ STRs are favored over VNTRs as genetic markers as they occur more abundantly throughout the genome.¹⁴¹ To improve the discrimination between individuals, multiple STRs must be analyzed. The difference in the length of the STRs enables an individual to be identified quite accurately, and the more STR regions analyzed, the more discriminating the test becomes. In the UK, forensic DNA analysis involves the characterization of 11 markers within the genome.¹⁴⁶ These markers can be amplified simultaneously in the same PCR reaction by adding primers which are complementary to the flanking region of each. The ability to amplify multiple regions of DNA simultaneously is termed multiplexing. The advantage of multiplexing is that the cost and time required for DNA fingerprinting can be greatly reduced. The DNA fragments which are produced are considerably smaller than those produced in RFLP, as a result, are more readily separated by electrophoretic methods. Separation and detection of the STR fragments is readily automated, which is an important benefit as the demand for DNA testing has increased. In practice, the STR fragments are labeled with multiple fluorescent dyes through the use of labeled primers. Separation of the fragments is commonly achieved by capillary electrophoresis with concomitant detection of the fluorescent emission signals from the attached dyes. The spectra produced show the fluorescent intensity of each of the fragments versus the time taken for the fragment to reach the detector. Genotyping software is then used to identify and size each of the fragments based on their peak heights and areas. The data is compared to that generated from a control sample, and a DNA fingerprint generated. Using this approach a DNA fingerprint can be generated within a few hours, which is vast improvement over RFLP.

Although PCR and fluorescence-based techniques have become the gold standard for DNA fingerprinting, in recent times there has been a demand for alternative technologies which enable point-of-care analysis. As current techniques require large optical components, DNA fingerprinting can only be performed by centralized laboratories. This

often results in a delay in obtaining results and hence a reduced response time. Furthermore, the current technologies often require a high start-up cost (more than \$50,000), a fact that prohibits its use by some laboratories.¹³⁹ The cost of analysis is also increased by the use of expensive reagents such as fluorophores. It is envisaged that the demand for portable or handheld devices can be satisfied through the electrical detection of STRs using nanopore technology. A nanopore-based strategy would not require costly or large optical components and is more amenable to miniaturisation.

3.3. Aim

The ultimate aim of this project was to apply nanopore analysis to the detection of forensically important DNA samples. A successful strategy could be used to pre-screen DNA samples at the scene of a crime. Due to rapid translocation speeds, it is not possible to discriminate between nucleotides as DNA moves through the pore. Our strategy makes use of chemically modified nucleotides to reduce the translocation speed of DNA strands and takes advantage of the repetitive nature of STRs to count their number. The nanopore sensing strategy is outlined in Figure 29.

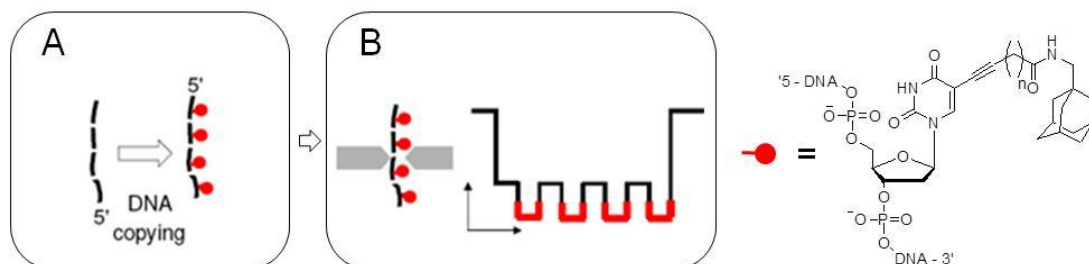


Figure 29: Proposed strategy to size nucleotide repeat sequences. (A) A nucleotide repeat region of DNA is enzymatically copied in the presence of adamantane-modified nucleotides. A single modified nucleotide is incorporated per repeat unit. (B) The resultant DNA strand gives rise to distinct step-like modifications as it translocates the nanopore. The number of repeat units is determined by simply counting the number of step-like modifications. In this example, four tags have been incorporated into the tetranucleotide repeat, resulting in four step-like modifications in the electrical trace.

Akin to the current technique, a copy of the STR will be made using PCR (Figure 29A). However, rather than using the naturally occurring nucleotides, PCR will be performed with an adamantane-modified nucleotide. As a result, the enzymatically copied DNA strands will also be modified or ‘tagged’. The PCR is performed in such a way that a single

modified nucleotide is incorporated per repeat unit. Essentially, the number of chemical tags will mirror the number of repeat units. The number of tags/repeat units is then determined using nanopore analysis by threading the tagged DNA through the pore (Figure 29B). Due to their size, the chemical tags should occlude the ionic current in the pore, giving rise to distinct step-like modifications in the electrical traces each time a tagged repeat translocates the inner constriction of the pore. Determining the number of repeat units will simply entail counting the number of step-like modifications observed in the electrical traces. The non-tagged DNA template will simply pass through the pore with minimal affect on the ionic pore current. With several pores working in parallel, each one analysing a different STR, the throughput will be increased and thus deliver a rapid turnaround time. Furthermore, as the strategy does not require modification of the pore, it can be readily applied to solid-state nanopores which are ideally suited for robust portable sensors. Support for this sensing strategy comes from preliminary work performed within our research group. In particular, a synthetic DNA strand bearing two peptide tags was shown to give rise to distinct step-like current blockades within the nanopore (Figure 30).⁹

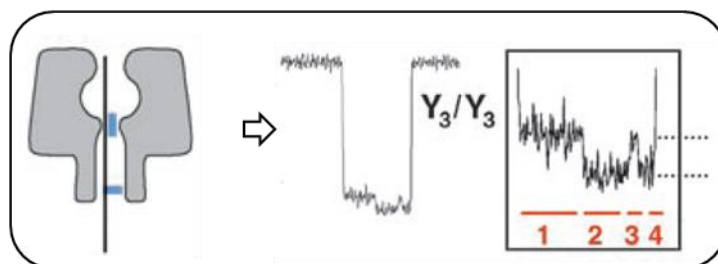


Figure 30: Electrical traces for an oligonucleotide modified with two tyrosine peptide tags. As each tag is sequentially pulled through the nanopore, step-like alterations in the ionic current are observed. Step 1 (red) corresponds to the presence of leading non-modified region of the oligonucleotide strand. As the first bulky tag passes through the inner constriction, the ionic current is reduced (step 2). The intermediate non-modified region of the oligonucleotide causes the ionic current to increase (step 3). Finally, as the second tag translocates the inner constriction, another reduction in the ionic current is observed (step 4). Adapted from [10].

It was anticipated that adamantane would be an ideal tag in our sensing strategy. With dimensions of approximately 4 Å cubed,¹⁴⁷ adamantane would effectively increase the diameter of single stranded DNA to 1.3 nm; close to the diameter of the inner constriction of the α HL pore. This, combined with the rigidity of the structure, would facilitate the

formation of a bulky oligo-tag segment large enough to temporarily block the flow of ions through the pore.

In summary, the first specific aim of this project was to synthesise adamantane-modified nucleotides and assess their incorporation into DNA strands by PCR-based techniques. Secondly, to synthesise synthetic oligonucleotides bearing adamantane-tags and subject them to nanopore analysis to determine whether step-like current blockades could be detected.

3.4. Results and Discussion

3.4.1. Synthesis of Chemically Modified Nucleosides

The structural and functional diversity of DNA can be greatly expanded by employing chemically modified nucleotides as monomer units. Chemical modifications can be introduced directly onto the base, through substitution of the hydroxyl groups on the sugar or substitution of the oxygen atoms of the phosphate groups.¹⁴⁸ In order to maintain the biological function of DNA, chemical modifications should be such that the fundamental structure of the monomer, for example the hydrogen-bonding capability or key functional groups, are unaltered. Yet, there are instances in which it has been advantageous to modify crucial positions of the monomer. For example, some chemotherapeutic agents used to alleviate the symptoms of AIDS are nucleoside/nucleotide analogues which lack or have a modified 3'-hydroxyls.^{149,150} Incorporation of these analogues into the growing 3'-end of a DNA stand prevents the formation of the 3'-5' phosphodiester bond with the next incoming nucleotide. These analogues effectively terminate DNA synthesis and prevent the HIV virus from replicating. If it is not the intention to disrupt the process of DNA replication, modifications are introduced into the base. The most attractive sites for modification are C5 and N7/C8 of pyrimidines and purines, respectively (Figure 26).¹⁵¹⁻¹⁵³ Both positions are located in the major groove of the double helix, thus chemical modification is unlikely to perturb base pairing which is essential in maintaining the structural integrity of the helix.¹⁵¹ Modification of these positions has vastly increased the number of applications of DNA in

molecular biology. For example DNA sequencing,¹⁵⁴ detection of single-nucleotide polymorphisms¹⁵⁵ and fluorescence *in situ* hybridization,¹⁵⁶ are all techniques which make use of chemically modified DNA. In many cases, chemically modified nucleotides bearing functional tags are enzymatically incorporated into DNA strands. This is termed pre-modification as the incorporated nucleotide is already functionalised. The approach is complemented by a two step post-modification strategy in which nucleotide precursors carrying linkers are first enzymatically introduced into the DNA then covalently derivatised with a functional tag. By employing the post-modification approach, a wide range of coupling chemistries can be exploited. For instance, click chemistry¹⁵⁷ Diels-Alder cycloaddition,¹⁵³ maleimide or amide¹⁵⁸ couplings. Furthermore, this approach enables DNA to be functionalised with the large tags which are not readily accepted by polymerase enzymes. With regard to the post-modification approach, the substituents must protrude from the major groove of the DNA duplex into solution to facilitate coupling between the linker and the tag.

Nucleotide analogues bearing non-biogenic substitutions at either C5 or N7/C8 can be chemically synthesised using palladium-catalysed cross-coupling reactions.¹⁵⁹ These reactions are favoured as the mild conditions employed enable the derivatisation of sensitive molecules. For the Suzuki-Miyaura and Sonagashira reactions, halogenated nucleosides are used in combination with substituents carrying terminal aryl/vinyl boronic acids and terminal alkynes, respectively.^{160,161}

The synthesis of the adamantane-derivatised nucleosides and their corresponding nucleotides is outlined in Figure 31.

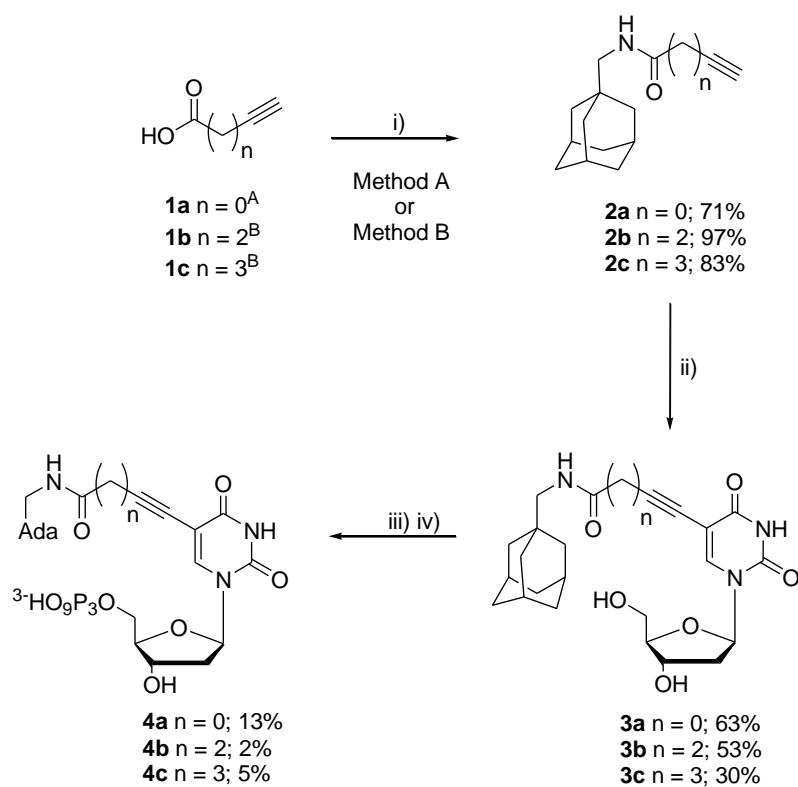


Figure 31: Synthesis of adamantane-modified nucleotides. (i) Adamantane methylamine, then Method A: EEDQ, DMF, 18 h or Method B: HBTU, Et₃N, cat. DMAP, DMF 18 h; (ii) 5-iodo-2'-deoxyuridine, CuI, Et₃N, Pd(PPh₃)₄, DMF, 16 h; (iii) POCl₃, proton sponge, 2 h; (iv) nBu₃NH⁺.H₃P₂O₇⁻, 0.1 M TEAB, 45 min.

In order to determine the effect of spacer lengths in our sensing approach, three different adamantane-modified nucleosides were synthesised (Figure 31, **3a-c**). The adamantane-acetylene moieties (Figure 31, **2a-c**) were generated by standard amide bond formation between adamantane methylamine and the appropriate acetylene-carboxylic acid. The synthesis of **2b** and **2c** was mediated by HBTU, whilst the synthesis of **2a**, was achieved using *N*-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline (EEDQ). The moieties' were isolated in good yields, 71%, 97% and 83%, for **2a**, **2b** and **2c**, respectively. Attempts to synthesize **2a** using HBTU often resulted in no reaction or poor yields; 38% was the highest yield obtained. The use of *N,N'*-dicyclohexylcarbodiimide (DCC), also resulted in similarly poor yields. Specifically, the desired product was isolated in 36% yield whilst the self-condensation by-product, formed as a result of a reaction between the *O*-acylisourea intermediate and another equivalent of the acid, was isolated in 34% yield. In our hands, EEDQ gave the most satisfactory yields with the greatest purity. The self-condensation

product was not detected and the desired product could be isolated in reasonable purity after a simple acidic wash.

The modified nucleosides (Figure 31, **3a-3c**) were prepared by Sonogashira cross-coupling reaction between 5-iodo-2'-deoxyuridine and adamantane-acetylene moieties for a pre-modification approach of DNA. 5-iodo-2'-deoxyuridine was chosen as the model nucleoside as it is considerably cheaper than other halogenated pyrimidines nucleosides. Furthermore, as it lacks an exocyclic amine (as found in cytosine or thymine, Figure 26), the synthetic strategy would be simpler as no protecting groups would be required. The Sonogashira cross-coupling reaction was performed according to the method of Hobbs.¹⁶² Specifically, tetrakis(triphenylphosphine)palladium(0) ($\text{Pd}(\text{PPh}_3)_4$) was employed as the catalyst and the reaction performed in DMF in the presence of triethylamine (Et_3N) and copper iodide (CuI). Importantly, a ratio of 2:1, $\text{Pd}(0)$ to $\text{Cu}(I)$ was employed to ensure that the reaction proceeded at maximum efficiency.¹⁶² Furthermore, when a higher ratio of Pd was used, the reaction mixture turned black within 1 h and lower yields were obtained. Purification of the crude material by flash chromatography was complicated and the product was often contaminated with catalyst. Conversely, when a ratio of 2:1 was used, the reaction mixture remained a yellow/green colour for almost the entire duration of the reaction (on average 16 h). As $\text{Pd}(\text{PPh}_3)_4$ is light and moisture sensitive, the reaction vessels were wrapped in aluminium foil and thoroughly dried, and the DMF degassed with argon prior to the reaction. Exposure of the reaction mixture to air or moisture caused the oxidation of the Pd which in turn resulted in a lower yield. Taking all of the above precautions, the modified nucleosides were obtained in yields of 63%, 53% and 30% for **3a**, **3b** and **3c**, respectively.

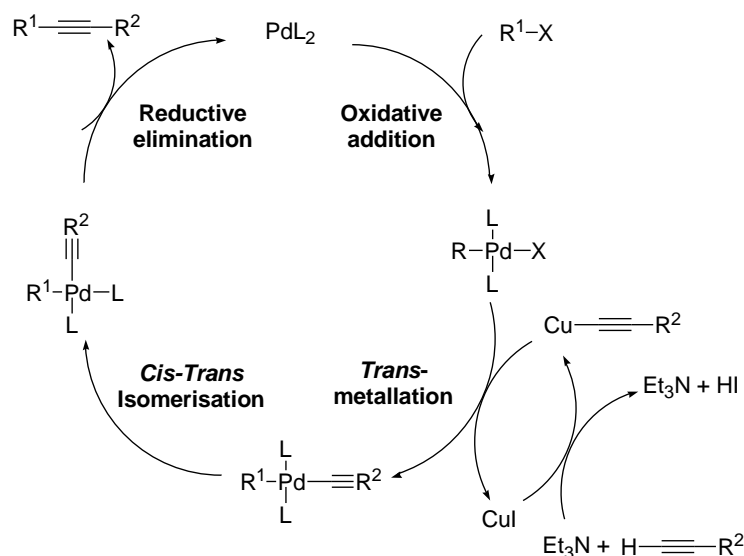


Figure 32: Mechanism of Sonogashira coupling. The Pd(0) undergoes oxidative addition to the nucleobase (the aryl halide) and in the process is oxidised to Pd(II). The addition of CuI as co-catalyst gives better results as CuI activates alkynes by forming a copper acetylide, which is more reactive and undergoes transmetalation with the palladium complex. The ligands attached to the Pd are converted from the *trans* to the *cis* isomers in an isomerism step. Finally, reductive elimination occurs to afford the product and regenerate the catalyst.

The identity and the purity of the nucleosides were confirmed by NMR analysis. The ¹H spectra of each of the products were similar to that of the starting materials minus the terminal alkynyl protons. The alkynyl protons gave signals between 2-4 ppm in the starting adamantane-acetylene moieties, however the corresponding signals were absent in NMR traces of the modified nucleosides. A similar observation was made in the ¹³C NMR spectra. The terminal alkynyl carbons were found between 69-76 ppm in the starting adamantane-acetylene moieties, but in the modified nucleosides the corresponding carbon was shifted further downfield to between 74-88 ppm. The C5 carbon was also used as a diagnostic signal. The C5 carbon in the starting iodouridine gave a signal at approximately 69 ppm. By comparison, the C5 carbons in the modified nucleosides shifted further down field to 97-101 ppm.

NOESY NMR analysis was able to give an indication of conformation of the nucleoside about the glycosidic bond. In solution, the base of a nucleoside is able to rotate about the glycosidic bond which enables it to occupy either of two principle orientations.¹³⁶ These are shown in Figure 33. In the *anti* conformation, the H6

(pyrimidine) or H8 (purine) atoms lie above the sugar ring, whereas in the *syn* the O2 (pyrimidine) or N3 (purines) lie above the sugar.¹³⁶ Typically, the naturally occurring nucleosides occupy the *anti* conformation which enables them to form hydrogen bonds to their complementary base-pair in the double helix.^{163,164} In pyrimidines, preference for the *anti* conformation has been attributed to favourable interactions between the molecular orbitals of the H5-H6 double bond and the lone pairs of electrons of the sugar O4'.¹⁶⁵ As the sugar ring is not planar, but puckered out of plane, the *syn-anti* conformation can be elucidated from the NOE interactions between H6, H1', H2' and H3' protons. A *syn* conformation is expected to result in an NOE interaction between H1' and H6, whilst an *anti* conformation is expected to result in an interaction between H6 and H2'/H3' (Figure 33).^{136,166}

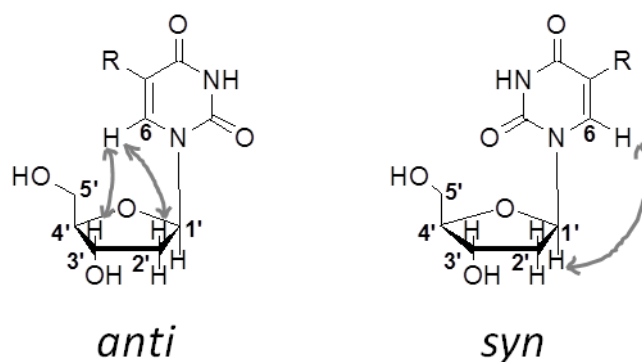


Figure 33: Possible conformation of nucleosides and diagnostic NOE interactions.

NOESY analysis for each of the modified nucleosides showed that interactions occurred between H6 and the H1', H2' and H3' protons (Appendix 2). The result suggested that, at times, H6 came in close proximity with each of the protons, reinforcing the notion of free rotation about the glycosidic bond. However, the major interaction was between H6 and H2'. This interaction was 4-timesⁱⁱⁱ larger than the interaction between H1', H2' and H3' suggesting that the modified nucleosides preferentially adopted the *anti* conformation, akin to the starting 5-iodo-2'-deoxyridine.¹⁶³ Consequently, it was assumed that if the nucleosides were well tolerated by the polymerase enzyme, they should be able to hydrogen bond to their complementary base-pair without significantly altering duplex

ⁱⁱⁱ NOESY analysis performed with the help of Dr. Abil Aliev, UCL Chemistry.

stability. Conclusive evidence about the conformation could be provided by circular dichroism or x-ray crystallography.

3.4.2. Phosphorylation of Nucleosides

To be of use in our sensing strategy, the modified nucleosides were converted to their corresponding triphosphate derivatives. This would enable them to serve as substrates for the DNA polymerase (DNA pol) enzyme used to replicate the forensically important DNA fragments (Figure 29). Phosphorylation of nucleosides is a non-trivial and complicated process for several reasons.^{167,168} Firstly, as the synthesis involves the use of a combination of ionic and hydrophobic reagents, finding the appropriate reaction media which can solubilise both classes of reagents, can be challenging. Secondly, purification procedures require the separation of ionic, hydrophilic and hydrophobic species. Additionally, both the glycosidic and the phosphodiester bonds are susceptible to hydrolytic cleavage; the rate of which is accelerated by acidic and basic media, and elevated temperatures.

Among the many published phosphorylation approaches, the adapted Yoshikawa procedure is a popular method for generating nucleotides.^{136,167} The original Yoshikawa procedure,^{169,170} involves reaction of the nucleoside with phosphorous oxychloride (POCl_3) to yield the highly reactive phosphorodichloridate intermediate (Figure 34, **6**). The intermediate is then hydrolysed using a trialkylammonium bicarbonate buffer to afford the monophosphate (Figure 34, **8**). In a slight adaptation, the phosphorodichloridate (Figure 34, **7**) intermediate can be used directly in reactions with a pyrophosphate salt to yield a nucleotide triphosphate.^{167,171} The Yoshikawa procedure typically makes use of trimethyl- or triethylphosphate as the solvent. These are favoured for their ability to solubilise nucleosides, nucleotides and reagents to give homogenous reaction mixtures.^{167,169} Furthermore, they have been shown to accelerate the rate of phosphorylation.¹⁶⁹ It is thought that this occurs through the formation of a highly reactive complex with POCl_3 .^{169,172} In a further modification of the original phosphorylation procedure, 1,8-bis(dimethylamino)naphthalene (Proton-Sponge™, Sigma-Aldrich)¹⁷³ is

often added to the reaction to neutralise the liberated hydrogen chloride (HCl) which ordinarily would facilitate degradation of the nucleotide.¹⁷⁴

Within our research group, the Yoshikawa procedure was investigated with the view of improving the regioselectivity and yield.¹⁵³ Published protocols typically called for the use of an unprotected nucleoside with POCl₃ for up to 14 h,^{169,174} followed by incubation with a pyrophosphate salt. Under these conditions, phosphorylation occurred predominantly at the 5'-hydroxyl, however the regioselectivity was not perfect and in our hands, significant amounts of the 3'-phosphate and 3',5'-diphosphates were observed. While 3'-protection strategies could be enlisted to address this issue, the use of unprotected starting material was seen as key to minimising the complexity of the synthesis. The unsatisfactory results prompted a detailed investigation into the phosphorylation of unprotected nucleosides as a function of reaction duration, temperature and molar excess of POCl₃.¹⁵³

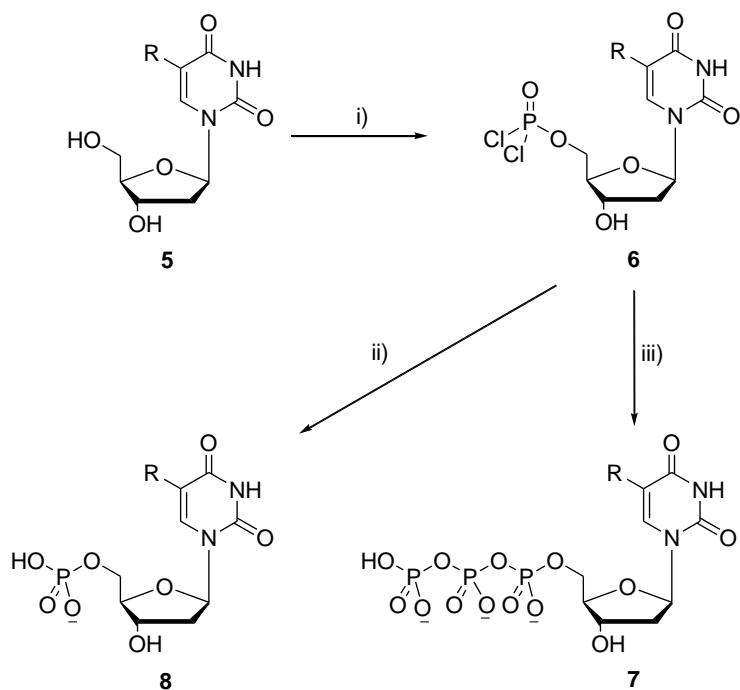


Figure 34: The Yoshikawa phosphorylation *via* the phosphorodichloridate intermediate yields mono- and triphosphorylated nucleotides. (i) proton sponge, POCl₃, (Me)₃PO; (ii) TEAB; (iii) *n*Bu₃NH⁺·H₃P₂O₇⁻, TEAB. R=H or acetylene linker. Where R = H or a chemical tag.

Using 5-iodo-2'-deoxyuridine as a model nucleoside, the formation of the monophosphate, which is crucial to the success of triphosphorylation, was monitored by

analytical HPLC. The nucleoside was reacted with 1.5 equivalents of POCl_3 at $0\text{ }^\circ\text{C}$, similar to published conditions.¹⁵² At regular intervals, aliquots of the reaction mixture were withdrawn, treated with triethylammonium bicarbonate buffer (TEAB) to hydrolyse the phosphorodichloridate intermediate, and then subjected to HPLC analysis. Surprisingly, the 5',3' diphosphate (Figure 35, peak A) and 3'-phosphate derivatives (Figure 35, peak C) were observed after just 3.5 min of reaction. Continued incubation led to further increases in the desired 5'-monophosphate and the diphosphate, but no significant changes in the 3'-phosphate. Kinetic analysis found that a reaction time of between 3 and 7 min was ideal to obtain high levels of the desired monophosphate and low levels of the by-products.^{iv} This short time scale was in stark contrast to published protocols.¹⁶⁹ Investigations into reaction temperature revealed that conducting the reaction at $-13\text{ }^\circ\text{C}$ as opposed to $0\text{ }^\circ\text{C}$, also significantly reduced by-product formation. Furthermore, employing a ratio of 1:1 POCl_3 to nucleoside, completely avoided the formation of any by-products but this was at the expense of decreased yield of the desired 5'-monophosphate. Conversely, increasing the ratio to 3:1 POCl_3 to nucleoside, neither affected by-product nor product formation. Accordingly, the optimum conditions were determined as 1.5 equivalents of POCl_3 , reaction time of less than 10 min and a reaction temperature below $0\text{ }^\circ\text{C}$. The optimised conditions were successfully applied to a range of modified nucleosides (Figure 37) to obtain their nucleotide derivatives in 33-73% yield.

^{iv} Experiment performed by Dr. Vinciane Borsenberger

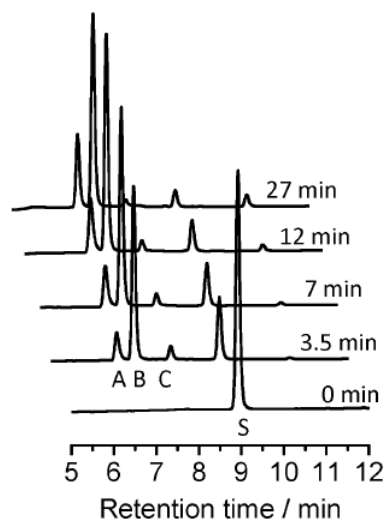


Figure 35: HPLC traces for the reaction of 5-iodo-2'-deoxyuridine with 1.5 equivalents of POCl_3 at 0°C . Peak A: 5',3'-diphosphate; Peak B: 5'-monophosphate; Peak C: 3'-phosphate; Peak S: Starting nucleoside. Y-axis: Absorbance at 260 nm. Adapted from [153].

The optimised Yoshikawa procedure could not be directly applied to the synthesis of the adamantane modified nucleotides (Figure 31, **4a-c**). In particular, formation of the initial phosphorodichloridate intermediates required careful monitoring by either TLC or analytical HPLC to ensure successful triphosphorylation. Figure 36A shows HPLC traces for the reaction of **3b** with POCl_3 . Aliquots of the reaction mixture were withdrawn at regular intervals, treated with 0.1 M TEAB to form the monophosphate, and then subjected to analytical HPLC analysis. The traces reveal that maximal formation of the monophosphate was achieved within 2 h as opposed to 10 min which was observed in the optimisation experiments (Figure 36A). A plausible explanation for this is that the 5'-hydroxyl of the nucleosides are sterically hindered by the adamantane functional groups and thus less available for reaction. As the other optimised conditions were adhered to (1.5 equiv. POCl_3 and temperature below 0°C), minimal amounts of the 5',3'-disphosphorylated or 3'-phosphate, were observed. Final transformation to the triphosphates (Figure 31, **4a-c**) was achieved by the addition of a pre-mixed solution of tributylamine and tributylammonium pyrophosphate in DMF. The reaction was stopped after approximately 45 min by the addition of 0.1 M TEAB buffer. After washing with organic solvent, the crude material was

initially purified by preparative HPLC to remove traces of the pyrophosphate salt and the proton sponge (Figure 36B).

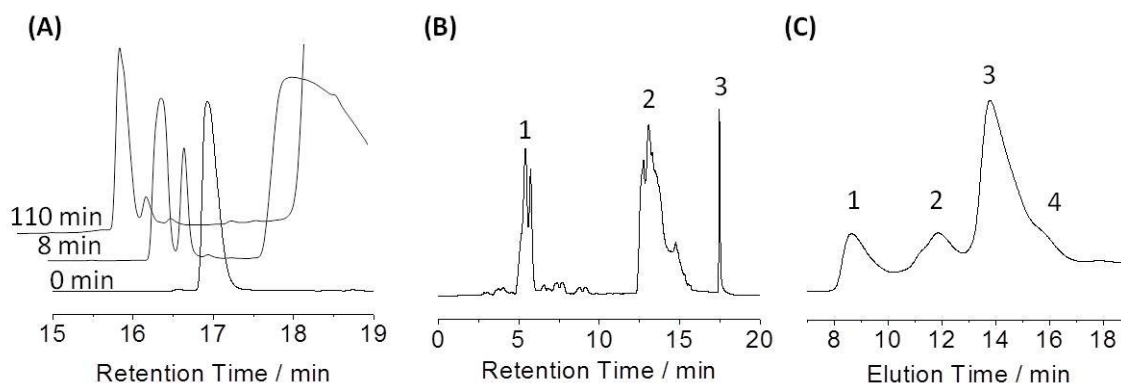


Figure 36: (A) HPLC traces showing the formation of the phosphorodichloridate intermediate of **3c**. Maximal formation occurs within 2 h. The peak between 17.5 and 19 min is due to the proton sponge. Analytical HPLC analysis was performed using a Varian Pursuit C18 column. See Materials and Methods for details. (B) HPLC trace showing the crude reaction mixture for the triphosphorylation of **3c**. Peak 1: pyrophosphate salts; peak 2: a mixture of mono-, di-, tri- and tetraphosphorylated derivatives of **2b**; peak 3: proton sponge. Preparative HPLC was performed using a Polaris C18 column. See Materials and Methods for details. (C) Representative ion exchange trace. Peak 1, 2, 3 and 4 correspond to the mono-, di-, tri- and tetraphosphorylated nucleosides, respectively. Performed using a Resource 5 anion exchange column packed with methyl sulfonate (strong cation exchanger). See Materials and Methods for details.

Further purification was achieved by ion exchange chromatography using a linear gradient of 0.1–0.8 M TEAB for elution. A typical ion exchange trace is shown in Figure 36C. Peaks 1, 2, 3 and 4 correspond to the mono-, di-, tri- and tetraphosphates species, respectively. Fractions containing the nucleotide triphosphates were pooled and concentrated *in vacuo* to give **4a**, **4b**, and **4c** (Figure 37) as triethylammonium salts in 12%, 2% and 5% yields, respectively. Given that for each nucleotide, formation of the phosphorodichloridate intermediates went to near completion, the isolated yields of the triphosphates were surprisingly low. This was attributed to loss of material during ion exchange chromatography.

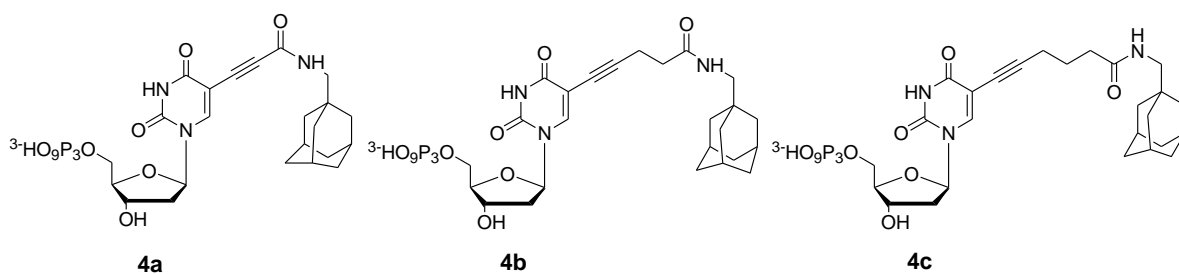


Figure 37: Structure of adamantane-modified nucleotides

In an attempt to improve on the isolated yields of **4a-c**, another approach towards the synthesis of the nucleotides was investigated. As formation of the initial phosphorodichloridate intermediates occurred with relative ease and near quantitative yields, it was hypothesised that an approach, in which the nucleoside monophosphates were isolated en route to the nucleoside triphosphate, would be higher yielding. The phosphorylation procedure based on the morpholidate chemistry of Khorana *et al.*, is one such approach.^{175,176} In this approach, the nucleoside monophosphate is isolated and converted to the phosphoromorpholidate. Triphosphorylation is achieved via a tetrazole-catalysed condensation between a pyrophosphate salt and the activated nucleoside.¹⁷⁷ The validity of the approach was assessed using 5-iodo-2'-deoxyridine as a model nucleoside. The synthetic scheme is outlined in

Figure 38.

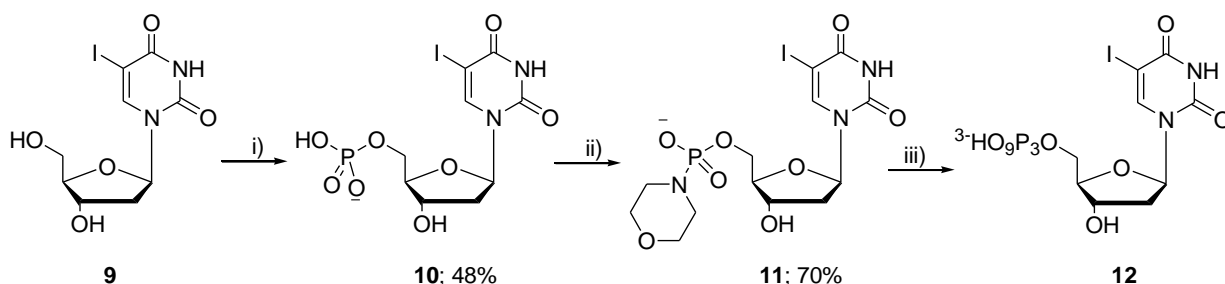


Figure 38: Triphosphorylation using morpholidate chemistry. (i) POCl_3 , proton sponge, $(\text{MeO})_3\text{PO}$; then 0.1 M TEAB; (ii) Morpholine, 2,2'-dipyridyl disulfide, PPh_3 , DMSO; then NaI (iii) tetrazole, $n\text{Bu}_3\text{NH}^+\cdot\text{H}_3\text{P}_2\text{O}_7^-$; then 0.1 M TEAB.

5-Iodo-2'-deoxyridine-5'-monophosphate, **10**, was synthesised in 48% yield using the Yoshikawa procedure described above. The progress of the reaction was monitored by analytical HPLC, which suggested that near complete conversion was achieved within 2 h.

