Selective Modification and Detection of the DNA Bases

A thesis submitted to the Board of the Faculty of Physical Sciences

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

at the

University of Oxford

by

Emma Victoria Bristowe Wallace

Linacre College

and

The Chemistry Research Laboratory

Michaelmas Term 2011

For Mum and Dad

Table of Contents

Declaration

The work described in this thesis is entirely my own, except where I have either acknowledged help from a named person or given reference to a published source. Figures or text adapted from another source have the appropriate reference given.

Emma Victoria Bristowe Wallace

University of Oxford, Michaelmas Term 2011

Acknowledgements

I would like to thank Professors Hagan Bayley and Tim Donohoe for the opportunity to undertake doctoral studies in their research groups. Their guidance and constant enthusiasm for my project throughout these past four years have been invaluable. I now feel I have a greater appreciation of the challenges and rewards of studying both chemical biology and organic chemistry.

The CRL staff are also thanked, particularly Tim Claridge, James McCullagh and Lingzhi Gong for their invaluable advice and assistance in running both NMR and MS of DNA oligonucleotides.

I would like to thank all past and present members of both the Bayley and Donohoe groups for their help and friendship during my DPhil. I owe a great deal of thanks to David Stoddart for his assistance with the electrophysiology experiments. He always answered my questions no matter how busy he was. I would like to thank Drs Gráinne Hargaden, Giovanni Maglia and Andrew Heron for their assistance in all aspects of my project. A big thank you goes to Ellina Mikhailova who prepared all of the αHL proteins for this project. I would also like to thank my team of proof readers without whom writing this thesis would have been a struggle. Particular thanks must go to all those with whom I have shared lab F8. Their unusual choices in music and lively discussions have kept me amused and made me smile over the last four years.

A special thank you goes to Paul whose continued love and support, throughout my DPhil and especially during the writing of this thesis, have been invaluable.

Finally, I would like to thank my parents for their never ending love and support, without which I could not have undertaken these studies. They have always encouraged and helped me with everything I have tried to do and for this I will always be grateful.

Abstract

Selective Chemical Modification of the DNA Bases

Emma Victoria Bristowe Wallace, Linacre College, Doctor of Philosophy, MT 2011

α-Hemolysin (αHL) is a biological nanopore, which is currently under investigation for implementation into a new method for DNA sequencing. It has been established that αHL is capable of discriminating the canonical bases (adenine, cytosine, guanine and thymine) when they are immobilised within a pore by means of a biotin•streptavidin complex. Work in this thesis develops this procedure for the discrimination of these standard nucleobases from the the epigenetic modifications of cytosine – 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (hmC). Strategies to selectively modify and detect the modified bases are also explored.

Introduction

DNA sequencing strategies from the initial methods employed by Sanger up to current techniques utilised for the sequencing of entire human genomes, are reviewed. The initial discoveries of the epigenetic modifications 5-methylcytosine and 5-hydroxymethylcytosine are discussed. The proposed effects these bases have *in vivo* and a number of methods for their detection are also covered. Finally, methods for the selective chemical modification of the DNA bases are reviewed.

Results and Discussion

Initially, the biotin•strepatvidin immobilisation strategy is implemented for the discrimination of the epigenetically modified analogues of cytosine (5mC and hmC), without prior chemical modification. It is subsequently observed that an α-hemolysin mutant is capable of discriminating all six bases – adenine, cytosine, guanine, thymine, 5-methylcytosine and 5-hydroxymethylcytosine. A number of different chemical reactions are then investigated on the *bis*-TBS protected deoxynucleoside model system, for their ability to selectively modify one of the DNA bases. Two selective oxidation reactions are then further optimised for use on ssDNA. Finally, electrical recording experiments were used to investigate two selective chemical reactions, as well as the modification of the bases, by the cancer drug temozolomide.

Abbreviations

A - adenine or Ampere(s)

Å - Angstrom

Ac - acetyl

AD - asymmetric dihydroxylation

ADP - adenosine diphosphate

αHL - α-hemolysin

AMI - active 5-methylene intermediate

AMP - adenosine monophosphate

ATP - adenosine triphosphate

BIPY - 2,2'-bipyridine

Boc - *tert*-butoxycarbonyl

borax - sodium tetraborate

br. - broad

btn - biotin

bn - benzyl

bu - butyl

C - cytosine

°C - degrees Celcius

caC - 5-carboxylcytosine

CEI - covalent enamine intermediate

CpG - cytosine-phosphate-guanine

cm^{-1} - wavenumbers

CMS - cytosine-5-methylenesulfonate

COSY - correlation spectroscopy

C5-MTases - 5-methyltransferases

CNS - central nervous system

CSI - bicyclic sulfonium intermediate

d - day(s) or deoxy

Da - Daltons

dATP - deoxyadenosine triphosphate

dCTP - deoxycytidine triphosphate

ddATP - 2',3'-didexoyadenosine triphosphate

ddCTP - 2',3'-didexoycytidine triphosphate

ddGTP - 2',3'-didexoyguanosine triphosphate

ddNTP - 2',3'-didexoynucleoside triphosphate

ddTTP - 2',3'-dideoxythymidine triphosphate

decomp. - decomposition

DEPT - distortionless enhancement by polarization transfer

dGTP - deoxyguanosine triphosphate

DHQ - dihydroquinyl

DHQD - dihydroquinidyl

DPhPC - 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine

DNMT(s) - deoxyribonucleic acid methyltransferase enzyme(s)

dNTP - deoxynucleoside triphosphate

DNA - deoxyribonucleic acid

d.r. - diastereomeric ratio

dsDNA - double-stranded deoxyribonucleic acid

dTTP - deoxythymidine triphosphate

eb - ethidium bromide

E. Coli - *Escherichia coli*

EDTA - ethylenediaminetetraacetic acid

ESC - embryonic stem cells

ESI - electrospray ionisation

ePCR - emulsion polymerase chain reaction

eq. - equivalent(s)

Et - ethyl

fC - 5-formylcytosine

G - guanine

g - gram

h - hour(s)

HIF - hypoxia-inducible factor(s)

hmC - 5-hydroxymethylcytosine

HMBC - heteronuclear multiple-bond correlation

HMQC - heteronuclear multiple-quantum correlation

HPLC - high-performance liquid chromatography

HRMS - high resolution mass spectrometry

HSQC - heteronuclear single-quantum coherence

hν - ultraviolet light

Hz - Hertz(s)

I_B - blocked pore current value (pA)

ICON - interstrand complexes formed by osmium and nucleic acids

I_O - open pore current value (pA)

IR - infra-red

I_{RES} - residual current (%) = ($I_O/I_B \times 100$)

ΔI_{RES} - difference in residual current levels

$\Delta I_{RES}^{OVERALL}$ - the difference between the most widely dispersed residual current levels in a histogram of residual current blocks

IUPAC - International Union of Pure and Applied Chemistry

k - kilo

kb - kilobase

m - milli

MALDI - Matrix-assisted laser desorption/ionization

M - molar

Me - methyl

mL - millilitre(s)

µL - microlitre(s)

µm - micrometer(s)

min - minute(s)

mol - mole

m.p. - melting point

MS - mass spectrometry

MspA - *Mycobacterium smegmatis* porin A

5mC - 5-methylcytosine

mv - methyl viologen

m/z - mass-to-charge ratio

N - generic nucleobase (A, C, G, T, 5mC or hmC)

n - normal

NBS - *N*-bromosuccinimide

NIH - National Institute of Health

NMO - 4-methylmorpholine-*N*-oxide

NMR - nuclear magnetic resonance

NN - E111N/K147N

NNX - E111N/K147N/M113X

NNY - E111N/K147N/M113Y

NQ - 2-methyl-1,4-napthaquinone

ODN - oligodeoxyribonucleotide

^{32}P - radio-labelled phosphorus

p - pico

PAGE - polyacrylamide gel electrophoresis

PCR - polymerase chain reaction

PHAL - phthalazinyl

PHEN - 2,10-phenanthroline

ppm - parts per million

pH - $-\log_{10}[H_3O^+]$

Ph - phenyl

PPi - inorganic pyrophosphate

PTFE - polytetrafluoroethylene

R - DNA base

R_1 - recognition site 1

R_2 - recognition site 2

R_3 - recognition site 3

RNA - ribonucleic acid

RSM - recovered starting material

rt - room temperature

s - second(s)

SAM - *S*-adenosylmethionine

SD - standard deviation

SDS - sodium dodecyl sulfate

Selectfluor® - 1-chloromethyl-4-fluoro-1,4-diazoniabicyclo[2.2.2]octane bis(tetrafluoroborate)

SiO_2 - silicon dioxide

SOLiD - support oligonucleotide ligation detection

Sp - spiroiminodihydantoin

S_N2 - bimolecular nucleophilic substitution

ssRNA - single-stranded ribonucleic acid

ssDNA - single-stranded deoxyribonucleic acid

T - thymine

TBATB - tetrabutylammonium tribromide

TBCD - 2,4,4,6-tetrabromocyclohexa-2,5-dienone

TBS - *tert*-butyldimethylsilyl

TBSCl - *tert*-butyldimethylsilyl chloride

TEG - triethyleneglycol

TLC - thin-layer chromatography

TMEDA - *N*,*N*,*N*',*N*'-tetramethylethylenediamine

TOF - time of flight

Tris - 2-amino-2-hydroxymethyl-propane-1,3-diol

tSMS™ - true single-molecule sequencing

U - uracil

UMS - uridine-5-methylenesulphonate

ν_{max} - infra-red absorption

V - volt(s)

ve - applied voltage

WT - wild-type

Chapter 1

Introduction

Chapter 1 – Introduction

1.1 – DNA Sequencing

1.1.1 – Introduction to DNA Sequencing

Deoxyribonucleic acid (DNA) contains the genetic information that allows all living organisms to function, grow and reproduce. To better understand how DNA has affected human evolution, the causation of disease and the interplay between heredity and the environment in defining the human condition, the determination of its sequence is necessary. Two methods were initially developed for DNA sequencing; 1) the chemical degradation method of Maxam and Gilbert[1] and 2) the dideoxy chain termination method of Sanger.[2]

In 1990 the Human Genome Project was launched. The project was a 15-year, $3 billion plan for completing the sequence of the three billion base pairs that makes up the human genome. By employing the Sanger method and either whole-genome shotgun sequencing approach[3, 4] or hierarchical shotgun sequencing[5] the initial sequencing and analysis of the first human genomes were completed in 2001.[4, 5] Since then work has continued to try and reduce the time and cost of sequencing, so that it will be possible to perform personal genome sequencing. In 2004, in an effort to encourage further research in this area, the National Institute of Health (NIH) set a challenge to sequence a complete human genome for $1000 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html). A second prize of $10 million known as the 'Archon X-prize for genomics', which was first launched in 2006, will go to the first team to build a device capable of sequencing 100 human genomes, to a high degree of accuracy, within 10 days, for less than $1000/genome – http://genomics.xprize.org/genomics/archon-x-prize-for-genomics. The current cost of sequencing an entire genome is $4400, provided by Complete Genomics in California,[6] which

is a considerable price reduction compared to the >$100 million that was necessary for the first human genomes.[4, 5] The cost of sequencing is continuing to fall and it is proposed that in time, personal genome sequencing could be carried out routinely, facilitating disease prevention, improving diagnosis and guiding better treatment for that individual.[7]

1.1.2 – Sanger Sequencing

In 1977, Sanger reported a new method for DNA sequencing that was used to sequence the DNA of the bacteriophage φX174.[2] This method relied on the fact that when 2',3'-dideoxythymine triphosphate (ddTTP) was incorporated into a growing oligonucleotide chain, in place of deoxythymine triphosphate (dTTP), it prevented further chain extension by DNA polymerase I. At positions where ddTTP had been incorporated the chain was terminated, as it could not be further extended, because the necessary 3'-hydroxyl group required for enzymatic chain extension was not present. This resulted in chain termination specifically at positions where ddTTP had been incorporated. If a mixture of ddTTP and dTTP were incubated with a primer, template, DNA polymerase I and the other three deoxynucleoside triphosphates (dNTP) (one of which was radio-labelled with ^{32}P), then a mixture of DNA fragments resulted. These fragments were of varying lengths, had identical 5' ends and were all terminated at the 3' end with a ddTTP. If the resultant mixture was then fractionated by electrophoresis on a denaturing polyacrylamide gel, and the gel exposed to film, then the subsequent radiograph produced a pattern of bands that indicated the location of the ddTTP's in the newly synthesised strand. If this process was repeated using ddATP, ddGTP and ddCTP, respectively, and all four sample mixtures were run on a single gel in parallel, then the pattern of bands could be used to manually read the DNA-template sequence (Figure 1 – 1).