Again, the moderate isolated yield was attributed to the loss of material during ion exchange chromatography. The phosphormorpholidate **11** was prepared in 70% yield via an oxidation-reduction condensation reaction. The proposed mechanism is outlined in Figure 39. It is thought that triphenylphosphine nucleophilically attacks the 2,2'-dipyridyl disulfide generating a arylthiophosphonium-thiolate salt **13** and releasing pyridine dithione.^{178,179} The salt is in turn attacked by the monophosphate resulting in intermediate **14**. The nucleoside monophosphate is subsequently primed for nucleophilic attack by morpholine, resulting in the phosphoromorpholidate **11**. The driving force for the production of **11** is the generation of triphenylphosphine oxide. The phosphoromorpholidate was precipitated from solution as the sodium salt by the addition of sodium iodide in acetone, and used without any further purification. The reaction was catalysed by 1*H*-tetrazole which is thought to activate the monophosphate by protonation of the leaving nitrogen and concomitant nucleophilic catalysis via the highly reactive phosphotetrazolide **15** (Figure 39). It is this species which reacts with the pyrophosphate salt.^{168,177} In this instance, tetrazole serves as both a Bronsted acid ($pK_a = 4.9$)¹³⁶ and a nucleophile. Studies have shown that both properties are required for effective catalysis, as the use of a nucleophile (e.g. dimethylaminopyridine) or acid (e.g. acetic acid, which has a similar pK_a), alone was inferior to the use of tetrazole.¹⁷⁷ Disappointingly, only trace amounts of the desired triphosphate, **12**, formed (detectable by ion-exchange chromatography and MS analysis). As a result, the morpholidate approach was not applied to the adamantane-modified nucleosides rather, the Yoshikawa procedure was used.

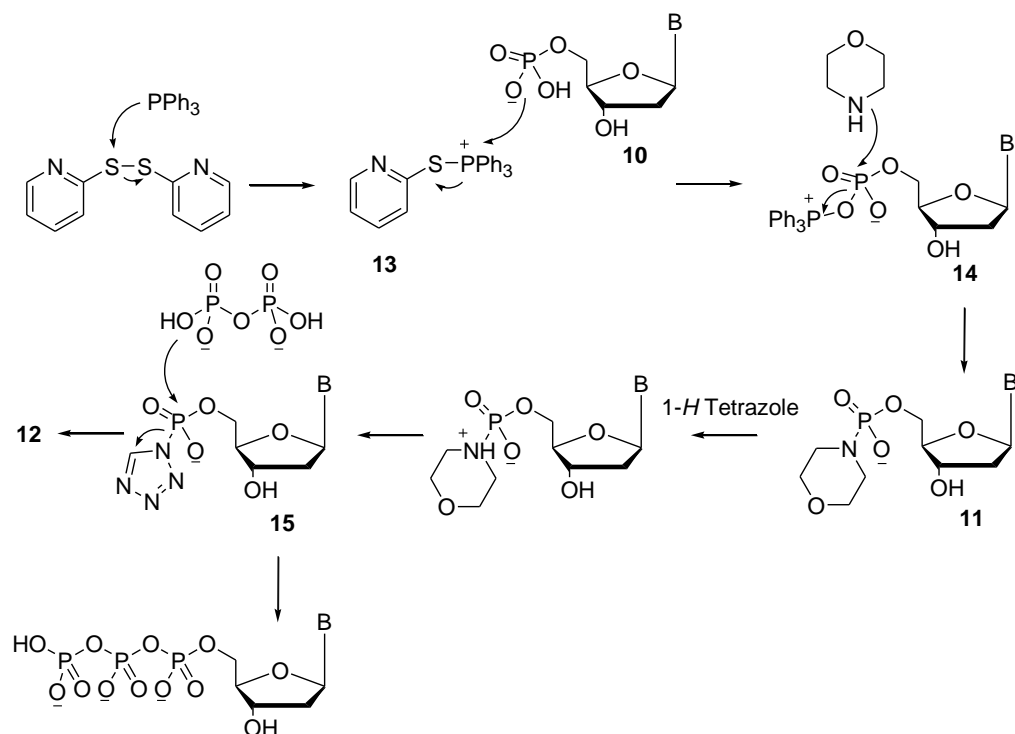


Figure 39: Proposed mechanism for triphosphorylation via a morpholidate chemistry. B: 5-iodouridine

3.4.3. Enzymatic Incorporation of Modified Nucleosides

The ability of the chemically modified nucleosides to serve as a substrate for DNA polymerase was evaluated using template-directed primer extension assays. The principle of the assay is schematically summarised in Figure 40A. A primer, a short DNA strand, is hybridised to complementary section of a longer DNA strand. The unhybridised section of the longer DNA strand then serves as a template to direct the extension of the primer. This extension is mediated by a DNA pol enzyme which sequentially adds nucleotides to the 3'-end of the primer. In the presence of all four nucleotides, the primer is extended to the same length as the template. However, if at least one of the bases is lacking, primer extension is terminated prematurely and giving rise to short fragments. The sequence of the unhybridised section of the template used in this investigation is shown in Figure 40.

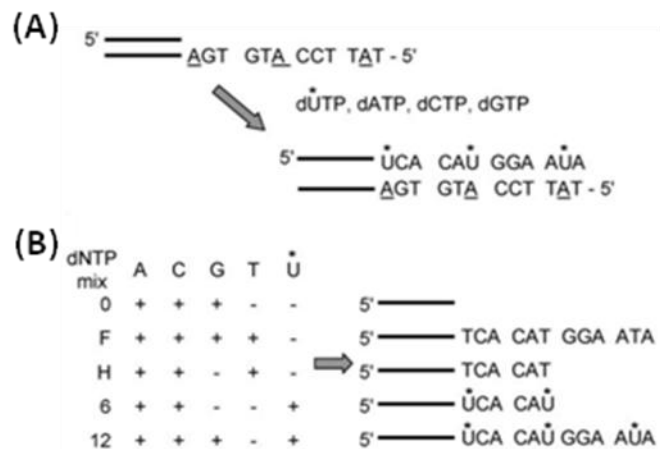


Figure 40: Template-directed primer extension. (A) The primer and template form a duplex, and the non-hybridised section of the template directs the polymerase-catalysed addition of nucleotides to the 3' of the primer. (B) dNTP mixes of varying composition determine the number of incorporated nucleotides comprising the nucleotide analogue dU*TP. Adapted from [153].

As the template contained three adenine bases (underlined in Figure 40A), the primer could be extended by incorporating up to three modified deoxyuridine residues, dU*TP. The sequence of the template was designed so that the extension could be controlled by the inclusion or omission of particular bases (Figure 40B). For instance, dNTP mix 0 lacked a uridine and thymine, either of which could be the first nucleosides incorporated into the primer. As a result, no extension was expected to occur for dNTP mix 0. Conversely, as dNTP mix F contained a full set of natural bases, the primer should be extended by 12 nucleotides to the same length as the template. By comparison, as dNTP mix H lacked 2'-deoxyguanosine triphosphate, dGTP, the primer should only be extended halfway i.e. by six nucleotides. In this fashion, the ability of the enzyme to catalyse the extension of the primers by six or 12 nucleotides, in the presence of the nucleotide analogues, was assessed. The primer extension products were evaluated by gel electrophoresis. In this type of analysis, the extension products are applied to wells at the top of a gel, and an electric field is applied. The fragments migrate towards the anode at different rates depending on their molecular weight, or specifically, according to how far they have been extended. Hence shorter fragments, which were not fully extended, move faster and further than those that are. The gels are visualised using a DNA-specific stain to determine the incorporation efficiency, that is, the ability of the enzyme to extend the primers. For

instance, if the enzyme is able extend the primer to the same length as the template; both stands should migrate to the same position in the gel giving rise to a single bright band. Thus it could be concluded that the enzyme has a good catalytic efficiency. In cases where short extension fragments are produced, two or more distinct bands can be observed. Typically, the higher band corresponds to the template whereas the lower bands correspond to the extension fragments, as these have lower molecular weights. In this instance, the bands would most likely be fainter in appearance as there are fewer oligonucleotides migrating at either position. In most cases, incorporation of the natural and modified nucleotides gives rise to clearly defined bands. However, if the extension products are poorly resolved the bands may appear as smears.¹⁸⁰

Prior to evaluating the substrate properties of the adamantane-modified nucleotides, a systematic study on the effect of C5 substituent on the enzymatic incorporation, was performed. A series of C5-modified 2'-deoxyuridine triphosphates, which possessed different functional substitutions, were synthesised via Sonogashira cross-coupling using the corresponding acetylene-based linkers and Yoshikawa reaction, as described above. The modifications included amine and carboxyl functionalities, alkyne and diene groups, as well as bulky Diels-Alder coupled peptide tags (Figure 41).

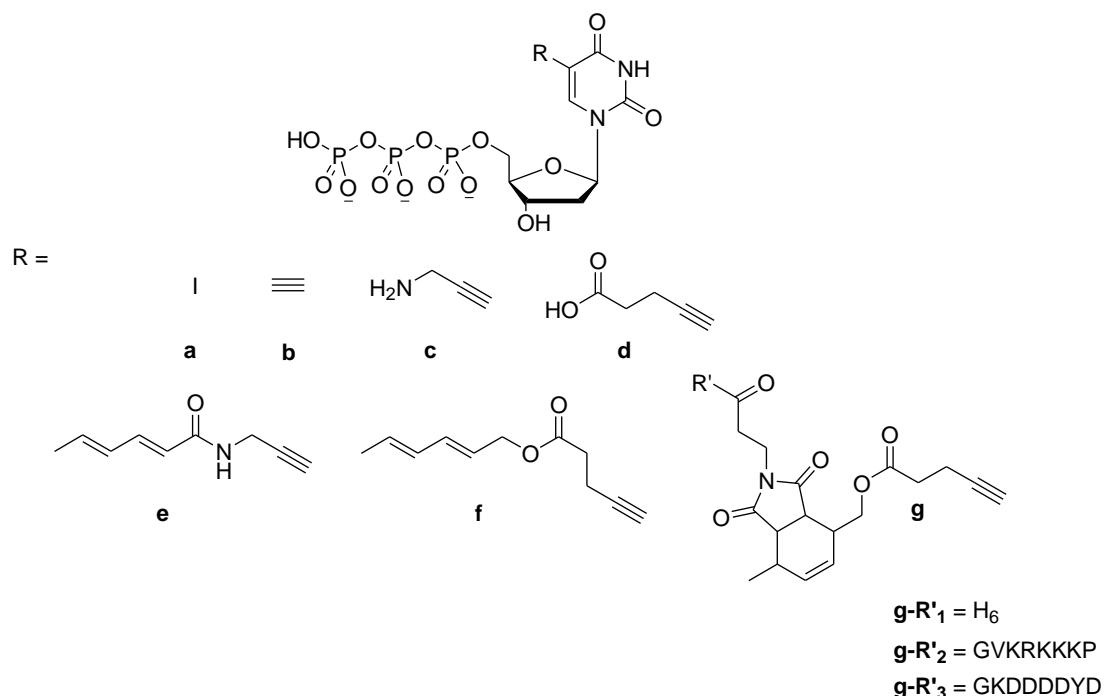


Figure 41: 2'-Deoxyuridine-5'-triphosphate bearing different substituents at position 5 of the pyrimidine. R' are peptide tags; R'₁=H₆, R'₂=GVKRKKKP, or R'₃=GKDDDDYD. The nucleotides were used to study the effect of substituent length, charge and bulkiness on enzyme catalysed incorporation into DNA. Adapted from [153].^v

Primer extension assays using the Deep Vent exo⁻ polymerase enzyme and subsequent gel electrophoresis showed that the enzyme was able to catalyse the incorporation of the majority of the nucleotides, with efficiency similar to that of the naturally occurring nucleotides. The gel analysis is summarised in Figure 42.

^v Nucleotides **b**, **e**, **f**, and **g** were synthesised by Dr. Vinciane Borsenberger.

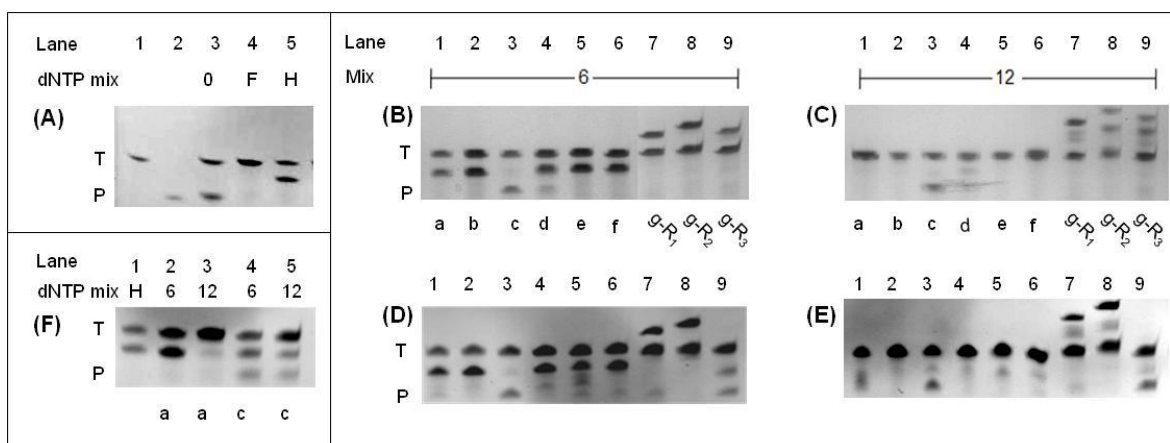


Figure 42: Denaturing PAGE analysis of extension reactions with modified nucleotides using Deep Vent exo- (B,C) and Taq (D, E) for dNTP mixes 6 and 12 as defined in Figure 40, conducted at 57 °C. (A) Control reactions with dNTP mixes 0, F, H, as defined in Figure 40 Lane 1 and 2 contain template, T, and primer, P, respectively. (F) Shows the extension conducted at 70 °C using mix H, and mixes 6 or 12 containing nucleotides a or c. Adapted from [153].

Panel A shows the results of the control experiments which were performed to confirm the viability of the primer extension assays. In Figure 42A, lane 1 and 2 correspond to the 36 nucleotide-long template and 24 nucleotide-long primer, respectively. By comparison, lane 4 shows the results of the positive control, dNTP mix F, where primer extension was performed using all four naturally occurring nucleotides. The primer was fully extended and thus co-migrated with the template resulting in a single band. Conversely, dNTP mix H, which was designed to direct the extension by 6 nucleotides, resulted in an up-shift of the primer band (Figure 42A, lane 5). In line with the expected results, no extension was observed for dNTP mix 0 (Figure 42A, lane 3). Both the modified nucleotides and 2'-deoxythymidine triphosphate, dTTP, were omitted from the primer extension assay, as result only the starting template and primer were observed. This control excludes the possibility of non-specific incorporation of a natural nucleotide in the place of a modified nucleotide.

The ability of the enzyme to extend the primer by six nucleotides and thus incorporate two modified nucleotides was evaluated. The results are shown in Figure 42B. The enzymatic reaction with a, b, d, e, f, and peptide-tagged analogues g-R₁, g-R₂, g-R₃ proceeded successfully, as indicated by a major up-shifted band stemming from the extension of the primer (Figure 42B, lanes 1, 2, 4 to 9). Enzymatic incorporation of 5-

ethynyl-2'-deoxyuridine, **b** and diene-modified nucleotide **f** has been observed in other studies.¹⁵⁷ These two nucleotides hence served as positive controls and references for the extension assay. The extended primers carrying the modified nucleotides migrated at about the same height as the strand from natural nucleotide mix H (compare Figure 42B, lanes 1, 2, 4 to 6 with Figure 42A, lane 5). The absence of any tag-induced gel shift attributed to the low mass of the chemical tags. This notion was supported by the results of primer extension assays using the peptide-modified nucleotides **g-R₁**, **g-R₂**, and **g-R₃**. The extended primer showed a significant gel shift (Figure 42B, lanes 7 to 9) reflecting the approximately ten-fold higher masses of the peptide tags compared to the other chemical tags. These results confirmed that in each case the enzyme was able to extend the primer by six nucleotides and thereby incorporate two modified nucleotides (Figure 40B, dNTP mix 6). Interestingly, almost no incorporation occurred for the amino-propargyl derivative **c** (Figure 42B, lane 3). It was hypothesised that the inability of **c** to serve as substrate was most likely related to the terminal amino group of the substituent at position 5 (see below for further explanation).

Similar results were observed for the extension of the primer by 12 nucleotides to incorporate three modified nucleotides (Figure 42C). The primer was fully extended using nucleotides **a**, **b**, **d**, **e** and **f** (Figure 42C, lanes 1, 2, 4 to 6), while the amino-propargyl **c** was not accepted as substrate for the enzyme (Figure 42C, lane 3). A lower incorporation extent was found for peptide-modified nucleotides **g-R₁**, **g-R₂**, and **g-R₃** (Figure 42C, lanes 7 to 9). In particular, it seemed that the addition of a third nucleotide into the partly extended primer occurred at a lower efficiency. This was most likely due to steric reasons. The first and the third tagged base were separated by 10 bases, which correspond to almost one complete helical turn.¹²⁴ Consequently, the peptide tags were most likely positioned on the same side of the duplex which is sterically crowded.

The incorporation efficiency was also assessed using the Taq pol under the same conditions as those used with Deep Vent exo⁻. The results of the primer extension are shown in Figure 42D and Figure 42E. Extension of the primer by six nucleotides was mostly complete for **a**, **b**, **d**, **e**, **f**, **g-R₁**, and **g-R₂** (Figure 42D, lanes 1, 2, 4 to 8). In comparison to

reactions with Deep Vent exo^- , the extent of incorporation by Taq was slightly lower. This was particularly evident in reactions involving **e** (Figure 42D, lane 5), where several minor bands stemming from the premature termination of the primer, were observed. As in the case of Deep Vent exo^- , amino-propargyl nucleotide **c** was a poor substrate for enzymatic polymerisation by Taq (Figure 42D, lane 3), even though another study reported good incorporation, albeit with a different enzyme, Tth.¹⁸¹ The low incorporation for **g-R₃** (Figure 42D, lane 9) was attributed to the partial hydrolysis of the triphosphate. This was most likely caused by the low pH of the nucleotide solution due to the acidic peptide tag (Figure 41). The assays for the extension of the primer by 12 nucleotides (Figure 42E) supported the previous findings. With the exception of nucleotides **c** (Figure 42E, lane 3), **g-R₁** and **g-R₂** (Figure 42E, lanes 7 and 8) the nucleotides analogues were suitable substrates for Taq pol. The poor incorporation of **g-R₃** (Figure 42E, lane 9) was also attributed to hydrolysis of the triphosphate.

The poor incorporation of amino-propargyl nucleotide **c** was investigated with a molecular model of a polymerase.^{vi} A model was constructed using the X-ray structure of the T7 DNA polymerase, which like Taq, belongs to the type A family of polymerases.¹⁸² The X-ray structure of T7 DNA polymerase is a ternary complex with the double stranded DNA template containing a single stranded overhang, and an incoming dTTP nucleotide complementary to adenine in the overhang.^{153,182} The model of the active site is shown in Figure 43. The protein is shown in grey and the DNA template in magenta. The nucleotide 5-(3-amino-propynyl)-2' deoxyuridine 5'-triphosphate **c** was placed in the same position as dTTP in the X-ray file. Visual inspection of the active site revealed that the primary amino group of the nucleobase is close to two charged amino acids. As measured from its nitrogen atom, the amino-propargyl group is 4.5 Å away from the nitrogen of lysine at position 522, and 4.7 Å away from the closest oxygen of aspartic acid at position 519. Based on the importance of substrate recognition by key residues in DNA polymerases, it was proposed that an electrostatic interaction between the amino group and the charged residues inhibited the activity of the polymerase activity. In particular, the electrostatic

^{vi} Modelling performed by Dr. Hugh Martin.

attraction of amino-propargyl to Asp, and the electrostatic repulsion to Lys, was most likely altering the position of the pyrimidine base within the active site thereby affecting the ability of the polymerase to accept the nucleotide as a substrate.

Several observations are consistent with the notion that electrostatic interactions cause the poor acceptance of amino-propargyl nucleotide **c**. Firstly, abolishing the ionic interaction by neutralising the amino group of **c** *via* acylation with hexa-2,4-dienoic acid (thereby yielding nucleotide **e**, Figure 41) resulted in an improved incorporation efficiency (lanes 5 in Figure 42B and C). In addition, weakening the interaction by increasing the temperature of polymerisation from 57 °C to 70 °C improved the acceptance of **c** by the enzyme, as seen for dNTP mix 6 (Figure 42F, lane 4) and dNTP mix 12 (Figure 42F, lane 5). While the ionic interaction between nucleotide and enzyme is a plausible explanation, other reasons cannot be completely ruled out. For example, a modified nucleotide added to a DNA strand could affect the interaction between primer and template strands and thereby affecting DNA duplex stability.¹⁸¹ Further work using point-mutated enzymes would be required to unequivocally clarify whether the charged residues in the active site are the key to the poor acceptance of amino-propargyl derivative **c**.

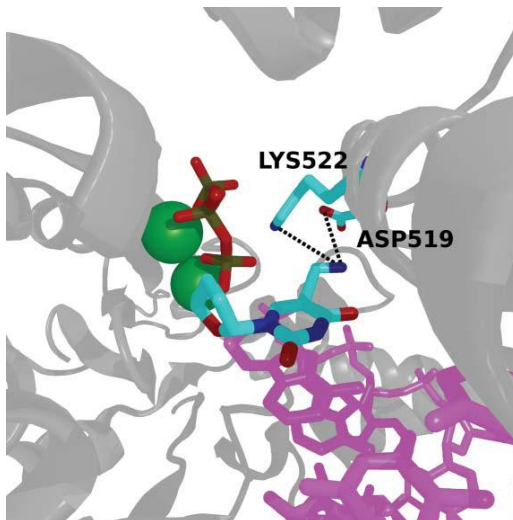


Figure 43: Molecular model of the active site of the T7 DNA pol from the A-type family. The model shows the electrostatic interactions between the amino group of 5-(3-amino-propynyl)-2'-deoxyuridine 5'-triphosphate **3c** and the side chains of aspartic acid and lysine. The protein is in grey, the DNA strand of the template is in magenta, and the two Mg^{2+} ions are represented as green spheres. The metal ions electrostatically interact with the triphosphate moiety of the nucleotide. The nucleotide **3c** was placed into the position of the incoming dTTP nucleotide in the crystal structure of the ternary protein·DNA·dTTP complex (PDB ID 1T7P) [¹⁵³, ¹⁸²].

Based on the primer extension assay results of the nucleotide analogues above, predictions could be made about the ability of the adamantane-modified nucleotides (**4a-c**, Figure 31) to serve as substrates for the DNA pol. As **4a-c** were uncharged and the adamantane functionality smaller than the Diels-Alder functionality (Figure 41, **g-R₁₋₃**), it was hypothesised that the nucleotides would be well tolerated by the enzyme. This was indeed found to be the case. Figure 44 shows the results of the primer extension assays using **4a-c**.

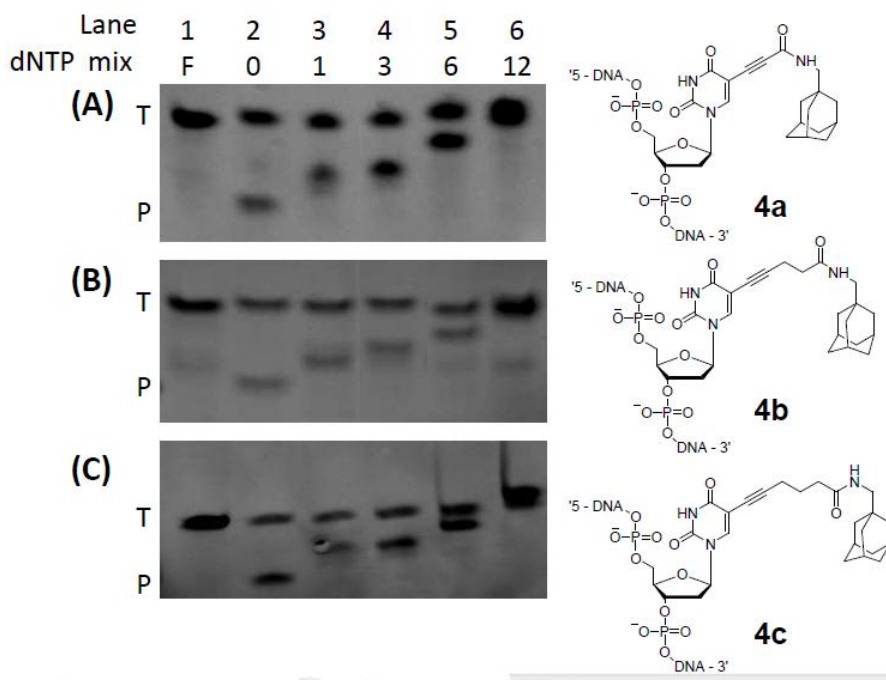


Figure 44: Denaturing PAGE analysis of extension reactions with adamantane-modified nucleotides (**4a-c**) using Deep Vent exo⁻. Lane 1: dNTP mix F; positive control in the presence of natural nucleotides. Lane 2: dNTP mix 0; negative control primer extension without dTTP or modified nucleotide. Lane 3: dNTP mix 1; extension of the primer by one modified nucleotide. Lane 4: dNTP mix 3; extension of the primer by 3 nucleotides. Lane 5: dNTP mix 6; extension of the primer by six nucleotides and concomitant incorporation of two modified nucleotides. Lane 6: dNTP mix 12; full extension of the primer and concomitant incorporation of three modified nucleotides. See Figure 40 for the sequence of the template and primer.

Akin to the primer extension assays described above, lanes 1 and 2 show the results of the control experiments, dNTP mix F and dNTP mix 0, respectively. The ability of the enzyme to catalyse the full extension of the primer and hence incorporate three modified nucleotides was assessed in a step-wise manor. Lane 3 (Figure 44) shows the extension of the primer by a single modified nucleotide (see Figure 40 for sequence of template and

primer). For each of the assays (Figure 44A-C), a slight gel shift was observed suggestive of single nucleotide incorporation. Greater gel shifts were observed as the primer was extended by three (Figure 44, Lane 4), six (Figure 44, Lane 5) and 12 (Figure 44, Lane 6) nucleotides. The latter two extensions resulted in the incorporation of two and three modified nucleotides, respectively (Figure 40). Similar to nucleotides **g-R₁₋₃**, the fully extended primer containing **4c** migrated above the template due to the increased molecular weight and thus slower electrophoretic mobility of the nucleotide compared to the other chemical tags (Figure 44C, Lane 6). Interestingly, shorter extension products were not observed, suggesting that DNA pol was able to incorporate **4c** with a greater efficiency compared to **g-R₁₋₃**. Hence incorporation of the third modified nucleotide was not sterically hindered by the first. The minor band observed in Figure 44B, lane 6, was attributed to partially hydrolysis of the nucleotide.

3.4.4. Oligonucleotide Synthesis

Having shown that the adamantane-modified nucleotides were readily incorporated into DNA by the action of DNA pol, efforts were made to chemically synthesise modified DNA strands that possessed nucleotide repeat sequences.^{vii} These modified DNA strands would subsequently be used to test the sensing approach. The sequence of the synthesised oligonucleotides is shown in Table 6. The sequences are based upon the complementary sequence of pentanucleotide STR called PentaD, which is used as a genetic marker for forensic DNA fingerprinting.¹⁸³ In each of the strands, a single modified nucleoside, U*, was incorporated per repeat unit.

^{vii} Oligonucleotide synthesis was performed at the University of Southampton with the help of Dr. Jonathan Burns.

Table 6: Sequence of PentaD STR and synthesised oligonucleotides.

Entry	Sequence
PentaD	5'-AAA AAA A <u>AAAGA</u> <u>AAAGA</u> <u>AAAGA</u> AAA AA-3'
1	3'-TTT TTT T <u>TU[*]TCT</u> <u>TTTCT</u> <u>TTTCT</u> TTT TT-5'
2	3'-TTT TTT T <u>TU[*]TCT</u> <u>TU[*]TCT</u> <u>TTTCT</u> TTT TT-5'
3	3'-TTT TTT T <u>TU[*]TCT</u> <u>TU[*]TCT</u> <u>TU[*]TCT</u> TTT TT-5'

Chemical synthesis of DNA was achieved using automated solid-phase chemistry. Reactive groups on the nucleosides were protected to prevent unwanted side reactions during synthesis. The 5'-hydroxyl was protected with dimethoxytrityl (DMT) ether, whilst the exocyclic amine of cytosine was protected with a benzoyl group (Figure 45A).¹⁸⁴ The reactive phosphate on the 3'-position was protected by β -cyanoethyl group and chain elongation occurred via phosphate-triester chemistry (Figure 45C). The four essential steps of solid-phase oligonucleotide synthesis have been summarised in Figure 45B.¹⁸⁵

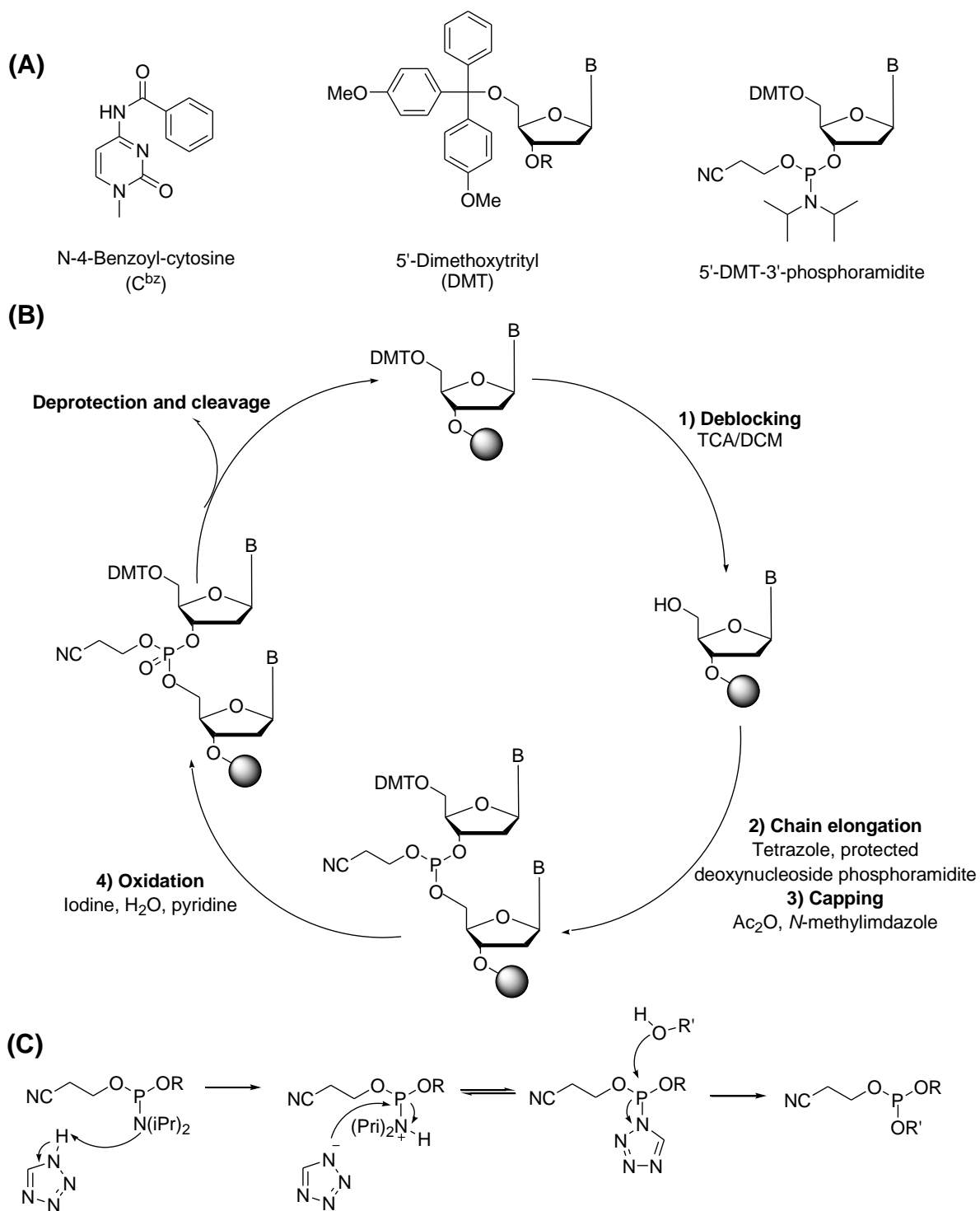


Figure 45: (A) From left to right, exocyclic protecting groups for cytosine for the 5' and 3' positions. (B) Solid-phase, four-step phosphoramidite oligonucleotide synthesis cycle. (C) Synthesis of phosphite triester. R and R' are nucleosides.

In brief, the 5'-*O*-DMT protecting group was removed from the resin-bound deoxynucleoside using a weak acid to yield an orange DMT cation (Figure 45B, step 1).¹⁸⁴ The highly chromophoric DMT cation was used to determine the coupling efficiencies throughout the synthesis. This initial deprotection step is termed deblocking. Chain elongation (Figure 45B, step 2) occurred via the formation of a phosphite triester internucleotide bond with the incoming nucleoside. The relatively stable and unreactive 3'-phosphoramidite was activated to the highly reactive phosphitylating agent by tetrazole. As tetrazole is a weak acid (pKa 4.6),¹³⁶ it was able to catalyse the reaction without compromising the DMT protecting groups. The mechanism of activation is shown in Figure 45C. Any unreacted 5'-hydroxyls were then capped or acetylated (Figure 45B, step 3) using acetic anhydride and *N*-methylimidazole. Finally, the phosphorous in the phosphite triester internucleotide linkage was oxidised from +3 oxidation state to +5 in the corresponding phosphate triester (Figure 45B, step 4). Oxidation was achieved with iodine and water in THF, and pyridine added to neutralise the liberated hydrogen iodide. Further repetitions of this cycle afforded the oligonucleotides with the desired length and sequence. At each step of the synthesis, the resin was washed with anhydrous ACN to prevent carryover of reagents. The final products were liberated from the resin and concomitantly deprotected by treating the resin with ammonium hydroxide at 50 °C overnight. The benzoyl protecting group on cytosine was removed via an addition-elimination reaction at the carbonyl group of the ester bond. Conversely, the 2-cyanoethyl group was removed from the phosphate by a β -elimination reaction. Lyophilisation of the ammonia solution afforded the crude oligonucleotides which were subsequently purified by HPLC and characterised by MS analysis.

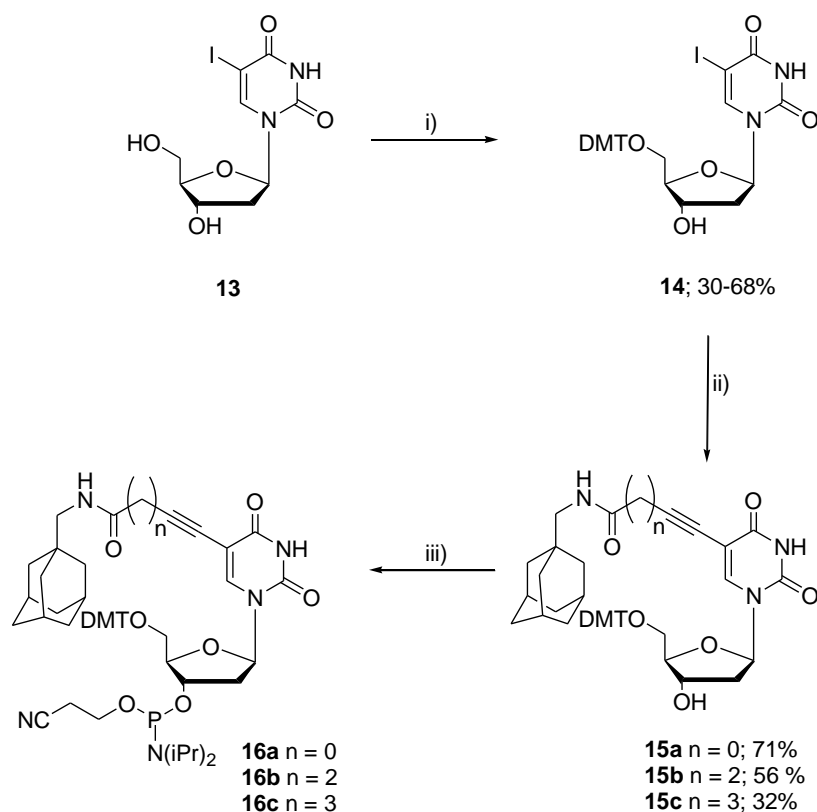


Figure 46: Synthesis of 5'-DMT-5-adamantane modified nucleoside phosphoramidites. (i) DMT-Cl, Et₃N, cat. DMAP, pyridine; (ii) CuI, Et₃N, Pd(PPh₃)₄, adamantane-acetylene linker, DMF, 16 h; (iii) CEP-Cl, DIPEA, DCM, 2 h.