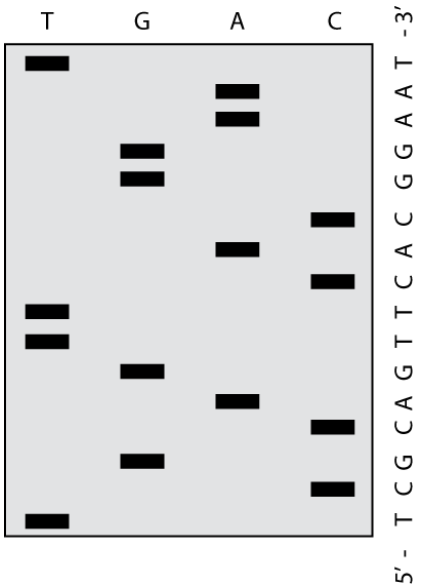


Figure 1 – 1 – *Schematic representation of a polyacrylamide gel produced by Sanger sequencing. In four separate reactions the template DNA, primers, DNA polymerase I, radio-labelled dNTP's and one type of dideoxynucleoside triphosphate (ddNTP) (which promoted chain termination) were incubated together. The enzymatic replication of the DNA template produced a mixture of radio-labelled fragments, which all terminated in a ddNTP. When the fragments were purified by polyacrylamide gel electrophoresis (PAGE), and the gel exposed to film, the radiography produced a pattern of bands from which the DNA sequence could be determined.*

Although this method was a great improvement over the techniques available at the time, it had a number of disadvantages: 1) it employed radioactively labelled dNTP's which were unstable and hazardous, 2) four separate reactions were necessary to determine the DNA sequence, 3) there was a delay in obtaining the sequencing data in order to allow for exposure and development of the autoradiographic patterns and 4) the sequencing data that was obtained required skilled interpretation and transcription. To combat these issues a new method, that incorporated fluorescently-tagged oligonucleotide primers, was developed by Smith *et al*.[8] Smith used Sanger's enzymatic method of DNA sequencing but, instead of using radioisotopes for detection, four different fluorescent DNA primers were used in each of the four sequencing reactions. For example when the ddNTP employed in the sequencing reaction was ddTTP then a fluorescent primer conjugated to a green dye was used, to initiate chain extension by DNA polymerase, and when ddGTP was utilised a red fluorescently labelled primer was used etc. After the four different

sequencing reactions were performed, they were combined and run on a single polyacrylamide gel. The identity of the ddNTP, used to terminate DNA chain extension by DNA polymerase, was identified by a different coloured band. Another significant improvement provided by this method was that the fluorescence data could be continuously acquired and stored. This eliminated the need to transfer information from gel to film and the tedious manual analysis of the previous Sanger procedure. This approach was so successful that it was further developed by Applied Biosystems, in 1987, to give the first automated DNA sequencer (Figure 1 – 2). Subsequently, the fluorescently-labelled primers were replaced with ddNTPs that were themselves fluorescently-tagged. This meant that only one reaction, that directly incorporated the fluorescently-labelled ddNTPs, was necessary to sequence a strand of DNA.[9]

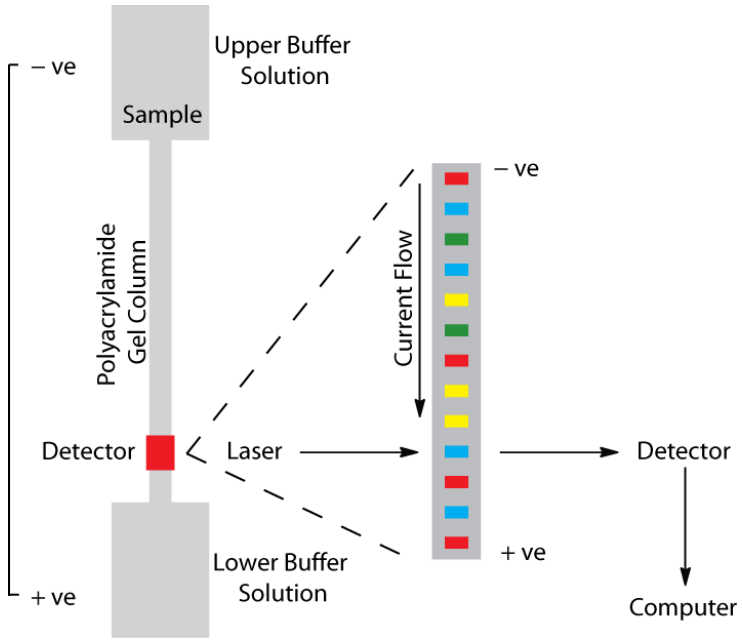


Figure 1 – 2 – *Simplified diagram, showing only one lane, of the Sanger automated DNA sequencer for detecting fluorescently-tagged DNA. The fluorescently-tagged DNA fragments were separated as they flowed through the polyacrylamide gel. They passed through a laser which excited the fluorophore and the fluorescence signal was detected automatically, in real-time, by the computer.*

DNA sequencing was significantly improved by modification of the original Sanger sequencing method. However, if this method were to be successfully implemented in the

sequencing of the human genome, it would be necessary to drastically increase the through-put of the technique. This was realised by the development of capillary array electrophoresis.[10] A low-viscosity separation matrix can be pumped through an array of capillaries, eliminating the need for PAGE. These systems provide a read length of 300-500 bases per capillary.[10, 11]

1.1.3 – Next Generation DNA Sequencing

Should the challenge of sequencing the human genome for less than $1000 be achieved, it would be necessary to develop high-throughput approaches that could process large quantities of data, more quickly and at a reduced cost. Since 2005, a number of next generation sequencing methods have been commercialised.[12-14] Several of these techniques use the emulsion polymerase chain reaction (ePCR),[15, 16] instead of bacterial cloning and traditional PCR, to amplify the DNA templates required for sequencing. The genomic DNA, to be sequenced, is prepared by fragmentation into ~1 kb sections which are then circularised with a universal linker. The circularised DNA is then fragmented in the centre of the universal linker, using restriction endonucleases, producing fragments of DNA (template) that contain complementary ends to the two distinct amplification primers.[12] Biotinylated complementary primers are then pre-incubated with streptavidin coated beads, creating beads which are coated in common primers (Figure 1 – 3). These beads are then mixed with the fragmented genomic DNA (containing the universal sequences) in a water/oil emulsion that contains all the necessary components for PCR (including the DNA amplification primer). This results in the formation of aqueous compartments in oil that each contain less than one primer-coated bead and one DNA template. This micro-emulsion is then subjected to a standard PCR protocol. If both a DNA template and a primer coated bead are present in a single aqueous compartment then the bead-bound oligonucleotide primers initiate replication of the DNA template. After the first PCR cycle,

the amplification primer, present in the emulsion can be used to replicate the newly synthesised template DNA resulting in an exponential increase in the number of DNA templates produced. The ePCR technique ensures that the products of PCR remain separate, producing multiple copies of the same fragment tethered to one bead. The resultant double-stranded DNA (dsDNA) is then denatured and the single-stranded DNA (ssDNA) purified ready for sequencing (Figure 1 – 3).[15]

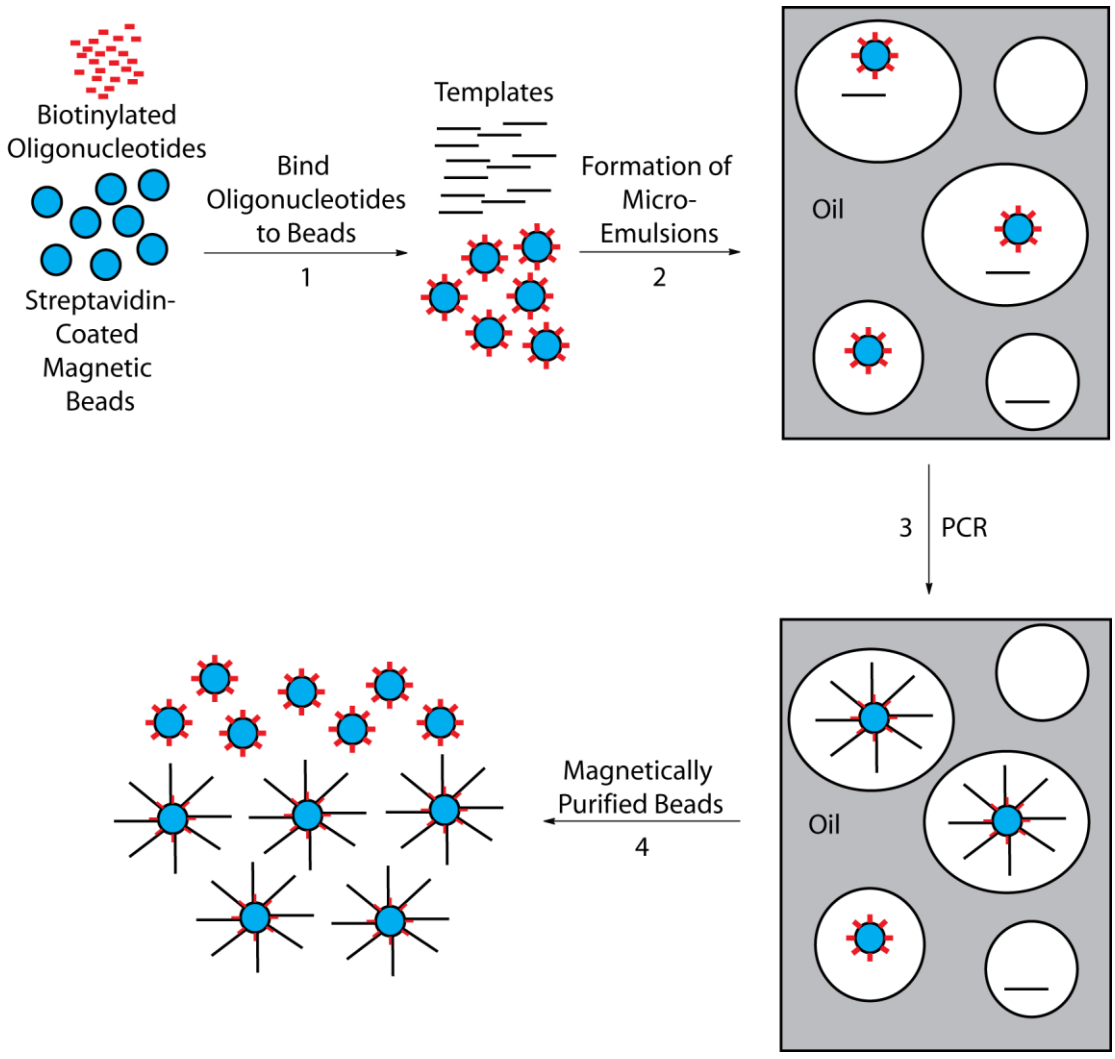


Figure 1 – 3 *– Schematic representation of the process of ePCR onto streptavidin-coated magnetic beads. 1) The biotinylated oligonucleotide amplification primers (red line) are bound to the magnetic streptavidin-coated beads (blue circles). 2) An aqueous mixture containing all the necessary components for PCR, plus the primer-bound beads and template DNA (which has the appropriate complementary amplification primers attached at either end) are stirred together in an oil/aqueous phase mixture to create a micro-emulsion. The aqueous droplets (white circles) contain on average one template DNA strand and less than one bead. 3) The microemulsion mixture undergoes standard PCR treatment, where the bead bound oligonucleotides act as primers for replication of the*

DNA template. 4) Finally, the dsDNA produced by ePCR is denatured and the ssDNA purified, in this example by using a magnet. Figure adapted from Dressman et al. 2003.[15]

1.1.3.1 – Sequencing by Ligation

Work carried out by Shendure *et al.* pioneered the use of a next generation sequencing method that employs four-colour sequencing by ligation.[12] ePCR beads containing a single DNA fragment are produced and then immobilised in a polyacrylamide gel in a disordered monolayer. First the appropriate 'anchor primer' is annealed to its complementary site on the DNA strand that is being sequenced. Four different fluorescently-labelled probes are then added to the solution. These probes have degenerate sequences except at the query point (Figure 1 – 4). An enzymatic ligation reaction is then performed to join the corresponding dye-labelled probe to the strand. Any unligated probes are then washed away. The incorporated probe is detected by fluorescent imaging and the base at that position determined. The anchor primer:fluorescent-probe complexes are then stripped from the target DNA and the cycle repeated. This procedure is designed to sequentially test 26 positions on the DNA fragment and report the identity of each base with its corresponding fluorophore. This sequencing method was used to determine the entire *Escherichia coli* (*E. coli*) genome, in two 60 hour instrument runs with 71% high-quality coverage.[12] The cost per base was $0.00011, roughly one-ninth of the cost of conventional Sanger sequencing. Life/APG has since commercialised this technique and named it support oligonucleotide ligation detection (SOLiD).[14]