Synthesis of the phosphoramidite derivatives of the adamantane-modified nucleosides (**3a-c**, Figure 31) is outlined in Figure 46. The 5'-hydroxyl of 5-iodo-2'-deoxyuridine, **13**, was protected as the DMT-ether using DMT-Cl and Et₃N in anhydrous pyridine. The resultant protected nucleoside, **14**, was isolated from the crude reaction mixture by flash chromatography in 30-68 % yield. In order to minimise loss of the DMT-group during purification, flash chromatography was performed using silica gel which had been neutralised with Et₃N. The residual base was removed from the product by co-evaporation with HPLC-grade MeOH. Using the Sonogashira cross-coupling reaction, the adamantane-acetylene linkers were coupled to the DMT-protected nucleoside. Purification of the resultant nucleosides (**15a-c**) was more cumbersome than had initially been anticipated. TLC analysis of the crude reaction mixtures showed good resolution between the desired nucleosides and by-products that had formed. Indeed, DMT-containing compounds were readily visualised by exposure of the TLC plates to concentrated HCl vapours. However,

despite using the same solvent system as that used for TLC analysis, flash chromatography resulted in mixed fractions with almost no resolution. Neither employing a gradient elution strategy, nor performing purification under gravity, improved the resolution. The purified nucleosides were finally isolated in clean fractions, by using a mixture of silica gels. Using 15% (w/w) silica gel Type H with silica gel Type G, and employing a gradient elution strategy under gravity, sufficiently reduced the migration rate of the crude material through the column and thus improved the resolution. Silica gel Type H has smaller particle size (10-40 μm) compared to silica Type G (40-60 μm), and essentially acted as a filter to reduce migration speed and increase resolution. Using the optimised conditions, nucleosides **15a**, **15b** and **15c** were isolated in 71%, 56% and 32%^{viii} yields, respectively. Formation of the phosphoramidites **16a-c**, was achieved according to standard phosphitylation methods using 2-cyanoethylchloro-*N,N*-diisopropylphosphoramidite (CEP-Cl) in dry DCM as the reagent. The phosphoramidites were precipitated from the crude reaction mixture, dried under a stream of nitrogen, and then dissolved to a final concentration of between 0.02-0.04 M in ACN/DCM mixture (1:1, v/v). Due to the extreme sensitivity of the phosphoramidites, they were synthesised immediately prior to commencement of DNA synthesis and used crude. No characterisation was performed bar TLC analysis which showed that formation of the phosphoramidites was complete within 2 h. DNA synthesis was performed with 100-fold molar excess of the phosphoramidites over the resin bound nucleoside. This was necessary to counteract the loss of active material through oxidative and hydrolytic degradation. Without a sufficient molar excess, degradation of phosphoramidite would ultimately have led to a reduced coupling efficiency and overall yield. Indeed, MS analysis of residual phosphoramidite solution of **15b** (Figure 47) showed traces of phosphonate and monophosphate hydrolysis products in addition to the starting phosphoramidite. These findings were corroborated by ³¹P NMR analysis in which signals were observed at 150.2 ppm, 35.5 ppm, 8.1 ppm and 3.3 ppm, which in accordance with literature

^{viii} **15c** was synthesised by Xiu Yao, a Master's student within our research group.

reports,^{186–188} correspond to the phosphoramidite, phosphonoamidate, the phosphonate and monophosphate, respectively.

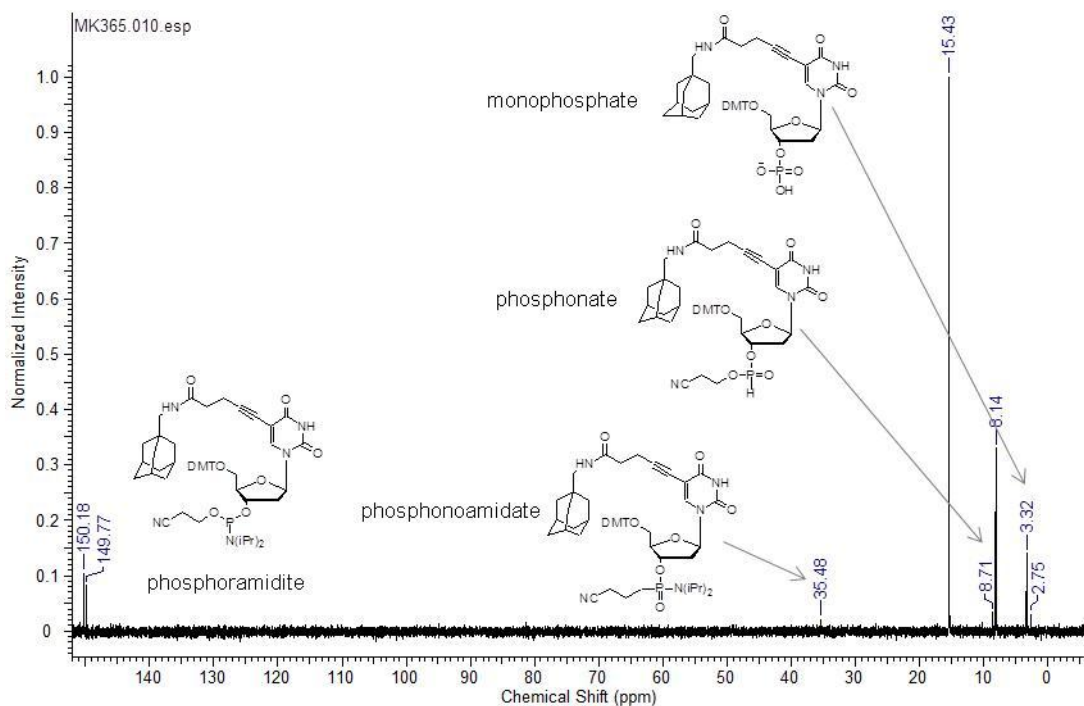


Figure 47: ³¹P NMR analysis of residual **15b**. NMR spectra recorded at 121.4 MHz in CDCl₄.

Using the nucleoside phosphoramidites **16a** and **16c**, a range of oligonucleotide strands were synthesised and subsequently analysed via nanopore analysis.

3.4.5. Nanopore Analysis

During nanopore analysis, the electric field applied to single-stranded oligonucleotides causes the DNA strands to rearrange from a randomly coiled state within the vestibule of the pore, to a linear extended state prior to translocation.³² Whilst in this linear state, it is thought that chemical tags placed at position C5 are held perpendicular to the oligonucleotide (Figure 48A).⁹ As the modified oligonucleotide translocates the pore, the tag is repositioned so that it lays parallel to, or flush against, the oligonucleotide (Figure 48B). In doing so, a bulky oligo-tag segment is formed which essentially increases the cross-sectional diameter of the oligonucleotide. ssDNA which has an average cross-sectional diameter of 0.9-1.2 nm⁹ is known to translocate the α HL pore, whose inner

constriction measures 1.3 nm in diameter,⁶ with minimal affect on the ionic current.³² Several hypotheses were made regarding the tranlocation of the adamantane-modified oligonucleotides. Firstly, the inclusion of an adamantane tag, which has dimensions of approximately 4 Å cubed,¹⁴⁷ would increase the cross-sectional diameter of the oligonucleotides to approximately 1.3-1.6 nm. As this is closer to the diameter of the inner constriction of the α -HL, the bulky oligo-tag segment was expected to occlude the ionic current during translocation, resulting in detectable disturbances in the electronic trace. Secondly, as the non-modified regions of the oligonucleotide, adjacent to the oligo-tag segments, translocated the pore, the flow of ions would be restored as these regions are equivalent to non-modified ssDNA (Figure 48C). In essence, it was hypothesised that step-like disturbances would be observed in the electrical traces as the tagged oliognucleotide translocated the pore. The number of steps observed should reflect the number of chemical tags in the oligonucleotide and hence the number of repeat units in the STR. Step-like behaviour has been observed previously in work by Mitchell *et al.*, who analysed peptide-tagged DNA strands.⁹

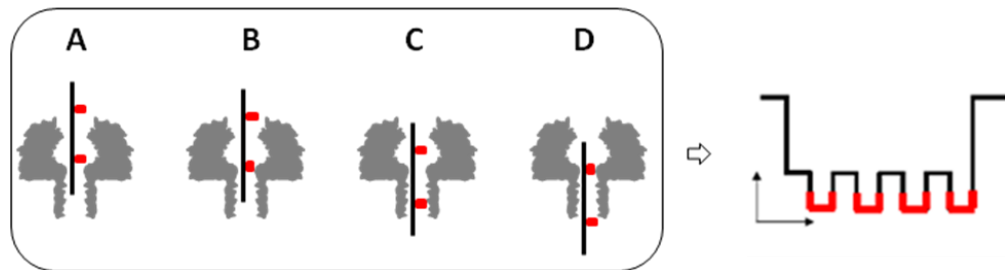


Figure 48: Translocation of tagged oligonucleotides and origin of step-like modifications in electrical traces. Adapted from [9].

Finally, it was hypothesised that the resolution of the ‘steps’ would be dependent upon the length of the linker between the oligonucleotide and the adamantane tag. An overly long linker could possibly span the entire STR repeat unit and interact with the preceding tag. This would ultimately reduce the resolution of the ‘steps’ in the electrical traces as the intermittent non-modified regions would not be detected. Conversely, if the linker was excessively short, the oligo-tag region may be insufficiently large for detection via the nanopore. In extended ssDNA, the distance between the bases is 0.5 nm.¹⁸⁹ Thus in a

pentanucleotide repeat sequence, the distance between a modified nucleotide in one repeat unit and the next is approximately 2.5 nm. It was hypothesised that nucleoside **4c** (Figure 44, longest linker) would offer the best resolution as the distance between C5 and the bottom of adamantane is approximately 1.74 nm. Hence the nucleoside should enable the detection of the intermittent non-modified region of the oligonucleotide in addition to forming an oligo-tag segment sufficiently large to block the flow of ions. For **4a** (Figure 44, shortest linker), the distance between C5 and the bottom of adamantane is 1.17 nm. It was hypothesised that linker would be too short of offer good resolution.

The modified oligonucleotides were subjected to nanopore analysis and the translocation parameters determined. The recordings were performed in 1 M KCl with 50 mM Tris, pH 8.0 and applied voltage of +100 mV at the *trans* side. In the absence of any oligonucleotide, α -HL displayed a constant open channel current of 110 ± 8.4 pA ($n=2$).

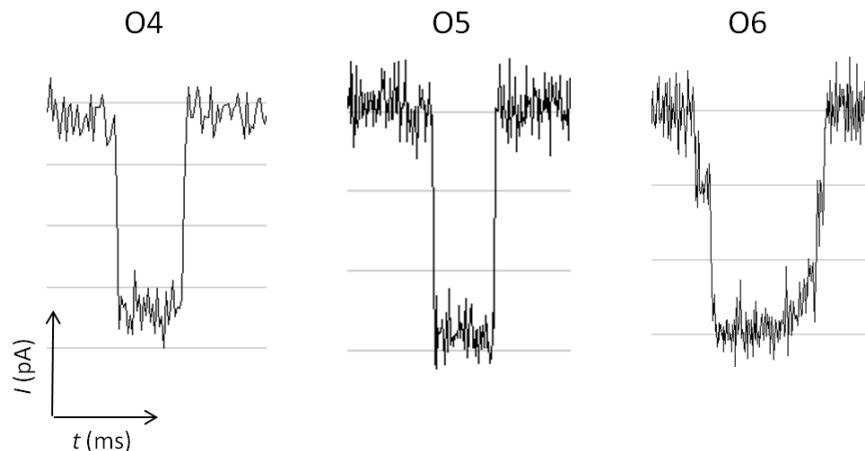


Figure 49: Representative traces for oligonucleotides modified with short adamantane linkers, **4C**. The electrical signals from oligonucleotides with one (O4) and two (O5) O4 modified nucleotides did not display any significant difference. Oligonucleotides with three modified nucleotides (O6) showed similar electrical signals as O4 and O6. However, several signals also possessed one or two ‘shoulders’.

Addition of the oligonucleotides to the *cis* side of the pore led to detectable disturbances of the ionic current which were attributed to the net negatively charged oligonucleotides translocating the pore, towards the positively charged *trans* side. Each oligonucleotide was characterised using data generated from 3 independent recordings with at least 1000 translocation events. Step-like modifications of the ionic current were not observed;

representative electrical traces are shown in Figure 49. The translocation parameters for each oligonucleotide are summarised in Table 7.

Table 7: Translocation Parameters of Adamantane-Modified Oligonucleosides.

Oligo Name	Modified nucleotide	Number of STR repeats ^[b]	Amplitude [%]	τ_{off} [ms]
O1	4c ^[c]	1	90.6 ± 0.01	5.0 ± 0.02
O2	4c	2	89.6 ± 3.49	12.3 ± 0.08
O3	4c	3	95.1 ± 1.68	8.9 ± 0.03
O4	4a ^[d]	1	85.9 ± 0.50	3.5 ± 0.16
O5	4a	2	86.8 ± 0.70	2.7 ± 0.17
O6	4a	3	92.3 ± 0.30	4.3 ± 0.14
ssDNA ^[a]	N/A	N/A	91.7 ± 1.1	0.18 ± 0.06

N/A: Not applicable.

^[a] 27mer ssDNA. Data taken from reference [9]

^[b] Oligonucleotide sequence: 1 STR, 3'-TTT TTT TTU*TCT TTTCT TTTCT TTT TT-5'; 2 STRs, 3'-TTT TTT TTU*TCT TU*TCT TTTCT TTT TT-5'; 3 STRs, 3'-TTT TTT TTU*TCT TU*TCT TU*TCT TTT TT-5'

^[c] Long linker

^[d] Short linker

It was observed that the event duration, τ_{off} , for O1, that is the oligonucleotide modified with a single long linker, was 5.0 ± 0.02 ms. Whilst that of O4, the oligonucleotide modified with a single short linker, was 3.5 ± 0.16 ms. Hence modification with the longer linker effectively increased the translocation time almost 1.5-fold. A similar observation was made for oligonucleotides O3 and O6. In general, the translocation speeds of oligonucleotides modified with short linkers (O4-O6) were faster than those modified with long linkers (O1-O3). For instance, the presence of three long linkers in O3 effectively increased the translocation time by almost 2-fold, versus three short linkers in O6. Furthermore, it was observed that the event duration of O3 and O6 were, 44% and 19% longer than those of O1 and O4, respectively. Therefore, the presence of three modified nucleotides always increased the event duration irrespective of the linker composition. This was attributed to the increased resistance experienced by the translocating oligonucleotide. These findings support the notion that, firstly, modulation of the ionic current is due to steric hindrance that occurs when the bulky oligo-tag segment translocates the pore. Secondly, that the tag is elongated and aligned parallel to the DNA

strand during translocation. Hence, when the tag is attached via a longer linker, a longer oligo-tag segment is formed resulting in a longer translocation event. In all cases the event durations were an order of magnitude longer than that of non-modified ssDNA which is typically 180 μ s.⁹ The average amplitude of the translocation events was also determined. It was expected that within each set of oligonucleotides (O1-O3 and O4-O6) the amplitude of blockades would be similar. This would indicate that the blockades were due to the action of a single bulky oligo-tag segment acting independently from the others. However, this was not the case; a near linear relationship was observed between the number of tags and the amplitude of blockade. The results suggested that as the number of tags increased, so did the cross-sectional diameter of the oligonucleotides, which would account for the increasing amplitude. There are two plausible explanations for this observation. Firstly, despite the length of the linker in comparison to the distance between the STR units, it is possible that the tags were completely spanning the STR and interacting with one another. Effectively, the tags acted as an insulator or shield to increase the overall diameter of the oligonucleotide. Alternatively, the DNA strands were being compressed as they translocated the pore, hence reducing the distances between the bases. Secondly, as the linkers are hydrophobic chains, they had no means of interacting with the oligonucleotide, which has a net negative charge. Consequently, the oligo-tag segment may not be compact enough to give reliable and comparable amplitudes of blockades. Furthermore, due to the hydrophobic nature of the linkers, they may have been more likely to interact with another linker rather than the oligonucleotide. These explanations are partially supported by the fact that the amplitudes of blockades were on a par with non-modified ssDNA (91.7%).⁹ This indicates the formed oligo-tag segment may not be much wider than non-modified DNA.

3.5. Conclusion

The aim of this study was to develop a nanopore-based strategy to size nucleotide repeat sequences, which are the basis for DNA fingerprinting. It was proposed that this could be achieved by incorporating a single adamantane-modified nucleotide per repeat unit. It

was hoped that subsequent nanopore analysis would result in electrical traces with distinct step-like modifications, where the number of steps corresponded to the number of modified nucleotides and hence the number of repeat units.

Three adamantane-modified nucleosides, with varying linker lengths, were successfully synthesized using Sonogashira cross-coupling. The corresponding nucleotides were prepared via the Yoshikawa phosphorylation procedure. The original procedure was optimised with regards to reaction time, temperature and number of equivalents of POCl_3 . The optimized procedure was successfully applied to a range of nucleoside analogues which were then subjected to primer extension assays. It was demonstrated that the majority of the nucleotides were suitable substrates for both DNA and Taq pol. However, poor incorporation was observed for the sterically demanding peptide-tagged nucleotides and the amino-propargyl derivative. Using a molecular model of the polymerase active site, it was proposed that the latter formed adverse electrostatic interactions within the enzyme. This notion was supported by experiments which aimed to minimise these interactions. In particular, performing the assay at elevated temperatures and acylating the amino-functional group resulted in increased incorporation. Despite the success of the optimized conditions, they could not be applied to all nucleoside analogues. Phosphorylation reactions of the adamantane-modified nucleoside analogues required careful monitoring for the formation of key intermediates, and as such, the reaction times were highly variable compared to other nucleoside analogues. Nonetheless, the nucleotides were successfully synthesised and incorporated into oligonucleotides.

Having shown that the adamantane-modified nucleosides were suitable substrates for DNA pol, the sensing approach was tested using chemically synthesised oligonucleotides which possessed STR sequences. A single adamantane-nucleoside was placed within each STR. A linear relationship was observed between the number of tags present in the oligonucleotide and the dwell time. This suggested that as the number of tags increased, the resistance experienced by the oligonucleotide increased, increasing the translocation time. In spite of this, step-like disturbances were not observed in the electrical traces. The most plausible explanation for this is that the tags did not sufficiently

increase the diameter of the DNA. Indeed, the amplitude of blockades of the modified oligonucleotides was on a par with non-modified ssDNA. This observation, combined with the τ_{off} data, indicated that the formed oligo-tag segments were not compact enough to occlude the flow of ions. As oligonucleotides possess a net negative charge, a more compact oligo-tag segment could be formed by employing a linker that possess some positive charge.

3.6. Materials and Methods

All reagents and anhydrous solvents were purchased from Sigma Aldrich unless otherwise stated. All organic solvents and water (HPLC grade) were purchased from Fischer Scientific. *O*-(benzotriazol-1-yl)-N, N, N', N'-tetramethyluronium hexafluorophosphate (HBTU) was purchased from NovaBiochem. 5-iodo-2'-deoxyuridine was purchased from CarboSynth. Deep Vent_R[®] (exo⁻) and Taq polymerases were purchased from New England Biolabs. DNA primers and templates were purchased from Integrated DNA Technologies.

Nanopore analysis was performed as described in the Materials and Methods section of Chapter 1 (section 2.11). HPLC analysis was performed using reverse phase Varian ProStar system with a Model 210 solvent delivery mode and a Model 320 dual wave-length detector. The mobile phase A consisted of 0.1 M triethylammonium bicarbonate buffer (pH 7) and mobile phase B consisted of neat acetonitrile. The UV absorbance was monitored at 260 nm. Analytical HPLC analysis was performed using a Varian Pursuit XRS C18 column (250 x 4.6 mm). The gradient started with 2% eluent B and rose linearly to 25% B over 10 min with a flow rate of 1 mL min⁻¹. Preparative purification was performed on the same system using a Polaris C18 column (100 x 21.2 mm, 5 mm beads). The gradient started with 2% eluent B and rose linearly to 25% B over 10 min with a flow rate of 10 mL min⁻¹. Fractions containing the desired product were pooled, the solvent removed under reduced pressure to approximately 2 mL and the solution lyophilised overnight. Ion exchange chromatography was performed on an ÄktaPurifier 10 FPLC system (GE Healthcare). Mobile phase A consisted of 0.1 M triethylammonium bicarbonate buffer (pH 7). Mobile phase B consisted of 0.8 M triethylammonium

bicarbonate buffer (pH 7). Analytical ion exchange was performed using a Resource Q column (1 mL, GE Healthcare). The gradient started with 0% B and rose linearly to 100% B over 10 min at 1 mL min⁻¹. Preparative ion exchange was performed using a Resource S column (6 mL, GE Healthcare). The gradient started at 0% B and rose linearly to 100% B over 10 min at 2 mL min⁻¹. ESI-MS analysis was performed using a Waters Aquity Ultra performance LC-MS system equipped with an Aquity UPLC BEH C18 column (50 x 2.1 mm, 1.7 mm beads). ¹H- and ¹³C-NMR spectra were recorded at 293K on a Bruker AVANCE500 or -600 MHz spectrometer. ³¹P-NMR spectra were recorded a Bruker AMX300 spectrometer. Coupling constants are reported in Hz. Reference NMR solvent signals CDCl₃: δH = 7.26 ppm, δC = 77.0 ppm; D₂O: δH = 4.79 ppm; (CD₃)₂SO: δH = 2.50 ppm, δC = 39.5 ppm; CD₃OD: δH = 3.31 ppm, δC = 49.0 ppm; unless otherwise stated. Multiplicities for ¹H coupling are shown as s (singlet), d (doublet), t (triplet), m (multiplet), or a combination of the above. Thin Layer Chromatography (TLC) was performed on aluminium backed Sigma-Aldrich TLC plates with F254 fluorescent indicator. Visualisation achieved by absorption of UV light or with the use of potassium permanganate solution [KMnO₄ (1.25 g), Na₂CO₃ (6.25 g), water (250 mL)]. DMT-protecting groups were visualised by exposing TLC plates to fumes of HCl. Normal phase flash chromatography was carried out using silica gel (43 – 60 μm) supplied by Merck. Oligonucleotide synthesis was performed at the University of Southampton according to [190]. In brief, the oligonucleotides were synthesised on an Applied Biosystems Expedite synthesiser on solid support (Glen Research, CPG, pore size 500 Å) on a 1.0 μmol scale. The synthesis was carried out in the standard mode using 2-cyanoethyl-*N,N*-diisopropylphosphoramidites. The coupling time for incorporation of the modified nucleotides (0.02-0.04 M in DCM/ACN, 1:1, v/v) was increased from 40 s to 650 s. The coupling efficiency was monitored by UV absorption of the cleaved dimethoxytrityl cation. After the synthesis, the solid supports were dried and incubated overnight with 1 mL aqueous ammonium hydroxide at 55 °C. The deprotected and cleaved oligonucleotides were then lyophilised. The crude oligonucleotides were redissolved into 0.1 M TEAB buffer, pH 7, and partially purified by size exclusion on NAP-25 columns (GE Healthcare). The column was equilibrated by passing through HPLC-grade

water (10 mL) under gravity. The DNA sample was loaded onto the column in a total volume of 0.5 mL. The first 1.3 mL was discarded thereby removing impurities formed during the deprotection step. Elution of the oligonucleotides was achieved by loading water (1 mL) onto the column. The collected oligonucleotides were further purified by preparative HPLC and subsequently characterised using MALDI-MS (performed by Julie Herniman, University of Southampton). The chemical characteristics of the oligonucleotides are summarised in Table 8.

Table 8: Chemical characterisation of adamantane-modified oligonucleotides

Oligo	Modified nucleotide ^[a]	Number of STR repeats	mass [m/z]	
			Expected	Observed ^[d]
O1	4c ^[b]	1	8308.5	8594.4
O2	4c	2	8524.6	8267.7
O3	4c	3	8740.7	9135.7
O4	4a ^[c]	1	8349.5	8334.4
O5	4a	2	8592.7	8727.5
O6	4a	3	8835.8	8822.1

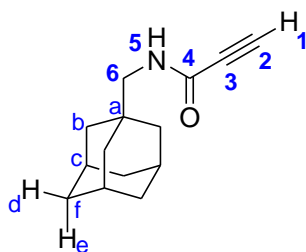
^[a]Oligonucleotide sequence: 1 STR, 3'-TTT TTT TTU*TCT TTTCT TTTCT TTT TT-5'; 2 STRs, 3'-TTT TTT TTU*TCT TU*TCT TTTCT TTT TT-5'; 3 STRs, 3'-TTT TTT TTU*TCT TU*TCT TU*TCT TTT TT-5'

^[b]Long linker

^[c]Short linker

^[d]Due to difficulties in desalting the samples, it was not possible to get very accurate masses.

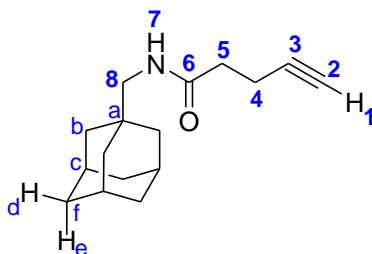
3.6.1. Synthesis of propynoic acid (adamantan-1-yl methyl)-amide, 2a



Propiolic acid (112 μ l, 1.8 mmol) was mixed with anhydrous DMF (4 mL) and the resultant solution cooled to 0 °C on an ice bath. To the chilled solution, EEDQ (540 mg, 2.2 mmol) was added followed by adamantane methylamine (321 μ L, 1.8 mmol) drop-wise. The resultant mixture was stirred at 0 °C for 30 min then at room temperature overnight. After 14 h, the reaction mixture was washed with 2.5 M HCl (3 x 20 mL) and dried over MgSO₄.

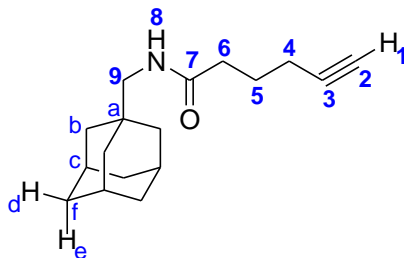
The residue was absorbed onto the silica and then purified by flash chromatography (35 % EtOAc/Hexane, $R_f = 0.50$) to afford a white solid (276 mg, 71%). ^1H NMR (600 MHz, MeOD) δ_{ppm} : 1.51 (d, $J=2.45$ Hz, 6H, H-b), 1.66-1.69 (m, 3H, H-d), 1.72-1.79 (m, 3H, H-e), 1.97 (br.s., 3H, H-c), 2.91 (s, 2H, H-6), 3.58 (s, 1H, H-1). ^{13}C NMR (600 MHz, MeOD) δ_{ppm} : 29.8 (C-c), 35.5 (C-a), 37.9 (C-f), 41.3 (C-b), 52.2 (C-6), 75.9 (C-2), 78.3 (C-3), 155.1 (C-4). m/z (ESI) 218.1540 [M], $\text{C}_{14}\text{H}_{20}\text{NO}$ requires 218.1545.

3.6.2. Synthesis of pent-4-ynoic acid (adamanta-1-ylmethyl)-amide, 2b



HBTU (688 mg, 1.8 mmol), 4-pentynoic acid (178 mg, 1.8 mmol), and DMAP (23 mg, 0.2 mmol) were placed in an oven dried flask and purged with argon then suspended in anhydrous DMF (5 mL). To the resultant solution, Et_3N (260 μL , 1.9 mmol) then adamantane methylamine (300 μL , 1.7 mmol) were added slowly drop wise. After 16 h stirring under argon, silica was added to the reaction mixture and the solvent was removed *in vacuo*. The crude material was absorbed onto the silica and purified by flash chromatography (35% EtOAc/Hexane, $R_f = 0.51$) to afford the title compound as a white solid (404 mg, 97%). ^1H NMR (500 MHz, CDCl_3) δ_{ppm} : 1.49 (d, $J=2.21$ Hz, 6H, H-b), 1.57 - 1.65 (m, 3H, H-d), 1.71 (d, $J=12.30$ Hz, 3H, H-e), 1.97 (br. s., 3H, H-c), 2.02 (t, $J=2.52$ Hz, 1H, H-1), 2.40 - 2.48 (m, 2H, H-4) 2.52 - 2.59 (m, 2H, H-5), 2.98 (d, $J=6.31$ Hz, 2H, H-8). ^{13}C NMR (500 MHz, CDCl_3) δ_{ppm} : 15.27 (C-5), 28.27 (C-c), 33.74 (C-a), 35.57 (C-4), 36.98 (C-f), 40.24 (C-b), 51.30 (C-8), 69.66 (C-2), 83.25 (C-3), 171.51 (C-6). m/z (CI) 246 [M+H] $^+$, $\text{C}_{16}\text{H}_{24}\text{NO}$ requires 245.35992.

3.6.3. Synthesis of hex-5-ynoic acid (adamantan-1-ylmethyl)-amide, 2c

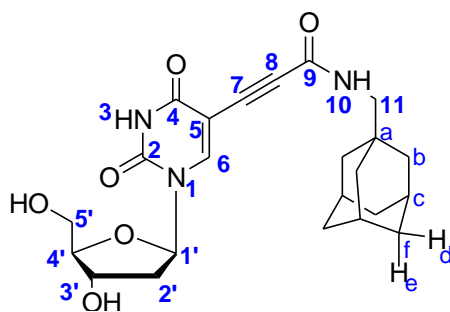


HBTU (1.2 g, 3.3 mmol) and DMAP (38 mg, 0.3 mmol) was placed in an oven dried flask, then purged with argon then suspended in anhydrous DMF (10 mL). To the resultant yellow solution, Et₃N (470 μ L, 3.3 mmol) was added followed by 5-hexynoic acid (370 μ L, 3.3 mmol) and adamantane methylamine (540 μ L, 3.0 mmol) dropwise. After 18 h stirring at room temperature, the crude material was concentrated *in vacuo* and the resulting oil purified by flash chromatography (35% EtOAc/Hexane, R_f = 0.5) to afford the title compound as a colourless oil which solidified on standing. The solid was co-evaporated with MeOH (3 x 10 mL) to remove traces of triethylamine the dried under high vacuum (688 mg, 83%). ¹H NMR (600 MHz, CDCl₃) δ_{ppm} : 1.48 (d, J =2.45 Hz, 6H, H-b), 1.63 (d, J =11.48 Hz, 3H, H-d), 1.72 (d, J =12.23 Hz, 3H, H-e), 1.85-1.91 (m, 2H, H-5), 1.97-1.99 (m, 3H, H-c), 2.00 (t, J =2.64 Hz, 1 H, H-1), 2.28 (td, J =6.87, 2.64 Hz, 2H, H-4), 2.36 (t, J =7.43 Hz, 2H, H-6), 2.97 (d, J =6.21 Hz, 2H, H-9). ¹³C NMR (600 MHz, CDCl₃) δ_{ppm} : 17.9 (C-4), 24.3 (C-5), 28.3 (C-c), 33.7 (C-a), 35.3 (C-6), 37.0 (C-f), 40.3 (C-b), 51.1 (C-9), 69.4 (C-2), 83.7 (C-3), 172.6 (C-7). m/z (CI) 260 [M+H]⁺, C₁₇H₂₅NO requires 259.19361.

3.6.4. General procedure for Sonogashira cross-coupling

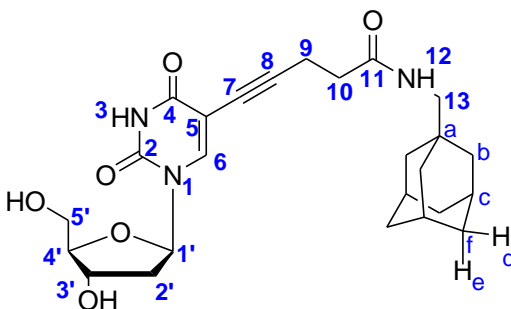
To the adamantane methylamide (2 eq), 5-iodo-2'-deoxyuridine (1 eq), palladium tetrakis (0.2 eq to nucleoside) and CuI (0.1 eq to nucleoside) were added. The solids were thoroughly purged with argon. Anhydrous DMF (5 mL) was added via syringe followed by triethylamine (1.5 eq to adamantane methylamide). After stirring at room temperature overnight, the solvent was removed *in vacuo*. The resulting orange/brown oils were purified by flash chromatography (7% MeOH/DCM). The purified products were co-evaporated with anhydrous MeOH and dried under high-vacuum to remove traces of triethylamine.

3.6.4.1. 5-(Propynoic acid (adamantan-1-ylmethyl)-amide)-2'-deoxyuridine, 3a



The product was isolated as a white solid, (74 mg, 63%, $R_f = 0.14$) ^1H NMR (600 MHz, MeOD) δ_{ppm} : 1.54 (br.s., 6H, H-b), 1.68 (d, $J=11.11$ Hz, 3H, H-d), 1.75 (d, $J=11.29$, 3H, H-e), 1.97 (br.s., 3H, H-c), 2.25 (dt, $J=13.32, 6.42$ Hz, 2H, H-2'), 2.95 (br.s., 2H, H-11), 3.70-3.88 (m, 2H, H-5'), 3.93-4.01 (m, 1H, H-4'), 4.36-4.46 (m, 1H, H-3'), 6.22 (t, $J=6.21$ Hz, 1H, H-1'), 8.60, (s, 1H, H-6). ^{13}C NMR (600 MHz, MeOD) δ_{ppm} : 29.77 (C-c), 35.31 (C-a), 38.00 (C-f), 41.22 (C-b), 41.99 (C-2'), 52.40 (C-11), 62.46 (C-5'), 71.89 (C-3'), 78.90 (C-8), 87.58 (C-1'), 87.97 (C-7), 89.36 (C-4'), 97.26 (C-5), 148.67 (C-6), 150.87 (C-2), 155.71 (C-9), 164.49 (C-4). m/z 442.1956 [M-H], $\text{C}_{23}\text{H}_{29}\text{N}_3\text{O}_6$ requires 443.49286.

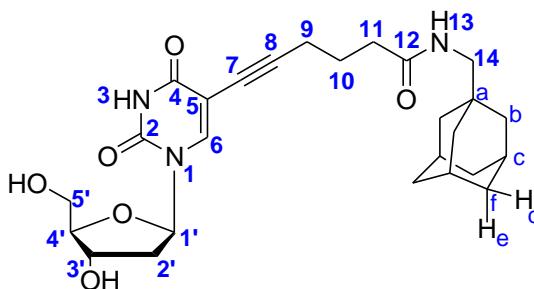
3.6.4.2. 5-(Pentynoic acid (adamantan-1-ylmethyl)-amide)-2'-deoxyuridine, 3b



The product was isolated as a white solid (44 mg, 53%, $R_f = 0.15$) ^1H NMR (600 MHz, MeOD) δ_{ppm} : 1.50 (d, $J=2.30$ Hz, 6H, H-b), 1.60 (d, $J=11.29$ Hz, 3H, H-d), 1.70 (d, $J=12.05$ Hz, 3H, H-e), 1.90 (br.s., 3H, H-c), 2.18 (m, 1H, H-2'), 2.29 (ddd, $J=13.55, 6.02, 3.39$ Hz, 1H, H-2''), 2.47 (t, $J=6.96$ Hz, 2H, H-9), 2.68 (t, $J=6.96$ Hz, 2H, H-10), 2.85-2.92 (m, 2H, H-13), 3.74 (dd, $J=11.95, 3.29$ Hz, 1H, H-5'), 3.80 (dd, $J=11.95, 3.11$ Hz, 1H, H-5''), 3.94 (q, $J=3.20$ Hz, 1H, H-4'), 4.40 (dt, $J=6.21, 3.29$ Hz, 1H, H-3'), 6.23 (t, $J=6.59$ Hz, 1H, H-1'), 8.25 (s, 1H, H-6). ^{13}C NMR (600 MHz, MeOD) δ_{ppm} : 17.29 (C-10), 29.78 (C-c), 35.16 (C-9), 36.16 (C-a), 38.00

(C-f), 41.24 (C-b), 41.75 (C-2'), 52.15 (C-13), 62.66 (C-5'), 72.15 (C-3'), 73.56 (C-7), 86.94 (C-1'), 89.14 (C-4'), 93.63 (C-8), 100.86 (C-5), 144.73 (C-6), 151.18 (C-2), 164.70 (C-4), 174.46 (C-11). m/z (ESI) 472.2467 $[M+H]^+$, $C_{25}H_{33}N_3O_6$ requires 471.5460.

3.6.4.3. 5-(Hexynoic Acid (adamantan-1-ylmethyl)-amide)-2'-deoxyuridine, 3c



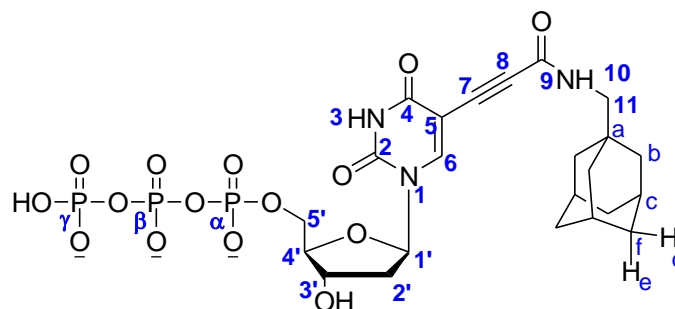
The product was isolated as a glassy colourless solid (83 mg, 30%, $R_f = 0.03$). 1H NMR (600 MHz, MeOD) δ_{ppm} : 1.47-1.53 (m, 6H, H-b), 1.65 (d, $J=11.48$ Hz, 3H, H-d), 1.73 (d, $J=12.05$ Hz, 3H, H-e), 1.86 (quin, $J=7.06$ Hz, 2H, H-10), 1.94 (br.s., 3H, H-c), 2.21 (dt, $J=13.60$, 6.66 Hz, 1H, H-2'), 2.30 (ddd, $J=13.60$, 6.16, 3.76 Hz, 1H, H-2''), 2.42 (m, 4H, H-9 & H-11), 2.86-2.90 (m, 2H, H-14), 3.74 (dd, $J=12.05$, 3.20 Hz, 1H, H-5'), 3.82 (dd, $J=12.05$, 3.01 Hz, 1H, H-5''), 3.93 (q, $J=3.20$ Hz, 1H, H-4'), 4.37-4.43 (m, 1H, H-3'), 6.24 (t, $J=6.49$ Hz, 1H, H-1'), 8.27 (s, 1H, H-6). ^{13}C NMR (600 MHz, MeOD) δ_{ppm} : 19.66 (C9), 25.87 (C10), 29.87 (C-c), 35.25 (C-a), 35.94 (C11), 38.15 (C-f), 41.42 (C-b), 41.84 (C-2'), 52.20 (C-14), 62.67 (C-5'), 72.07 (C-3'), 74.01 (C-7), 86.99 (C-1'), 89.18 (C-4'), 94.27 (C-8), 101.10 (C-5), 144.46 (C-6), 151.31 (C-2), 164.98 (C-4), 175.81 (C-12). m/z (ESI) 508.2433 $[M+Na]^+$, $C_{26}H_{35}N_3O_6$ requires 485.5726.

3.6.5. General procedure for Yoshikawa phosphorylation

The nucleoside (0.2 mmol) and proton sponge (71 mg, 0.3 mmol) were placed in an oven-dried flask then purged with argon before the addition of trimethyl phosphate (2.0 mL). Once the solids had dissolved, the resultant orange/brown clear solution was cooled to between -10 and $-15^\circ C$ on a salt ice bath and $POCl_3$ (30 μL , 0.3 mmol) added via syringe. The disappearance of the nucleoside was followed by TLC (7% MeOH/DCM) or by analytical HPLC. Once the nucleoside was no longer detected, typically after 2-3 h, tributylammonium pyrophosphate solution (0.5 M, 2 mL anhydrous DMF) was added

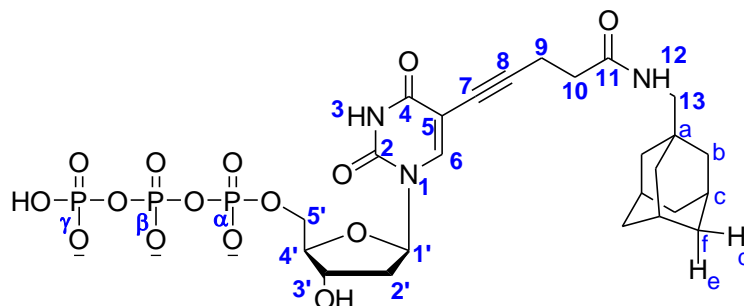
followed by tributylamine (0.16 mL). After 45 min, the reaction was quenched by the addition of ice cold TEAB (0.1 M, 20 mL) and stirred at 0 °C for 30 min. Once warmed to room temperature, the aqueous solution was washed with ethylacetate (3 x 30 mL) then concentrated *in vacuo*. The crude material was purified by preparative HPLC to remove traces of the proton sponge. Fractions containing the phosphorylated material were pooled and the volume reduced to approximately 2 mL. The solution was filtered through using a 0.2 micron spin column the purified further by preparative ion exchange chromatography. Fractions containing the desired triphosphate were pooled, concentrated *in vacuo* then lyophilised overnight to give the triphosphates as triethylammonium salts.

3.6.5.1. 5-(Propynoic acid adamantan-1-ylmethyl)-amide)-2'-deoxyuridine 5'-triphosphate triethylammonium salt, 4a



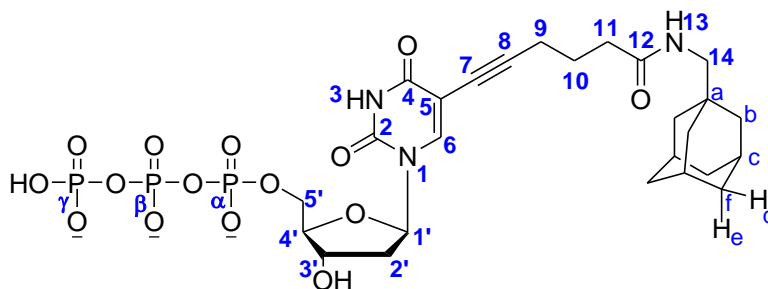
The product was isolated as a white solid (121 mg, 13%). ^1H NMR (600 MHz, D_2O) δ_{ppm} : 1.26-1.31 (m, 9H, $(\text{CH}_3\text{CH}_2)_3\text{N}$), 1.53 (br.s., 6H, H-b), 1.62-1.68 (m, 3H, H-d), 1.68-1.75 (m, 3H, H-e), 1.97 (br.s., 3H, H-c), 2.39-2.51 (m, 2H, H-2'), 2.92-3.02 (m, 2H, H-11), 3.04-3.12 (m, 2H, H-5'), 3.21 (q, $J=7.34$ Hz, 6H, $(\text{CH}_3\text{CH}_2)_3\text{N}$), 4.25 (m, 1H, H-4'), 4.66 (m, 1H, H-3'), 6.27 (t, $J=6.59$ Hz, 1H, H-1'), 8.43 (s, 1H, H-6). ^{13}C NMR (600 MHz, D_2O) δ_{ppm} : 28.61 (C-c), 34.55 (C-a), 36.96 (C-f), 39.74 (C-2'), 40.19 (C-b), 42.89 (C-5'), 52.16 (C-11), 71.12 (C-3'), 87.07 (C-4' & C1'), 97.31 (C-5), 148.40 (C-6), 150.98 (C-2), 155.81 (C-4), 164.83 (C-9). ^{31}P NMR (300 MHz, D_2O) δ_{ppm} : -22.00 (m, 1P, P- β), -10.45 (m, 1P, P- α), 1.26 (m, 1P, P- γ). m/z (ESI) 684.1129 $[\text{M}+\text{H}]^+$, $\text{C}_{23}\text{H}_{32}\text{N}_3\text{O}_{15}\text{P}_3$ requires 683.43256.

3.6.5.2. 5-(Pentynoic acid (adamantan-1-ylmethyl)-amide)-2'-deoxyuridine-5'-triphosphate triethylammonium Salt, 4b



The product was isolated as a white solid (3 mg, 2%). ^1H NMR (600 MHz, D_2O) δ_{ppm} : 1.34 (t, $J=7.15$ Hz, 9H, $(\text{CH}_3\text{CH}_2)_3\text{N}$), 2.37-2.41 (m, 2H, H-2'), 2.48-2.54 (m, 1H, H-2''), 2.56-2.62 (m, 2H, H-9), 2.71-2.76 (m, 2H, H-10), 2.80-2.87 (m, 2H, H-13), 3.19-3.25 (m, 6H, $(\text{CH}_3\text{CH}_2)_3\text{N}$), 3.68-3.72 (m, 2H, H-5'), 3.85-3.90 (m, 1H, H-4'), 4.24-4.30 (m, 1H, H-3'), 6.32 - 6.36 (m, 1H, H-1'), 8.21 - 8.25 (m, 1H, H-6). m/z (ESI) 712.1418 $[\text{M}+\text{H}]^+$, $\text{C}_{25}\text{H}_{36}\text{N}_3\text{O}_{15}\text{P}_3$ requires 711.13593.

3.6.5.3. 5-(Hexynoic Acid (adamantan-1-ylmethyl)-amide)-2'-deoxyuridine-5'-triphosphate triethylammonium salt, 4c

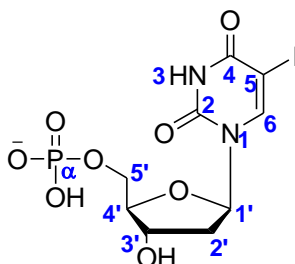


The product was isolated as a white solid (9 mg, 5%). ^1H NMR (600 MHz, D_2O) δ_{ppm} : 1.34 (t, $J=7.25$ Hz, 9H, $(\text{CH}_3\text{CH}_2)_3\text{N}$), 1.47 (m, 6H, H-b), 1.60 (d, $J=12.05$ Hz, 3H, H-d), 1.70 (d, $J=12.05$ Hz, 3H, H-e), 1.8 - 1.91 (m, 2H, H-10), 1.92 (br. s., 3H, H-c), 2.40 (m, 2H, H-2'), 2.47-2.54 (m, 4H, H-9 & H-11), 3.00 (m, 2H, H-14), 3.08 (m, 2H, H-5'), 3.54 (q, $J=7.20$ Hz, 6H, $(\text{CH}_3\text{CH}_2)_3\text{N}$), 4.22 (m, 1H, H-4'), 4.65 (m, 1H, H-3'), 6.31 (t, $J=6.87$ Hz, 2H, H-1'), 8.12 (s, 1H, H-6). ^{13}C NMR (600 MHz, D_2O) δ_{ppm} : 18.62 (C-9), 24.09 (C-10), 28.61 (C-c), 34.12 (C-a),

35.11 (C-11), 36.96 (C-f), 39.27 (C-2'), 40.23 (C-b), 42.91 (C-5'), 49.39 (C-14), 71.36 (C-3'), 72.51 (C-7), 86.24 (C-4'), 95.80 (C-8), 100.97 (C-5), 144.11 (C-6), 151.21 (C-2), 165.56 (C-4), 176.76 (C-12). ^{31}P NMR (300 MHz, D_2O) δ_{ppm} : -21.53 (br.s., 1P, P- β), -9.93 (br.s., 1P, P- α), 1.28 (br.s., 1P, P- γ). m/z (ESI) 724.1418 $[\text{M}-\text{H}]^-$, $\text{C}_{26}\text{H}_{38}\text{N}_3\text{O}_{15}\text{P}_3$ requires 725.15158

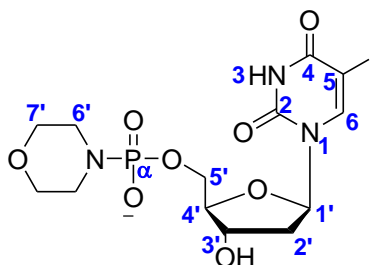
3.6.6. Phosphorylation via the Moffatt-Khorana approach

3.6.6.1. Synthesis of 5-iodo-2'-deoxyuridine-5'-monophosphate, 10



To a chilled ($-3\text{ }^\circ\text{C}$) suspension of iodouridine (250 mg, 0.7 mmol) and proton sponge (908 mg, 4.2 mmol) in anhydrous acetonitrile (5 mL), POCl_3 (260 μL , 2.6 mmol) was added. The resultant pale orange solution was stirred for 6 h at $-3\text{ }^\circ\text{C}$ at which point, TLC analysis indicated that the starting material had been consumed. The reaction mixture was quenched with 50 mL ice cold 0.1 M TEAB buffer. After stirring at room temperature for 1 h, the solution was washed with ethylacetate (4 x 60 mL) and concentrated *in vacuo*. The crude material was purified by preparative HPLC followed by ion exchange purification to yield the title compound as a colourless glassy solid (63 mg, 21%); (IPA: NH_4OH : H_2O , 6 : 1 : 3; R_f starting material = 0.47; R_f product = 0.12). ^1H NMR (600 MHz, D_2O) δ_{ppm} : 1.29 (t, $J=7.25$ Hz, 9H, (CH_3CH_2) $_3\text{N}$), 2.34-2.44 (m, 2H, H-2'), 3.21 (q, $J=7.30$ Hz, 6H, (CH_3CH_2) $_3\text{N}$), 4.04-4.11 (m, 2H, H-5'), 4.19-4.22 (m, 1H, H-4'), 4.57 (dt, $J=5.98, 3.13$ Hz, 1H, H-3'), 6.29 (t, $J=6.87$ Hz, 1H, H-1'), 8.28 (s, 1H, H-6). ^{13}C NMR (600 MHz, D_2O) δ_{ppm} : 39.67 (C-2'), 65.29 (C-5'), 68.34 (C-5), 71.68 (C-3'), 86.39-86.47 (C-1' & C-4'), 146.84 (C-6), 152.05 (C-2), 163.87 (C-4). ^{31}P NMR (300 MHz, D_2O) δ_{ppm} : 1.76 (s, 1P, P α). m/z (ESI) 432. 9298 $[\text{M}-\text{H}]^-$, $\text{C}_9\text{H}_{11}\text{N}_2\text{O}_8\text{P}^-$ requires 433.07102.

3.6.6.2. Synthesis of 5-iodo-2'-deoxyuridine-5'-phosphoromorpholidate, **11**



10 was coevaporated with anhydrous DMSO (3 X 2 mL) then suspended in anhydrous DMSO (2 mL). To the solution, morpholine (30 μ L, 308 μ mol) was added followed by 2, 2'-dithiopyridine (34 mg, 153 μ mol) 5 min later. After a further 5 min, triphenylphosphine (41 mg, 153 μ mol) was added causing the solution to turn from colourless to bright yellow. The solution was stirred at room temperature for 1 h after which TLC analysis suggested that the reaction has gone to completion. The title compound was precipitated from solution by the addition of NaI (0.1 M in acetone, 10 mL). The suspension was centrifuged at 4 $^{\circ}$ C, 3220 rcf for 10 min. The supernatant was discarded and replaced with fresh NaI solution. The pellet was resuspended into solution then recentrifuged to the pellet. This purification process was performed a total of three times. The pellet was subsequently dried under vacuum to remove traces of the solvent, giving rise to the title compound, a white solid, as the sodium salt. (14 mg, 70%); (IPA:NH₄OH:H₂O, 6 : 1 : 3; R_f 0.41). ¹H NMR (600 MHz, D₂O) δ_{ppm} : 2.38 (ddd, $J=14.16, 8.23, 6.21$ Hz, 1H, H-2'), 2.58 (ddd, $J=14.07, 5.98, 1.98$ Hz, 1H, H-2''), 3.05-3.12 (m, 4 H, H-6'), 3.73 (t, $J=4.42$ Hz, 4H, H-7'), 4.00-4.09 (m, 2H, H-5'), 4.36 (m, 1H, H-4'), 4.77 - 4.83 (m, 1H, H-3'), 6.32 (t, $J=8.40$ Hz, 1H, H-1'), 8.21 (s, 1H, H-6'). ¹³C NMR (600 MHz, D₂O) δ_{ppm} : 39.34 (C-2'), 45.41 (C-6'), 65.01 (C-5'), 67.55 (C-7') 70.44 (C-3'), 85.74-85.83 (C-1' & C-4'), 146.11 (C-6). ³¹P NMR (300 MHz, D₂O) δ_{ppm} : 8.43 (br.s., 1P, P- α). m/z (ESI) 501.9987 [M-H]⁻, C₁₃H₁₈IN₃O₈P⁻ requires 502.17610.

3.6.7. Primer extension assay

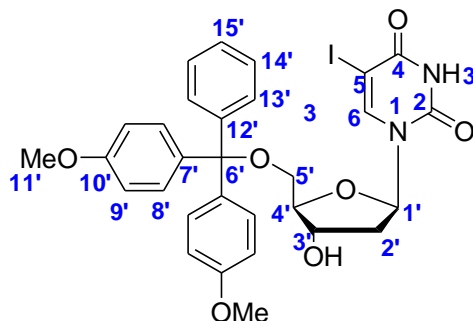
Samples containing 150 pmol of DNA oligonucleotide primer of sequence 5'-ATG GGA CTA ACT AAT CTT TGC TTA-3' and 170 pmol of template 5'-TAT TCC ATG TGA TAA GCA AAG ATT

AGT TAG TCC CAT-3' were added polymerase extension mix. The mix contained 1U Deep Vent_R exo⁻, 1X ThermopolTM buffer (from New England Biolabs; provided with the enzyme) and 40 mM of natural dNTP nucleotides or dUTP derivatives in a total volume of 30 µL. The reaction mixtures were heated to 85 °C for 2 min to denature DNA duplexes, and incubated at 57 °C for 1.5 h to anneal and extend the primers, followed by immediate chilling on ice. The extension products were visualised by denaturing PAGE (18%) using 4–5 µL aliquots of the mixtures denatured by addition of formamide and subsequent heating at 95 °C. The formamide denaturing solution was composed of: aqueous bromophenol blue (0.01%, 100 µL), ThermopolTM buffer (1 mL of 10X solution), formamide (8.8 mL) and aqueous EDTA (100 mM, 100 µL).

3.6.8. Preparation of 18% polyacrylamide gels for electrophoresis

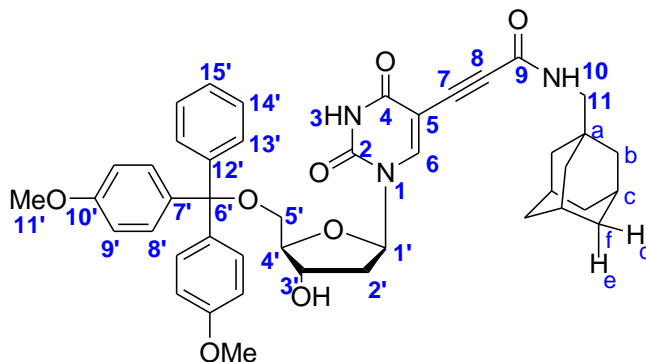
To urea (6.3 g), 10x TBE buffer (Tris-Borate-EDTA buffer, 1.5 ml), formamide (2.75 ml) and bis-acrylamide (6 ml, 40%) were added. The resultant solution was heated at 37 °C and shaken at regular intervals over a 15 minute period. Once the urea had dissolved, ammonium persulfate solution (75 µL) and TEMED (N,N,N',N'-tetramethyl ethylenediamine, 7.5 µL) were added to the preparation and thoroughly mixed. The gel solution was pipetted between glass plates and a comb added. The plates were maintained at 37 °C for 20 min to facilitate polymerisation. After this time they were allowed to cool to room temperature before being refrigerated. Primer extension products were visualised using ethidium bromide and recorded with a Ultra-Violet CCD camera.

3.6.9. Synthesis of 5'-(4,4'-dimethoxytrityl)-5-iodo-2'-deoxyuridine, 14



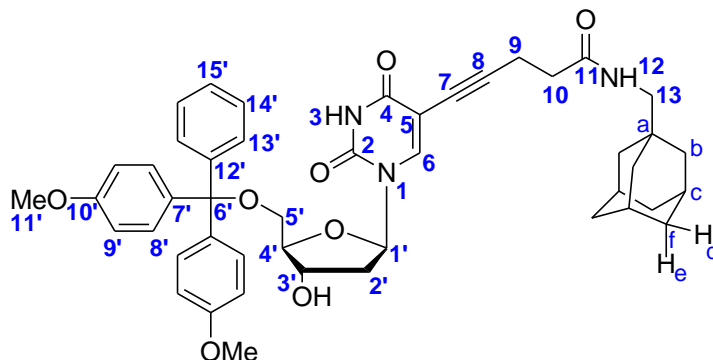
5-Iodo-2'-deoxyuridine (540 mg, 1.5 mmol) was co-evaporated with anhydrous pyridine (3 x 4 mL). To the resultant residue, DMAP (9 mg, 0.07 mmol, 0.05 eq) was added and the solids purged with argon. The solids were dissolved in anhydrous pyridine (4 mL) then Et₃N added (280 μ L, 2.0 mmol, 1.4 eq). The mixture was allowed to stir for 10 min before the addition of DMT-Cl (671 mg, 2.0 mmol, 1.4 eq) in three equal portions over three hours. After an additional 1 h stirring, MeOH (1 mL) was added to consume any excess DMT-Cl. After stirring for an additional 15 min, the reaction mixture was concentrated *in vacuo*. The resultant residue was re-suspended in DCM (20 mL), washed with sat. NaHCO₃ (3 x 20 mL) and dried over MgSO₄ then concentrated to give the crude material. Purification by flash chromatography (silica deactivated with Et₃N; eluent 4% MeOH/DCM + 1% Et₃N, R_f = 0.46) afforded an off-white solid. The solid was co-evaporated with HPLC grade MeOH (2 x 20 mL) and dried *in vacuo* overnight to remove any residual Et₃N (629 mg, 68%). ¹H NMR (600 MHz, MeOD) δ_{ppm} : 2.29-2.41 (m, 2H, H-2'), 3.34-3.35 (m, 2H, H-5'), 3.78 (s, 6H, OMe), 4.04 (q, *J*=3.20 Hz, 1H, H-4'), 4.46 (dt, *J*=5.88, 2.80 Hz, 1H, H-3'), 6.22 (t, *J*=7.62 Hz, 1H, H-1'), 6.86-6.89 (m, 4H, H-9'), 7.20-7.23 (m, 1H, H-15'), 7.30 (t, *J*=7.72 Hz, 2H, H-13'), 7.34-7.37 (m, 4H, H-8'), 7.44-7.47 (m, 2H, H-14'), 8.19 (s, 1H, H-6). ¹³C NMR (600 MHz, MeOD) δ_{ppm} : 41.61 (C-2'), 55.41 (OMe), 63.58 (C-5'), 68.74 (C-5), 72.56 (C-3'), 85.63 (C-1'), 86.52 (C-4'), 87.17 (C-6'), 113.50 (C-9'), 127.22-130.21 (C-8', C-13', C-14' & C-15'), 135.35 (C-7' & C-12'), 144.40 (C-6), 149.95 (C-2), 158.78 (C-10'), 160.181 (C-4). *m/z* (ESI) 679.0895 [M+Na]⁺, C₃₀H₂₉IN₂O₇ requires 679.4649.

3.6.10. Synthesis of 5'-(4,4'-dimethoxytrityl)-5-(propynoic acid-(Adamantan-1-ylmethyl))-amide-2'-deoxyuridine, 15a



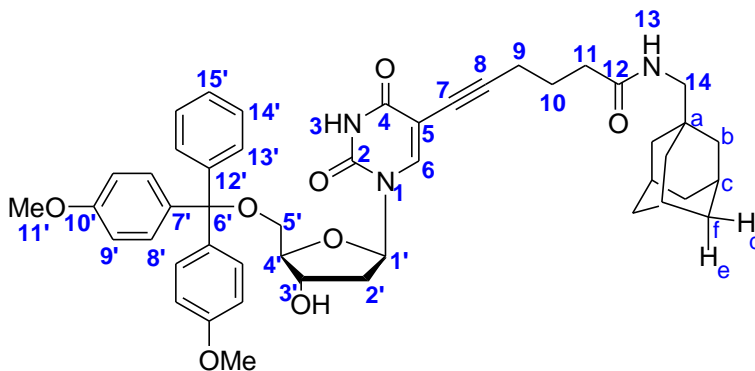
Prepared via the Sonogashira cross-coupling reaction described above. The crude material was purified twice by flash chromatography (silica-gel type-G/type-H, 15:1, w/w). The silica was neutralised with Et₃N (1 mL per 5 g silica). Eluent 0-2.5% MeOH/DCM + 0.1% Et₃N. The isolated product was co-evaporated with MeOH (3 x 15 mL) then dried *in vacuo* to remove the residual triethylamine. The title compound was isolated as an off-white solid (53 mg, 71%, (4% MeOH/DCM) R_f = 0.25). ¹H NMR (600 MHz, MeOD) δ_{ppm}: 1.47 (d, J=2.26 Hz, 6H, H-b), 1.63 (d, J=11.48 Hz, 3H, H-d), 1.72 (d, J=12.05 Hz, 3H, H-e), 1.93 (br. s., 3H, H-c), 2.41 (dd, J=13.65, 6.87 Hz, 1H, H-2'), 2.52 (ddd, J=13.74, 6.21, 3.01 Hz, 1H, H-2''), 2.85-2.92 (m, 2H, H-11), 3.27 (dd, J=10.73, 3.01 Hz, 1H, H-5'), 3.38 (dd, J=10.82, 3.86 Hz, 1H, H-5''), 3.75 (d, J=3.01 Hz, 6H, OMe), 4.10 (q, J=3.33 Hz, 1H, H-4'), 4.46 (dt, J=5.88, 2.99 Hz, 1H, H-3'), 6.13 (t, J=6.49 Hz, 1H, H-1'), 6.83-6.89 (m, 4H, H-9'), 7.18 (t, J=7.30 Hz, 1H, H-15'), 7.26-7.29 (m, 2H, H-13'), 7.30-7.34 (m, 4H, H-8'), 7.39-7.43 (m, 2H, H-14'), 8.25 (s, 1H, H-6). ¹³C NMR (600 MHz, MeOD) δ_{ppm}: 29.73 (C-c), 35.29 (C-a), 37.96 (C-f), 41.21 (C-b), 42.44 (C-2'), 47.86 (C-11), 52.32 (s), 53.59 (s), 55.77-55.79 (OMe), 64.93 (C-5'), 72.94 (C-3'), 78.35 (C-7), 88.25-88.70 (C-1', C-4', C-6', & C-8), 97.20 (C-5), 114.32 (C-9'), 127.98-131.41 (C-8', C-13', C-14' & C-15'), 136.62-137.13 (C7' & C-12'), 146.14 (C-6), 147.68 (C-2'), 155.30 (C-9), 160.22 (C-10'), 164.47 (C-4). *m/z* (ESI) 746.3385 [M+H]⁺, C₄₄H₄₇N₃O₈ requires 745.85928.

3.6.11. Synthesis of 5'-(4,4'-dimethoxytrityl)-5-(pentynoic acid-(adamantan-1-ylmethyl))-amide-2'-deoxyuridine, **15b**



Prepared as **15a**. The product was isolated as a white solid. (84 mg, 56% isolated yield, (4% MeOH/DCM) $R_f = 0.28$). ^1H NMR (600 MHz, MeOD) δ_{ppm} : 1.43-1.48 (m, 6H, H-b), 1.56 (d, $J=11.67$ Hz, 3H, H-d), 1.66 (d, $J=12.23$ Hz, 3H, H-e), 1.87 (br.s., 3H, H-c), 2.25 (td, $J=7.25$, 3.01 Hz, 2H, H-2'), 2.36-2.43 (m, 2H, H-13), 2.79 (m, 2H, H-9), 2.92 (m, 2H, H-10), 3.28 (dd, $J=10.73$, 2.82 Hz, 1H, H-5'), 3.33-3.37 (m, 1H, H-5''), 3.70-3.76 (s, 6H, OMe), 4.01-4.06 (m, 1H, H-4'), 4.49-4.54 (m, 1H, H-3'), 6.21 (t, $J=6.59$ Hz, 1H, H-1'), 6.81-6.87 (m, 4H, H-9'), 7.17-7.22 (m, 1H, H-15'), 7.25-7.31 (m, 2H, H-13'), 7.35 (d, $J=8.47$, 4H, H-8'), 7.44-7.48 (m, 2H, H-14'), 8.09 (s, 1H, H-6). ^{13}C NMR (600 MHz, MeOD) δ_{ppm} : 29.75 (C-c), 35.16 (C-a), 38.00 (C-f), 41.21 (C-b), 42.18 (C-2'), 52.21-53.57 (C-9, C-10, & C-13), 55.85 (OMe), 64.84 (C-5'), 72.74 (C-3'), 73.19 (C-7), 87.23 (C-1'), 89.68 (C-4'), 94.12 (C-8), 101.64 (C-5), 114.38 (C-9'), 127.98-131.43 (C-8', C-13', C-14' & C-15'), 136.87-137.18 (C-7' & C-12'), 143.97 (C-6), 151.05 (C-2), 160.20 (C-10'), 164.74 (C-4), 174.19 (C-11). m/z (ESI) 796.3518 $[\text{M}+\text{Na}]^+$, $\text{C}_{46}\text{H}_{51}\text{N}_3\text{O}_8$ requires 773.36762.

3.6.12. Synthesis of 5'-(4,4'-dimethoxytrityl)-5-(hexynoic Acid-(adamantan-1-ylmethyl))-amide-2'-deoxyuridine, **15C**^{ix}



Prepared as **15a**. (145 mg, 34% isolated yield). ¹H NMR (600 MHz, CDCl₃) δ_{ppm}: 1.43-1.48 (m, 6H, H-b), 1.56 (d, *J*=11.67 Hz, 3H, H-e), 1.66 (d, *J*=12.23 Hz, 3H, H-d), 1.87 (br.s., 3H, H-c), 2.25 (td, *J*=7.25, 3.01 Hz, 2H, H-2'), 2.36-2.43 (m, 2H, H-13), 2.79 (dd, *J*=13.36, 6.59 Hz, 2H, H-9), 2.92 (dd, *J*=13.36, 6.59 Hz, 2H, H-10), 3.28 (dd, *J*=10.73, 2.82 Hz, 1H, H-5'), 3.33-3.37 (m, 1H, H-5''), 3.70-3.76 (s, 6H, OMe), 4.01-4.06 (m, 1H, H-4'), 4.49-4.54 (m, 1H, H-3'), 6.21 (t, *J*=6.59 Hz, 1H, H-1'), 6.81-6.87 (m, 4H, H-9'), 7.17-7.22 (m, 1H, H-15'), 7.25-7.31 (m, 2H, H-13'), 7.35 (d, *J*=8.47, 4H, H-8'), 7.44-7.48 (m, 2H, H-14'), 8.09 (s, 1H, H-6).

^{ix} Synthesised by Xiu Yao.

4 PHOTOCAGING OF 5-AZACYTIDINE FOR DRUG DELIVERY PURPOSES

4.1. Summary

This chapter is concerned with the synthesis of photocaged derivatives of azacytidine. Azanucleosides are a class of cytosine analogues which are used to treat cancers and bone-marrow disorders that arise due to aberrant methylation of particular regions of the genome. Hypermethylation of these regions abnormally enhances the function of some genes, the oncogenes, or cause others, tumour-suppressor genes, to lose function or be 'silenced'.⁴⁵ This can enable proliferation of certain cancers and the onset of disease. Azanucleosides are able to reverse aberrant hypermethylation through two mechanisms. Firstly, by inducing a hypomethylated state through inhibition of DNA methyltransferase (DNMT), the enzyme which catalyses DNA methylation. Secondly, through direct incorporation into replicating DNA and RNA triggering cell death. To date, two drugs have been licensed for clinical use.^{191,192} However, they are associated with extreme toxicity and side-effects. It is envisaged that adverse effects could be mitigated with greater spatial and temporal control over drug release. Here, the synthesis of azacytidine analogues, whose biological activity is masked by photolabile protecting groups, is described. In cellular assays it was observed that biological activity of azacytidine was restored upon irradiation with UV light.

4.2. Introduction

4.2.1. DNA Methyltransferases and their Inhibition by Azanucleosides

DNA methylation is one of several mechanisms through which the activity or function of a gene can be altered without change to its primary nucleotide sequence.¹⁹³ This form of modification is termed epigenetic, and plays a crucial role in the development and function of many organisms. In humans and other mammals, DNA methylation occurs only on cytosines that precede a guanosine in the DNA sequence i.e. at CpG dinucleotides (where the 'p' stands for the phosphodiester bond between the cytosine and the guanine,

and indicates that they are next to each other in the DNA strand).⁴⁵ The distribution of these dinucleotides is far less random than one would expect. Within the genome, they are typically found clustered in small stretches of DNA termed CpG islands.^{45,194} CpG islands are defined as regions of at least 500 base pairs in length with a GC content greater than 55%,^{43,195} and are often found within the promoter regions of a gene. These are regions where transcription begins.¹⁹⁴ In a normal cell, CpG islands which are associated with a promoter region, are unmethylated which enables the DNA to be recognised by transcription factors and associated proteins.^{194,196} Conversely, CpG dinucleotides which are found outside of promoter regions are heavily methylated.¹⁹⁶ It is thought that the methyl groups prevent the binding of proteins and cofactors necessary for transcription, and thus renders the DNA transcriptionally inert or 'silent'.¹⁹⁴ Heavy methylation of these regions serves as a protective measure to prevent the transcription of regions which may contain nucleotide repeat sequences (as described in chapter 3), inserted viral sequences, and transposons (DNA sequences that have moved from their usual location into a new region of the genome), all of which could prove lethal to a cell.^{45,197} Consequently, failure to regulate DNA methylation and the occurrence of aberrant methylation patterns, can lead to the over- or under-expression of particular proteins and is associated with several human diseases including psychiatric diseases, diseases of the immune system and the onset of cancer.^{44,198}

The link between DNA methylation patterns and cancer was established over 20 years ago. Seminal studies showed that the occurrence of methylated cytosine throughout the genome of tumour cells was significantly lower than levels observed in normal cells.^{199,200} It was soon demonstrated, and is now well accepted, that many cancers are characterised by hypomethylation in regions where most CpG dinucleotides should be methylated.¹⁹⁹ In addition, hypermethylation of CpG islands in gene promoter regions is also a common trait.^{45,201} Loss of DNA methylation can ultimately lead to the transcription of normally silent regions of the genome and the expression of potentially harmful proteins. Conversely, methylation of CpG islands is associated with silencing of

transcription and the inactivation of genes, including tumour suppressor genes which would ordinarily counteract the proliferation of cancer cells.²⁰¹

Methylation of cytosine in a CpG context is catalysed by DNA methyltransferase (DNMT) enzyme. There are five DNMT isoforms: DNMT1, DNMT2, DNMT3a, DNMT3b and DNMT3L.¹⁹² However, only DNMT1, 3a and 3b are involved in the methylation of DNA.^{44,198,192} DNMT1 is a maintenance methyltransferase, methylating newly synthesised DNA strands after each round of DNA replication.^{44,198,192} This isoform is thought to be responsible for maintaining abnormal promoter methylation.⁴⁵ DNMT3a and DNMT3b are *de novo* methyltransferases, initiating methylation of unmethylated DNA in early embryonic development.^{44,198,192,194} The catalytic mechanism of DNMT is schematically represented in Figure 50A.

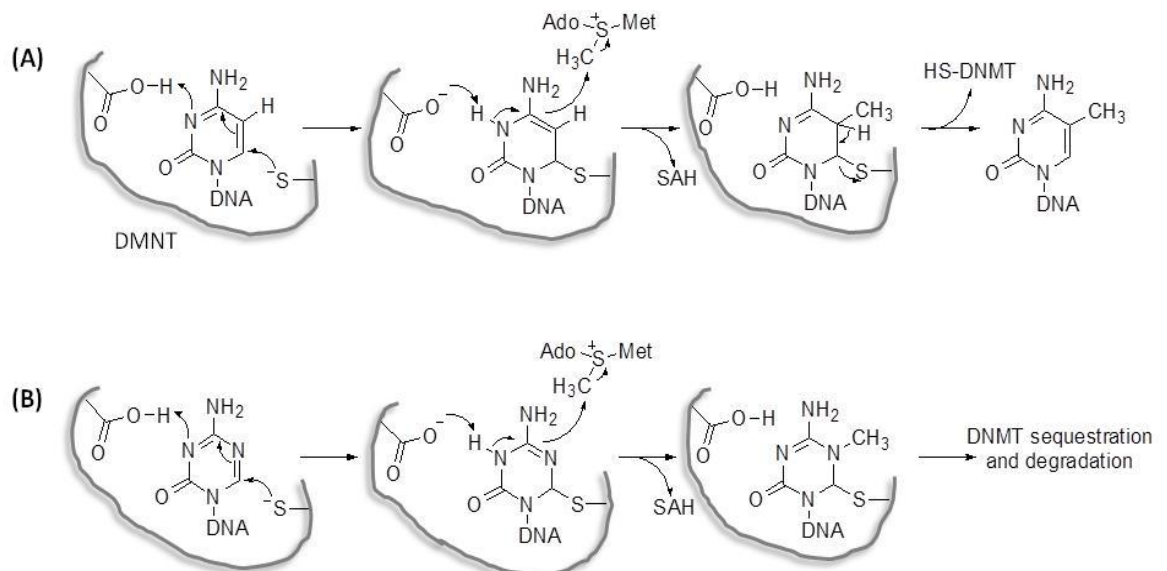


Figure 50: (A) Mechanism of DNA methylation by DNA methyltransferase (DNMT). (B) Mechanism of DNMT inhibition by azanucleosides. The additional nitrogen at C5 prevents β -elimination and reformation of the C5-C6 double bond. This results in a permanent covalent bond between the nucleoside and the enzyme. The DNA-DNMT adduct triggers apoptosis. Abbreviations: Ado, adenosyl; Met, methionine; SAH, S-adenosyl-L-homocysteine. Adapted from [191].