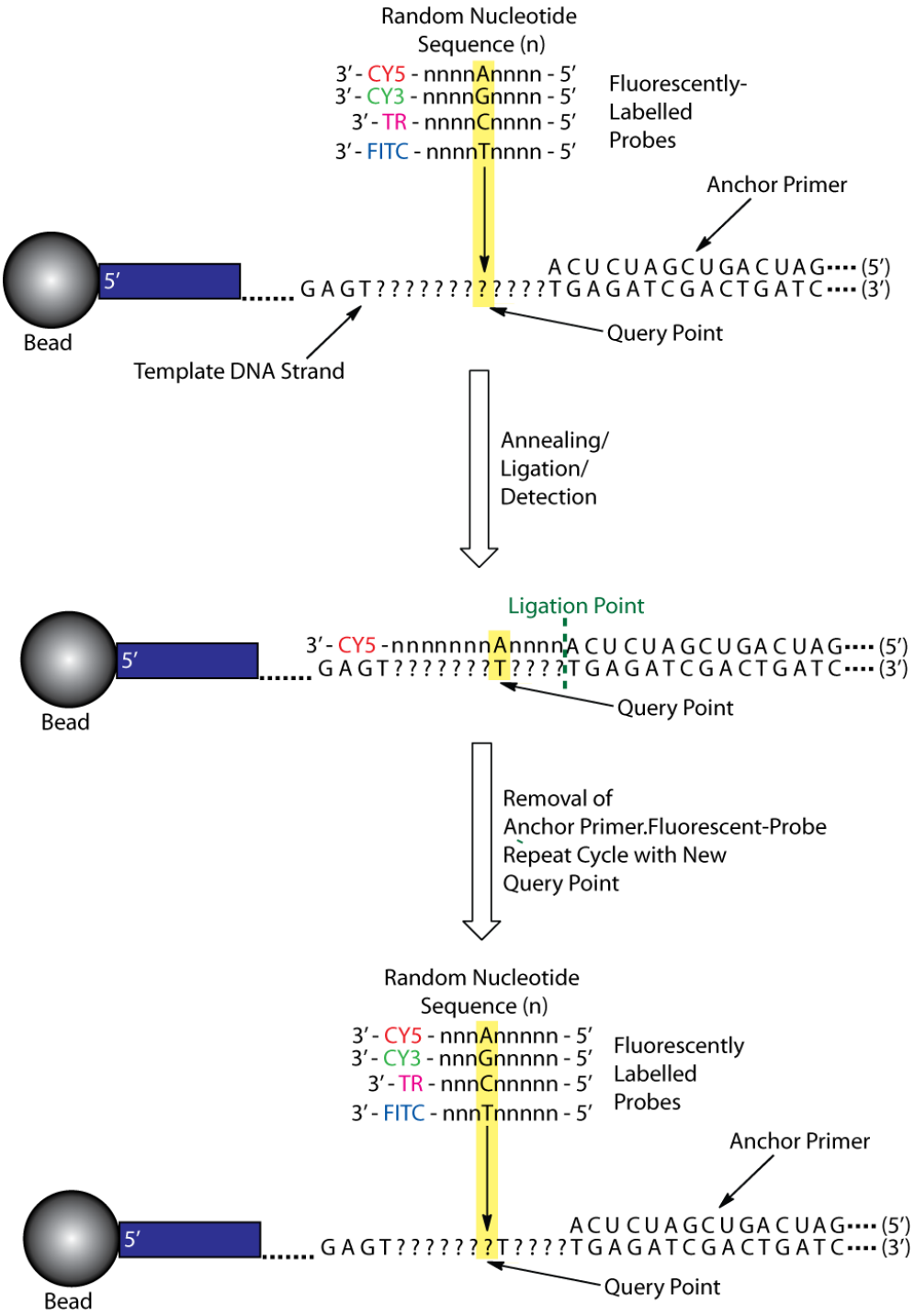


Figure 1 – 4 – *Sequencing by ligation. ePCR produces beads which contain multiple copies of a single DNA template strand (only one copy is shown in the figure above for illustration purposes), which are immobilised in a polyacrylamide gel. Anchor primers are annealed to the DNA template and the ligation reaction performed in the presence of four fluorescently-labelled probe strands. The probe strands are identical in sequence apart from at the query point. This means only the fluorescently-labelled probe containing the complementary base to the query point is incorporated. The unbound probes are then washed away and the incorporated base detected by fluorescent imaging. The anchor primer:fluorescent-probe complexes are then stripped from the target DNA and the cycle repeated. Figure adapted from Shendure et al. 2005.*[12]

1.1.3.2 – Sequencing by Synthesis

This method of sequencing employs the use of a DNA polymerase to synthesise a complementary strand of DNA, whereby each base is identified as it is incorporated into the growing DNA oligonucleotide. Two different methods, both of which have been commercialised, will be discussed further – cyclic reversible termination (commercial Genome Analyzer, Solexa, subsequently acquired by Illumina)[14] and pyrosequencing (454 Life Sciences Corp., subsequently acquired by Roche).[13]

Cyclic reversible termination does not use ePCR to amplify DNA, it instead employs solid-phase amplification methods. The solid surface is covered in primer adaptors that complement the forward and reverse primers of the DNA template. The DNA template is hybridised to the primers on the solid surface and the strand replicated by DNA polymerase. The original template strand is then denatured from the newly replicated strand and washed away. This produces a newly replicated strand, which has forward and reverse primers at either end, and is covalently attached to the solid surface (Figure 1 – 5, step 2).[14] The previously unbound primer, at the 3' end of the ssDNA template, then forms a bridge with another complementary primer immobilised on the surface. The primer acts as an initiator for the DNA polymerase enzyme which synthesises a complementary strand resulting in dsDNA bound to the surface at either end. If the DNA duplex is denatured it produces two complementary ssDNA fragments. This cycle is repeated many times to amplify the DNA template on the solid support. After the generation of the amplified DNA, sequencing is initiated by the addition of primers and the four dNTPs, which are each labelled with a different chain-terminating fluorescent group. A single complementary nucleotide is incorporated into the strand adjacent to the primer and the unbound nucleotides washed away (Figure 1 – 5, step 9). Fluorescent imaging is performed to identify the base and then the chain terminating group and fluorophore are

removed. It is necessary to wash the solid surface again before the incorporation of the next nucleotide.[14] Due to the prior amplification of the DNA sequence, the fluorescent signal detected is a summation of fluorescent signals from each base that is incorporated into each bound strand.[14]

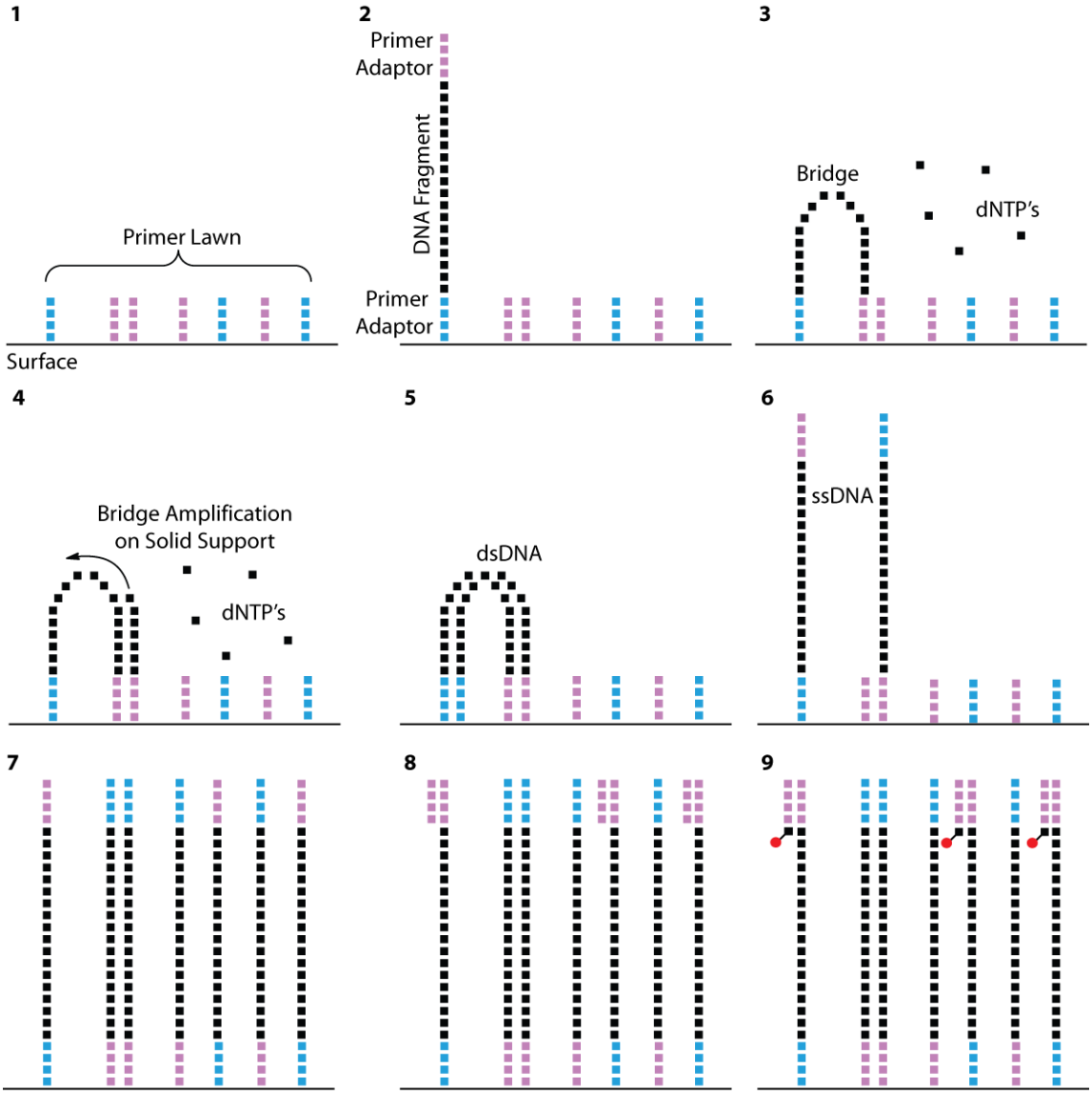


Figure 1 – 5 – *Sequencing by synthesis – cyclic reversible termination. 1) Forward and reverse primers are attached to a solid surface. 2) The DNA template is hybridised to the primer adaptor on the surface, and the template replicated by DNA polymerase. The resultant dsDNA is denatured and the original template, which is not covalently attached to the surface, is washed away. This produces a DNA fragment covalently attached to the solid surface with amplification primers attached to either end. 3) The unbound primer at the 3'-end of the DNA fragment (in this case purple) binds to its complementary primer on the solid surface forming a bridge. 4) The complementary primer acts as an initiator for DNA polymerase and it starts to synthesize a complementary strand. 5) Synthesis of the complementary strand is completed with formation of dsDNA. 6) Denaturation of the dsDNA duplex gives ssDNA. 7) Steps 3-7 are repeated until amplification of the DNA*

fragment has occurred. 8) Sequencing primers are added which anneal to the primer adaptor. 9) Four fluorescently labelled dNTP's each of which contains a different chain terminating fluorescent group are added onto the solid surface (red circle). DNA polymerase incorporates the complementary dNTP and the rest of the fluorescently labelled dNTP's are washed away. Fluorescent imaging is then performed to identify the base, the chain terminating group and fluorophore are removed and the cycle repeated. Figure adapted from www.illumina.com.

Another method for sequencing by synthesis is pyrosequencing. This procedure uses ePCR to generate amplified DNA sequences that are then immobilised onto beads.[13] These beads are positioned into individual wells on a fibre-optic slide which is mounted in a flow chamber. One dNTP is allowed to flow over the slide at a time. If that base is complementary to the next position in the sequence, then DNA polymerase incorporates it into the growing complementary strand. This results in the release of inorganic pyrophosphate (PPi). The pyrophosphate is detected as a photon release, owing to the action of the enzymes adenosine triphosphate (ATP) sulfurylase and luciferase (Figure 1 − 6). The first of these enzymes converts pyrophosphate to ATP which is subsequently converted by luciferase to adenosine monophosphate (AMP) and visible light. As each bead contains multiple copies of the same DNA sequence, many photons are generated per correct base incorporation. In the case where there are homopolymeric repeats the amount of light detected is proportional to the number of nucleotides incorporated. However, this is only the case in the incorporation of up to six nucleotides. This method was used to determine the sequence of the *Mycoplasma genitalium* genome with 96% coverage and 99.96% accuracy.[13] This technique would be ideal for high through-put sequencing as one slide has approximately 1.6 million wells, each of which holds one bead with multiple copies of the same DNA fragment attached to it.

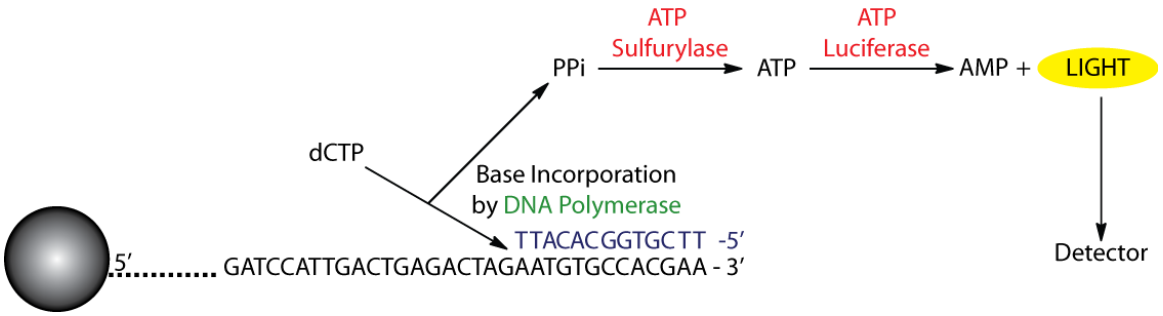


Figure 1 – 6 – *Sequencing by synthesis – pyrosequencing. ePCR is used to generate multiple copies of a fragment of DNA immobilised on a bead. The bead is placed in a well, on a slide, mounted in a flow chamber. One dNTP flows across the cell at a time and if that nucleotide is the complementary base for the next position in the growing strand (blue), then it is incorporated by DNA polymerase. This produces pyrophosphate which is converted to AMP plus light by the action of ATP sulfurylase and ATP luciferase. The light intensity is measured by the detector. Figure adapted from Bayley 2006.*[17]

In 2010 Ion Torrent Systems Inc. (who have recently been acquired by Life Technologies) launched a DNA sequencer based on ion detection. It works on a similar principle to pyrosequencing but instead of detecting light it measures the release of protons. As each dNTP is incorporated into the complementary DNA strand, by DNA polymerase, a single proton is released from the 3'-hydroxyl group of the growing strand, as well as a molecule of pyrophosphate. The released proton changes the pH of the solution, which is detected by an ion sensor. This method has a read length of approximately 100 base pairs and has been termed 'a pH meter that sequences'.[18]

1.1.4 – Single Molecule Next Generation Sequencing

Single molecule sequencing is defined as a technique capable of sequencing one single DNA molecule individually. There are some distinct advantages of this sequencing method over the techniques which have been previously described. Firstly, it requires much less genomic DNA material (<1 µg in comparison to 3-20 µg for methods that necessitate DNA amplification).[14] The amount of DNA required is not an issue if it comes from an abundant source (e.g. blood sample or a buccal swab), however, if the amount of DNA is limited, such as in applications in forensic science, techniques which require less DNA

provide a distinct advantage. As less DNA is required these techniques generally do not require amplification by PCR. This means that by using single molecule methods sequence biasing and mutation insertion, which can occur as a result of amplification, is reduced. In data processing of sequencing techniques which require clonal amplification, the observed signal is a consensus of all the nucleotides or probes being added to identical copies of the same strand of DNA for a given cycle. This means that if either multiple nucleotides are incorporated or if one nucleotide is not incorporated at all during one complete cycle, then the growing primers move out of synchronicity with each other, which is known as signal dephasing. This results in an increase in fluorescent signal noise which causes base-calling errors and ultimately limits the maximum-read length of DNA that can be sequenced by a given method.[14] By using single molecule techniques one is monitoring only the incorporation of a single base onto one individual strand of DNA and, therefore, signal dephasing is not an issue.