DNMTs transfer a methyl group from S-adenosyl-methionine (SAM), an enzyme co-factor, to the C5 position of the pyrimidine ring of cytosine. The reaction proceeds via a nucleophilic attack on position C6 of the pyrimidine ring by a sulphur atom of a cysteine residue located within the active site of the enzyme. This results in an increase in the

electron density at position C5 and concomitant attack on the methyl group of SAM.²⁰² Abstraction of the proton from N3 followed by β -elimination then allows reformation of the C5-C6 double bond and release of the enzyme and DNA with methylated cytosine.

Given that aberrant DNA methylation is associated with many diseases and cancers,²⁰³ DNMT represents a valid therapeutic target. The main difference between genetic mutations which involve alteration of the nucleotide sequence of the genome, and epigenetic processes, is that the former is irreversible whilst the latter is potentially reversible.^{43,45} Indeed several chemical agents have been shown to inhibit or reverse DNA methylation and effectively reactivate genes which have undergone epigenetic silencing.^{6,9,14,17} DNMT inhibitors fall into two categories; nucleoside analogues which are typically derivatives of cytidine and thus can be incorporated into genetic material, and non-nucleoside analogues which are structurally different from cytidine and elicit their effects through alternative mechanisms. To date, two nucleoside analogues have been approved for clinical use against certain bone-marrow disorders and leukaemia: 5-azacytidine (Vidaza™, azacytidine) and 5-aza-2'-deoxycytidine (Dacogen™, decitabine). These compounds are nucleoside analogues in which the carbon atom at position 5 of the pyrimidine ring is replaced by a nitrogen atom and linked to a ribose and deoxyribose sugar, respectively (Figure 50B). Once administered, these analogues are transported into cells by the same facilitated nucleoside transport system that carries uridine and cytidine.²⁰⁵ The intracellular nucleosides undergo three successive phosphorylation reactions to their triphosphorylated active forms. Azacytidine is incorporated into both RNA and DNA (after deoxy-conversion), whilst the decitabine analogue is only incorporated into DNA.⁴⁴ The nucleosides replace cytidine during DNA and RNA replication mediated by polymerase enzymes. As a result, they are only effective in cells which are actively dividing and have little effect on cells which are non-dividing or in the resting phase of their life cycle.^{44,205} This imparts some specificity towards cancerous cells which typically proliferate at a greater rate than normal cells.²⁰⁶ Once incorporated into DNA, the analogues which are located within the CpG islands are recognised by DNMT as though they were the natural nucleoside and react in a similar fashion to cytidine. However, due

to the presence of the nitrogen at C5, the β -elimination step cannot occur leaving an irreversible covalent bond between C6 and the catalytic cysteine of the enzyme (Figure 50B). Sequestration of the enzyme by the DNA results in a rapid loss of activity and a reduction in methylation with each cycle of replication.²⁰² This ultimately leads to a hypomethylated state and reactivation of silenced genes including those which were selected for during tumorigenesis. Furthermore, the presence of the DNMT-DNA adduct itself is thought to be cytotoxic and mutagenic triggering DNA repair with subsequent degradation of the DNMT or apoptosis.^{43,44} Azacytidine is also able to elicit cytotoxic effects through incorporation into RNA and interfering with protein and nucleic acid synthesis.^{202,205}

Although azanucleosides exhibit some specificity towards cancerous cells, they are associated with some severe side-effects, indicating that they adversely affect other cellular processes. These non-specific effects have been attributed to the incorporation of the nucleoside into rapidly dividing, normal cells of the body. For instance, myelotoxicity, the most acute side-effect, is due to the adverse effects on bone marrow and blood cell production. Such side effects have limited the clinical use of azanucleosides.⁴⁵

It is envisaged that the toxicity of azanucleosides could be mitigated by having greater spatial and temporal control over drug release. Essentially, by constraining drug release to cancerous tissue and minimising the exposure of healthy tissue, the side-effects associated with azanucleosides could be reduced. Such control could be achieved by administration of a 'caged' derivative of azanucleosides with subsequent release of the active drug being mediated by UV light.

4.2.2. Principles of Photocaging

Caged compounds are those which have been rendered biologically inert by chemical modification of a key functional group with a photolabile protecting group.^{207,208} Restoration of the biological activity, or 'uncaging', is achieved by irradiation with light of a particular wavelength resulting in photolysis and removal of the photolabile group.^{207,208}

This results in a concentration spike of the unmodified active molecule which can proceed to exert its biological affect. Within the biological arena, light is often considered to be the ideal trigger as it is orthogonal to the activity of the cell.^{207,209} With the exception of highly specialised cells such as those of the eye or specific plant cells, cells do not react to light.²⁰⁸ Consequently, uncaging of bioactive molecules can be achieved specifically. Furthermore, as light is able to penetrate cell membranes, cellular processes can be directly regulated by uncaging within intracellular compartments. Similarly, uncaging can also be triggered in the extracellular space.²¹⁰ With the use of specialised equipment, such as lasers, it is possible to manipulate light very precisely.^{208,211} For instance, light can be emitted as a focused beam and used to illuminate very specific areas of the cytosol or tissue, or conversely, applied to entire cells and larger areas of the body. Light can also be modulated with respect to time and amplitude.²⁰⁸ It can be applied in a single burst resulting in rapid uncaging, or pulsed for gradual uncaging of a compound. In essence, using light to uncage an azanucleosides would enable one to control the location, time and dosage of the drug once it has entered the cell.

The classical strategy towards the design and synthesis of caged bioactive compounds begins with the identification of functional groups which are crucial to their activity.^{24,212} Modification of these positions with a photolabile protecting group should prevent the formation of key interactions and thus render them biologically inert. Typically, the activity of a bioactive molecule is dependent upon an amino, hydroxyl, carboxylate or phosphate group.²¹² Therefore, the most common conjugation strategies exploit the nucleophilicity of these groups towards an electrophilic aliphatic centre. Such strategies are straightforward, have fair to excellent yields and can be applied in mild conditions; important factors when dealing with sensitive bioactive compounds.²¹² The solubility of the caged compound is also a factor which must be considered. In order to be used in an *in vivo* study, the caged compound must be soluble in aqueous conditions. This is traditionally quite difficult to achieve as photolabile protecting groups are generally hydrophobic chromophores.^{17,208,213} Thus when conjugated to a bioactive compound which is only moderately soluble to begin with, the resultant caged compound may only

be sparingly soluble.²⁰⁸ If the caged compound is to be used intracellularly, then its solubility will also dictate the ease with which it is taken up by the cell. For instance, in order to diffuse through the cell membrane passively, it must be sufficiently hydrophobic. Thus there is a fine line between the solubility required for use in physiological conditions and the hydrophobicity required for passive cellular uptake. This hurdle may potentially be overcome if the caged compound is able to exploit protein channels that exist in the cell membrane and be taken up by active or facilitated transport. The photochemical properties of the photolabel are also important. Ideally, uncaging should occur at wavelengths which are not too short (>300 nm) in order to prevent damage to the cell.²¹² Additionally, the by-products of uncaging, which are released stoichiometrically,²⁰⁹ should be non-toxic. Several photolabile protecting groups meet these criteria.

o-Nitrobenzyl (NB) alcohol derivatives are the most widely used photo-labile protecting groups for bioactive molecules.^{207,212} The structure of *o*-nitrobenzyl and some commonly used derivatives are shown in Figure 51. Their popularity is attributed to their commercial availability and the ease with which they can be introduced into bioactive molecules.

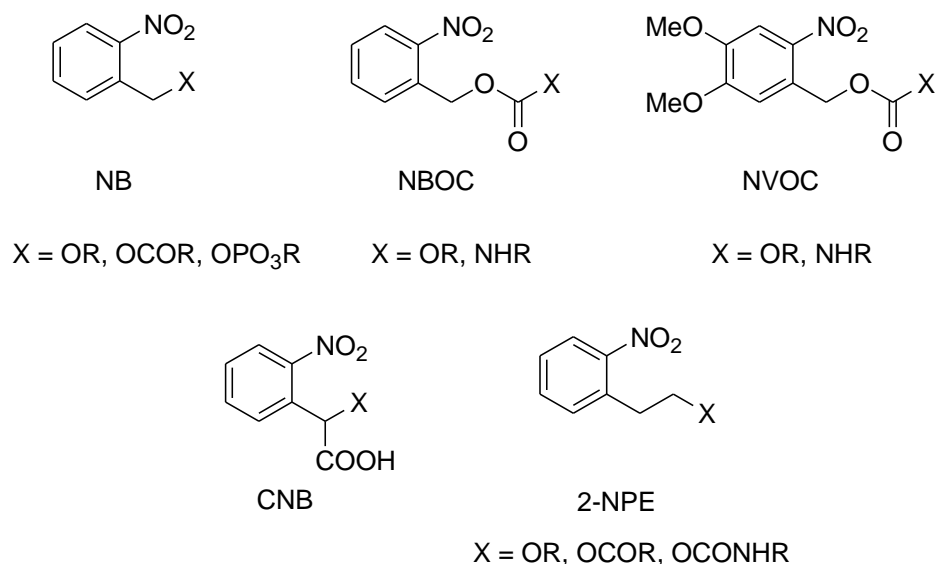


Figure 51: Commonly used *o*-nitrobenzyl photolabile protecting groups.

These are typically introduced as the chloroformate thus can be used to protect amines or alcohols as carbamates, carbonates and esters. Furthermore, their physical and

photochemical properties can be readily tuned. For instance, adding two methoxy groups to the aromatic ring of nitrobenzyloxycarbonyl (NBOC) gives rise to a 6-nitroveratryloxycarbonyl (NOVC,) derivative, which absorbs at longer wavelengths thus is of more use in biological applications.²⁰⁹ Indeed, NOVC is undoubtedly the most popular photolabel in this class of compounds. The water solubility of NB derivatives can be increased by the introduction of a carboxylic acid functional group to the benzylic carbon giving rise to α -carboxy-3-nitrobenzyl (CNB). Although preparation of CNB is marginally more complex (the synthesis is achieved in four synthetic steps),^{214,215} it is particularly valuable for the caging of carboxylate functionalities, especially those of neurotransmitters.^{216,210,212} Indeed, several CNB-caged neurotransmitters are now commercially available. Furthermore, CNB-caged neurotransmitters have been shown to have accelerated rates of uncaging and higher quantum yields compared to their NB- and NPE-caged derivatives.^{212,210,208} The [2-(2-nitrophenyl)ethoxy]carbonyl (NPE) photolabel is the α -substituted derivative of NB. NPE caged compounds have also been shown to have higher rates of uncaging compared to NB caged compounds.^{212,217}

4.2.3. Photodynamic Therapy

Precedence for the activation of prodrugs by UV light has been demonstrated in the field of photodynamic therapy (PDT). PDT is a photochemical approach in which a light-activatable chemical, termed a photosensitizer (PS), and light of the appropriate wavelength are utilised to impart cytotoxicity.²¹⁸ A schematic representation of how PDT might be used in a clinical setting is shown in Figure 52.

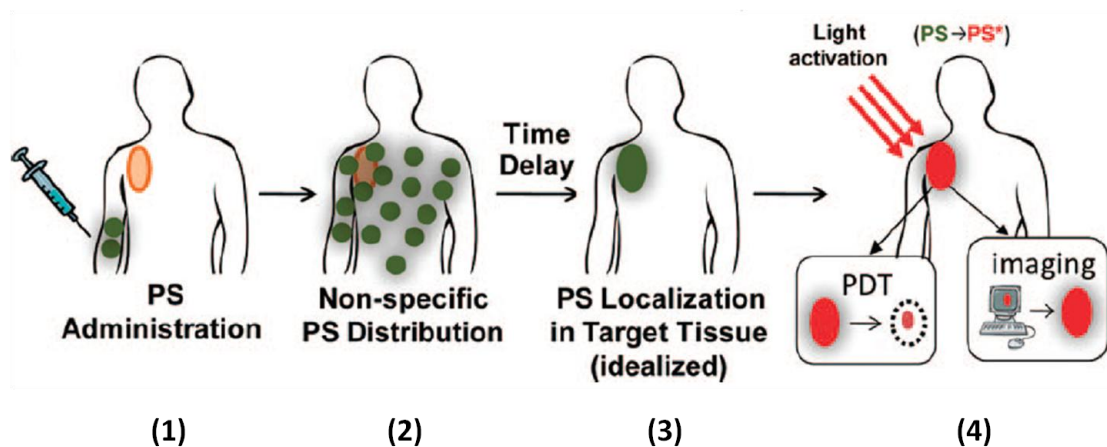


Figure 52: Schematic representation of photodynamic therapy (PDT). The photosensitizer (PS, green) is a photoactivatable chemical which can act as both an imaging and chemotherapeutic agent. Upon administration (1), the PS enters the systemic circulation and is widely distributed (2). The PS preferentially accumulates within the tumour tissue (3). Alternatively, the PS can be formulated for topical administration to enable direct application to diseased tissue. Irradiation of the tumour or a specific area results in fluorescence emission which enables the tumour to be clearly located. Additionally, a singlet oxygen (PS^*) is generated which results in cytotoxicity. Adapted from [218].

In essence, PDT involves the administration of a PS which localises to a target cell or tissue. Illumination of a specific anatomical region with UV light results in the activation of the PS and localised cytotoxicity. The putative cytotoxic agent in PDT is singlet oxygen.²¹⁹ The photochemical reaction through which the singlet oxygen is generated is schematically represented in Figure 53. In brief, following absorption of light of an appropriate wavelength, the PS is excited from its ground state to an excited singlet state (A). From here, the PS can either relax back down to the ground state by emitting a fluorescent photon (B) or enter the excited triplet state (C). Again, there is the opportunity for PS to relax back to the ground state by emitting a phosphorescent photon (D). Alternatively, the PS can transfer energy to other molecules via a radiationless transition.^{219,220} The transfer of an electron to a substrate molecule such as the cell membrane, results in the formation of free radicals. These radicals can in turn react with oxygen to produce oxygenated byproducts (Type I PDT) (F).²²¹ Alternatively, the energy can be transferred through a direct reaction with oxygen (Type II PDT) (E).²²¹ In doing so, the spin of one of the outermost electrons is inverted resulting in the highly reactive singlet oxygen which reacts with different substrates resulting in toxicity (F).^{219,221} Due to the high

reactivity and extremely short half-life ($<0.04 \mu\text{s}$) of singlet oxygen, cytotoxicity is limited to cells within a reaction sphere of approximately $0.02 \mu\text{m}$.²²² Hence a photodynamic reaction will only occur in cells or tissues which contain photosensitizer, molecular oxygen and are illuminated with light of the appropriate wavelength.

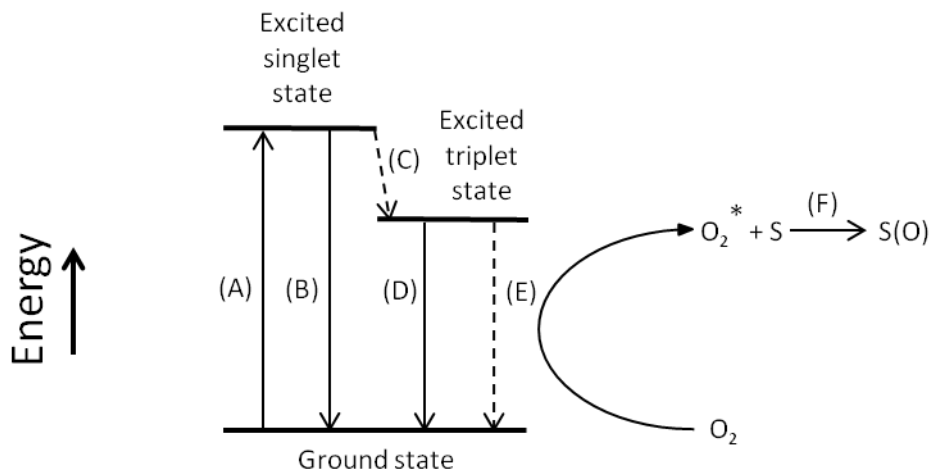


Figure 53: Simplified Jablonski diagram showing the mechanism of action of photodynamic therapy (PDT). When the PS is illuminated by light of a particular wavelength, it is excited from the ground state to the singlet state (A). Energy can be released in the form of fluorescence emission (B) returning it to the ground state. Alternatively, it can enter the excited triplet state (C). From here, energy can be released in the form of phosphorescence (D) or transferred to other molecules (E). The energy can be transferred to oxygen resulting in singlet oxygen which is highly cytotoxic (F). Adapted from [219].

Clinically, PDT has been used to treat a variety of conditions. These include the treatment of age-related macular degeneration;²²¹ the leading cause of blindness in the elderly,²²³ and a number of different dermatological conditions.²²⁴ The most notable application of PDT is in the treatment of malignant tumours.²²¹

The chemical structures of some PSs which have been legalised for the treatment of cancer are shown in Figure 54. The drugs are typically administered intravenously directly into systemic blood supply. However, some such as 5-aminolevulinic acid (Figure 54), have been formulated to enable topical administration directly onto the skin.²²⁵ There is evidence that the drugs preferentially accumulate in tumour cells and are cleared from normal cells at a faster rate.²²⁰ The reason for this selectivity is not fully understood, but has been attributed to the reduced lymphatic drainage and increased vascular permeability of tumours.^{209,212} During surgical interventions, the fluorescent properties of

PSs can be used to help differentiate between healthy and cancerous tissue, and provide an additional means of achieving selectivity.

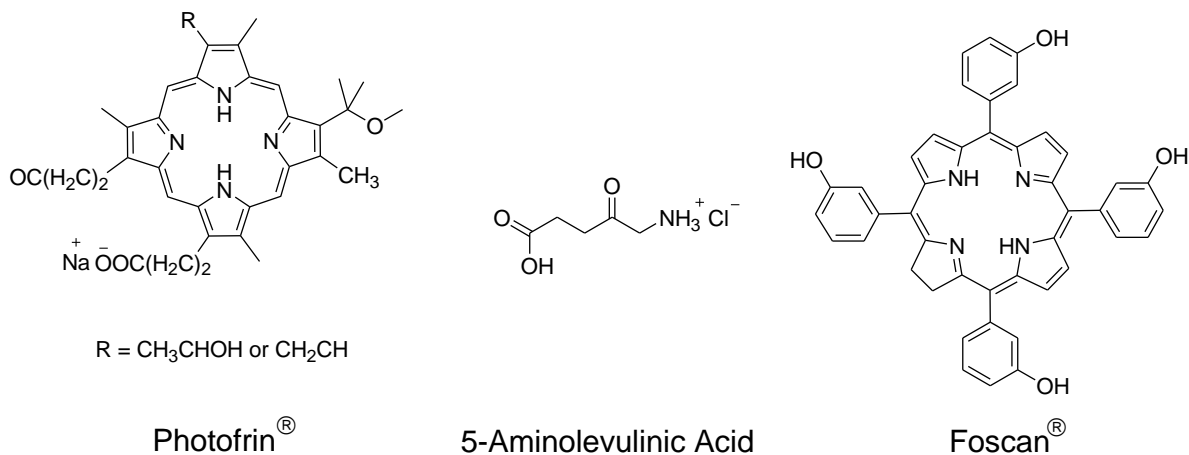


Figure 54: Molecular structure of some photosensitizers used in clinical photodynamic therapy. [²²⁵, ²²⁶]

On a cellular level, cytotoxicity is attributed to damage caused to a number of different organelles. The chemical structure of a PS dictates which organelle is targeted.²²⁰ For instance, lipophilic and anionic PSs typically localise to membranes (including plasma, mitochondrial, endoplasmic reticulum and nuclear) where they alter membrane stability.²²⁰ Conversely, hydrophilic PSs preferentially target lysosomes to promote the release of hydrolytic enzymes into the cytosol, which are lethal to cells.²²⁰ Cationic PSs accumulate within the mitochondria due to the electrochemical gradient across the mitochondrial membrane.^{219,220} The net effect of these actions is the promotion of cell death via apoptosis or necrosis. On a broader level, tumour destruction can also be achieved by damaging the vascular supply of the tumour. The viability of a tumour is dependent on its ability to establish and maintain a blood supply which can provide it with nutrients.²²¹ PDT has been shown to trigger constriction of the blood vessels which ultimately leads to tumour hypoxia and restricts growth.

The efficacy of PDT is largely determined by the efficiency with which light penetrates the target tissue.²²⁷ This is in turn determined by the wavelength of the light. In general, longer wavelengths of light penetrate tissue more efficiently than shorter wavelengths.²²⁷ A dramatic decrease in light penetration is observed at approximately 580 nm which corresponds to the absorbance band of haemoglobin.²²⁸ This is followed by a rapid rise in

light penetration between 600-680 nm to a gradual plateau between 700 and 800 nm.^{227,228} In essence, the greatest amount of light penetration is observed within the red region of the electromagnetic spectrum. As a result, red light is typically used for most PDT applications. The exact wavelength is determined by the absorption maxima of the PS.

The depth with which light is able to penetrate tissue also instrumental in PDT efficacy. This parameter is determined by the absorption and scattering properties of the tissue.²²⁰ As photons pass through tissue, a proportion is reflected from surface of the tissue whilst the remainder scattered within the tissue until they can escape or are absorbed by a chromophore.²¹⁹ The extent of absorption or scattering is tissue dependent.²²⁹ For instance, liver tissue is characterised by poor light penetration due to its high haemoglobin content.²²⁰ Conversely, brain tissue has been shown to be more light scattering.^{220,228} Wavelengths of light less than 800 nm are scattered with greater efficiency as they are equal to or less than the size of many macromolecules.²²⁰ Hence at longer wavelengths, light penetration increases due to decreased scattering. The relationship between the wavelength of light and penetration depth is described by the attenuation coefficient. Attenuation is the loss of mechanical energy of waves as they pass through a tissue.²³⁰ An inverse relationship exists between frequency (or wavelength) of a wave and penetration depth; the higher the frequency of a wavelength the greater the attenuation, that is, loss of mechanical energy. Hence higher frequency light has less penetration due to greater attenuation. As a result, high frequency light cannot be used to probe deep tissue or structures.²³⁰ A penetration depth of 2-4 mm is considered deep light penetration and this is typically achieved using wavelengths of 600-1000 nm.²³¹

4.2.4. Design of Photo-Caged Azacytidine Derivatives

Based on the classical design strategy described above, it was hypothesised that the most effective way to cage azanucleosides would be to place the photolabile protecting group directly on the nucleobase. In terms of chemical synthesis, it was thought that the simplest way to achieve this would be to place the photolabile group on the exocyclic

amine at C4, by taking advantage of its nucleophilic character. Whilst modification in this manner would almost certainly render azanucleosides biologically inert, other key cellular process might also be affected.

The cytotoxicity and activity of azanucleosides is dependent upon their ability to translocate the plasma membrane of a cell. In human cells, the influx and efflux of nucleosides is mediated by nucleoside transporter channels. These can be subdivided into two classes depending on their mode of action: sodium-dependent (concentrative) or sodium-independent (equilibrative). The former, which are limited to specialised cells including leukemic cells, generally mediate the influx of nucleosides via an active process that requires ATP; hence can carry nucleosides against a concentration gradient.^{232,233} The latter, which are found in all cells, mediate both the influx and efflux of nucleosides in a process that is independent of ATP but dependant on a transmembrane concentration gradient.^{233,234} Azanucleosides, which are themselves nucleoside analogues, are readily transported by concentrative channels, and to a lesser extent, the equilibrative channels.²³⁵ Hence, it has been shown that the additional nitrogen in the pyrimidine ring has little effect on the ability of the nucleoside to act as a substrate for the channel. To the best of our knowledge, there have been no nucleoside transporter studies which utilise nucleosides with a sizeable base-modification. The majority of the research is centred on the transport of antiretroviral nucleoside analogues for the treatment of HIV. This class of drugs are characterised by the absence of the 3'-hydroxyl on the deoxyribose sugar.²³⁶ Whilst this is only a slight modification, studies have shown that it is significant enough to reduce or completely preclude the transport of some analogues and thus reduce their efficacy. For instance, a study by Pastor-Anglada *et al.*,²³⁷ compared the abilities of cytidine, 2'-deoxycytidine and 2',3'-dideoxycytidine (ddC, Zalcitabine; antiretroviral drug), to interact with and translocate concentrative transporters. It was observed that the interaction of ddC with the transporter was reduced by two orders of magnitude compared to cytidine and 2'-deoxycytidine. This study and others illustrated the importance of the 2'-hydroxyl functionality in influencing the transportability of nucleoside analogues.^{237,238} Interestingly, the absence of the 2' and 3' hydroxyl group can

render some nucleoside analogues sufficiently hydrophobic to enable simple passive diffusion across the plasma membrane. This form of diffusion occurs without the aid of a protein channel or the expenditure of ATP. Simple diffusion has been observed for AZT,²³⁹ d4T,²⁴⁰ and ABC²⁴¹ (Figure 55). Based on the studies described above, it was hypothesised that photocaged derivatives of azanucleosides would be able to translocate the membrane via a transporter provided the sugar moieties were unaltered. Furthermore, the photolabile protecting group may render azanucleosides sufficiently hydrophobic to translocate the membrane passively.

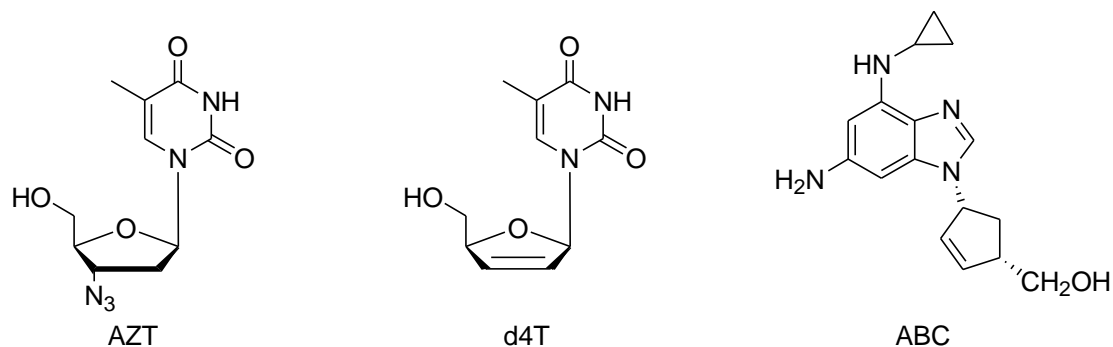


Figure 55: Antiretroviral drugs which are able to translocate the plasma membrane by simple passive diffusion. Abbreviations: AZT, Zidovudine, 3'-azido-3'-deoxythymidine; d4T, Stavudine, 2',3'-dideoxy-3'-deoxythymidine; ABC, Abacavir, [4R-(2-Amino-6-cyclopropylamino-purin-9-yl)-cyclopent-2-en-1-yl]-methanol.

The effect of the photolabile protecting group on intracellular metabolism of the azanucleosides was also considered. All nucleoside analogue drugs are essentially inactive prodrugs which are dependent upon intracellular phosphorylation for their activity.^{242,243} The nucleosides must be metabolised to the active triphosphate in order to be incorporated into cellular DNA or RNA by the action of polymerase enzymes. Once incorporated into the genetic material they can interfere with DNA replication or protein synthesis. Hence the efficacy of nucleoside analogue drugs is partly determined by the efficiency with which they are phosphorylated. Nucleoside triphosphorylation occurs in three sequential steps, each of which is catalysed by a different enzyme (Figure 56). Activation to the monophosphate, the initial rate limiting step, is mediated by nucleoside kinase enzymes.²⁴³ The exact nature of the enzyme is determined by whether a nucleoside possesses a ribose or deoxyribose sugar.²⁴⁴ For azacytidine, this step is catalysed by

uridine-cytidine kinase (UCK),²⁴⁵ whilst for decitabine this step is catalysed by deoxycytidine kinase.²⁴⁶

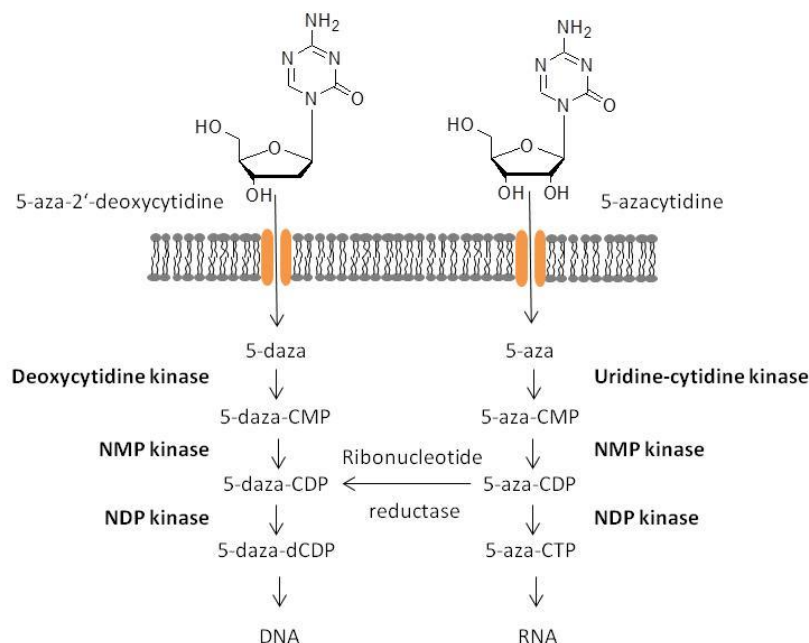


Figure 56: Intracellular metabolism of azanucleosides. The nucleosides enter the cell through protein pores (shown in orange). After cellular uptake, azacytidine is monophosphorylated by uridine-cytidine kinase, whilst 5-aza-2'-deoxycytidine is phosphorylated by deoxycytidine kinase. The monophosphates are subsequently phosphorylated to the active triphosphate, which is then incorporated into either DNA or RNA. Abbreviations: NMP, nucleoside monophosphate kinase; NDP, nucleoside diphosphate.

UCK, which belongs to the nucleoside monophosphate kinase family of enzymes, exclusively catalyses the phosphorylation of pyrimidine ribonucleosides.²⁴² It is responsible for the phosphorylation of uridine and cytidine to their respective monophosphates, however, can also phosphorylate cytotoxic ribonucleoside analogues that are used as chemotherapeutic agents.²⁴² The specificity of UCK has been attributed to an induced fit enzyme-substrate interaction which is triggered by the binding of the sugar moiety to the active site. The hydrogen-bonding network of a ribose sugar and pyrimidine ring of cytidine bound to human UCK are shown in Figure 57. It has been proposed that the sugar moiety binds first resulting in several hydrogen-bonds between the hydroxyls and the amino acids that line the active site. It is thought that the hydrogen-bonds between the 2'-hydroxyl and the guanidinium proton ($\eta 1$) of arginine and the oxygen atom ($\delta 2$) of aspartic acid (Figure 57A), endow UCK with specificity for ribose

sugars. Once these interactions have been established, a conformational change occurs which enables the base moiety to be recognised and bound. Any alterations at the 2'-hydroxyl which result in a loss of hydrogen-bonding are likely to prevent the binding of the nucleoside as the conformational changes would not occur.

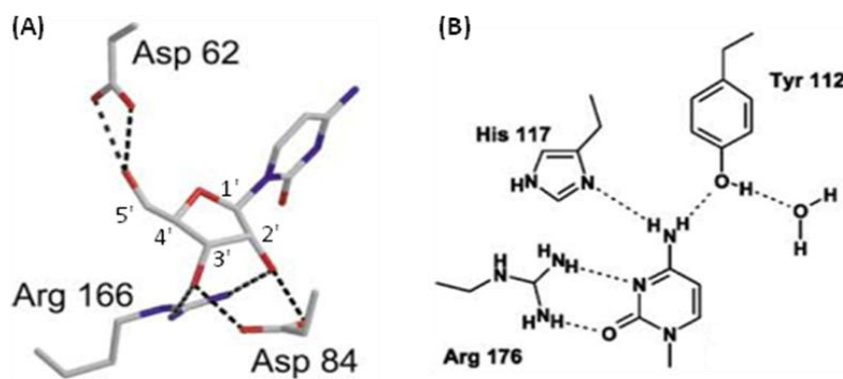


Figure 57: The specificity of uridine-cytosine kinase for pyrimidine ribonucleosides. (A) Sugar recognition by UCK. The carbon atoms are coloured in grey, nitrogen atoms are coloured in blue and the oxygen atoms are coloured in red. (B) Hydrogen-bonding network of the substrate binding site of UCL and cytosine. Hydrogen-bonding interactions are shown as dashed lines. Adapted from [242].

The cytosine base forms four hydrogen bonds in the active site (Figure 57B). Molecular modelling has shown that the binding pocket is deep within the active site of the enzyme and is too small to accommodate larger purine bases. However, *in vivo* studies have shown that cytidine analogues with bulky substituents at the N4-position, such as a benzoyl group, are efficiently phosphorylated.²⁴⁷ This would suggest that provided the 2'-hydroxyl is unaltered and that at least one hydrogen-bonding interaction exists between N4 and the active site, the UCK will be able to phosphorylate the caged azanucleosides to their respective monophosphates.

4.3. Aim

The aim of this project was to synthesise a photocaged derivative of azacytidine to enable greater spatial and temporal control over drug release *in vivo*. The intention was to mitigate the side effects and toxicity associated with the drug. Azacytidine was used as the model nucleoside as it is readily and cheaply available. The effects of any structural changes on the biological activity of the nucleoside and any synthetic methods could also

be applied to deoxyazacytidine. At present, azacytidine has been licenced for the treatment of leukemic cancers. However, research is underway into its use for treating solid malignant tumours.²⁴⁸ Hence this work should be viewed as a proof-of-concept study.

A schematic representation of the aim is shown in Figure 58. While the scheme shows photocaged nucleoside is incorporated into the RNA, the desired outcome of tissue-specific hypomethylation could also be achieved if uncaging occurs at an earlier stage in the cascade of steps, that is, prior to incorporation.

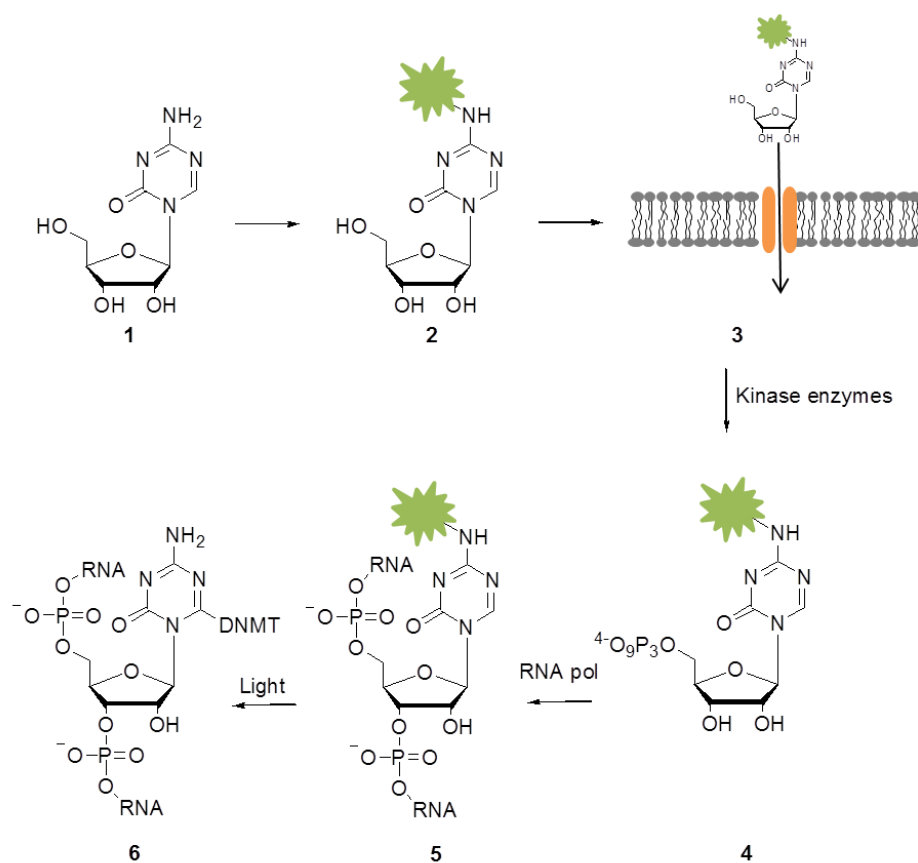


Figure 58: Schematic representation of the objectives of this project. Objective 1: attach a photolabile protecting group at N4 of azacytidine, **2**. It is hypothesised that the modified nucleoside will be able to enter the cell through a protein pore or passively diffuse across the membrane, **3**. Objective 2: determine whether the nucleoside is triphosphorylated by intracellular kinase enzymes, **4** and subsequently incorporated into RNA or DNA, **5**. Irradiation of the cell with light of a particular wavelength should remove the protecting group, releasing azacytidine which can form a covalent bond to DNMT, **6**.