1.1.4.1 – True Single-Molecule Sequencing (tSMS™)

Helicos Biosciences commercialised a true single-molecule sequencing method, which is based upon work by S. Quake and co-workers.[19] In this method a short poly(dA) sequence, containing a single fluorescently labelled adenine, is attached to one end of the DNA template. The poly(dA) tail pairs up with a corresponding poly(dT) oligonucleotide which is immobilised onto a glass cover slip. The immobilised oligonucleotides are then imaged, in order to determine their location on the glass clover slip. The fluorescently labelled adenine is then cleaved and washed away. A fluorescently labelled dNTP is then attached to the end of the poly(dT) primer, by a DNA polymerase enzyme, in order to synthesise the complementary strand (Figure 1 – 7). The fluorescent group is subsequently imaged, by total internal reflection fluorescence, cleaved and washed away before the cycle is repeated. Four different dNTPs, each labelled with the same fluorescent group and

dispensed individually in a predetermined order, are used to generate the complementary DNA strand with a read length of 25 bases (www.helicosbio.com). Using this method, each immobilised strand is sequenced individually, by detection of fluorescence as each base is incorporated into the complementary strand. This sequencing technology was used to sequence an entire human genome in 2009.[19]

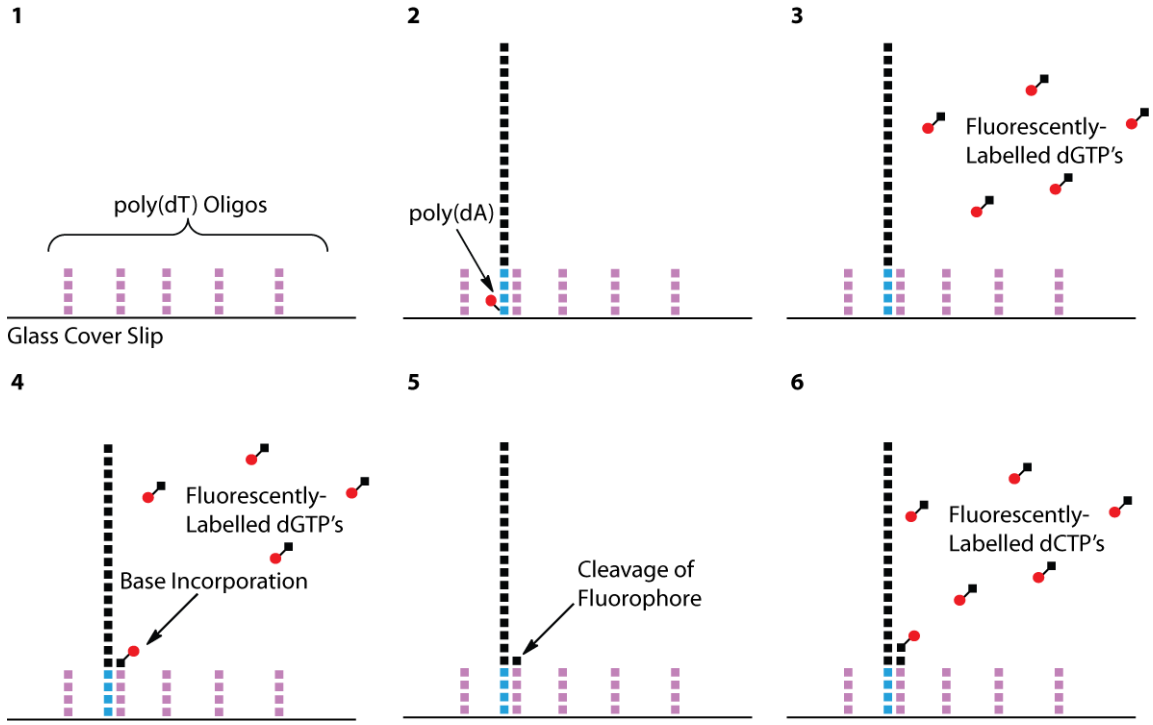


Figure 1 – 7 *– True single-molecule sequencing. 1) Short strands of poly(dT) oligonucleotides are immobilised on a glass cover slip. 2) The poly(dA) strand (containing a single fluorescently labelled base) which has been attached to the end of the DNA template is hybridised to the poly(dT), immobilising the DNA template on the cover slip. The location of the strand is then determined by imaging of the fluorescently labelled adenine base, before the fluorophore is cleaved and washed away. 3) One of the four different dNTPs (each of which is labelled with the same fluorescent group) is added to the surrounding solution (in this case dGTP). 4) DNA polymerase incorporates the fluorescently labelled base into the growing complementary strand and any excess un-incorporated fluorescently labelled base is washed away. The fluorescent group is then imaged by total internal reflection fluorescence. 5) The fluorescent tag is then cleaved and washed away. 6) The cycle begins again by the incorporation of another fluorescently labelled dNTP (in this instance dCTP).*

1.1.4.2 – Single-Molecule Real-Time Sequencing

Initial work by Webb and co-workers has been further developed by Pacific Biosciences for the purpose of single molecule sequencing in zeptolitre observation volumes.[20, 21] The

ability to perform PCR, for the incorporation of fluorescently-tagged nucleotide analogues, and then measure the fluorescent signal in real-time has proven to be challenging in the past. This is owing to the fact that the enzyme requires a micromolar concentration of fluorescently-tagged nucleotide analogue, for efficient incorporation into the growing DNA strand, and the fluorescent detection system requires a concentration in the pico- to nanomolar range, so that it can accurately measure the incorporation of one single fluorophore (and it not be obscured by background signal). This new method immobilizes a DNA polymerase enzyme at the bottom of a nanostructure, called a zero-mode waveguide, which provides a zeptolitre-effective observation volume. Under these conditions, the concentration of the fluorescently tagged nucleotide analogue is high enough for the enzyme to work effectively and the dimensions of the waveguide are so small, that the number of observable diffusing fluorescently-labelled analogues is reduced. Under these conditions, it is possible to detect the signal for the incorporation of a single fluorescently-tagged nucleotide analogue from the background.

In this single-molecule, real-time DNA sequencing method the zero-mode waveguides consist of small holes in a metal film which is deposited on a microscope cover slip. Millions of these holes can be made on one cover slip which makes this technique easily adaptable for massive parallelism. DNA polymerase is adsorbed onto the bottom of the waveguides, in the presence of a solution that contains all four fluorescently-tagged nucleotide analogues and the DNA template. The polymerase begins to synthesize the complementary strand to the DNA template and as each fluorescently-tagged nucleotide analogue is incorporated the fluorophore is excited by a laser and a fluorescent pulse detected. The duration of the pulse is governed by the catalytic rate and ends when the fluorophore-linker-pyrophosphate group is released and diffuses out of the well. Translocation of the DNA template then prepares the enzyme active site for incorporation

of the next fluorescently-tagged nucleotide analogue and the beginning of the next pulse (Figure 1 – 8). To test this sequencing technique a 158-base linear template was sequenced using a four-colour sequencing setup. Of the 158 bases, 131 were correctly identified by the automated base caller.[21] This method utilises the processivity and high reaction rates of DNA polymerases, which improve read lengths and throughput. It also eliminates the need for complex and time-consuming flow cycles to remove chain-terminating fluorescent tags from the surrounding solution. However, the enzyme is constantly being irradiated by the laser, which is used to excite the fluorophores on the nucleotide analogues, and is susceptible to photo damage which reduces the maximum read-length.

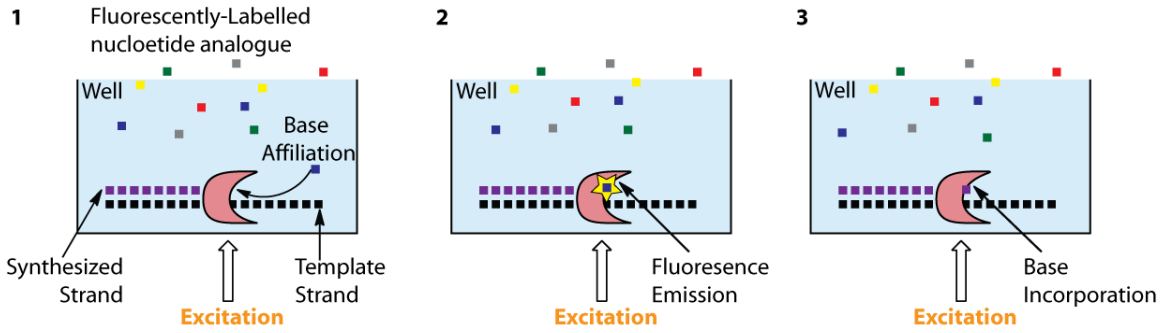


***Figure 1 – 8** – Single-molecule real-time sequencing. 1) A DNA polymerase (pink arc) is immobilised at the base of a zeptolitre well which is constantly being excited by a laser. It binds to the ssDNA template, in a solution which contains four fluorescently tagged nucleotide analogues each with a different fluorescent label. The fluorescently labelled nucleotide analogue, which corresponds to the next base to be incorporated into the synthesised strand, associates with the DNA template by Watson-Crick base pairing. 2) As the corresponding fluorescently tagged nucleotide analogue associates with the polymerase it is excited by the laser and detected. 3) The incorporation of the base, and subsequent phosphodiester bond formation, liberates the fluorophore, therefore, allowing it to diffuse from the well. The cycle can then be repeated.*

1.1.5 – Nanopore Sequencing

1.1.5.1 – Introduction to Nanopore Sequencing

Nanopore sequencing is another single molecule sequencing technique which has undergone heavy exploration over the last fifteen years.[22, 23] It is based upon the principle that an ionic current, which flows through a nanometer-scale pore, can be partially blocked

when the pore is occupied by a target molecule. In order to gain information about the concentration and identity of a particular analyte (e.g. ssDNA), the amount of current that flows through the chamber when that analyte is present, the duration for which the analyte is held in the pore, the frequency of the block events observed and the current pattern it produces when residing in the pore can be monitored. Kasianowicz *et al.* showed that it was possible for DNA to be electrophoretically driven through a biological nanopore (α-hemolysin, αHL) and observe current blocks which were dependent upon the length of the ssDNA.[24] Subsequently, it was proposed that a single nucleobase passing through a recognition element, within a nanopore, could modulate the current in such a manner that a distinctive current reading for each base cytosine, thymine, adenine and guanine could be determined.[24] This illustrated the possibility that electrical recording experiments could be utilised to determine the sequence of ssDNA (Figure 1 – 9).

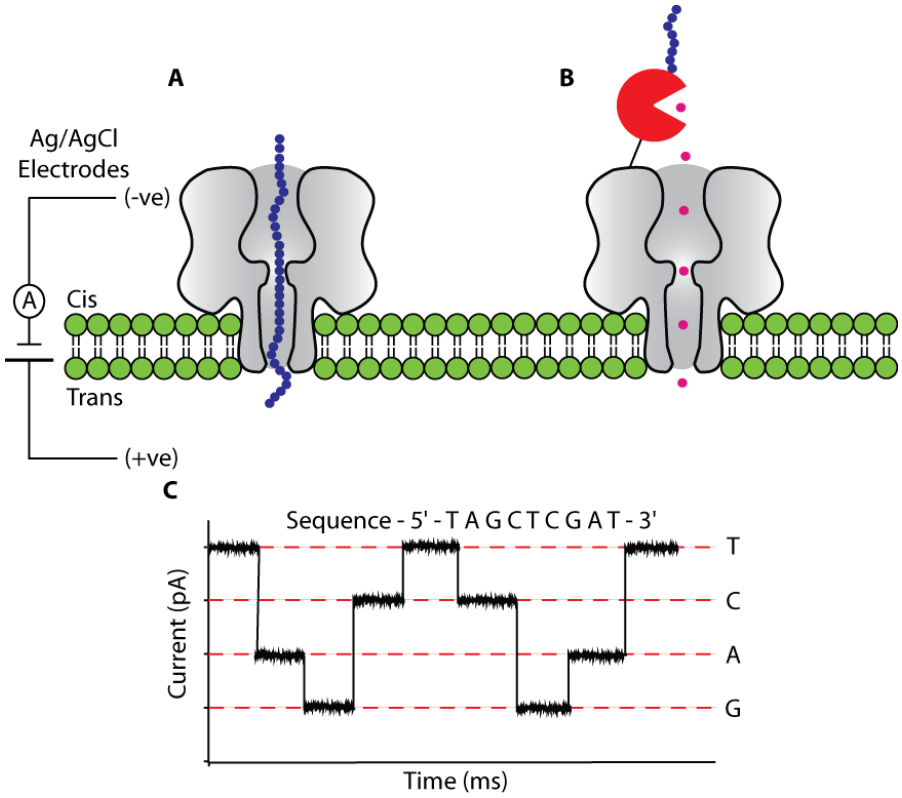


Figure 1 – 9 *– Two possible methods for employing nanopores to sequence DNA A) direct strand sequencing and B) sequencing modulated by a covalently attached exonuclease. A) In direct strand sequencing the ssDNA would be driven through the pore, under an applied potential, and as each individual base (blue circle) passed through the recognition*

site the current would be modulated to give a direct read-out of the DNA sequence. B) In exonuclease sequencing, the enzyme (red shape) would processively break down DNA into its component nucleoside monophosphates (pink circle), each if which would be identified as it was released and translocated through the pore. C) An ideal electrical recording read-out from which it would be possible to identify each base in the sequence from its characteristic block level.

There are several major advantages of nanopore sequencing which make it a worthy technique for further investigation.[25] Due to the concentration-limited rate at which nanopores capture DNA from the surrounding volume,[26] it was estimated by Branton *et al.* that nanopore sequencing will require ~10^8 copies of the target genome to provide adequate sequencing throughput, from the 25-50 μL volume of material necessary to feed an array of nanopores.[25] This can be directly obtained, using commercially available kits, without the need for amplification. This will mean the errors normally associated with PCR amplification will not be a problem for this sequencing method. It would also allow for the direct detection of epigenetic modifications, such as 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (hmC) (See Chapter 1, section 1.2), which would normally be lost during amplification or by using sequencing by synthesis methods. Once the necessary quantity of DNA has been obtained, in an ideal nanopore sequencing system, the DNA itself would require minimal sample processing. This would eliminate the need for expensive fluorescently-labelled nucleotides, DNA polymerases and ligases that are employed in other sequencing procedures.

It has been possible to produce both solid state[27] and biological[28-31] nanopore arrays on a disposable detector chip. One such technique, used to form lipid bilayers across apertures on a glass chip, employs giant unilamellar vesicles (GUVs).[28, 29] By the application of negative pressure (~10-40 mbar) the GUVs are forced onto the surface of the glass chip, where they burst and form bilayers, into which biological nanopores can be subsequently inserted e.g. αHL[28] or OmpF.[29] Work carried out by Fertig and co-workers, obtained a 60% functional insertion of OmpF into bilayer arrays using this approach.[29] There have

also been various techniques which have investigated the use of both biological and solid-state nanopores in one system, by either forming lipid bilayers across solid-state nanopores[32, 33] or more recently by inserting biological nanopores directly into solid state nanopores (see Chapter 1, section 1.1.5.3).[34] Therefore, if a practicable DNA sequencing method could be achieved, then sequencing could be carried out by many pores in parallel. This would make nanopore sequencing a very high-throughput technique, for which the major costs would be preparation of adequate quantities of DNA (~700 μg of human diploid genomic material for an entire human genome),[25] the cost of running the instrument itself and a single disposable chip.

One of the major difficulties of current sequencing methods is the maximum read-length that these techniques can process (approximately several hundred bases at a time). This requires segmentation of the entire human genome into small fragments which are sequenced and pieced back together in the correct alignment. For direct strand sequencing using nanopores, it should be possible to fragment a mammalian genome into sections of ssDNA consisting of 50,000 nucleotides.[25, 35] Meller and co-workers have demonstrated that by employing a salt gradient, it has been possible to observe translocation of dsDNA (consisting of 48, 000 base pairs) through solid-state nanopores.[35] The extremely long read lengths that may be possible with nanopore methods would drastically simplify the genome assembly process. All of these factors combine to make nanopore sequencing a much cheaper method which has the potential to meet the challenge of the $1000 per mammalian genome target.[25]

There are two main varieties of nanopore used in this sequencing approach – biological nanopores and solid-state nanopores (see Chapter 1, section 1.1.5.3). Currently, the two main biological pores which are under investigation for implementation in DNA sequencing are α-hemolysin (αHL)[36] and *Mycobacterium smegmatis* porin A (MspA).[37]

1.1.5.2 – Nanopore Sequencing with Biological Nanopores

α-Hemolysin is a transmembrane pore found in *Staphylococcus aureus*. It is a self-assembling monomer toxin (33.2 kDa) which binds to the plasma membranes of susceptible cells, where it oligomerises into its heptameric form (232.4 kDa), producing a water-filled transmembrane channel.[36] It is a mushroom-shaped pore that is 100 Å in height and 100 Å in diameter, consisting of cap and stem domains (Figure 1 – 10). The pore has a solvent-filled channel running along its centre, which narrows at the constriction point (between the cap and the stem domains) to a diameter of only 14 Å (1.4 nm).[36] The narrow dimensions of the constriction site only allow translocation of ssDNA as dsDNA is too wide (2.2 nm).[38]

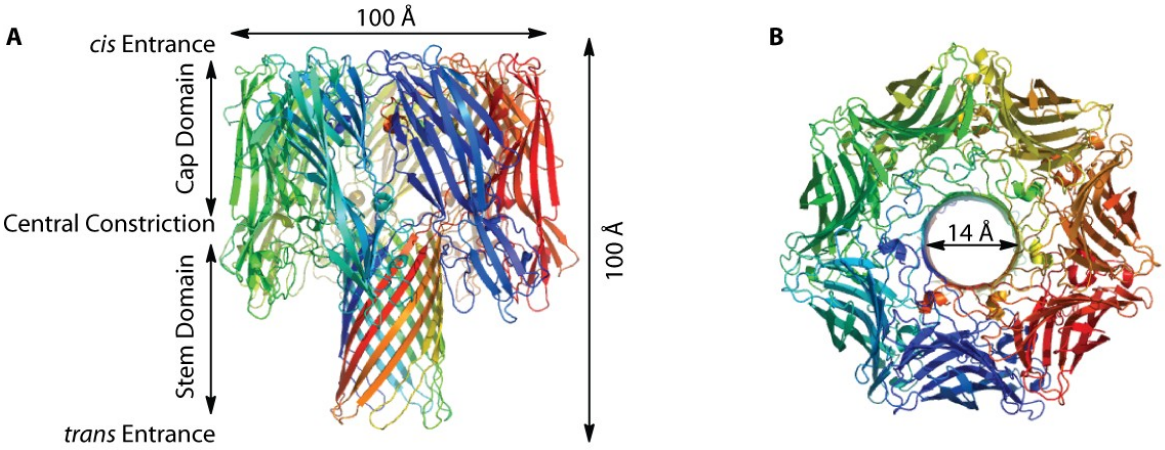


Figure 1 – 10 – *Cartoon representations of the α-hemolysin heptamer – made up of a cap and a stem domain. A) Side view of the αHL WT pore (created in Pymol from the pdb file 7AHL). The wild-type channel is known to exhibit partial rectification and modest anion selectivity. The cap domain, which protrudes from the membrane by ~65 Å, is composed of seven β sandwiches (each highlighted in a different colour) and the amino latch of each monomer unit. The stem domain is composed of 14 antiparallel β-strands, two of which are contributed by each monomer, that form a right-handed barrel. Between these two domains lies the constriction site – the narrowest part of the channel with a diameter of 14 Å.*[36] *The αHL cap domain protrudes out of the cis side of the bilayer (see Chapter 7 for the definition of cis and trans) and stem domain forms the transmembrane channel which opens out into the trans side.*[36, 39] *B) View from the cis side looking down through the αHL WT pore (created in Pymol from the pdb file 7AHL).*

αHL has shown great promise for DNA sequencing applications.[22] The pore can be chemically engineered using molecular biology techniques such as site-directed

mutagenesis, making it possible to produce well defined local changes in its structure. Several studies have made drastically mutated forms of the pore which were still hemolytically active (the mutant pores will insert into the membranes of red blood cells, causing cell lysis) and inserted into artificial bilayers forming stable channels, illustrating that mutations do not impair pore functionality.[39] This has allowed αHL to be specifically engineered for DNA sequencing methods e.g. alteration of the charge distribution inside the pore to aid DNA capture.[40] Such manipulations could not be performed with solid-state nanopores. It is also a common misconception that biological nanopores are not particularly stable. Extensive work by Bayley and co-workers has shown αHL to be stable and functional at temperatures approaching 100 °C,[41] at pH values as high as 11.7[42] and in conditions that are known to prevent ssDNA from forming any secondary structure (4 M urea).[43] α-Hemolysin has already been targeted as a stochastic sensor for analytes other than DNA.[44] These included the detection of a divalent cation using a modified αHL mutant which contained covalently attached metal chelators[45] and the detection of organic molecules by non-covalent adaption of the lumen of αHL with β-cyclodextrin.[46] It has also been utilised for the measurement of the hydrogen-deuterium isotope effect at the single molecule level.[47]

In 1996, Deamer and co-workers determined that ssDNA and ssRNA translocated through wild-type (WT) α-hemolysin and that the duration of the translocation events gave a measure of the length of the strand.[24] Since this initial publication, there has been great interest in determining the base discriminating abilities of αHL. It was observed that αHL could be used to distinguish between homopolymers of poly(dA) and poly(dC), when they translocated through the pore, by comparison of the block level, event duration and overall event dispersion.[48] However, methods to detect single nucleotides in a fully translocating

strand were unsuccessful as the DNA translocated too quickly for detection of individual bases (~1-20 μs per base).[48, 49]

To overcome this issue, and establish if it was indeed possible to observe full base discrimination using nanopores, two methods were targeted: 1) immobilisation of ssDNA in the pore[50-61] and 2) detection of individual nucleoside monophosphates by utilising a cyclodextrin-adapted αHL.[62, 63] There are several methods for immobilisation of DNA these include a 5' or 3'-terminal hairpin,[50, 51, 64] a short complementary dsDNA region attached to the 5' or 3' end of a ssDNA oligonucleotide[52-55] or a biotin•streptavidin complex.[52, 53, 55-61, 65] The main advantage of immobilising DNA is that it increases the residence time (e.g. >100 ms) of the base in the pore thus increasing the signal to noise ratio and improving resolution.[50, 58, 59] Work by Stoddart *et al.* has shown that by employing the biotin-streptavidin immobilisation approach it was possible to distinguish all four canonical bases when they were immobilised in the αHL pore (this approach will be discussed in more detail in Chapter 2.1).[58-61] Jetha *et al.* have also shown that cytosine can be discriminated from its epigenetically modified base 5mC by comparison of the current block levels produced, upon immobilisation of oligonucleotides containing 5mC and C by the biotin•streptavidin approach.[65] It was also important to investigate αHL's ability to detect individual nucleoside monophosphates because if an exonuclease was to be utilised in tandem with αHL it could provide another method of sequencing DNA (Figure 1 – 9, see Chapter 1, section 1.1.6). To this end in 2009, workers at Oxford Nanopore Technologies Ltd. illustrated that it was possible to distinguish the nucleoside 5'-monophosphates of C, T, A, G and the epigenetic modification 5-methylcytosine (5mC) using an αHL mutant with a covalently attached cyclodextrin-adapter.[63] Therefore, it has been demonstrated, using both these methods that αHL has the ability to discriminate the standard DNA bases – A, C, G and T.

MspA is an octameric porin isolated from *Mycobacterium smegmatis*.[37] It has a goblet shape with a large barrel at the top (40 Å in diameter) and a smaller 16-stranded β barrel, that forms the channel constriction point towards the base (28 Å in diameter, Figure 1 − 10). The goblet is 96 Å in height with a largest outer diameter of 88 Å. Like αHL, MspA is a robust channel that retains its activity at high temperature, within the pH range 1-14 and in the presence of denaturants such as urea and sodium dodecyl sulfate.[66] The inside of the MspA channel contains several negatively charged residues, which preclude translocation of ssDNA in the WT pore. However, mutation of the three negatively charged aspartate residues, located near the constriction (at positions 90, 91 and 93), to neutral asparagines resulted in ssDNA translocation at a rate of ~2-10 base/μs, which is much faster than for αHL (~0.5-1 base/μs).[64] MspA has also shown the ability to discriminate C, T, A and G, when immobilised in the pore by means of a DNA hairpin.[67]

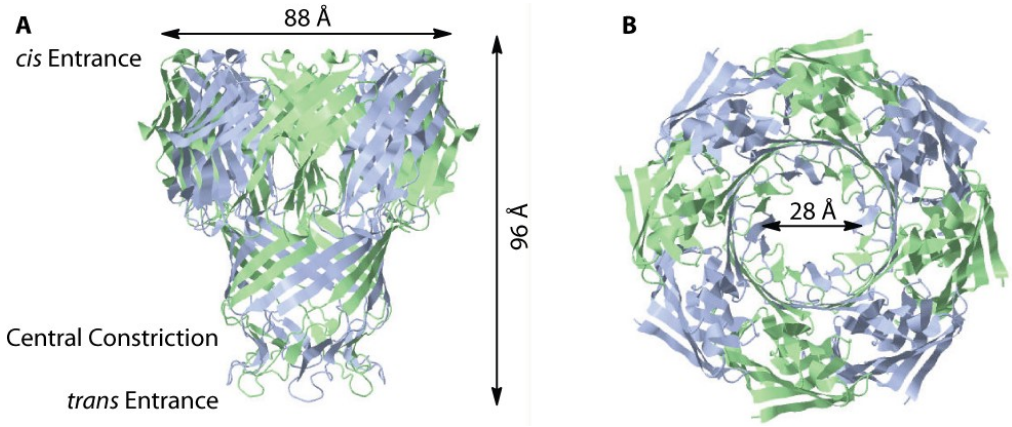


Figure 1 – 11 – *Cartoon representations of the MspA biological nanopore. A) Side view of the MspA WT pore (created in Jymol from the pdb file 1UUN).The nanopore is goblet-like in shape, with a single constriction site (1.2 nm in diameter) towards the base at the trans side. B) View from the cis side looking down through the MspA WT pore (created in Jymol from the pdb file 1UUN).*