The first specific objective was to place an *o*-nitrobenzyl-based photolabile protecting group on the N4 position of azacytidine and assess its photochemical properties. The second specific objective was to determine whether the caged nucleoside would be tolerated by cancerous cells. Based on previous studies, it was hypothesised that the caged nucleoside would be able to enter the cell and would be non-toxic. Once inside the cell, the caged nucleoside was expected to be phosphorylated by UCK and incorporated into genetic material. Uncaging with light of the correct wavelength was expected to release the bioactive azacytidine, which would then inhibit DNMT.

4.4. Results and Discussion

In this study, two different caging groups were introduced into the N4 exocyclic amine of azacytidine as carbamates (Figure 59, **1** and **2**). A photocaged derivative of cytosine was also synthesised as a control to test the conjugation procedure (Figure 59, **3**). Nucleosides **1** and **3** were isolated with sufficient purity and yield to enable their photochemical properties to be determined. Furthermore, nucleoside **1** was used in cellular assays to determine its effects on cell viability once uncaged.

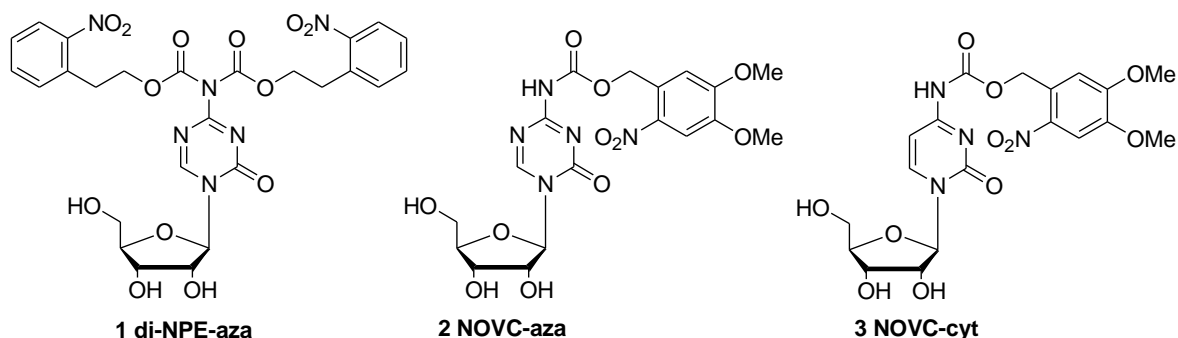


Figure 59: Caged nucleosides synthesised in this project.

4.4.1. Synthesis of Caged Nucleosides

The initial synthetic approach towards the synthesis of the caged azacytidine derivatives is shown in Figure 60. It was envisaged that the hydroxyl groups of the sugar moiety could be protected as silyl ethers (Figure 60A, **5**) then the caging group introduced into the exocyclic amine in the form of a chloroformate (Figure 60B, **7**). The resultant carbamate could then be desilylated to afford the photocaged azacytidine (Figure 60A, **2**).

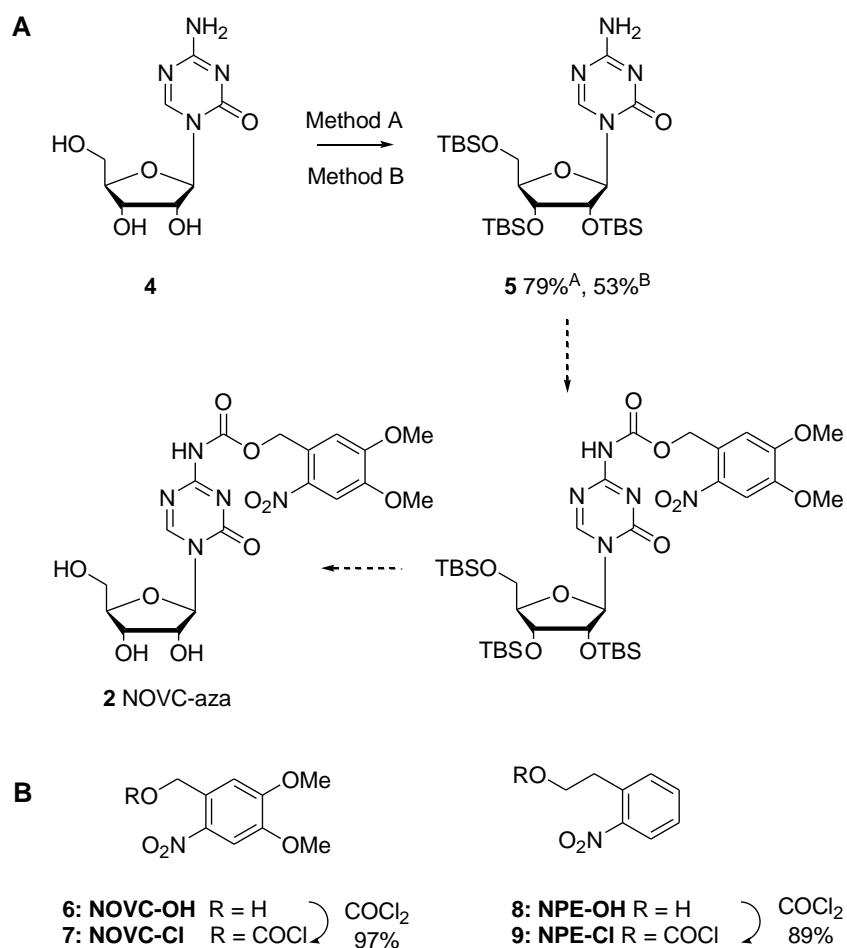


Figure 60: (A) Initial strategy towards the synthesis of caged nucleoside, **2**, NOVC-aza. Regents and Conditions; Method A: TBSOTf, pyridine, 0°C 20 min then 6 h at rt; Method B: TBDMS-Cl, imidazole, DMF, 60-65 °C 8h. (B) Synthesis of photolabels. Regents and Conditions: Phosgene, THF, 0°C for 1 h, then rt overnight.

Silyl protection of azacytidine proceeded smoothly in moderate to high yields depending on the conditions employed. Using TBSOTf in pyridine (Figure 60B, Method A) the trisilylated product, **5**, was isolated in 79% yield. Purification of the crude material was complicated by the presence of the disilylated nucleoside and other by-products, most likely the O2 or N4 silylated products. Under milder silylation conditions (Figure 60A, Method B), TBDMS-Cl with imidazole, the trisilylated product, **5**, was isolated in a moderate 53 % yield. Although the mono- and disilylated products were also formed, the other by-products were not detected.

It was initially sought to couple 6-nitroveratryloxycarbonyl chloride, NOVC-Cl, **7** (Figure 60B) to the exocyclic amine. NOVC-Cl, was prepared in 97% yield by reaction with phosgene and 4,5-dimethoxy-2-nitrobenzyl alcohol, **6** (Figure 60B), and used without any further purification. Formation of the carbamate bond proved more challenging than had initially been anticipated. Reactions in neat anhydrous pyridine, at room temperature overnight, resulted in only trace amounts of the desired product. The product was detected by LCMS analysis, but not via TLC or NMR. Increasing the temperature steadily from 55 °C to 80 °C over the course of several hours did not facilitate formation of the product. Rather, degradation of the chloroformate was observed. Employing DIPEA and DBU as bases, marginally increased the yield of the product, as indicated by the peak area of the LC chromatogram. However, again, the product could not be detected by TLC or NMR. In each of these failed reactions, TLC analysis indicated that hydrolysis of the chloroformate occurred within 1 h of the reaction. Whilst efforts were made to work under strictly anhydrous conditions (all solvents and vessels were purged with argon, equipment dried in an oven then cooled in a dessicator), hydrolysis continued to be a problem. The poor yields and the failure of some reactions could have been attributed to a reduction in the amount of active NOVC-Cl. Consequently, alternative methods of activating **6** were investigated.

Three alternative acylating agents which possessed greater reactivity or stability were investigated. Their structures are shown in Figure 61. NOVC-tetrazolide, **10**, was investigated for its increased reactivity compared to its chloroformate counterpart.²⁴⁹ The tetrazolide was generated *in situ* by the addition of 1*H*-tetrazole to NOVC-Cl. The trisilylated azacytidine, **5**, and Et₃N were subsequently added to this solution. Disappointingly, after 24 h, LCMS analysis showed that the reaction had not yielded a significant amount of the desired product. A similar result was obtained with the NOVC-carbonylimidazolium chloride acylating agent, **11**. This was synthesised in 85% yield by reaction of NOVC-Cl with 1-methylimidazole.²⁵⁰ The inability of both of these reagents to acylate the exocyclic amine of azacytidine was surprising as they had both been successfully used to acylate the exocyclic amine of adenosine.^{249,250} The final acylating

agent to be investigated was *N*-hydroxysuccinimidyl 4,5-dimethoxy-2-nitrobenzyl carbonate, **12**. It was hypothesised that this reagent would display an increased stability in organic solvent compared to **10**, **11** and the chloroformate (NOVC-Cl) counterparts. It was synthesised in 34% yield, in a reaction between NOVC-Cl, and *N,N*-disuccinimidyl carbonate.^{251,252} TLC of analysis of the conjugation reaction with azacytidine, did not show any traces of the 4,5-dimethoxy-2-nitrobenzyl alcohol, **6** (Figure 60B), suggesting that **12** was indeed more stable. However, yet again, only small amounts of **2** were detected by LCMS analysis.

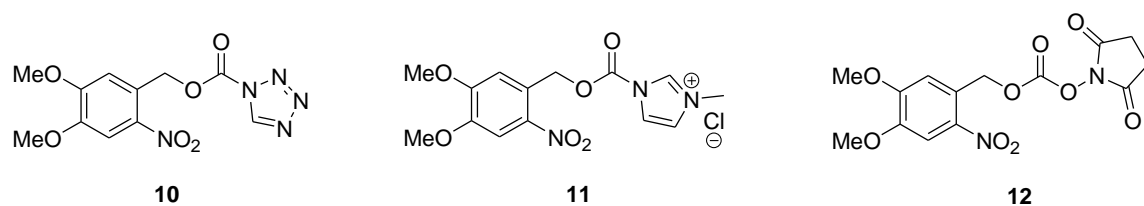


Figure 61: Alternative NOVC-based acylating agents.

Acylation of azacytidine was eventually achieved using a one-pot, transient-protection strategy, developed by Alvarez *et al.*²⁴⁹ The synthetic route, which is outlined in Figure 62, was initially tested using cytidine, **13** and NOVC-Cl. The hydroxyl groups of cytidine were silylated using trimethylsilyl chloride in pyridine (Figure 62, Method A). Once the reaction was complete, as indicated by TLC analysis, NOVC-Cl was added. The reaction was allowed to proceed overnight, after which silyl deprotection was effected by the addition of methanol/water mix. Interestingly, LCMS analysis of the crude reaction mixture, prior to the deprotection step showed that the silyl groups were often already absent. Using this approach, NOVC-cytidine, **3**, was obtained in 31% yield and good purity (as determined by ¹H NMR, Appendix 3). The synthesis was modified slightly to achieve acylation of azacytidine (Figure 62, Method B).²⁵³ The hydroxyl groups were silylated using hexamethyldisilazane in DMF. Acylation was affected by the addition of NOVC-Cl in pyridine, and subsequent silyl deprotection achieved using tris(dimethylamino) sulfonium difluorotrimethylsilicate (TAS-F) in DMF. Using this procedure, NOVC-azacytidine **2** and di-NPE-azacytidine **1** were obtained in 2% and 10% yields, respectively.

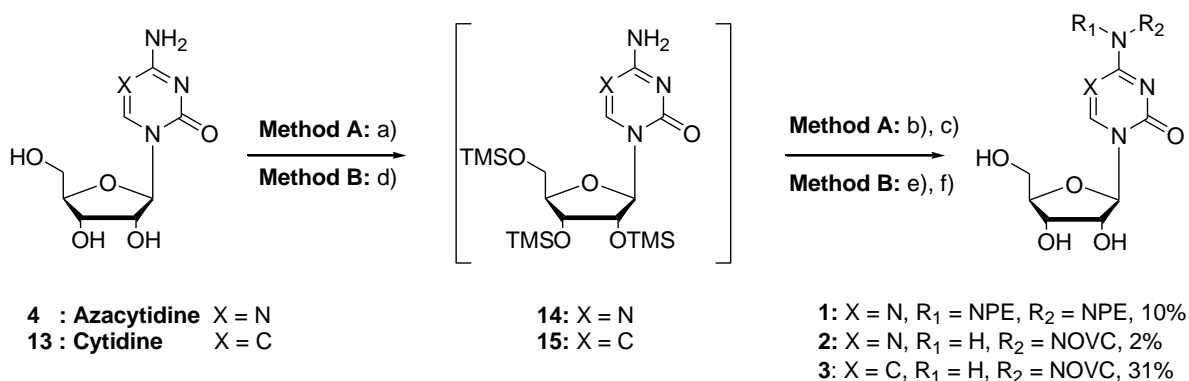


Figure 62: One-point transient protection strategy. Photocaged derivatives of azacytidine (X = N) were synthesised via Method A. Reagents and conditions: a) HMDS, DMF, 2 h, rt; b) chloroformate **7** or **9**, pyridine, rt overnight; c) TAS-F, DMF, 2 h, rt. Photocaged derivative of cytidine (X = C) was synthesised via Method B. Reagents and conditions: d) TMS-Cl, pyridine, 1 h, rt; e) NOVC-Cl, pyridine, overnight, rt; f) MeOH/H₂O (1:1, v/v), 30 min, rt.

One of the most significant differences between Methods A, B and the two-step procedures described previously was the stringency of the anhydrous conditions. Applying Method A to azacytidine resulted in only trace amounts of the desired product or no reaction at all. One plausible explanation for this is that the product degraded during the silyl deprotection step which employed a methanol/water mix. Azanucleosides are known to be unstable in aqueous solutions and in the presence of strong nucleophiles.^{254–256} The nucleosides have been shown to decompose into a number of different products via the opening and deformylation of the triazine ring (Figure 63).²⁵⁵

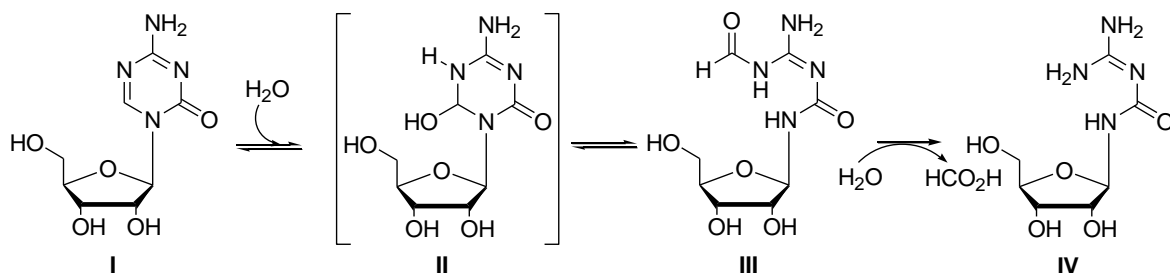


Figure 63: Ring opening and hydrolysis of azacytidine. I is in dynamic equilibrium with its ring-opened formylated derivative, III. Irreversible deformylation results in the formation of a guanylurea derivative, IV.

In Method B, anhydrous conditions were maintained by co-evaporating the reaction mixture with anhydrous pyridine and toluene prior to each synthetic step. Furthermore, to minimise exposure of the product to water, TAS-F, an anhydrous source of fluoride ions,

was used to achieve silyl deprotection.²⁵⁷ In TAS-F the fluoride is present as a silicate, rather than the 'naked' ion, such as in the commonly used tetra-*n*-butylammonium fluoride (TBAF).²⁵⁸ As a result, TAS-F is considerably less basic and hygroscopic compared to TBAF, which is often contaminated with tetrabutylammonium hydroxide and is extremely difficult to dry.²⁵⁸⁻²⁶⁰ Indeed, it has been reported that TBAF, ammonium fluoride, sodium ethoxide, *p*-toluenesulfonic acid and acetic acid can effectively remove the silyl groups from azacytidine, however, they also cause degradation of the triazine ring.²⁵³ As an additional precaution, aqueous work-up of the crude reaction mixture was not performed and the crude material was directly submitted to flash chromatography. Analogues **2** and **3** were extremely polar; requiring up to 13% organic solvent for elution. Nucleoside **3** was marginally more hydrophobic than nucleoside **2** which facilitated its resolution during flash chromatography. This, and its greater stability, could also account for the improved isolated yield (31% versus 2% for nucleosides **3** and **2**, respectively). The difference in the isolated yields of nucleosides **1** and **2** could also be attributed to the difference in hydrophobicity (10% versus 3% for nucleosides **1** and **2**, respectively).

4.4.2. Uncaging of Caged Cytidine and Azacytidine

The absorbance and photochemical properties of analogues **1** and **3** were measured as they were isolated with sufficient purity. Specifically, analogue **1** was obtained in 71% yield and analogue **2** was obtained in 79% yield, as determined by analytical HPLC analysis. Their absorbance spectra are shown in Figure 64. Both nucleosides possessed two absorption maxima. Analogue **3** possessed maxima at 293 nm and 346 nm corresponding to the cytosine base and the NOVC-protecting group, respectively.²⁶¹ Whilst nucleoside **1** displayed maxima at 237 nm and 260 nm which corresponded to triazine base of azacytidine and the NPE protecting group, respectively.^{217,262}

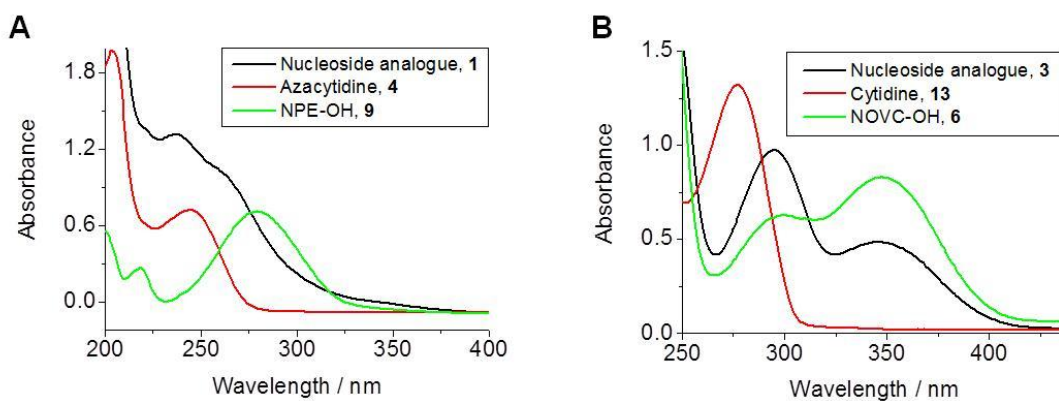


Figure 64: (A) Absorbance spectra of di-NPE-Aza, **1**. The absorbance spectra of azacytidine, **4**, and NPE-OH, **8**, are superimposed. (B) Absorbance spectra of NOVC-cytidine, **3**. The absorbance spectra of cytidine, **13**, and NOVC-OH, **6** are superimposed. All components at 100 μM , $\text{ACN}:\text{H}_2\text{O}$, 1:1, v/v, with 1% acetic acid.

Uncaging of nucleosides was studied by irradiating the samples at 365 nm using a LED photoreactor. Photolysis of *o*-nitrobenzyl-based protecting groups typically occurs via an intramolecular hydrogen abstraction from the exocyclic α -position (with respect to the aryl group).^{207,263} In the case of NOVC, the resultant radical undergoes photoisomerisation to a cyclic acetal derivative which spontaneously collapses to *o*-nitrosobenzaldehyde and the uncaged compound. The uncaged compound undergoes a further decarboxylation reaction to release the free amine. The mechanism is outlined in Figure 65.

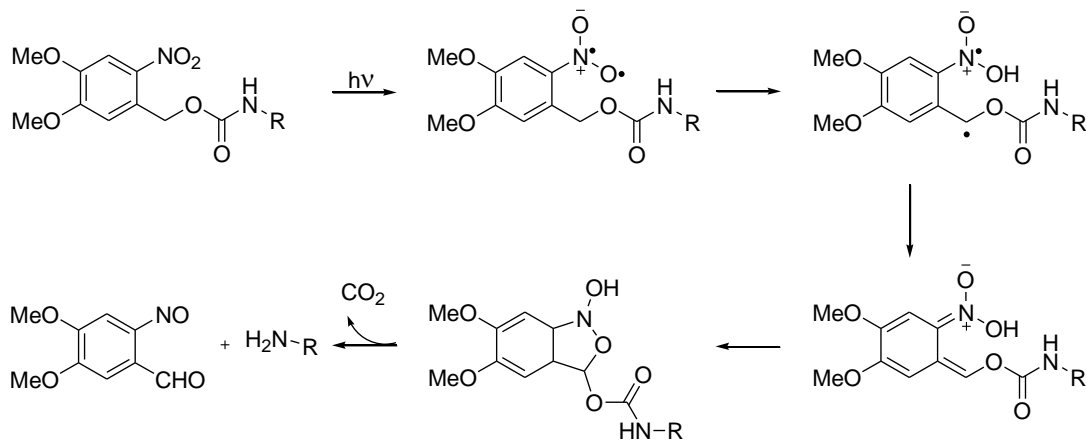


Figure 65: Mechanism of uncaging nucleoside analogue **3**.

Decarboxylation of the uncaged nucleoside with concomitant release of the amine presented a potential complication. It was possible that the released amine could have

formed an imine with the *o*-nitrosobenzaldehyde photo-byproduct.²⁰⁷ In order to avoid this, uncaging of analogue **3** was performed under slightly acidic conditions, to reduce the nucleophilicity of the amine and hydrolyse any imine that may have formed.^{264,265} Such precautions were not required for the uncaging of analogue **1**, as the photo-by-product of NPE is an *o*-nitrostyrene (Figure 66). Akin to NOVC, photodeprotection begins with intramolecular hydrogen abstraction at the α -position, but is followed by β -elimination of the protected group with concomitant liberation of CO₂ and the amine (Figure 66).^{266,267} However, in order to directly compare the photochemical properties of the analogues, all experiments were performed in the same reaction medium; ACN:H₂O, 1:1, v/v, with 1% acetic acid.

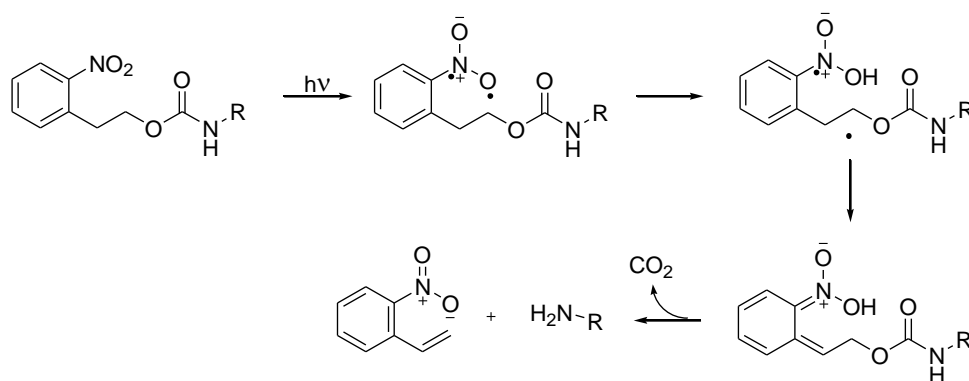


Figure 66: Mechanism of uncaging nucleoside analogue **1**.

Photolysis of the caged nucleosides was monitored by analytical HPLC over several hours (Figure 67B). The HPLC traces indicated that the caged nucleosides were uncaged to the parent nucleoside and photo-by-products in almost quantitative yields. Representative traces for the photolysis of analogue **1** are shown in Figure 67A. Similar results were obtained for analogue **3**.

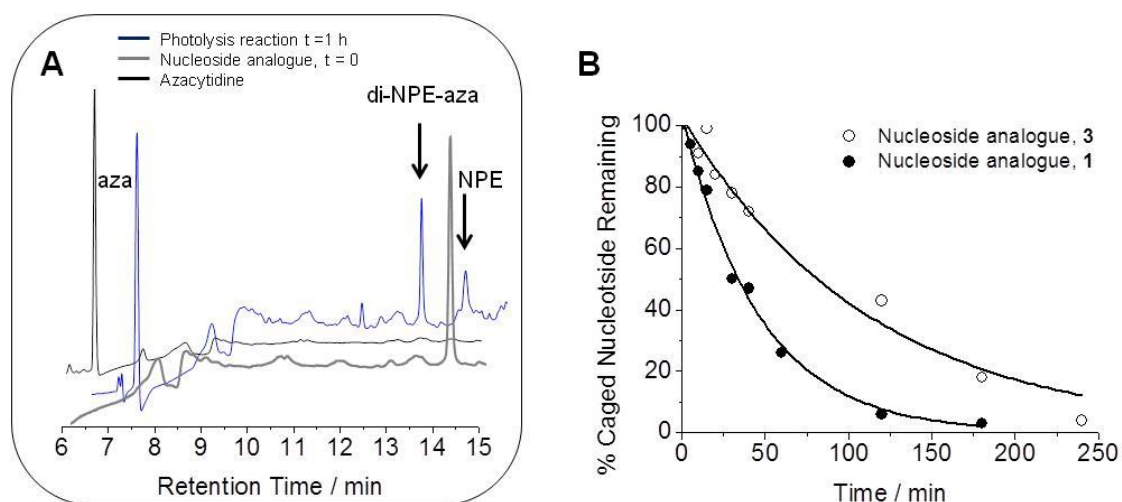


Figure 67: (A) HPLC traces for the photolysis of analogue **1** (measured at 260 nm). (B) Time-course for the photolysis of analogues **1** (filled circles) and **2** (empty circles). Samples (100 μM , ACN/ H_2O , 1:1, v/v, 1% acetic acid) were irradiated at 350 nm. Aliquots were taken at regular intervals then analysed by HPLC.

Table 9: Chemical Properties of Caged Nucleosides.

Nucleoside analogue	λ_{max} [nm]	ϵ_{max} ($\times 10^4$) ^a	$t_{1/2}$ [min]
1	293, 346	1.03, 0.53	34
3	237, 260	1.25, 0.96	77

^a Molar absorptivity ($\text{M}^{-1} \text{cm}^{-1}$).

^b Quantum yields for the disappearance of the starting nucleosides

The rate of photolysis for both nucleosides could be approximated to a first-order exponential decay curve. This suggested that undesired secondary effects such as further photolysis of the parent nucleoside, were not occurring.²⁶¹ In agreement with literature reports, analogue **1** was photolyzed almost twice as fast as analogue **3**. Specifically, the half-life (Table 9) for uncaging was 34 min and 77 min for **1** and **3**, respectively.

4.4.3. Preliminary Cellular Assays ^x

Cellular assays were performed to determine the effect, if any, of analogue **1** on cell viability. Central to the drug delivery approach is the notion that the caged nucleoside must not have any toxic effects on the cell, and that toxicity must only occur once the

^x These experiments were performed by Sabrina Stewart, UCL Cancer Institute.

nucleoside is uncaged to the biologically active parent nucleoside. Based on literature reports, it was hypothesised that analogue **1** would be readily transported into cells via a nucleoside transporter or through passive diffusion across the plasma membrane. As the two photolabile-protecting groups eliminated the hydrogen-bonding capability of the exocyclic amine, it was believed that the nucleoside would not be metabolised to the active triphosphate or hence incorporated into RNA or DNA.²⁴² Thus any effects of cell viability would solely be due to uptake of the nucleoside.

For this study, human osteosarcoma cells (SaOS-2) were used. The cells were seeded at a density of approximately 1×10^4 cells per well, in a 24-well dish. This was performed 24 h prior to exposing the cells to either azacytidine or analogue **1**. The cells were treated with nucleosides at three different concentrations; 0.5, 1 and 1.5 μM . Due to solubility problems, analogue **1** was dissolved in ACN rather than distilled water. Cells were also treated with a solution of analogue **1** which had previously been irradiated at 365 nm for 24 h. HPLC and MS analysis confirmed that the uncaged azacytidine was present in this solution (as discussed above). Additionally, untreated controls of acetonitrile and growth media alone were also performed. Analogue **1** and azacytidine were added every 24 h in fresh growth media over a 96-hour period. The density or confluence of the cells, which gives an indication of cell growth, was determined approximately every 2 h. The results are summarised in Figure 68.

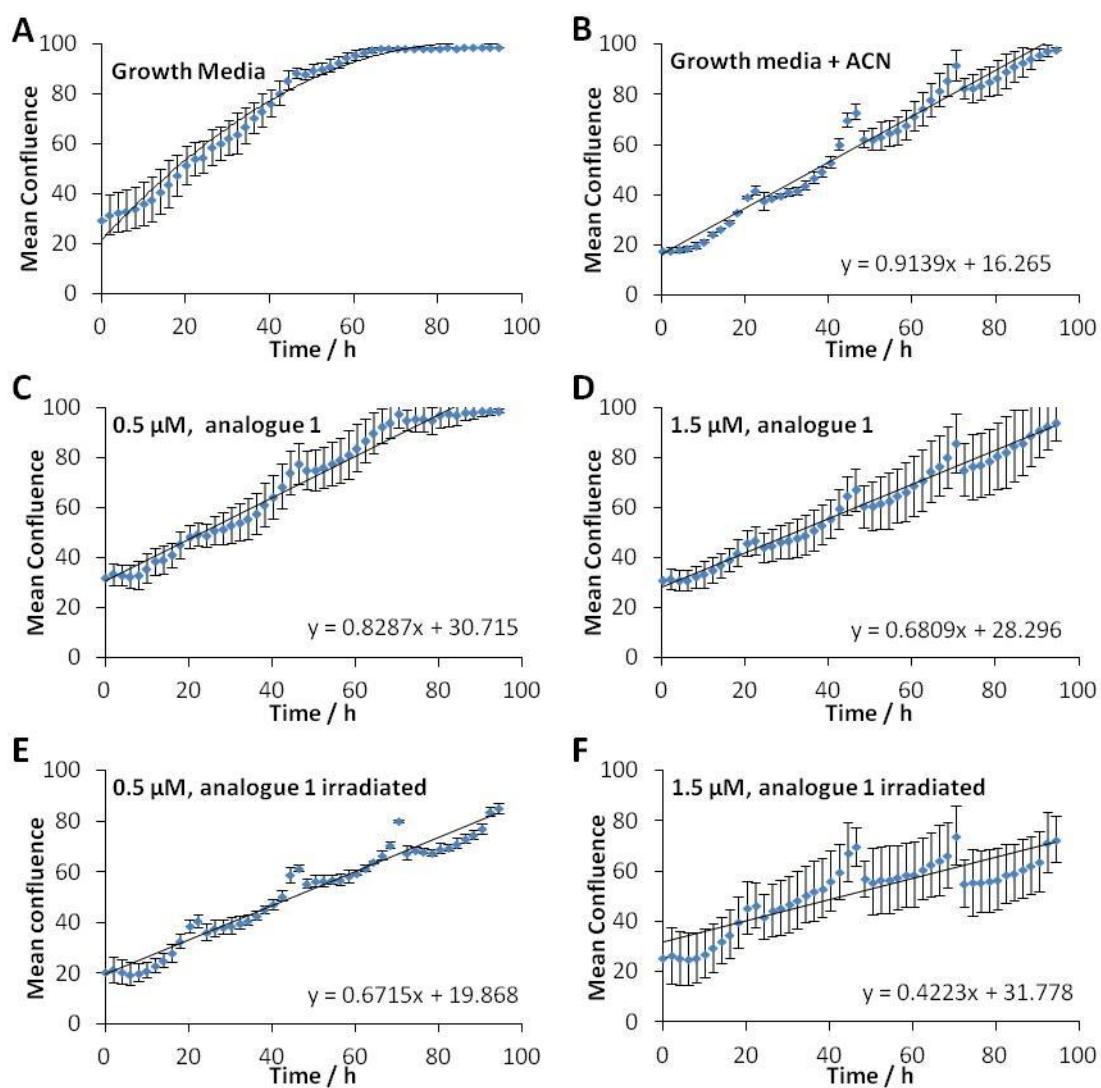


Figure 68: Growth curves for SaOS-2 cells under different conditions over 96 h period. Analogue **1** was dissolved into HPLC-grade acetonitrile. (A) Growth media alone. (B) Growth media with acetonitrile. (C) 0.5 μM **1**. (D) 1.5 μM **1**. (E) 0.5 μM **1** irradiated. (F) 1.5 μM **1** irradiated. For (E) and (F) a 1 mM stock solution of analogue **1** was irradiated for 24 h at 365 nm. The sample was diluted to the appropriate concentration then applied to the cells. All treatment conditions were performed in triplicate. The mean confluence has been plotted per treatment condition, with the error bars representing the standard deviation.

In the presence of growth media alone (Figure 68A), the cells displayed a characteristic growth pattern.²⁶⁸ The cells possessed a base-line confluence of 30% and the number of cells increased almost 3-fold/h within 96 h. The cells were able to grow in the presence of analogue **1**, albeit at a lower rate compared to the untreated control (Figure 68C and D). In particular at the lowest concentration, 0.5 μM , the growth pattern was similar to that of

the untreated control. As the concentration of the analogue **1** increased, the rate of cell growth decreased slightly. In the presence of 1.5 μM analogue **1**, less than 95% confluence was observed within 96 h. At each concentration of the analogue, a slight dip in the confluence was observed every 24 h, that is, after the addition of fresh media (which contained analogue **1**). This initially suggested that **1** elicited some toxic effects on cell growth. However, treating the cells with acetonitrile alone appeared to inhibit cell growth considerably (Figure 68B). The dips in confluence at each media change were larger than those observed during treatment with analogue **1** (Figure 68D). This suggested that acetonitrile may account for a significant proportion of the cell growth inhibition observed during the analogue **1** treatment. As expected, treating the cells with an irradiated sample of analogue **1** reduced the rate of cell growth. Even at the lowest concentration, 0.5 μM , only 85% confluence was achieved at 96 h (Figure 68E). In the presence of 1.5 μM irradiated sample, only approximately 70% confluence was achieved at 96 h (Figure 68F). Furthermore, the rate of cell growth was 2-fold slower than in the presence of the caged nucleoside at the same concentration (Figure 68D).

4.5. Conclusion

The aim of this study was to introduce a photolabile protecting group into the exocyclic amine of azacytidine to gain greater spatial and temporal control over drug release *in vivo*. Two photo-labelled nucleoside analogues were synthesised. Their synthesis was somewhat complicated by the hydrolytic sensitivity of azacytidine. The presence of small amounts of water resulted in poor yields or no reaction at all. This was overcome by employing a one-pot, two-step protocol in which the reaction vessel was co-evaporated with anhydrous pyridine prior to the addition of each reagent, to remove residual water. Analogue **1** was isolated in sufficient yield and purity for cellular assays in which its effect on cell viability was determined.

Cell growth was measured in the presence of analogue **1** and a sample which had previously been irradiated at 365 nm for 24 h. HPLC and MS analysis confirmed that the irradiated sample contained azacytidine and photo-by products. Both treatments resulted in a reduction in the rate of cell growth compared to the untreated control. A significant

proportion of this was attributed to adverse effects of ACN. However, a greater rate of reduction was observed in cells treated with the irradiated sample. The results suggested that the released azacytidine and photo by-products were inhibiting cell growth.

In order to determine the exact cause and mechanism of this inhibition, further experiments are required. A comparison of the methylation levels in treated and untreated cells will give conclusive evidence of the mechanism of inhibition. A reduction in methylation levels compared to untreated controls, would suggest that the uncaged azacytidine is being incorporated into DNA and subsequently sequestering DNMT. However, if the methylation levels in the treated and untreated cells are the same, then this would suggest that inhibition is occurring through another mechanism; most likely involving the released photo-by products.

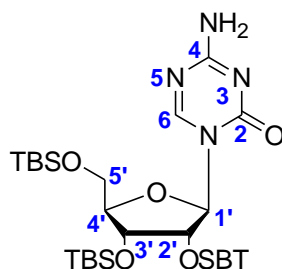
In summary, these preliminary studies suggest by employing a photolabile protecting group, it is possible to have temporal and spatial control over drug release.

4.6. Materials and Methods

All reagents and anhydrous solvents were purchased from Sigma Aldrich unless otherwise stated. All other solvents and water were purchased from Fischer Scientific. Cytidine and 5-azacytidine was purchased from Carbosynth Limited. HPLC analysis was performed using reverse phase Varian ProStar system with a Model 210 solvent delivery mode and a Model 320 dual wave-length detector. The mobile phase A was HPLC-grade water and mobile phase B was neat HPLC-grade acetonitrile. The UV absorbance was monitored at 260 nm. Analytical analysis was performed using a Varian Pursuit XRS C18 column (250 x 4.6 mm). The gradient started with 5% mobile phase B and rose linearly to 50% B over 10 min with a flow rate of 1 mL min⁻¹. ESI-MS analysis was performed using a Waters Aquity Ultra performance LC-MS system equipped with an Aquity UPLC BEH C18 column (50 x 2.1 mm, 1.7 mm beads). CI-MS analysis was performed by Dr. Lisa Haigh, UCL Chemistry. ¹H and ¹³C NMR spectra were recorded at 293K on a Bruker Avance III 600 MHz spectrometer. Coupling constants are reported in Hz. Reference NMR solvent signals CDCl₃: δH = 7.26 ppm, δC = 77.0 ppm; D₂O: δH = 4.79 ppm; (CD₃)₂SO: δH = 2.50 ppm, δC = 39.5 ppm;

CD₃OD: δ H = 3.31 ppm, δ C = 49.0 ppm. Multiplicities for ¹H coupling are shown as s (singlet), d (doublet), t (triplet), m (multiplet), or a combination of the above. FT-IR analysis was performed on a computer driven Shimadzu 8700, range: 7800-350 cm⁻¹, resolution: up to 0.5 cm⁻¹. Thin Layer Chromatography (TLC) was performed on aluminium backed Sigma-Aldrich TLC plates with F254 fluorescent indicator. Visualisation achieved by absorption of UV light or with the use of potassium permanganate solution [KMnO₄ (1.25 g), Na₂CO₃ (6.25 g), water (250 mL)]. Normal phase flash chromatography was carried out using silica gel (43 – 60 μ m) supplied by Merck. UV analysis was performed using Varian Cary 300 Bio UV-vis Spectrophotometer and a quartz cuvette with a path length of 1 cm. Photolyses was conducted using a LED Photoreactor (University of York) at 365 nm. Nucleoside samples were at 100 μ M, ACN/H₂O (1:1, v/v) with 1% acetic acid, unless otherwise stated.

4.6.1. Synthesis of 2', 3', 5'-tri-*O*-(*tert*-butyldimethylsilyl)-azacytosine, 5, via Method A



Anhydrous pyridine (5 mL) was added to aza-cytidine (103 mg, 0.4 mmol) and the resultant suspension cooled to 0 °C. TBDMSOTf (470 μ L) was added drop wise over the course of 15 min. The resultant mixture was stirred at 0 °C for an additional 5 min then at room temperature for 6 h after which TLC (5% MeOH/DCM) analysis indicated that the reaction was complete. The resultant colourless solution was partitioned between DCM (50 mL) and H₂O, the organic phase washed with H₂O, sat. NaHCO₃, and brine (50 mL each); then dried over anhydrous Na₂SO₃. The dried solution was purified by flash chromatography (5% MeOH/DCM, R_f = 0.5) giving rise to the title compound as colourless glassy solid (195 mg, 79%). ¹H NMR (600 MHz, DMSO-*d*₆) δ _{ppm}: -0.02-0.16 (m, 18H, Si-CH₃), 0.77-0.92 (m, 27H, C-CH₃), 3.68-3.77 (m, 1H, H-4'), 3.91-4.00 (m, 2H, H-5'), 4.12 (t, *J*=4.41

Hz, 1H, H-3'), 4.29 (t, $J=3.94$ Hz, 1H, H-2'), 5.67 (d, $J=4.10$ Hz, 1H, H-1'), 7.55 (d, $J=8.20$ Hz, 1H, NH₂), 8.44 (s, 1H, H-6). ¹³C NMR (600 MHz, MeOD) δ_{ppm} : -4.83--3.17 (Si-CH₃), 17.69-18.18 (C-CH₃), 25.69-25.96 (C-C-CH₃), 61.45 (C-4'), 70.26 (C-3'), 75.01 (C-2'), 83.56 (C-5'), 88.84 (C-1'), 153.09 (C-2), 155.60 (C-6), 165.73 (C-4). m/z (ESI) 586.7528 [M+H]⁺, C₂₇H₅₄N₄O₅Si₃ requires 586.9873.

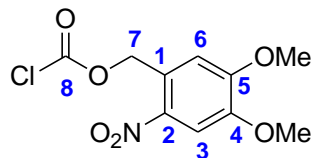
4.6.2. Synthesis of 2', 3', 5'-tri-*O*-(*tert*-butyldimethylsilyl)-azacytosine, 5, via Method B

To a solution of the azacytidine (200 mg, 0.8 mmol) in anhydrous DMF (10 mL), imidazole (556 mg, 8.2 mmol) and TBDMSCl (614 mg, 4 mmol) were added. The resultant mixture was heated at 60 °C for 3 h. TLC analysis indicated that the reaction was not yet complete. Consequently, the temperature was increased to 65 °C and the reaction mixture stirred for an additional 5 h. The reaction mixture was partitioned between DCM (50 mL) and H₂O. The organic phase was then washed with H₂O, sat. NaHCO₃, and brine (3 x50 mL each), and dried over anhydrous Na₂SO₃. The dried organic phase was concentrated *in vacuo* giving rise to colourless oil and subsequently purified by flash chromatography (0-3% MeOH/DCM, 5% MeOH/DCM $R_f = 0.42$). The title compound was isolated as colourless glassy solid (254 mg, 53%). NMR and MS as above.

4.6.3. General Synthesis of Photolabile-Protecting Groups, 7 and 9

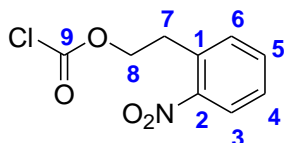
A solution of phosgene in toluene (2.3 mL, 20% w/w, 10.8 mmol) was added to a chilled suspension of the corresponding alcohol in THF (15 mL). The suspension was stirred at 0 °C for 1 h, then at room temperature for 18 h. In each case, TLC (5% MeOH/DCM) analysis suggested complete conversion of the starting material to the chloroformate. The excess phosgene was removed under low vacuum and trapped with aqueous 0.1 M NaOH. Owing to the instability of the resultant products, they were immediately used for the subsequent reactions without any further purification.

4.6.3.1. (4,5-Dimethoxy-2-nitrophenyl)methyl Chloroformate, 7



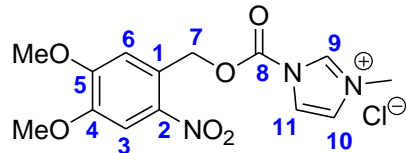
Isolated as solid yellow foam (627 mg, 97 %). 5% MeOH/DCM, R_f starting material = 0.35, R_f product = 0.91. ^1H NMR (600 MHz, CDCl_3) δ_{ppm} : 3.90-3.92 (m, 3H, OMe), 3.95-3.97 (m, 3H, OMe), 4.98 (s, 2H, H-7), 7.05 (s, 1H, H-3), 7.54-7.73 (m, 1H, H-6). ^{13}C NMR (600 MHz, CDCl_3) δ_{ppm} : 44.13 (C-7), 56.77 (OMe), 108.79 (C-6), 113.40 (C-3), 127.36 (C-1), 140.62 (C-2), 149.35 (C-4), 153.91 (C-5), 165.56 or 133.79 (C8). IR (neat): 1764.5 cm^{-1} (C=O), 1514.7 (C=C, Ar.). m/z (CI) 275.01889 [M], $\text{C}_{10}\text{H}_{10}\text{ClNO}_6$ requires 275.01911.

4.6.3.2. 2-Nitrobenzyl Chloroformate, 9



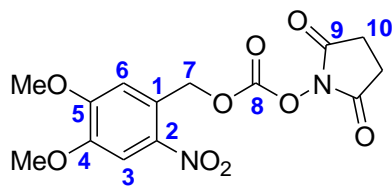
Isolated as an off-white solid (1.22 g, 88.9%). 5% MeOH/DCM, R_f starting material = 0.46, R_f product = 0.95. ^1H NMR (600 MHz, $\text{DMSO}-d_6$) δ_{ppm} : 3.15 (t, $J=6.49$ Hz, 2H, H-7), 4.53-4.67 (m, 2H, H-8), 7.36-7.42 (m, 2H, H-4 & H-6), 7.50-7.53 (m, 1H, H-3), 7.89 (t, $J=8.66$ Hz, 1H, H-5). ^{13}C NMR (600 MHz, CDCl_3) δ_{ppm} : 34.97 (C-7), 75.10 (C-8), 123.16 (C-3), 130.22 (C-4), 133.21-133.26 (C-1, C-5 & C-6), 148.55 (C-2). IR (neat): 1767.8 (C=O), 1516.4 (C=C, Ar.), 1343.0 (NO_2). m/z (CI) 229.61692 [M], $\text{C}_9\text{H}_8\text{ClNO}_4$ requires 229.61712.

4.6.4. 3-(4,5-Dimethoxy-2-Nitro-Benzoyloxycarbonyl)-1-Methyl-Imidazolium Chloride, **11**²⁵⁰



To a stirred solution of 1-methylimidazole (90 μ L, 1.1 mmol) in anhydrous DCM (15 mL), **7** (300.5 mg, 1.1 mmol) was added giving rise to a clear yellow solution. After 2 h at room temperature, a thick yellow precipitate formed. The suspension was filtered and the precipitate washed with ice cold DCM (20 mL) to afford the title compound as a pale yellow solid (389 mg, 85%). 5% MeOH/DCM $R_f = 0.87$. ^1H NMR (600 MHz, CDCl_3) δ_{ppm} : 3.95 (s, 3H, OMe), 4.06 (s, 3H, OMe), 5.61 (s, 2H, H-7), 7.04 (s, 1H, H-3), 7.34 (s, 1H, H-6), 7.70-7.73 (m, 2H, H10 & H11), 9.47 (H-9). ^{13}C NMR (600 MHz, CDCl_3) δ_{ppm} : 56.52-56.73 (OMe), 67.03 (C-7), 108.18 - 108.36 (C-10 & C-11), 120.22 (C-3), 121.76 (C-6), 132.58 (C-9), 139.80 (C-2), 148.02 (C-8), 153.78 - 154.48 (C-4 & C-5). m/z (CI) 322.10336 $[\text{M}-\text{Cl}]^+$, $\text{C}_{14}\text{H}_{16}\text{ClN}_3\text{O}_6$ requires 357.07276.

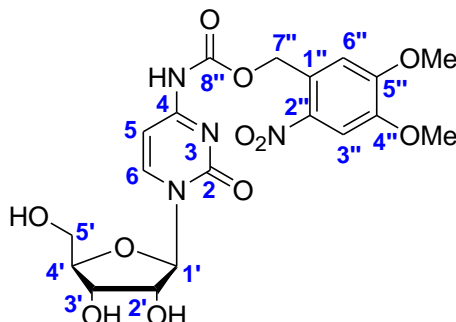
4.6.5. Synthesis of 4,5-Dimethoxy-2-Nitrobenzyl Succinimidyl Carbonate, **12**



To a stirred suspension of 4,5-dimethoxy-2-nitrobenzyl alcohol (302 mg, 1.4 mmol) and *N,N*-disuccinimidyl carbonate (430 mg, 1.7 mmol) in anhydrous ACN (15 mL), Et_3N (580 μ L, 4.2 mmol) was added. The resultant colourless solution was stirred at room temp for 3 h after which the solvent was removed under reduced pressure and the crude material purified by flash chromatography (50% EtOAc/Hexane, 5% MeOH/DCM $R_f = 0.7$) to yield the title compound as a solid yellow foam (202 mg, 34%). ^1H NMR (600 MHz, $\text{DMSO}-d_6$) δ_{ppm} : 2.86 (s, 4H, H-10), 3.97 (s, 3H, OMe), 4.06 (s, 3H, OMe), 5.79 (s, 2H, H-7), 7.05 (s, 1H, H-3), 7.77 (s, 1H, H-6). ^{13}C NMR (600 MHz, CDCl_3) δ_{ppm} : 25.41 (C-10), 56.00-56.34 (OMe),

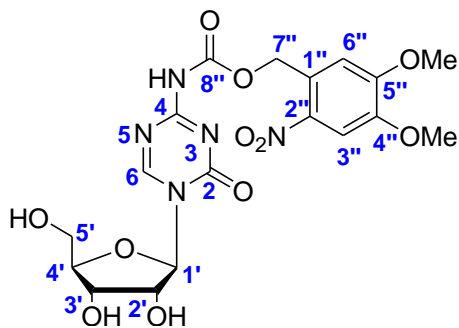
69.14 (C-7), 108.45 (C-6), 112.21 (C-3), 123.51 (C-1), 140.05 (C-2), 146.94 (C-4), 148.65 (C-8), 153.52 (C-5), 169.93 (C-9). m/z (CI) 452.10063 [M], $C_{19}H_{20}N_2O_{11}$ requires 452.10616.

4.6.6. Synthesis of 4-*N*-(6-Nitroveratryloxycarbonyl)-Cytidine, **3**²⁴⁹



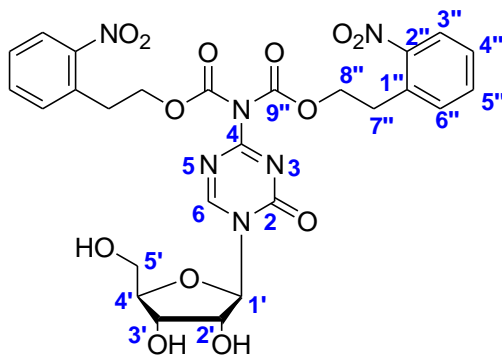
To a suspension of cytidine (299 mg, 1.2 mmol) in anhydrous pyridine (10 mL), TMS-Cl was added (790 μ L, 6.2 mmol) was added over the course of 15 min. The solution was stirred at room temperature for 1 h at which point TLC analysis suggested that protection of the nucleoside was complete. To the resultant colourless solution, NOVC-Cl (400 mg, 1.4 mmol) was added and the resultant solution stirred at room temperature overnight. Water/methanol mix (1:1, v/v, 5 mL) was added and the mixture stirred for 30 min then concentrated *in vacuo*. The residual pyridine was removed by co-evaporation with toluene (3 x 20 mL) and the resultant yellow opaque oil purified by flash chromatography (0-10% MeOH/DCM, R_f = 0.43) to afford the title compound as yellow foam (182 mg, 31%). ^1H NMR (600 MHz, DMSO- d_6) δ_{ppm} : ^1H NMR (600 MHz, DMSO- d_6) δ_{ppm} : 3.60 (ddd, $J=12.19$, 5.03, 3.11 Hz, 1H, H-5'a), 3.74 (ddd, $J=12.14$, 4.99, 2.82 Hz, 1H, H-5'b), 3.88 (s, 3H, OMe), 3.90 (dt, $J=5.98$, 2.94 Hz, 1H, C-4'), 3.95 (m, 4H, C-3' & OMe), 3.96 - 4.01 (m, 1H, C-2'), 5.08 (d, $J=5.83$ Hz, 1H, 3'-OH), 5.18 (t, $J=5.18$ Hz, 1H, 5'-OH), 5.50 (s, 2H, H-7''), 5.78 (d, $J=3.20$ Hz, 1H, H-1'), 7.03 - 7.07 (m, 1H, H-5), 7.41 (s, 1H, H-6''), 7.74 (s, 1H, H-3''), 8.45 (d, $J=7.34$ Hz, 1H, H-6), 11.15 (s, 1H, NH). ^{13}C NMR (600 MHz, DMSO- d_6) δ_{ppm} : 56.55-56.10 (OMe), 59.97 (C-5'), 63.82 (C-7''), 68.72 (C-3'), 74.52 (C-2'), 84.23 (C-4'), 90.11 (C-1'), 94.04 (C-5), 108.10 (C-3''), 110.30 (C-6''), 126.81 (C-1''), 138.83 (C-2''), 152.61-153.65 (C-4'' & C-5''), 145.39 (C-6) 153.66 (C-8''), 154.52 (C-2), 162.76 (C-4). m/z (ESI) 505.1195 [M+Na] $^+$, $C_{19}H_{22}N_4O_{11}$ requires 482.3982.

4.6.7. Synthesis of 4-*N*-(6-Nitroveratryloxycarbonyl)-5-Azacytidine, 2



5-Azacytidine (246 mg, 1.0 mmol) was coevaporated with anhydrous pyridine (3 x 5 mL) and subsequently suspended in anhydrous DMF (10 mL) before adding HMDS (730 μ L, 4.5 mmol). After stirring at room temperature for 2 h, TLC analysis indicated that protection of the nucleoside was complete. The solvent was removed *in vacuo* and the resultant residue co-evaporated with anhydrous toluene (3 x 5 mL) followed by anhydrous pyridine (3 x 5 mL). The dried residue was subsequently suspended in anhydrous pyridine (7 mL) and solution of NOVC-Cl in anhydrous DCM (553 mg, 2 mmol) added. The mixture stirred at room temperature overnight. The solvent was removed under reduced pressure and the residue coevaporated with toluene (3 x 10 mL) then suspended in anhydrous DMF (10 mL). Silyl drprotection was achieved with the addition of tris(dimethylamino) sulfonium difluorotrimethylsilicate (TAS-F) (422 mg, 1.5 mmol). After 2 h the solvent was removed and the residue coveporated with anhydrous toluene (3 x 5 mL). The crude material was purified by flash chromatography (0-12.5 % MeOH/DCM, 10% MeOH/DCM R_f = 0.44) to yield the title compound as an off-white solid (11 mg, 2%). ^1H NMR (600 MHz, DMSO- d_6) δ_{ppm} : 3.64 - 3.68 (m, 1H, H-4'), 3.70 (d, J =5.27 Hz, 1H, H-2'), 3.82 - 3.85 (m, 1H, H-3'), 3.86 (m, 4H, OMe & H-5'a), 3.88 (m, 4H, OMe & H-5'b), 4.94 - 5.00 (m, 1H, 5'OH), 5.09 - 5.18 (m, 1H, H-1'), 5.32 - 5.38 (m, 1H, H-7''), 7.14 - 7.22 (m, 1H, H-3''), 7.66 - 7.71 (m, 1H, H-6''), 8.64 - 8.75 (m, 1H, H-6). ^{13}C NMR (600 MHz, DMSO- d_6) δ_{ppm} : 56.10 - 56.59 (OMe), 60.98 - 62.91 (C-5' & C-7''), 70.31 (C-3'), 74.18 (C-2'), 83.69 - 84.01 (C-1' & C-4'), 108.12 (C-6''), 110.57 (C-3''), 139.47 (C-2''), 147.71 (C-4''), 153.24 (C-5''). m/z (ESI) 484.0941 $[\text{M}+\text{H}]^+$, $\text{C}_{18}\text{H}_{21}\text{N}_5\text{O}_{11}$ requires 483.38624.

4.6.8. Synthesis of 4-*N*-[di-(2-Nitrophenyl)ethoxycarbonyl]-Azacytidine, 1



5-Azacytidine (122 mg, 0.5 mmol) was coevaporated with anhydrous pyridine (3 x 5 mL) and subsequently suspended in anhydrous DMF (10 mL) before adding HMDS (542.0 μ L, 2.5 mmol). After stirring at room temperature for 2 h, TLC analysis indicated that protection of the nucleoside was complete. The solvent was removed *in vacuo* and the resultant residue co-evaporated with anhydrous toluene (3 x 5 mL) followed by anhydrous pyridine (3 x 5 mL). The resultant residue was subsequently suspended in anhydrous pyridine (30 mL) and solution of 2-NPE-Cl in anhydrous DCM (1.07 g, 5 mmol) added and the resultant mixture stirred at room temperature overnight. The solvent was removed under reduced pressure and the resultant residue coevaporated with toluene (3 x 10 mL) then suspended in anhydrous DMF (10 mL). Silyl deprotection was achieved with the addition of tris(dimethylamino) sulfonium difluorotrimethylsilicate (TAS-F) (486 mg, 2 mmol). After 2 h the solvent was removed and the resultant residue coevaporated with anhydrous toluene (3 x 5 mL). The crude material was purified by flash chromatography (0-10% EtOH/DCM, 10% EtOH/DCM R_f = 0.63) to yield the title compound as a colourless film (62.3 mg, 10%). ^1H NMR (600 MHz, DMSO- d_6) δ_{ppm} : 3.13 - 3.20 (m, 4H, H-7''), 3.62-3.67 (m, 1H, H-5'a), 3.83 - 3.88 (m, 1H, H-5'b), 3.95-4.00 (m, 1H, H-4'), 4.07-4.12 (m, 1H, H-3'), 4.12-4.16 (m, 1H, H-2'), 4.39-4.49 (m, 4H, H-8''), 5.11 (d, J =6.78 Hz, 1H, 2'-OH), 5.35 (t, J =4.71 Hz, 1H, 5'-OH), 5.68 (d, J =1.51 Hz, 1H, 3'-OH), 5.74 (d, J =4.89 Hz, 1H, H-1'), 7.42-7.54 (m, 4H, H-4'' & H-6''), 7.60-7.68 (m, 2H, H-5''), 7.89 - 7.95 (m, 2H, H-3''), 9.10 (s, 1H, H-6). ^{13}C NMR (600 MHz, DMSO- d_6) δ_{ppm} : 30.63 (C-7''), 58.87 (C-5'), 67.67 (C-3' & C-8''), 74.03 (C-2'), 83.98 (C-4'), 91.26 (C-1'), 118.16 (C-1''), 124.54 - 133.42 (C-3'', C-4'', C-5'' & C-

6"), 149.15 (C-2"), 150.18 (C-2), 152.68 (C-9"), 159.05 (C-6), 163.61 (C-4). *m/z* (ESI) 631.09 [M+H]⁺, C₂₆H₂₆N₆O₁₃ requires 630.51704.

4.7. Cellular Assays

A human osteosarcoma cell line, SaOS-2, was used for the cellular assays. The cells were grown in a humidified 37 °C incubator containing 5% CO₂. A 24-well plate format was used; each well capable of holding a maximum volume of 0.5 mL growth medium. Cells were seeded at 1 X 10⁴ cells per well 24 h prior to treatment with azacytidine. Two forms of the nucleoside analogue were used for the study; caged **1** and un-caged **1** (100 μM stock solutions). The un-caged sample had been irradiated at 365 nm for 1 h. HPLC and MS analysis confirmed the presence of the released azacytidine, **4**. Four treatment conditions were investigated; cells with caged **1**, cells with irradiated **1**, cells with acetonitrile and cells in growth media alone. Three concentrations of caged-azacytidine, **1** were tested; 0.5 μM, 1 μM and 1.5 μM. Caged azacytidine and acetonitrile were added every 24 h in fresh growth media over a 96 h period. The confluence of the cells was measured every 2 h over this period.

5 OUTLOOK

The overarching aim of this project was to develop novel sensing approaches which made use of native α HL protein pores and thus could be readily applied to solid-state nanopores. Although the α HL protein pore is the mainstay of nanopore technology, it is associated with some inherent drawbacks which have limited its use. Most notably, instability of the lipid bilayer which supports the nanopore. Solid-state nanopores, which are typically fabricated from inorganic materials or organic polymers² are more robust alternatives to biological pores. As result of their increased stability, they can be more readily integrated into sensing devices and thus are more amenable to commercialisation. However, they cannot be modified to the same atomic-scale precision as a biological pore, which limits their scope. Additionally, they are far less sensitive than biological pores. In this project, attempts were made to overcome this inherent discrepancy by developing sensing approaches where selectivity and sensitivity are not achieved via pore modification, but through modification of the analyte.

Chapter 2 describes efforts to develop a nanopore-based approach to detect the enzymatic activity of a protease enzyme. Previously reported strategies¹⁴ have been based upon the detection of multiple peptide fragments generated as the enzyme acted upon a substrate in bulk solution, or required engineering of the pore to improve sensitivity. In the presented approach, multiple copies of a peptide substrate were tethered via a flexible linker to PEGA beads. The beads prevented the substrate from translocating the pore. Enzymatic cleavage of the immobilised substrate released a soluble peptide fragments. These were subsequently electrically detected as they traversed the nanopore. The release of a single substrate greatly simplified data analysis. Initial proof-of-concept experiments were performed with trypsin and a simple peptide substrate. Using HPLC analysis, it was observed that the rate of fragment release increased with reaction time. However, a disparity was observed between the rates of peptide release in homogenous (solution-phase) and heterogeneous (on-bead) reactions, with the latter occurring 35%

slower than the former. This difference was attributed to the enzyme having difficulty diffusing into the PEGA beads to digest the substrate. Although PEGA beads are permeable to enzymes with molecular weights up to 45 kD¹¹⁶ its mere presence retarded the action of the enzyme. In future experiments, this could be overcome by employing solid supports that possess higher surface-to-volume ratios or by simply placing a longer linker between the substrate and the resin.¹¹³ Alternatively, studies have shown that introducing positive charges into the PEGA resin improve its swelling properties and facilitate biotransformation.

Further experiments were performed in the using renin as a model enzyme and a His₆-tagged analogue of its natural substrate, angiotensinogen. The experiments were performed in the presence of human serum proteins in order to mimic biological conditions. It was envisaged that placing a His₆-tag at the *N*-terminus of the substrate would enable enzymatic digestion and downstream purification to be performed in a single reaction vessel. In the course of the investigation it was demonstrated that the presence of the His₆-tag did not interfere with the catalytic ability of the enzyme. Specifically, the His₆-tagged fragment, angiotensin I-His₆, was released. In the presence of Ni²⁺-NTA beads, the digested fragment could be concomitantly released, bound and eluted. Furthermore, the enzymatically released fragment could be isolated from His-rich serum proteins. However, only approximately 48% of the enzymatically released peptide was recovered. This would need to be vastly improved in order to develop commercially viable technology. Utilising a shallower imidazole elution gradient may aid this. Interestingly, non-specific release of the intact substrate was also observed. This was attributed to hydrolysis of the ester bond between the peptide and the linker. Although the hydrolytically released peptide only accounted for 5% of the amount released enzymatically, this may still be too high for commercial purposes. While removing the linker and directly coupling the peptide to resin via a permanent amide bond would resolve this issue, it would be impossible to assess the purity of the peptide during synthesis. Alternative synthetic protocols would be required. Attempts to synthesise the intact peptide prior to conjugation to the resin gave poor yields. An alternative method

would be to include a biotin molecule onto the C-terminus of the peptide. This would enable conjugation to streptavidin coated beads. This could also potentially solve the problems concerning diffusion of the enzyme to the bead surface and removal of residual His-rich plasma proteins. The next step in this investigation would be to determine whether the nanopore is sensitive enough to detect biologically relevant concentrations of the enzyme.

Chapter 3 continues in the same vein as Chapter 2, describing efforts to develop a nanopore-based strategy to characterise forensically important nucleotide repeat sequences, which are the basis of DNA fingerprinting. Nanopore analysis of ssDNA has been hampered by rapid translocation speeds and subsequent poor resolution of individual bases. It was envisaged that a nanopore-based sensing strategy could be realised by using DNA strands modified with nucleotides bearing bulky adamantane tags. It was hoped that these tags would increase the steric bulk and resistance of the DNA stands as they translocated the pore. The net effect would be a reduction in the translocation speed and improved resolution. Incorporation of one adamantane tag per repeat unit was expected to yield step-like blockades which could easily be counted and measured to determine the number and length of the repeats.

Three adamantane-bearing nucleosides, each possessing different linker lengths, and their corresponding triphosphate derivatives were synthesised. The nucleotides were successfully incorporated into a model DNA template using PCR. Hence they were shown to be suitable substrates for DNA pol. The corresponding phosphoramidite derivatives were also synthesised and utilised to generate model oligonucleotides bearing STR repeat units. However, nanopore analysis of the DNA strands did not yield the expected results. The oligonucleotides gave rise to fleeting reductions in ionic conductivity rather than the expected step-like blockades. In spite of this, a linear correlation was observed between the number of tags present in the oligonucleotides and the dwell time. This suggested that the presence of the adamantane tags increased the resistance experienced by the oligonucleotides. The lack of step-like blockades suggested that the adamantane was not

sufficiently large to occlude the movement of ions through the pore. In the future, this could be overcome by utilising a larger tag or introducing a positive charge into the linker or tag itself. As oligonucleotides possess a net negative charge, a positively charged linker may facilitate the formation of a more compact oligo-tag segment. A positive charge could be introduced in the form of a quaternary ammonium species after reduction of the amide bond.

In the course of this project, three novel modified nucleosides, and their corresponding triphosphate derivatives were synthesised. Furthermore, the versatility of the Yoshikawa reaction was demonstrated.

The motivation of the work presented in Chapter 4 deviates from that of the previous chapters, however it builds on the notion of employing nucleoside analogues to overcome challenges encountered in the arena of chemical biology. The specific aim of this project was to synthesise photocaged derivatives of the nucleoside drug azacytidine, which is used to treat cancers that arise due to aberrant methylation of particular regions of the genome. Hypermethylation of these regions can abnormally enhance the function of some genes, such as the oncogenes, or cause other, such as tumour-suppressor genes, to be silenced. This can enable the proliferation of certain cancers and trigger the onset of some diseases. Azacytidine acts as a potent anticancer agent by inducing a hypomethylated state through inhibition of DNA methyltransferase, the enzyme that catalyses DNA methylation. However, the drug is associated with extreme toxicity and side-effects due to its non-specific actions. It was envisaged that a photocaged derivative of azacytidine would enable greater temporal and spatial controls over drug release.

Two photocaged derivatives of azacytidine were synthesised, di-NPE-aza and NOVC-aza. However, only di-NPE-aza was isolated in sufficient purity (71%) to enable investigation of its photochemical properties and activity in cellular assays. In accordance with literature reports,^{217,262} the nucleoside possessed two absorption maxima; one at 237 nm and the other at 260 nm. These corresponded to the triazine base of azacytidine and the NPE protecting group, respectively. Uncaging of the nucleoside was effected by

irradiation at 365 nm using a LED photoreactor. Photolysis was monitored by analytical HPLC and MS analysis over several hours. The results indicated that the caged nucleoside was uncaged to the parent nucleoside and photo-by-products in near quantitative yields within 3 h.

Cellular assays were performed to determine the effect of the nucleoside analogue on cell viability. Central to this drug delivery approach was the notion that the caged nucleoside must not impart any cytotoxicity, and that toxicity must only occur once the nucleoside is uncaged to the biologically active parent nucleoside. The caged nucleoside conformed to this notion somewhat. A reduction in the rate of cell growth was observed in the presence of di-NPE-aza. However, a more drastic reduction was observed in the presence of di-NPE-aza which had previously been irradiated at 365 nm. The results suggested that the released azacytidine was having a greater affect on cell viability. Additional control experiments also indicated that the ACN present in the nucleoside solution also had a significant negative impact on cell growth. ACN was added to aid dissolution of the nucleoside. Whilst the conditions were not physiologically relevant, the results can be viewed as proof-of-concept. The next logical step in this investigation would be to synthesise a nucleoside analogue which is water soluble. This would enable its photochemical properties and cytotoxicity to be determined under physiologically relevant conditions. This could potentially be achieved by placing a carboxylic functionality directly on the aromatic ring or the methylenes.^{269,270} Alternatively, the monophosphate derivative of di-NPE-aza could be a suitable candidate. The phosphate group would likely render the analogue water soluble. However, the negative charge could hinder cellular uptake by passive and active diffusion. A suitable protecting group, which could be removed intracellularly without causing toxicity, would be required. Akin to PDT, it would also be desirable to have a nucleoside analogue which was uncaged at longer wavelengths, preferably in the 600-700 nm range. This would facilitate uncaging deep within tissue. Light at 365 nm will not penetrate tissue sufficiently.

REFERENCES

1. J. J. Nakane, M. Akesson, and A. Marziali, *J. Phys.: Condens. Matter.*, 2003, **15**, R1365–R1393.
2. S. Howorka and Z. Siwy, *Chem. Soc. Rev.*, 2009, **38**, 2360–2384.
3. B. M. Venkatesan and R. Bashir, *Nat. Nanotechnol.*, 2011, **6**, 615–624.
4. R. D. Maitra, J. Kim, and W. B. Dunbar, *Electrophoresis.*, 2012, **33**, 3418–3428.
5. G. Menestrina, M. Dalla Serra, and G. Prévost, *Toxicon*, 2001, **39**, 1661–1672.
6. L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J. E. Gouaux, *Science*, 1996, **274**, 1859–1865.
7. H. Bayley and C. R. Martin, *Chem. Rev.*, 2000, **100**, 2575–2594.
8. R. Stefureac, Y. Long, H.-B. Kraatz, P. Howard, and J. S. Lee, *Biochemistry*, 2006, **45**, 9172–9179.
9. N. Mitchell and S. Howorka, *Angew. Chem., Int. Ed.*, 2008, **47**, 5565–5568.
10. C. Dekker, *Nat. Nanotechnol.*, 2007, **2**, 209–215.
11. O. Braha, B. Walker, S. Cheley, J. J. Kasianowicz, L. Song, J. E. Gouaux, and H. Bayley, *Chem. Biol.*, 1997, **4**, 497–505.
12. J. E. Gouaux, O. Braha, M. R. Hobaugh, L. Song, S. Cheley, C. Shustak, and H. Bayley, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12828–12831.
13. X. Guan, L.-Q. Gu, S. Cheley, O. Braha, and H. Bayley, *ChemBioChem*, 2005, **6**, 1875–1881.
14. Q. Zhao, R. S. S. de Zoysa, D. Wang, D. A. Jayawardhana, and X. Guan, *J. Am. Chem. Soc.*, 2009, **131**, 6324–6325.
15. J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley, *Nat. Nanotechnol.*, 2009, **4**, 265–270.
16. L.-Q. Gu, O. Braha, S. Conlan, S. Cheley, and H. Bayley, *Nature*, 1999, **398**, 686–690.
17. X. Kang, L.-Q. Gu, S. Cheley, and H. Bayley, *Angew. Chem., Int. Ed.*, 2005, **44**, 1495–1499.
18. D. K. Shenoy, W. R. Barger, A. Singh, R. G. Panchal, M. Misakian, V. M. Stanford, and J. J. Kasianowicz, *Nano Lett.*, 2005, **5**, 1181–1185.
19. T.-J. Jeon, N. Malmstadt, and J. J. Schmidt, *J. Am. Chem. Soc.*, 2005, **128**, 42–43.
20. B. M. Venkatesan, A. B. Shah, J.-M. Zuo, and R. Bashir, *Adv. Funct. Mater.*, 2010, **20**, 1266–1275.
21. P. Chen, T. Mitsui, D. B. Farmer, J. Golovchenko, R. G. Gordon, and D. Branton, *Nano Lett.*, 2004, **4**, 1333–1337.
22. M. Wanunu and A. Meller, *Nano Lett.*, 2007, **7**, 1580–1585.
23. F. S. Collins, *Science*, 2003, **300**, 286–290.
24. E. S. Lander et al, *Nature*, 2001, **409**, 860–921.
25. W. G. Feero, A. E. Guttmacher, and F. S. Collins, *N. Engl. J. Med.*, 2010, **362**, 2001–2011.
26. *The Economist*, 2011.
27. F. Sanger, S. Nicklen, and A. R. Coulson, *Proc. Natl. Acad. Sci.*, 1977, **74**, 5463–5467.
28. J. Shendure and H. Ji, *Nat. Biotechnol.*, 2008, **26**, 1135–1145.
29. E. S. Lander et al., *Nature*, 2001, **409**, 860–921.
30. E. E. Schadt, S. Turner, and A. Kasarskis, *Hum. Mol. Genet.*, 2010, **19**, R227–R240.

31. D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, and J. A. Schloss, *Nat. Biotechnol.*, 2008, **26**, 1146–1153.
32. J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13770–13773.
33. M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, and D. W. Deamer, *Biophys. J.*, 1999, **77**, 3227–3233.
34. A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 1079–1084.
35. D. G. Wang, *Science*, 1998, **280**, 1077–1082.
36. J. N. Hirschhorn, K. Lohmueller, E. Byrne, and K. Hirschhorn, *Genet. Med.*, 2002, **4**, 45–61.
37. V. Borsenberger, N. Mitchell, and S. Howorka, *J. Am. Chem. Soc.*, 2009, **131**, 7530–7531.
38. M. S. Cooke, R. Olinski, and M. D. Evans, *Clin. Chim. Acta.*, 2006, **365**, 30–49.
39. A. E. P. Schibel, N. An, Q. Jin, A. M. Fleming, C. J. Burrows, and H. S. White, *J. Am. Chem. Soc.*, 2010, **132**, 17992–17995.
40. J. Nakane, M. Wiggin, and A. Marziali, *Biophys. J.*, 2004, **87**, 615–621.
41. Y. Astier, O. Braha, and H. Bayley, *J. Am. Chem. Soc.*, 2006, **128**, 1705–1710.
42. H.-C. Wu, Y. Astier, G. Maglia, E. Mikhailova, and H. Bayley, *J. Am. Chem. Soc.*, 2007, **129**, 16142–16148.
43. G. Egger, G. Liang, A. Aparicio, and P. A. Jones, *Nature*, 2004, **429**, 457–463.
44. J. Fahy, A. Jeltsch, and P. B. Arimondo, *Expert Opin. Ther. Pat.*, 2012, **22**, 1427–1442.
45. J. G. Herman and S. B. Baylin, *N. Engl. J. Med.*, 2003, **349**, 2042–2054.
46. Available from: <http://www.nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/dna-exonuclease-sequencing->
47. D. Deamer, *Annu. Rev. Biophys.*, 2010, **39**, 79–90.
48. K. R. Lieberman, G. M. Cherf, M. J. Doody, F. Olasagasti, Y. Kolodji, and M. Akeson, *J. Am. Chem. Soc.*, 2010, **132**, 17961–17972.
49. L. Blanco and M. Salas, *J. Biol. Chem.*, 1996, **271**, 8509–8512.
50. M. Salas, L. Blanco, J. M. Lázaro, and M. de Vega, *IUBMB Life*, 2008, **60**, 82–85.
51. E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, and J. H. Gundlach, *Nat. Biotechnol.*, 2012, **30**, 349–353.
52. M. X. Macrae, S. Blake, X. Jiang, R. Capone, D. J. Estes, M. Mayer, and J. Yang, *ACS Nano*, 2009, **3**, 3567–3580.
53. I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, and J. H. Gundlach, *Proc. Natl. Acad. Sci.*, 2010, **107**, 16060–16065.
54. E. A. Manrao, I. M. Derrington, M. Pavlenok, M. Niederweis, and J. H. Gundlach, *PLoS ONE*, 2011, **6**, e25723.
55. X. S. Xie and H. P. Lu, *J. Biol. Chem.*, 1999, **274**, 15967–15970.
56. R. Eisenthal and M. J. Danson, *Enzyme Assays: A Practical Approach*, Oxford University Press, 2002.
57. D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Nat. Rev. Genet.*, 2013, **14**, 9–22.
58. S. L. Cockroft, J. Chu, M. Amorin, and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2008, **130**, 818–820.