It was proposed that the structure of MspA might make it a superior porin for DNA sequencing. This is because it may be able to resolve single nucleobases owing to it possessing a single narrow constriction site (diameter ~1.2 nm and length 0.5 nm). However, theoretical calculations have proposed, that even with an infinitely short

channel, it would not be possible to obtain single base resolution.[68-70] This is due to the high electric field region, which determines the electrical read-out region of the pore, being observed to extend approximately one channel diameter either side of the pore.[68-70] This would give MspA a calculated read region of approximately ~2.9 nm, which would include multiple bases (~6-7 nucleotides). This is supported by initial experimental work by Gundlach and co-workers which demonstrated that the read-region of MspA included at least 3 nucleotides, when immobilised at the constriction site.[67]

1.1.5.3 – Nanopore Sequencing Using Solid-State Nanopores

Solid-state nanopores provide an alternative means of nanopore sequencing.[38] Very stable and functionally useful nanopores can be formed (in silicon nitride (SiN), silicon oxide, metal oxides or graphene[71]) using a variety of techniques such as ion-bean sculpting,[72] e-beam drilling[73] and atomic layer deposition.[74] As these types of nanopores can be manufactured to have a specific internal diameter, it is possible to study the translocation of both double and single stranded DNA.[72, 74, 75] The main advantages of solid-state nanopores are that they do not require a lipid bilayer support, which is the major source of instability of biological nanopore experimental design. However, collaborative work between the Bayley and Dekker labs has also shown that it is possible to form a hybrid pore by inserting a biological nanopore (αHL) into a solid state nanopore (SiN). This system could be used to combine the improved stability of solid state pores with the increased potential for genetic engineering provided by the biological nanopores.[34] These hybrid structures may also provide a new approach for the implementation of biological nanopores into array based systems that do not employ lipid bilayers.

1.1.6 – Challenges of Nanopore Sequencing

Before nanopore sequencing can be actively applied to the sequencing of entire human genomes there are a number of challenges that must be overcome.[23] The first of these would be achieving the potential of long DNA read lengths, which would offer a massive advantage over current procedures. Nanopores read DNA, base by base, as it is threaded through the pore and the accuracy of a base call at one particular moment in time does not depend on the prior history of the system.[25] This would mean that in theory the maximum length of DNA that could be read is limited only by the practicalities of fragmentation of the original strand during sample preparation. In reality, however, there a number of other factors which need consideration. As DNA's genomic form is a double-stranded helix, it would be necessary to denature this duplex for sequencing using αHL or MspA. One method of doing this would be to perform sequencing experiments at high pH where dsDNA would be fully denatured. Work carried out by Maglia *et al.* illustrated that αHL remained stable and functional in the alkaline conditions required to denature DNA.[42] Work by Fologea *et al.* showed that in solid state nanopores, it was possible to denature 3 kb paired dsDNA at pH's greater than 12 and observe ssDNA translocation events.[76] Another possible method would be electromechanically un-zip DNA using the nanopores themselves. It has been shown that both solid state and biological nanopores are capable of un-zipping DNA hairpins, 50 base-paired and 10 based-paired duplex regions respectively.[54, 77, 78]

The next problem to be addressed would be the attainment of efficient capture and translocation of long ssDNA through the pore. To reduce the chance of ssDNA forming secondary structure by mediated base pairing, single-channel recording experiments have been carried out in the presence of 4 M urea (a denaturant that eliminates the secondary structure of ssDNA).[43] Strands of ssDNA and ssRNA (95 nucleotides in length), that were

deliberately designed to have significant secondary structure, were successfully translocated through WT αHL in the presence of 4 M urea at + 180 mV. No translocation events were observed in the absence of urea.[43] Several experiments have also been carried out in both solid-state[35] and biological nanopores[40] to improve the capture rate of DNA. In solid-state nanopores, capture rate can be increased by a factor of more than 30 by applying a 20-fold salt gradient across the pore.[35] Whereas, with biological nanopores the most effective way of increasing capture rate is by the site-directed mutagenesis of charged amino acid residues in the lumen of αHL. By increasing the positive charge near the internal constriction, the DNA translocation frequency (at + 120 mV) increased by almost an order of magnitude and decreased the threshold voltage for which DNA translocation occurred.[40]

At present, the rate of DNA translocation is observed to be ~1-20 μs per base,[48, 49] this highlights the potential for the use of this method for ultra-fast sequencing. However, owing to the current limitation of patch clamp amplifiers, it is not possible to resolve individual nucleotides travelling at these rates because it pushes the required detector bandwidth into the MHz region which, therefore, precludes the measurement of pico-ampere steps in ion current.[25] This issue is further complicated by the fact that the motion of DNA through nanopores is also not uniform owing to DNA interacting with the walls of both solid-state and biological nanopores.[25, 26, 49, 79-83] This would mean that a DNA strand could exhibit stick-slip transitions[80] as it translocates through the pore which could result in bases not being detected by the probing element of the nanopore. Altering experimental conditions such as decreasing temperature,[48] increasing solvent viscosity and salt concentration[84] have not been able to produce the required reduction in the translocation rate.

A procedure is, therefore, required that will effect unidirectional motion of DNA through a nanopore in single nucleotide steps and at appropriate speeds so that the residence time for each base in the channel is sufficient to enable its unequivocal determination. The most likely method for achieving such processive motion of DNA through a nanopore is by the use of a DNA processing enzyme. The most likely candidates would be either DNA helicases,[85, 86] polymerases[53-55] or exonucleases.[62, 63] Work has been carried out in real-time, by Ghadiri and co-workers, that demonstrates that DNA polymerase can be used to control both the speed and ratcheted motion of ssDNA translocation through αHL.[53] Another approach would be to covalently attach a DNA exonuclease to the nanopore, so that as it processively breaks down DNA into its component nucleoside monophosphates,[87] each would be identified as it is released and translocated through the pore (Figure 1 – 9). Towards this goal, work has been carried out to show that a mutant of αHL that contains a covalently attached cyclodextrin can discriminate all four canonical bases and the epigenetic modification 5mC under conditions in which the exonuclease functions (asymmetric salt).[63] The next significant challenge for this approach is the alignment of the active site of the exonuclease with the pore entrance, to ensure that as each nucleoside monophosphate is released it enters the nanopore. Although both these techniques have shown great promise, they still require significant optimisation before they will be able to challenge current sequencing techniques.

1.2 – Epigenetic Modifications

Epigenetics is defined as 'the study of heritable changes in gene expression that occur without a change in DNA sequence'.[88] It is a mechanism by which certain genes in a cell are switched on and off, depending on whether or not they are necessary for that particular cell's normal development. There are several epigenetic modifications that occur *in vivo*, to the DNA bases, these include 5-methylcytosine (5mC)[89] and 5-hydroxymethylcytosine

(hmC).[90] It has also been shown that epigenetic modifications that silence gene expression can occur on amino acids (e.g. histone deacetylation and methylation).[89]

1.2.1 – 5-Methylcytosine (5mC)

5-Methylcytosine is the most commonly modified DNA base in the eukaryotic genome. It was first discovered in 1948 and is produced by DNA methyltransferase enzymes (DNMTs) which catalyse the transfer of a methyl group from the cofactor *S*-adenosylmethionine (SAM) to the 5'-carbon of the pyrimidine ring.[91, 92] Approximately 5% of cytosines in humans were found to be methylated, however, the exact percentage depends on the cell tissue type.[93, 94] Methylation occurs almost exclusively on cytosines that precede a guanosine in the DNA sequence, these sites are known as CpGs (cytosine-phosphate-guanine). CpG dinucleotides are generally found in clusters in small stretches of DNA termed 'CpG islands'. These islands are often associated with promoter regions which are areas of a gene where DNA transcription is believed to begin. In the human genome around 80% of the CpG dinucleotides, that are not associated with these CpG islands, were found to be methylated.[89] The CpG islands, themselves, are generally found to be unmethylated (Figure 1 – 12).[95]

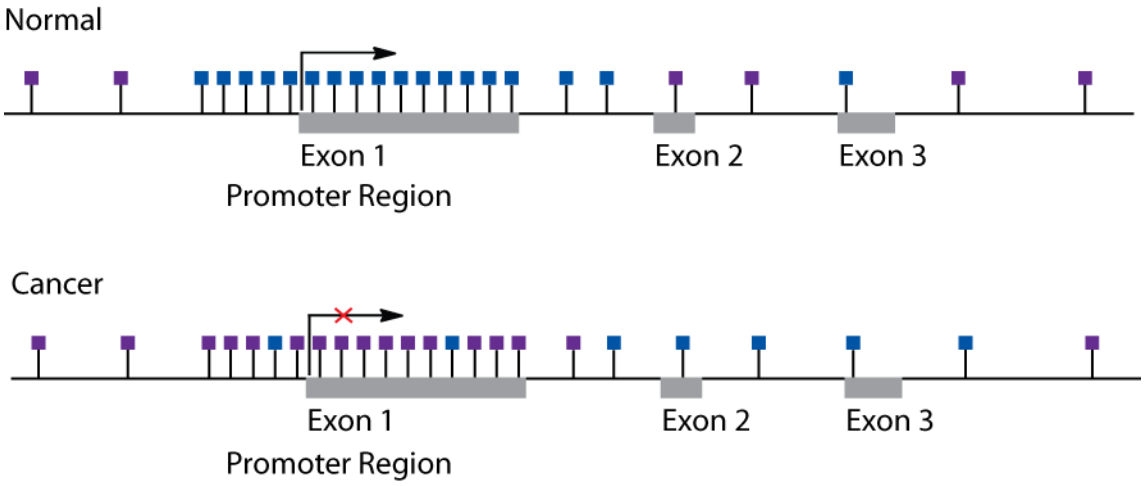


Figure 1 – 12 *– A small section of the mammalian genome is depicted showing exons 1, 2 and 3 (grey boxes) of a sample gene, introns of that gene (black line between the exons) and regions outside that gene which contain a small number of CpG sites (purple and blue squares). Sections which are associated with gene promoter regions show a high level of*

CpG sites which are un-methylated (blue squares) in normal cells, these are known as CpG islands. It is at these sites that gene transcription begins (black arrow). In normal cells, most CpG sites outside of the CpG islands are methylated (purple squares). However, in cancerous cells the promoter regions are found to be methylated preventing transcription (red cross on the black arrow) of tumour suppressor genes. Other regions of the gene, which are methylated in normal cells, are un-methylated in cancerous cells leading to weakening of transcriptionally repressed genes and expression of potentially harmful genes. This figure has been adapted from Herman et al. 2010.[89]

DNA methylation is required for normal development because it maintains a large amount of the noncoding DNA in the cell in a transcriptionally inert state. This prevents the replication of areas of the genome such as repeat elements, inserted viral sequences and transposons (DNA sequences that have moved from their normal location into a different section of the genome) which, if transcripted, may harm cells.[89] However, abnormal patterns of DNA methylation have also been linked with various types of cancer and diseases.[89, 92, 96] Virtually all varieties of cancer have shown both gains in methylation of CpG islands, located at gene promoter regions, and losses of methylation in CpG depleted regions, where most CpG dinucleotides would normally be methylated (Figure 1 – 12). In the areas where methylation is lost this is likely to weaken transcriptional repression in normally silent regions of the genome which could result in the expression of potentially harmful genes. Losses in methylation could also result in functional instability of the chromosomes, whereas, hypermethylation is associated with silencing of genes which are known to act as tumour suppressors.[89]

A key difference between cancers caused by epigenetic mutations versus those caused by genetic mutations is that the later are irreversible but the former could potentially be reversed. If the genes which code for the three different DMNT's were disabled then this would prevent methylation. However, this was not a viable approach as these enzymes are vital for embryonic development and disabling them in mice resulted in embryonic and early postnatal death.[97, 98] This has lead to the investigation of possible demethylating agents that could cause re-expression of silenced genes.[89] Two such inhibitors are

5-azacytidine and deoxy-5-azacytidine which inhibit the DNMTs through covalent binding. These inhibitors are thought to promote their effect through the reversal of gene silencing. Despite their specificity for the DNMTs they caused toxic side effects. When treatment was stopped the effects of these drugs were reversed, so prolonged low-dose drug regimes would be necessary. Further investigation is required if these are to become useful in cancer treatment, although they have shown promise in the low-dose treatment of myelodysplastic syndromes.[99, 100]

Owing to the fact that DNA methylation occurs at the early stages of cancer development, it could be possible to use it as a biomarker for early diagnosis and for the detection and assessment of high-risk individuals. This would be extremely useful in types of cancer which currently do not have a method of premature detection e.g. lung cancer.[96] However, it is very important to be able to establish which hypermethylated genes are responsible for cancer because some CpG islands (located in genes mostly in the colon) have been observed to be increasingly methylated as individuals aged.[89] This increase in methylation could account for the increased risk of developing certain cancers with age (such as colorectal cancer)[101] but it could also complicate the implementation of DNA methylation as a suitable biomarker. However, if DNA methylation changes could be reliably detected in body fluid such as plasma, sputum and urine, then it could provide an extremely useful, non-invasive method of early cancer detection.[102]

Over the past few years, there has been widespread interest in developing techniques towards the analysis of DNA methylation – both the location of sites and the overall percentage of methylation in the genome.[92, 103, 104] One of the first methods used to provide large-scale genome-wide analysis of cytosine methylation was DNA hydrolysis and subsequent detection of the deoxyribonucleosides by chromatography-based methods. DNA was initially quantitatively hydrolysed, using either DNA I and nuclease P1 or snake

venom phosphodiesterase, and then treated with alkaline phosphatase. The resultant deoxyribonucleosides were then separated and detected by various chromatographic techniques e.g. thin layer chromatography or reverse-phase high-performance liquid chromatography coupled with mass spectrometry.[103, 104] This method is useful for whole genome analysis, however, it does not provide any information as to the location of the sites of methylation.

Another technique that has been used to detect methylation in a genome-wide approach is immunoprecipitation. This method involves a highly-specific reaction between a monoclonal antibody and 5mC. DNA is initially immobilised onto membranes and then incubated with an antibody directed against 5mC. The antibody can be detected in a number of different ways including fluorescent labelling. In this instance the intensity of the measured fluorescence was proportional to the degree of methylation. The main limitation of this procedure was that it must be performed on ssDNA.[104] This technique was used to determine the methylation profile of the *Arabidopsis thaliana* genome in 2006.[105]

Methods that use restriction endonucleases are also employed to measure the extent and location of methylation sites in a specific gene. Restriction endonuclease isoschizomers, that have different sensitivities to cytosine methylation, provide a mechanism for the study of methylation changes within their recognition sites.[104] Methylation of a site renders it sensitive to cleavage by one isoschizomer but not the other. There have been a number of different methods developed for the detection of 5mC from C, once they had been successfully discriminated by endonucleases. These include southern blots[104] and detection of photocurrents.[106, 107] This is a popular laboratory technique for methylation detection as it is simple to use and cheap to perform on a large number of samples. However, this technique is only useful for probing a limited number of methylation sites, all of which

must contain the restriction site of interest, making it impractical for whole-genome wide profiling.[103]

The most commonly implemented technique for the detection of 5-methylcytosine is the bisulfite sequencing method.[108] When ssDNA is treated with a high concentration of sodium bisulfite it preferentially deaminates cytosine residues to uracil, compared with a very slow rate of deamination of 5-methylcytosine (Figure 1 – 13).[109] It is this difference in reactivity that is used to discriminate cytosine from the epigenetically modified 5mC. The reacted DNA is then amplified by PCR yielding a DNA fragment in which all the uracils (which were formally cytosines) and thymine residues have been amplified as thymine and only 5mC has been amplified as cytosine. The amplified DNA is most commonly interpreted by DNA sequencing techniques – where the presence of a band in the cytosine-track indicates a 5mC. Using the bisulfite method in tandem with modern DNA sequencing techniques (such as the Illumina's Genome Analyzer, which uses cyclic reversible termination sequencing by synthesis) it has been possible to sequence the methylome (the DNA methylation profile of an entire genome) of plants,[110, 111] animals[112] and humans.[93, 113] There are, however, a number of disadvantages to this technique: 1) the bisulfite reaction causes severe damage to the original DNA template (Okamoto and co-workers measured the amount of DNA present before and after a 16 hour bisulfite assay and found only 0.1% of the original DNA was available, which was a significant level of degradation);[114] 2) it can take up to half a day for complete modification of cytosine and a shorter reaction time would be desirable; 3) it can suffer from incomplete conversion of cytosines to uracil which results in false positive readings. Despite these issues the bisulfite method is currently the most reliable and practical method of distinguishing 5-methylcytosine from cytosine.

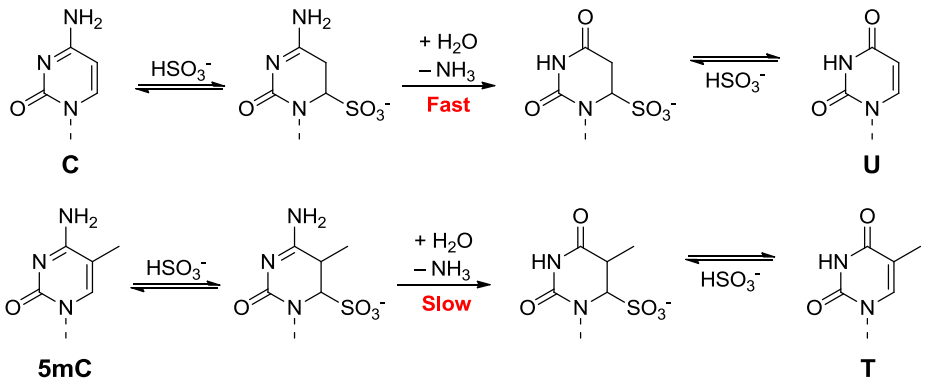


Figure 1 – 13 – *The reaction of bisulfite with cytosine and 5-methylcytosine.*

1.2.2 – 5-Hydroxymethylcytosine (hmC)

5-Hydroxymethylcytosine was first discovered in 1953, in *E. coli* infected with bacteriophage T2, T4 or T6.[115] It was only in the last two years that it has been discovered in mammalian DNA.[116, 117] Since the back-to-back publications in Science, where Kriaucionis *et al.* discovered that hmC was present in 0.59 ± 0.05% of Purkinje cells from the brain of mice[116] and Rao and co-workers noted that the enzyme TET1 converted 5mC to hmC in mammalian DNA,[117] there has been immense interest in the precise location and function of the 'sixth-base'.[90]

In order to determine more about the function of hmC *in vivo*, it is necessary to develop techniques to accurately determine its distribution throughout the body and exact location on individual genes. The most common method of determining the global percentage of hmC found in a specific tissue type, is to enzymatically digest the DNA into its component nucleosides and then determine the amount of hmC nucleoside present. hmC has been detected by a number of different methods including chromatography[116] and liquid chromatography combined with mass spectrometry (MS) analysis.[118, 119] These studies showed that hmC was present in all mouse tissue-types tested but in varying amounts (ranging from 0.03% up to 0.69% of deoxyguanosine) with the highest percentages being isolated from the central nervous system (CNS).[119] As the brain was found to contain a

large percentage of hmC, it was further investigated to determine hmC's distribution within its different regions. In this study, the highest levels of hmC (>0.6% of deoxyguanosine) were observed in the hippocampus and cortex – areas of the brain which have higher cognitive function.[118] This provided valuable information as to the distribution of hmC, however, to determine its function it would be necessary to be able to pin-point hmC's location on individual genes.

Initially, the methods currently employed to distinguish 5mC from C were investigated, to determine if it they could also be used to identify hmC. An affinity-based detection of hmC was investigated using a monoclonal antibody directed against 5mC. When DNA strands, exclusively containing either C, 5mC or hmC, were subjected to immunoprecipitation with an anti-5mC antibody high affinity for the 5mC strand was observed. The strands containing only C or hmC exhibited no affinity.[120] This meant that immunoprecipitation, employing an anti-5mC antibody, could not be used to distinguish C, 5mC and hmC.[120] Recently, anti-hmC antibodies have been developed which can be used to determine its location. This technique was used to show that hmC was found to be present, along with 5mC, in mitochondrial DNA in humans.[121] An antibody which can target cytosine-5-methylenesulfonate (CMS), which is produced by the bisulfite modification of hmC, can also be employed to detect the location of hmC (see below).[122, 123]

The implementation of restriction enzymes that are specific for 5mC were investigated. By using restriction enzymes that are either methylation-sensitive or methylation-insensitive, it was not possible to clearly distinguish C, hmC and 5mC from each other.[120, 124, 125] However, recent progress has developed a restriction enzyme that is able to selectively cleave DNA containing hmC that has been glycosylated at the free hydroxyl.[126] This technique has been employed to establish the genome-wide distribution of

5-hydroxymethylcytosine in both mice[127] and humans.[123] In both examples the glucose moiety was installed using a β-glucosyltransferase enzyme and then a biotin tag was attached onto the sugar (see Chapter 1, section 1.3.6 for further details).

Bisulfite sequencing is the most common method for distinguishing 5mC and C from each other. It converts all cytosines to uracils whilst leaving 5mC unchanged. When hmC-containing DNA was treated with bisulfite it reacted to form cytosine-5-methylenesulfonate (CMS) (Figure 1 – 14).[120, 124, 128] This was found to have a deamination half-life to uridine-5-methylenesulfonate (UMS) of 28 h (at pH 5.5 at 37 °C) which was significantly longer than C's (half-life of 6 minutes) and 5mC's (half-life of 8 h).[129] DNA strands which exclusively contained either C, 5mC or hmC have been subjected to bisulfite modification, PCR amplification and sequencing.[120] In the strand which contained only C, all the cytosines were converted to uracils and were read as thymines in the sequencing reads. The strand which contained 5mC did not react with bisulfite and was, therefore, read as cytosine in the sequencing reads. The hmC containing strand reacted with bisulfite forming the CMS adduct which was again read as a cytosine. This meant that the bisulfite modification could be used to distinguish cytosine from its epigenetic modifications but 5mC and hmC could not be distinguished from each other.[120, 124, 128, 130] Recent methods have attempted to target the product of the bisulfite modification of hmC in order to aid its identification.[122, 123] This process relied on an antibody which acts against CMS (see Chapter 1, section 1.3.6). This technique has been used to map genome-wide hydroxymethylation in human embryonic stem cells (ESC)[123] and for investigating the connection between hmC and myeloid cancers.[122]

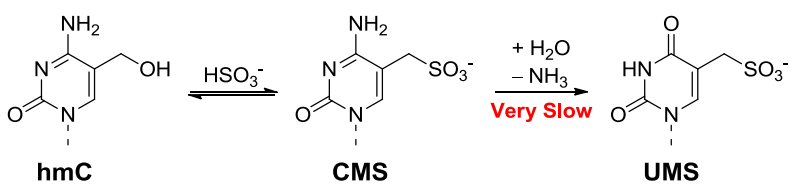


***Figure 1 – 14** – The mechanism of reaction of bisulfite with 5-hydroxymethylcytosine.*

There have been several attempts to develop other sequencing methods for the detection of cytosine and its epigenetically modified bases.[131, 132] One technique used solid-state nanopores to try to distinguish hmC from 5mC.[131] This procedure relied on differences in current-block levels, transport times and current signatures to distinguish the bases. In this case it was possible to distinguish hmC from 5mC and C but the signals produced by 5mC and C were too similar for discrimination between all three bases.[131] Another approach used single-molecule real-time sequencing to attempt to distinguish C, 5mC and hmC.[132] This technique used a DNA polymerase to catalyse the incorporation of fluorescently labelled nucleotides. It was differences in the kinetics of incorporation of each base that distinguished them from one another. Flusberg *et al.* demonstrated that the incorporation kinetics for 5mC and hmC were sufficiently altered from that of C that it was possible to directly distinguish cytosine from its epigenetic modifications. However, enhancement of kinetic sensitivity would be required to discriminate 5mC and hmC from each other.[132] The final example that will be discussed employed an artificial phosphopeptide which interacted with dsDNA.[133] This phosphopeptide, which contained a phosphotyrosine incorporated into the zinc finger peptide, was originally found to selectively bind methylated DNA ten-times more strongly than unmethylated DNA.[133] When it was used to distinguish between hmC and 5mC, it was shown to bind 5mC much more tightly because the CH_2OH functionality of hmC sterically inhibited the approach of the phosphopeptide.[134] However, it bound cytosine and hmC with a similar affinity and, therefore, could not be used to distinguish C, 5mC and hmC.[134]

5-Hydroxymethylcytosine is produced by the action of TET enzymes on 5-methylcytosine.[117, 135] There are three different TET enzymes TET1, TET2 and TET3 all of which require 2-oxoglutarate and Fe(II) to function. Although all three enzymes produce hmC they have differing functions *in vivo*.[135, 136] TET1 and TET2 are thought to

be linked to the maintenance of pluripotency in ESC (a pluripotent stem cell is one which has the potential to differentiate into any of the three germ layers), whereas, TET3 has been linked to the promotion of cell differentiation. This means that hmC is thought to be involved with the formation and maintenance of pluripotent stem cells and their differentiation into germ cell lines. TET enzymes have also been linked to cancer because mutations in the gene which encoded for TET2 have resulted in myeloid malignancies.[122] These mutations were found to reduce the catalytic activity of TET2 and thus displayed low levels of hmC in genomic DNA, compared to healthy controls.[122] The accurate measurement of genomic hmC levels could be employed as a useful indicator for the early diagnosis of these types of cancer. In an attempt to establish if low levels of hmC were purely a consequence of cancer or if low levels of hmC caused it, Koh *et al.* injected stem cells which had depleted TET1 levels into mice. Injection of non TET1-depleted cells resulted in the production of benign teratomas (encapsulated tumour), whereas, TET1-depleted cells formed very aggressive tumours that were in a state of high-cell division.[136] This indicated that hmC levels are extremely important for the correct function of cells and any perturbation of hmC levels and location could promote serious side-effects.