59. F. Olasagasti, K. R. Lieberman, S. Benner, G. M. Cherf, J. M. Dahl, D. W. Deamer, and M. Akeson, *Nat. Nanotechnol.*, 2010, **5**, 798–806.
60. D. R. Garalde, C. A. Simon, J. M. Dahl, H. Wang, M. Akeson, and K. R. Lieberman, *J. Biol. Chem.*, 2011, **286**, 14480–14492.
61. R. D. Kuchta, V. Mizrahi, P. A. Benkovic, K. A. Johnson, and S. J. Benkovic, *Biochemistry*, 1987, **26**, 8410–8417.
62. S. Cheley, H. Xie, and H. Bayley, *ChemBioChem*, 2006, **7**, 1923–1927.
63. H. Xie, O. Braha, L.-Q. Gu, S. Cheley, and H. Bayley, *Chem. Biol.*, 2005, **12**, 109–120.
64. L. Movileanu, *Trends. Biotechnol.*, 2009, **27**, 333–341.
65. L. Movileanu, *Soft Matter*, 2008, **4**, 925–931.
66. B. Domon and R. Aebersold, *Science*, 2006, **312**, 212–217.
67. H. J. Issaq, T. P. Conrads, G. M. Janini, and T. D. Veenstra, *Electrophoresis*, 2002, **23**, 3048–3061.
68. D. S. Wishart and B. D. Sykes, in *Methods in Enzymology*, ed. N. J. O. Thomas L. James, Academic Press, 1994, vol. Volume 239, pp. 363–392.
69. A. J. Wolfe, Mohammad, S. Cheley, H. Bayley, and L. Movileanu, *J. Am. Chem. Soc.*, 2007, **129**, 14034–14041.
70. L. Movileanu, J. P. Schmittschmitt, J. M. Scholtz, and H. Bayley, *Biophys. J.*, 2005, **89**, 1030–1045.
71. A. S. Panwar and M. Muthukumar, *J. Am. Chem. Soc.*, 2009, **131**, 18563–18570.
72. Q. Zhao, D. A. Jayawardhana, D. Wang, and X. Guan, *J. Phys. Chem. B*, 2009, **113**, 3572–3578.
73. H. Castrop, K. Hocherl, A. Kurtz, F. Schweda, and V. Todorov, *Physiol. Rev.*, 2010, **90**, 607–673.
74. M. Paul, A. P. Mehr, and R. Kreuz, *Physiol Rev*, 2006, **86**, 747–803.
75. *Human Physiology*, McGraw-Hill International Edition, 12th edn., 2011.
76. M. A. Ondetti and D. W. Cushman, *Annu. Rev.*, 1982, **51**, 283–308.
77. T. Unger and B. A. Scholkens, *Angiotensin*, Springer, 2004, vol. 163/1.
78. World Health Organisation, *World Health Statistics 2012: A Snapshot of Global Health*.
79. D. R. Davies, *Annu. Rev. Biophys. Biophys. Chem.*, 1990, **19**, 189–215.
80. M. J. Peach, *Physiol. Rev.*, 1977, **57**, 313–370.
81. J. A. Johnson and R. R. Anderson, *The Renin-Angiotensin System*, Plenum Press, 1980, vol. 130.
82. L. T. Skeggs, J. R. Kahn, K. Lentz, and N. P. Shumway, *J. Exp. Med.*, 1957, **106**, 439–453.
83. D. A. Tewsbury, R. A. Dart, and J. Travis, *Biochem. Biophys. Res. Commun*, 1981, **99**, 1311–1315.
84. K. Arakawa, M. Nakatani, A. Minohara, and M. Nakamura, *Biochem. J.*, 1967, **104**, 900–906.
85. J. Burton and T. Quinn, *Biochem. Biophys. Acta.*, 1988, **952**, 8–12.
86. M. Poe, J. K. Wu, T.-Y. Lin, K. Hoogsteen, H. G. Bull, and E. E. Slater, *Anal. Biochem.*, 1984, **140**, 459–467.
87. W. Wang and T. C. Liang, *Biochemistry*, 1994, **33**, 14636–14641.
88. I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun*, 1967, **27**, 157–162.

89. M. Volpe, A. Battistoni, D. Chin, S. Rubattu, and G. Tocci, *Nutr., Metab. Cardiovasc. Dis.*, 2012, **22**, 312–317.
90. G. Remuzzi, N. Perico, M. Macia, and P. Ruggenenti, *Kidney Int.*, 2005, **68**, S57–S65.
91. A. H. Danser, *Curr. Opin. Nephrol. Hypertens.*, 2012, **21**, 508–514.
92. D. Hartman, G. A. Sagnella, C. A. Chesters, and G. A. MacGregor, *Clin. Chem.*, 2005, **50**, 2159–2161.
93. J. E. Sealey, J. Gerten-Banes, and J. H. Laragh, *Kidney Int.*, 1972, **1**, 240–253.
94. J. Brossaud and J.-B. Corcuff, *Clin. Chem. Acta.*, 2009, **410**, 90–92.
95. D. J. Campbell, J. Nussberger, M. Stowasser, A. H. J. Danser, A. Morganti, E. Frandsen, and J. Menard, *Clin. Chem.*, 2009, **55**, 867–877.
96. M. Azizi and J. Menard, *J. Renin-Angio-Aldo. S.*, 2010, **11**, 89–90.
97. T. Chard, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier, 3rd edn., 1987, vol. 6.
98. D. Gruson, D. Maisin, P. Lison, D. Maiter, and A. Persu, *Biomarkers.*, 2011, **16**, 605–609.
99. A. Morganti, *J. Hypertens.*, 2010, **28**, 1307–1312.
100. N. Nakamura-Imajo, S. Satomura, S. Maturra, and K. Murakami, *Clin. Chim. Acta.*, 1992, **211**, 47–57.
101. K. Paschalidou, U. Neumann, B. Gerhartz, and C. Tzougraki, *Biochem. J.*, 2004, **382**, 1031–1038.
102. G. T. Wang, C. C. Chung, T. F. Holzman, and G. A. Krafft, *Anal. Biochem.*, 1993, **210**, 351–359.
103. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154.
104. A. Meister, Ed., in *Advances in Enzymology and Related Areas of Molecular Biology*, 1969, vol. 32, pp. 221–297.
105. G. B. Fields and R. L. Noble, *Int. J. Pept. Protein Res.*, 1990, **35**, 161–214.
106. W. C. Chan and P. D. White, *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*, Oxford University Press, 2000.
107. C. A. G. N. Montalbetti and V. Falque, *Tetrahedron.*, 2005, **61**, 10827–10852.
108. S. Y. Han and Y.-A. Kim, *Tetrahedron*, 2004, **60**, 2447–2467.
109. M. Kukwikila and S. Howorka, *J. Phys.: Condes. Matter.*, 2010, **22**, 454103.
110. V. Scellenberger, C. W. Turck, L. Hedstrom, and W. J. Rutter, *Biochemistry*, 1993, **32**, 4349–4353.
111. V. Scellenberger, C. W. Turck, and W. J. Rutter, *Biochemistry*, **33**, 4251–4257.
112. T. Kurth, D. Ullmann, H.-D. Jakubke, and L. Hedstrom, *Biochemistry*, 1997, **36**, 10098–10104.
113. J. Deere, R. F. De Oliveira, B. Tomaszewski, S. Millar, A. Lalaoui, L. F. Solares, S. Flitsch, and P. J. Halling, *Langmuir*, 2008, **24**, 11762–11769.
114. M. Meldal, *Tetrahedron Lett.*, 1992, **33**, 3077–3080.
115. A. Basso, P. Braiuca, C. Ebert, L. Gardossi, and P. Linda, *J. Chem. Technol. Biotechnol.*, 2006, **81**, 1626–1640.
116. J. Kress, R. Zanaletti, A. Amour, M. Ladlow, J. G. Frey, and M. Bradley, *Chem. Eur. J.*, 2002, **8**, 3769–3772.
117. M. Karas and F. Hillenkamp, *Anal. Chem.*, 1988, **60**, 2301–2303.
118. I. Coin, M. Beyermann, and M. Bienert, *Nat. Protoc.*, 2007, **2**, 3247–3256.

119. P. Lloyd-Williams, F. Albericio, and E. Giralt, *Chemical Approaches to the Synthesis of Peptides and Proteins*, CRC Press Taylor & Francis Group, 1997.
120. A. Basso, R. V. Ulijn, S. L. Flitsch, G. Margetts, I. Brazendale, C. Ebert, L. De Martin, P. Linda, S. Verdelli, and L. Gardossi, *Tetrahedron*, 2004, **60**, 589–594.
121. Novabiochem Letters No. 2/08, .
122. P. J. Halling, R. V. Ulijn, and S. L. Flitsch, *Curr. Opin. Biotech.*, 2005, **16**, 385–392.
123. N. Mitchell, Thesis, University College London, 2009.
124. H. R. Horton, L. A. Moran, R. S. Ochs, J. D. Rawn, and K. G. Scrimgeour, *Principles of Biochemistry*, Pearson Education International, 3rd edn., 2002.
125. T. C. Sutherland, Y.-T. Long, R.-I. Stefureac, I. Bediako-Amoa, H.-B. Kraatz, and J. S. Lee, *Nano Lett.*, 2004, **4**, 1273–1277.
126. S. Leon, R. Quarrell, and G. Lowe, *Bioorgan. Med. Chem. Lett.*, 1998, **8**, 2997–3002.
127. B. A. Maltman, M. Bejugam, and S. L. Flitsch, *Org. Biomol. Chem.*, **3**, 2505–2507.
128. *A handbook for high-level expression and purification of 6xHis-tagged proteins. Fifth Edition.*, Qiagen, 2003.
129. M. De, S. Rana, H. Akpinar, O. R. Miranda, R. R. Arvizo, U. H. F. Bunz, and V. M. Rotello, *Nat. Chem.*, 2009, **1**, 461–465.
130. S. Mori, H. K. Takahashi, K. Yamaoka, M. Okamoto, and M. Nishibori, *Life Sci.*, 2003, **73**, 93–102.
131. G. Candiano, M. Bruschi, L. Musante, L. Santucci, G. M. Ghiggeri, B. Carnemolla, P. Orecchia, L. Zardi, and P. G. Righetti, *Electrophoresis.*, 2004, **25**, 1327–1333.
132. M. Gude, J. Ryf, and P. D. White, *Lett. Pept. Sci.*, 2002, **9**, 203–206.
133. O. Braha, B. Walker, S. Cheley, J. J. Kasianowicz, L. Song, J. E. Gouaux, and H. Bayley, *Chem. Biol.*, 1997, **4**, 497–505.
134. L. Movileanu, S. Howorka, O. Braha, and H. Bayley, *Nat. Biotechnol.*, 2000, **18**, 1091–1095.
135. L. Berti and G. A. Burley, *Nat. Nanotechnol.*, 2008, **3**, 81–87.
136. M. G. Blackburn, M. J. Gait, D. Loakes, and D. M. Williams, *Nucleic Acids in Chemistry and Biology*, The Royal Society of Chemistry, 3rd edn., 2006.
137. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
138. J. C. Venter et al., *Science*, 2001, **291**, 1304–1351.
139. J. M. Butler, *Forensic DNA Typing: Biology, Technology and Genetics of STR markers*, Elsevier Academic Press, 2nd edn., 2005.
140. H. Ellegren, *Nat. Rev. Genet.*, 2004, **5**, 435–445.
141. G. K. Chambers and E. S. MacAvoy, *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.*, 2000, **126**, 455–476.
142. A. J. Jeffreys, V. Wilson, and S. L. Thein, *Nature*, 1985, **314**, 67–73.
143. A. J. Jeffreys, V. Wilson, and S. L. Thein, *Nature*, 1985, **316**, 76–79.
144. H. C. Lee, C. Ladd, M. T. Bourke, E. Pagliaro, and F. Tirnady, *Am. J. Forensic. Med. Pathol.*, 1994, **15**, 269–282.
145. R. K. G. Saiki, *Science*, 1988, **239**, 487–491.
146. E. Giardina, *Nanomedicine*, 2011, **6**, 257–270.
147. D. Lenoir, R. E. Hall, and P. V. R. Schleyer, *J. Am. Chem. Soc.*, 1974, **96**, 2138–2148.
148. M. Kuwahara and N. Sugimoto, *Molecules.*, 2010, **15**, 5423–5444.
149. J. J. Lipsky, *Lancet.*, 1996, **1996**, 800–803.

150. E. De Clercq, *J. Clin. Virol.*, 2004, **30**, 115–133.
151. T. Kottysch, C. Ahlborn, F. Brotzel, and C. Richert, *Chem. Eur. J.*, 2004, **10**, 4017–4028.
152. G. Giller, T. Tasara, B. Angerer, K. Muhlegger, M. Amacker, and H. Winter, *Nucleic Acid Res.*, 2003, **31**, 2630–2635.
153. V. Borsenberger, M. Kukwikila, and S. Howorka, *Org. Biomol. Chem.*, 2009, **7**, 3826–3835.
154. L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood, *Nature.*, 1986, **321**, 674–679.
155. X. Chen and P. Y. Kwok, *Nucleic Acids Research*, 1997, **25**, 347–353.
156. J. A. McNeil, C. V. Johnson, K. C. Carter, R. H. Singer, and J. B. Lawrence, *Gent. Ana.: Biomol. Eng.*, 1991, **8**, 41–58.
157. G. A. Burley, J. Gierlich, M. R. Mofid, H. Nir, S. Tal, Y. Eichen, and T. Carell, *J. Am. Chem. Soc.*, 2006, **128**, 1398–1399.
158. J. Matulic-Adamic, A. T. Daniher, A. Karpeisky, P. Haeberli, D. Sweedler, and L. Beigelman, *Bioorgan. Med. Chem.*, 2000, **10**, 1299–1302.
159. L. A. Agrofoglio, I. Gillaizeau, and Y. Saito, *Chem. Rev.*, 2003, **103**, 1875–1916.
160. K. Sonogashira, Y. Tohda, and N. Hagihara, *Tetrahedron Lett.*, 1975, **16**, 4467–4470.
161. N. Miyaura and A. Suzuki, *Chem. Rev.*, 1995, **95**, 2457–2483.
162. F. W. Hobbs, *J. Org. Chem.*, 1989, **54**, 3420–3422.
163. J. M. Neumann, S. Tran-Dinh, J. M. Bernassau, and M. Gueron, *Eur. J. Biochem.*, 1980, **108**, 457–463.
164. A. E. V. Haschemeyer and A. Rich, *J. Mol. Biol.*, 1967, **27**, 369–384.
165. Y.-C. Chang, J. Herath, T. H.-H. Wang, and C. S. Chow, *Bioorg. Med. Chem.*, 2008, **16**, 2676–2686.
166. H. Rosemeyer, G. Toth, B. Golankiewicz, Z. Kazimierczuk, W. Bourgeois, U. Kretschmer, H. P. Muth, and F. Seela, *J. Org. Chem.*, 1990, **55**, 5784–5790.
167. K. Burgess and D. Crook, *Chem. Rev.*, 2000, **2000**, 2047–2059.
168. G. K. Wagner, T. Pesnot, and R. A. Field, *Nat. Prod. Rep.*, 2009, **26**, 1172–1194.
169. M. Yoshikawa, T. Kato, and T. Takenishi, *Bull. Chem. Soc. Jpn.*, 1969, **42**, 3505–&.
170. M. Yoshikawa, T. Kato, and T. Takenishi, *Tetrahedron Lett.*, 1967, **8**, 5065–5068.
171. A. R. Kore, M. Shanmugasundaram, A. Senthilvelan, and B. Srinivasan, *Nucleosides, Nucleotides Nucleic Acids.*, 2012, **31**, 423–431.
172. F. Cramer and M. Winter, *Chem. Ber.*, 1961, **94**, 989–996.
173. R. W. Alder, P. S. Bowman, W. R. S. Steele, and D. R. Winterman, *Chem. Commun. (London)*, 1968, 723–724.
174. T. Kovács and L. Ötvös, *Tetrahedron Lett.*, 1988, **29**, 4525–4528.
175. J. G. Moffatt and H. G. Khorana, *J. Am. Chem. Soc.*, 1961, **83**, 649–658.
176. J. G. Moffatt, *Can. J. Chem.*, 1964, **42**, 599–604.
177. V. Wittmann and C.-H. Wong, *J. Org. Chem.*, 1997, **62**, 2144–2147.
178. T. Mukaiyama and M. Hashimoto, *Bull. Chem. Soc. Jpn.*, 1971, **44**, 2284.
179. T. Mukaiyama, *Angew. Chem., Int. Ed.*, 1976, **15**, 94–103.
180. T. Tasara, B. Angerer, M. Damond, H. Winter, S. Dorhofer, U. Hubscher, and M. Amacker, *Nucleic Acid Res.*, 2003, **31**, 2636–2646.
181. O. Thum, S. Jäger, and M. Famulok, *Angew. Chem., Int. Ed.*, 2001, **40**, 3990–3993.

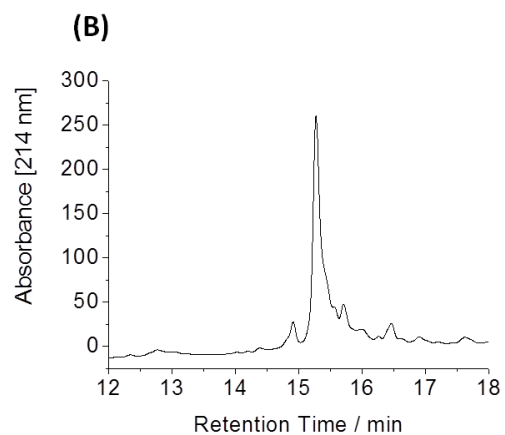
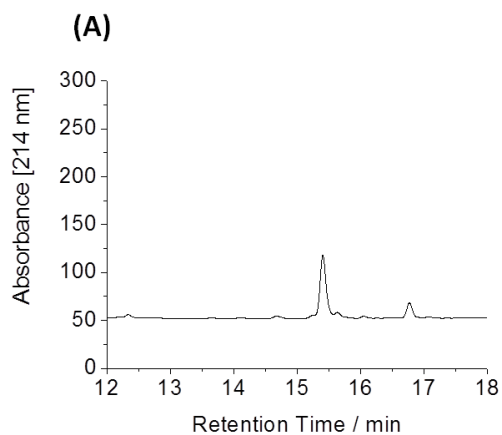
182. S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, and T. Ellenberger, *Nature*, 1998, **391**, 251–258.
183. J. M. Butler, *J. Forensic Sci.*, 2006, **51**, 253–265.
184. M. Gait, *Oligonucleotide Synthesis: A practical Approach*, IRL Press, Oxford, 1984.
185. P. Herdewijn, *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, Totowa, New Jersey, 2005, vol. 288.
186. L. J. McBride and M. H. Caruthers, *Tetrahedron Lett.*, 1983, **24**, 245–248.
187. V. T. Ravikumar and R. K. Kumar, *Org. Process Res. Dev.*, 2004, **8**, 603–608.
188. A. H. Krotz, C. Rentel, D. Gorman, P. Olsen, H. J. Gaus, J. V. McArdle, and A. N. Scozzari, *Nucleosides Nucleotides Nucleic Acids*, 2004, **23**, 767–775.
189. J. Mathé, A. Aksimentiev, D. R. Nelson, K. Schulten, and A. Meller, *PNAS*, 2005, **102**, 12377–12382.
190. L.-A. Fendt, I. Bouamaied, S. Thöni, N. Amiot, and E. Stulz, *J. Am. Chem. Soc.*, 2007, **129**, 15319–15329.
191. E. Kaminskas, A. T. Farrell, Y.-C. Wang, R. Sridhara, and R. Pazdur, *The Oncologist*, 2005, **10**, 176–182.
192. J. M. Foulks, K. M. Parnell, R. N. Nix, S. Chau, K. Swierczek, M. Saunders, K. Wright, T. F. Hendrickson, K.-K. Ho, M. V. McCullar, and S. B. Kanner, *J. Biomol. Screening.*, 2012, **17**, 2–17.
193. S. L. Berger, T. Kouzarides, R. Shiekhattar, and A. Shilatifard, *Genes Dev.*, 2009, **23**, 781–783.
194. A. Bird, *Genes Dev.*, 2002, **16**, 6–21.
195. D. Takai and P. A. Jones, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 3740–3745.
196. C. De Smet, C. Lurquin, B. Lethé, V. Martelange, and T. Boon, *Mol. Cell. Biol.*, 1999, **19**, 7327–7335.
197. T. H. Bestor, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 1990, **326**, 179–187.
198. R. Z. Jurkowska, T. P. Jurkowski, and A. Jeltsch, *ChemBioChem*, 2011, **12**, 206–222.
199. A. P. Feinberg and B. Vogelstein, *Nature*, 1983, **301**, 89–92.
200. M. A. Gama-Sosa, V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke, and M. Ehrlich, *Nucleic Acids Res.*, 1983, **11**, 6883–6894.
201. P. A. Jones and S. B. Baylin, *Nat. Rev. Genet.*, 2002, **3**, 415–428.
202. J. Christman, *Oncogene*, 2002, **21**, 5483–5495.
203. K. D. Robertson, *Nat. Rev. Genet.*, 2005, **6**, 597–610.
204. J.-P. J. Issa, H. M. Kantarjian, and P. Kirkpatrick, *Nat. Rev. Drug Discovery.*, 2005, **4**, 275–276.
205. A. Glover and B. Leyland-Jones, *Cancer. Treat. Rep.*, 1987, **71**, 959–964.
206. S. Preston-Martin, M. C. Pike, R. K. Ross, P. A. Jones, and B. E. Henderson, *Cancer Res.*, 1990, **50**, 7415–7421.
207. C. G. Bochet, *J. Chem. Soc., Perkin Trans. 1*, 2002, **0**, 125–142.
208. G. C. R. Ellis-Davies, *Nat. Methods.*, 2007, **4**, 619–628.
209. G. Mayer and A. Heckel, *Angew. Chem., Int. Ed.*, 2006, **45**, 4900–4921.
210. R. Wieboldt, K. R. Gee, L. Niu, D. Ramesh, B. K. Carpenter, and G. P. Hess, *Proc. Natl. Acad. Sci.*, 1994, **91**, 8752–8756.

211. H.-M. Lee, D. R. Larson, and D. S. Lawrence, *ACS Chem. Biol.*, 2009, **4**, 409–427.
212. H. Yu, J. Li, D. Wu, Z. Qiu, and Y. Zhang, *Chem. Soc. Rev.*, 2010, **39**, 464–473.
213. L. Wang, J. E. T. Corrie, and J. F. Wootton, *J. Org. Chem.*, 2002, **67**, 3474–3478.
214. K. R. Gee, L. Niu, K. Schaper, and G. P. Hess, *J. Org. Chem.*, 1995, **60**, 4260–4263.
215. K. R. Gee, L. Niu, K. Schaper, V. Jayaraman, and G. P. Hess, *Biochemistry*, 1999, **38**, 3140–3147.
216. K. Schaper, S. A. Madani Mobarekeh, P. Doro, and D. Maydt, *Photochem. Photobiol*, 2010, **86**, 1247–1254.
217. S. Walbert, W. Pfeleiderer, and U. E. Steiner, *HCA*, 2001, **84**, 1601–1611.
218. J. P. Celli, B. Q. Spring, I. Rizvi, C. L. Evans, K. S. Samkoe, S. Verma, B. W. Pogue, and T. Hasan, *Chem. Rev.*, 2010, **110**, 2795–2838.
219. I. J. Macdonald and T. J. Dougherty, *J. Porphyr. Phtalocya.*, 2001, **5**, 105–129.
220. B. W. Henderson and T. J. Dougherty, *Photochemistry and Photobiology*, 1992, **55**, 145–157.
221. D. E. J. G. J. Dolmans, D. Fukumura, and R. K. Jain, *Nat. Rev. Cancer.*, 2003, **3**, 380–387.
222. J. Moan and K. Berg, *Photochem. Photobiol.*, 1991, **53**, 549–553.
223. R. A. Mitra and L. J. Singerman, *Optom. Vis. Sci.*, 2002, **79**, 218–224.
224. Fritsch C, Goerz G, and Ruzicka T, *Arch. Dermatol.*, 1998, **134**, 207–214.
225. R. R. Allison, G. H. Downie, R. Cuenca, X.-H. Hu, C. J. Childs, and C. H. Sibata, *Photodiagn. Photodyn.*, 2004, **1**, 27–42.
226. W. M. Sharman, C. M. Allen, and J. E. van Lier, *DDT*, 1999, **4**, 507–517.
227. J. V. Moore, C. M. L. West, and C. Whitehurst, *Phys. Med. Biol.*, 1997, **42**, 913.
228. B. V. Eggert. H. R, *Neurosurgery*, 1987, **21**, 459–64.
229. W.-F. Cheong, S. A. Prahl, and A. J. Welch, *IEEE Journal of Quantum Electronics*, 1990, **26**, 2166–2185.
230. F. L. Arbona, B. Khabiri, and J. A. Norton, *Ultrasound-Guided Regional Anesthesia: A Practical Approach to Peripheral Nerve Blocks and Perineural Catheters*, Cambridge University Press, 2011.
231. G. S. Keller and V. G. Lacombe, *Lasers in Aesthetic Surgery*, Thieme, 2001.
232. C. E. Cass, J. D. Young, and S. A. Baldwin, *Biochem. Cell Biol.*, 1998, **76**, 761–770.
233. C. A. Koczor, R. A. Torres, and W. Lewis, *Expert Opin. Drug Metab. Toxicol.*, 2012, **8**, 665–676.
234. D. R. Rauchwerger, P. S. Firby, D. W. Hedley, and M. J. Moore, *Cancer Res.*, 2000, **60**, 6075–6079.
235. V. L. Damaraju, D. Mowles, S. Yao, A. Ng, J. D. Young, C. E. Cass, and Z. Tong, *Nucleosides, Nucleotides and Nucleic Acids*, 2012, **31**, 236–255.
236. M. K. Kukhanova, *Molecular biology*, 2012, **46**, 768–779.
237. M. Pastor-Anglada, P. Cano-Soldado, M. Molina-Arcas, M. P. Lostao, I. Larráyoz, J. Martínez-Picado, and F. J. Casado, *Virus Res.*, 2005, **107**, 151–164.
238. C. E. Cass, L. A. Gaudette, and A. R. P. Paterson, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1974, **345**, 1–10.
239. T. P. Zimmerman, W. B. Mahony, and K. L. Prus, *J. Biol. Chem.*, 1987, **262**, 5748–5754.
240. A. Q. Siddiqui, C. Ballatore, C. McGuigan, E. De Clercq, and J. Balzarini, *J. Med. Chem.*, 1999, **42**, 393–399.

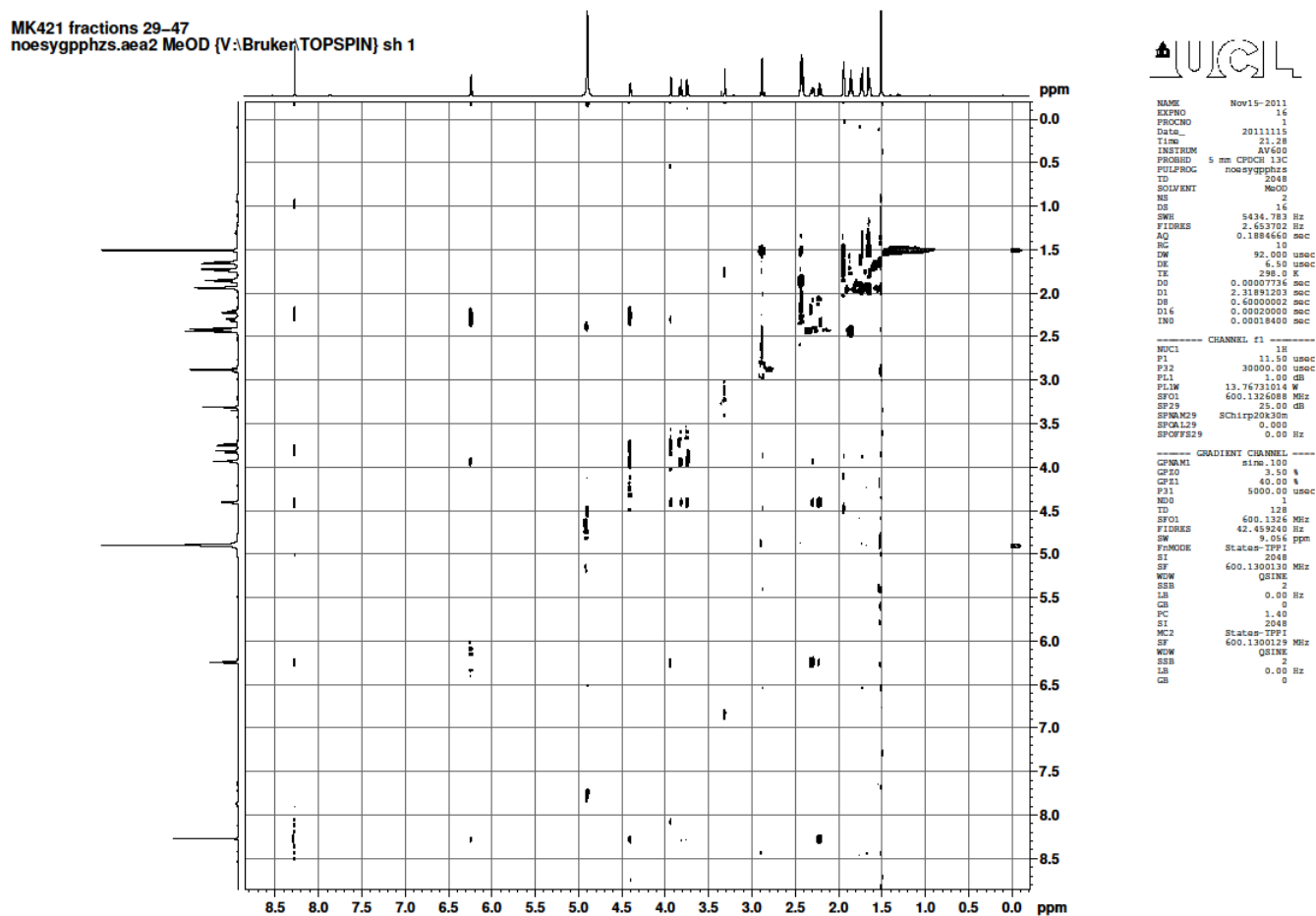
241. S. A. Thomas, A. Bye, and M. B. Segal, *J. Pharmacol. Exp. Ther.*, 2001, **298**, 947–953.
242. N. N. Suzuki, K. Koizumi, M. Fukushima, A. Matsuda, and F. Inagaki, *Structure*, 2004, **12**, 751–764.
243. A. R. Van Rompay, M. Johansson, and A. Karlsson, *Pharmacol. Ther.*, 2003, **100**, 119–139.
244. C. Stresemann and F. Lyko, *Int. J. Cancer*, 2008, **123**, 8–13.
245. T. Lee, M. Karon, and R. L. Momparler, *Cancer Res.*, 1974, **34**, 2482–2488.
246. L. Richard, Momparler, and D. Derse, *Biochem. Pharmacol.*, 1979, **28**, 1443–1444.
247. A. R. V. Rompay, A. Norda, K. Lindén, M. Johansson, and A. Karlsson, *Mol. Pharmacol.*, 2001, **59**, 1181–1186.
248. J. Bauman, C. Verschraegen, S. Belinsky, C. Muller, T. Rutledge, M. Fekrazad, M. Ravindranathan, S.-J. Lee, and D. Jones, *Cancer. Chemother. Pharmacol.*, 2011, **69**, 547–554.
249. K. Alvarez, J.-J. Vasseur, T. Beltran, and J.-L. Imbach, *J. Org. Chem.*, 1999, **64**, 6319–6328.
250. F. Himmelsbach, B. S. Schulz, T. Trichtinger, R. Charubala, and W. Pfeleiderer, *Tetrahedron*, 1984, **40**, 59–72.
251. M. I. Sanchez, O. Vazquez, M. E. Vazquez, and J. L. Mascareñas, *Chem. Commun.*, 2011, **47**, 11107–11109.
252. J. W. Højfeldt and K. V. Gothelf, *J. Org. Chem.*, 2006, **71**, 9556–9559.
253. R. G. García, A. S. Brank, J. K. Christman, V. E. Marquez, and R. Eritja, *Antisense Nucleic Acid Drug Dev.*, 2001, **11**, 369–378.
254. P. J. Hrdlicka, J. S. Jepsen, C. Nielsen, and J. Wengel, *Bioorg. Med. Chem.*, 2005, **13**, 1249–1260.
255. D. K. Rogstad, J. L. Herring, J. A. Theruvathu, A. Burdzy, C. C. Perry, J. W. Neidigh, and L. C. Sowers, *Chem. Res. Toxicol.*, 2009, **22**, 1194–1204.
256. J. Ben-Hattar and J. Jiricny, *J. Org. Chem.*, 1986, **51**, 3211–3213.
257. T. Hiyama, *Organofluorine: Chemistry and Applications*, Springer, Germany, 2000.
258. R. Noyori, I. Nishida, and J. Sakata, *Tetrahedron Lett.*, 1980, **21**, 2085–2088.
259. D. P. Cox, J. Terpinski, and W. Lawrynnowicz, *J. Org. Chem.*, 1984, **49**, 3216–3219.
260. R. K. Sharma and J. L. Fry, *J. Org. Chem.*, 1983, **48**, 2112–2114.
261. T. Furuta, T. Watanabe, S. Tanabe, J. Sakyo, and C. Matsuba, *Org. Lett.*, 2007, **9**, 4717–4720.
262. W. S. Zielinski and M. Sprinzl, *Nucleic Acids Res.*, 1984, **12**, 5025–5036.
263. W. M. Horspool, *Aspects of Organic Photochemistry*, Academic Press, London, 1976.
264. D. E. Metzler, *Biochemistry: The Chemical Reactions of Living Cells*, Harcourt/Academic Press, 2nd edn., 2001, vol. 1.
265. S. A. Robertson, J. A. Ellman, and P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 2722–2729.
266. S. Walbert, W. Pfeleiderer, and U. E. Steiner, *HCA*, 2001, **84**, 1601–1611.
267. A. Hasan, K.-P. Stengele, H. Giegrich, P. Cornwell, K. R. Isham, R. A. Sachleben, W. Pfeleiderer, and R. S. Foote, *Tetrahedron*, 1997, **53**, 4247–4264.
268. J. M. Davis, *Basic of Cell Culture: A Practical Approach*, Oxford University Press, 2nd edn., 2002.
269. K. R. Gee, L. Niu, K. Schaper, and G. P. Hess, *J. Org. Chem.*, 1995, **60**, 4260–4263.
270. K. Schaper, S. A. Madani Mobarekeh, P. Doro, and D. Maydt, *Photochem. Photobiol.*, 2010, **86**, 1247–1254.

APPENDIX

1. (A) Enzymatically released AngI-His₆ released after 4 h incubation with renin ($t_R = 15.36$ min). (B) The enzymatically released peptide co-migrating with the synthesised peptide. An aliquot of the digest solution was spiked with 5 μg of the synthesised peptide.



2. NOESY Spectrum for nucleoside **3c** (Chapter 3). Similar spectra were obtained for nucleosides **3a** and **3b**.



3. ^1H NMR spectrum for nucleoside 3 (NOVC-cyt, Chapter 4).