At present the precise function of 5-hydroxymethylcytosine is not known. However, there are several hypotheses as to what hmC's biological role *in vivo* could be. It was proposed that hmC could be a stable end-product that influences chromatin structure and local transcriptional activity. It could do this by recruiting selective hmC binding proteins or by excluding specific 5mC binding proteins which do not recognise or bind hmC.[137] Although no hmC specific binding proteins have been discovered there is evidence to support this hypothesis because 5mC-CpG binding protein 2 has been found to have a binding affinity for hmC that is an order of magnitude less than for 5mC.[137]

Another possibility could be that hmC is an intermediate in an oxidative demethylation process (Figure 1 – 15). Evidence to support this pathway was obtained during genome-wide distribution studies.[123, 127] Both studies noted differences in the distribution of hmC and 5mC – hmC was found to be located at CpG and CAG sequences, whereas, 5mC was mainly found at CpG sites only.[123] If hmC and 5mC are found at sites of gene promotion then this is generally predictive of low gene expression. However, hmC was found at gene promoter regions, which were upregulated upon cell differentiation, suggesting that hmC could be involved in priming a site for rapid activation in response to a specific signal.[123] This could offset the gene repression effect of 5mC which would support the idea of an oxidative demethylation process. Carell and co-workers have also carried out several experiments to try and determine if any 5-formylcytosine (fC) or 5-carboxylcytosine (caC) were present in digested mammalian DNA.[119, 138] In an initial study of a variety of different tissue types they were unable to detect any fC or caC. However, a more recent publication has shown the presence of fC at a surprisingly high level in mice ESC (0.02% of deoxyguanosine).[138] In this cell type around every tenth or twentieth hmC base is oxidised to fC. No caC was detected during the experiments which meant, if this base is an intermediate in oxidative demethylation, its levels were below the detection limit of the current procedure.[138] Carell also illustrated, that in DNMT double knock-out mice ESC, practically no hmC (0.02% in comparison to 0.39% in normal cells) and no fC were detected.[138] This further supported the hypothesis that fC is likely to be produced from 5mC via hmC through a further oxidation mechanism.

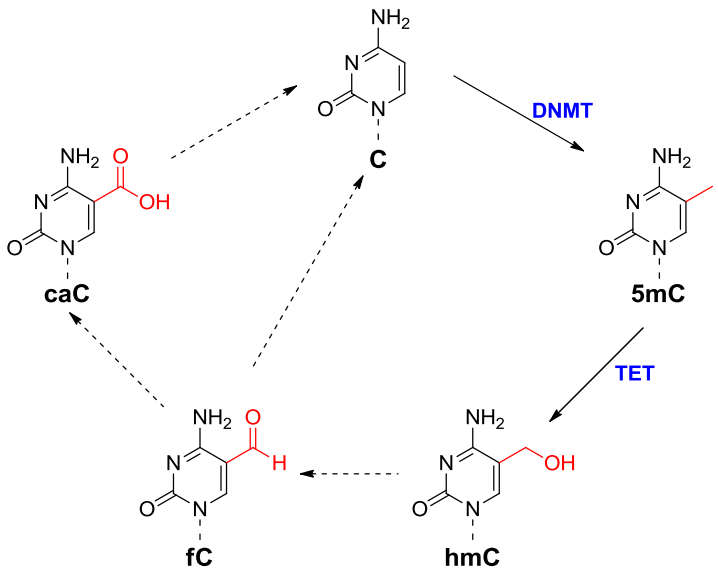


Figure 1 – 15 – *The proposed cycle of methylation and oxidative demethylation of cytosine. DNMT is known to methylate the 5-position of cytosine to give 5-methylcytosine. 5mC is then oxidised by TET enzymes to produce hmC. 5-formylcytosine (fC) and 5-carboylcytosine (caC) are potential intermediates for the oxidative demethylation to cytosine. fC has been detected in ESC, however, caC is yet to be found at measurable concentrations in DNA. This figure has been adapted from Pfaffeneder et al. 2011.*[138]

The genome-wide hydroxymethylation study in mice performed by Song *et al.*, also proposed two additional functions for hmC.[127] They observed enrichment in the level of hmC in genes linked to hypoxia and angiogenesis (the process by which new blood vessels grow from pre-existing vessels). This could mean that the oxidation of 5mC to hmC by TET proteins could constitute a new oxygen-sensing and regulation pathway in mammals. This was further supported by the fact that the TET enzyme belongs to the same mononuclear iron-containing dioxygenase super family as hypoxia-inducible factors (HIF), a protein known to have links with hypoxia and angiogenesis.[127] This work also highlighted the fact that the level of hmC in mouse cerebellum was observed to gradually increase from 0.1% of total nucleotides in the genome at postnatal day 7 to 0.4% of total nucleotides in the genome when in the adult stage.[127] This indicated that hmC could play a role in neurological development. As hmC has also been associated with genes that have been implicated in neurodegenerative disorders, this suggests that hmC could be associated with the cause of some neurological diseases.[127]

Despite the huge amount of progress that has been made in the last few years on identifying the distribution and location of hmC, its precise function has yet to be determined. An efficient process that could easily distinguish hmC, from C and 5mC, would be hugely beneficial in identifying hmC's role *in vivo*.

1.3 – Selective Chemical Modification of the DNA Bases

There has been a great deal of interest in chemical reactions and modifications of the DNA bases over the past 40 years.[1] Some of these investigations were prompted because researchers were trying to understand the various reaction mechanisms which occur *in vivo* e.g. what modified bases result from exposure of 5-methylcytosine to ultraviolet light.[139, 140] Others were designed for the modification of one base selectively, in order to aid its detection for DNA sequencing purposes.[1] A number of these reactions will be discussed below, highlighting those which have proven to be selective for each of the bases – adenine, guanine, thymine, cytosine, 5-methylcytosine and 5-hydroxymethylcytosine (Figure 1 – 16).

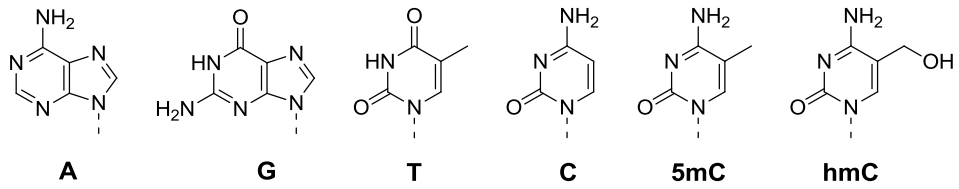


Figure 1 – 16 – *DNA bases adenine, guanine, thymine, cytosine, 5-methylcytosine and 5-hydroxymethylcytosine.*

1.3.1 – Adenine

There have been a number of chemical methods investigated for the detection of adenine in DNA. In 1972 a fluorometric assay was developed to selectively react adenine with glyoxal hydrate trimer.[141] When reacted under acidic conditions, adenine produced a fluorescent response when excited with 328 nm wavelength light, whereas, thymine, cytosine and uracil did not. Guanine, however, produced a fluorescent signal when excited

by a different wavelength of light (285 nm), but when excited at 328 nm no fluorescence was observed. The mechanism of reaction is thought to involve the free amine of adenine because when this nitrogen was bis-methylated no fluorescent signal was noted. These reaction conditions, although selective, are harsh (heating at ~100 °C for 3 h at pH 3.8) and were not tested on ss or dsDNA.[141]

A variety of methylating agents have been used to selectively methylate the N-3 and N-1 positions of adenine.[1, 142-144] Maxam and Gilbert's initial conditions for the identification of adenine were also found to lead to a reaction with guanine. Dimethyl sulfate methylated adenine at the N-3 position and guanine at the N-7 position, leaving the glycosidic bond of both bases susceptible to cleavage upon heating at neutral pH. Subsequent treatment of the DNA with 0.1 M NaOH at 90 °C resulted in cleavage of the sugar phosphate which was then visualised by PAGE.[1] Guanine was five-fold more reactive than adenine so it was possible to distinguish the two bases from each other by their band intensities.[142] It was also noted that the methylated adenine was less stable than methylated guanine when treated with dilute acid for the cleavage of the glycosidic bond. This produced cleavage bands which were strong for adenine and weak for guanine. By inspection of the band intensities of the two sets of reaction conditions it was possible to indentify the locations of adenine and guanine residues in a strand of DNA.[1] Accurate identification of the bases sometimes proved difficult, by comparison of differences in band intensities only, therefore, these conditions were improved by Maxam and Gilbert several years later. They established two new sets of conditions where treatment with dimethyl sulfate and then piperidine reacted selectively with guanine and treatment with acid caused reactions with both adenine and guanine.[143] Therefore, cleavage bands produced by both reactions corresponded to G and those that were only visible after acid treatment corresponded to A. This allowed for the selective chemical distinction between adenine and guanine bases in

DNA.[143] A number of synthetic drugs have been found to selectively methylate adenine bases in dsDNA e.g. AS-I-145 a novel antitumor analogue of duocarmycin (Figure 1 − 17).[144] This family of drugs are known to specifically methylate adenines, however, they could not be used in a sequencing context, to locate all adenines in a strand of DNA, because methylation at the N-3 position is sequence dependent (for example only one adenine was found to be methylated, by AS-I-145, in this sequence – 5'-AT<u>A</u>A-3').[144]

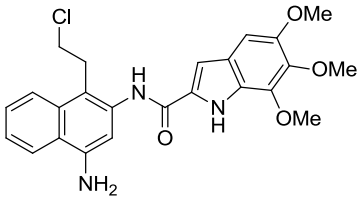


Figure 1 – 17 – *The novel antitumor analogue of duocarmycin – AS-I-145.*

Iverson *et al.* utilised the complexation of K_2PdCl_4 with adenine at pH 2.0 followed by piperidine cleavage to selectively modify and cleave this particular base in DNA.[145] This reaction resulted in high cleavage efficiency at adenine residues and produced DNA fragments with similar electrophoretic mobilities to other DNA sequencing reactions. This meant that this reaction could be used in tandem with other base-specific DNA sequencing reactions for complete DNA sequence determination. Both adenine and guanine are known to bind palladium at the N-1 and N-7 positions.[145] At low pH, binding through N-7 predominates because for guanine this is the most reactive nitrogen and for adenine the N-1 is protonated at this pH (see Figure 1 – 18). It is this increase in positive charge on the adenine ring due to protonation, that increases the lability of its glycosidic bond towards cleavage. As the N-1 of guanine is not protonated at pH 2, it makes this base more stable to glycosidic bond cleavage, resulting in complete selectivity for adenine.[145]

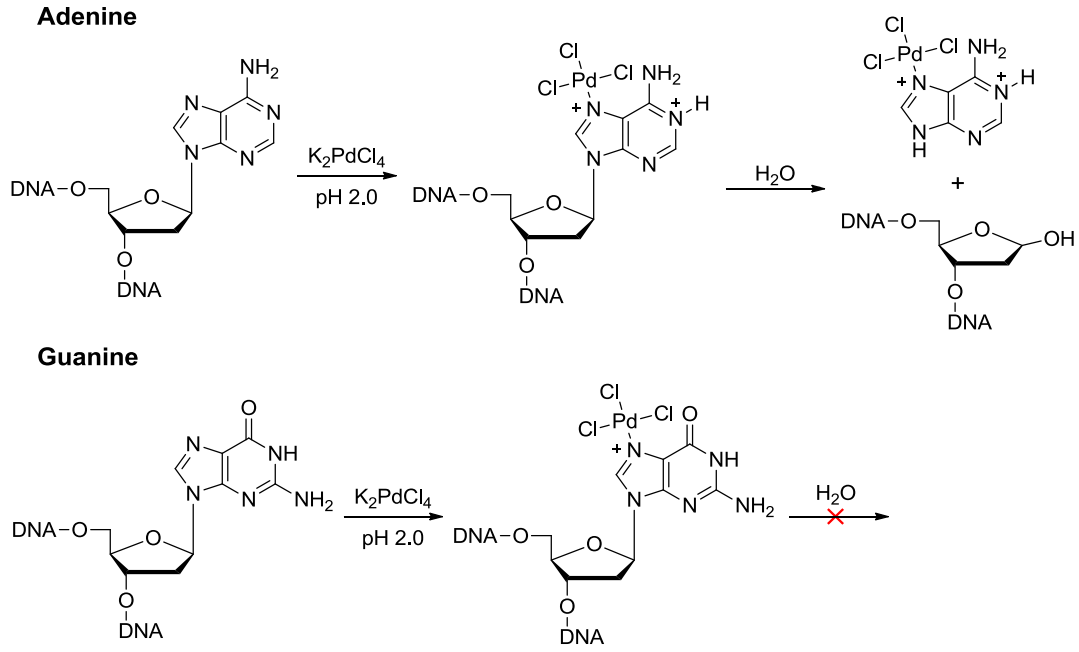


Figure 1 – 18 – *The proposed mechanism for the complexation of K_2PdCl_4 to adenine and guanine at pH 2.0. It is the positive charge on the six-membered adenine ring that is thought to increase its susceptibility towards strand cleavage. This figure has been adapted from Iverson et al. 1987.*[145]

1.3.2 – Guanine

There are two main types of selective chemical modification that have been extensively investigated for reaction with guanine 1) reactions to aid in the detection of guanine residues in DNA including methylation and reaction with phenylglyoxal, 2) investigation of the products produced by one- and multiple-electron oxidations of guanine by various reagents. [1, 143, 146] As discussed previously (see Chapter 1, section 1.3.1) Maxam and Gilbert used a variety of methylation conditions to selectively modify guanine residues in the presence of adenine.[1, 143] After optimisation they found conditions that methylated only guanine residues by the use of dimethyl sulfate and piperidine (Figure 1 – 19).[143] Work by Kai *et al.* has shown that by the selective reaction of phenylglyoxal it was possible to detect guanine bases as low as 47-210 pmol mL^{-1} by fluorescent means.[146] The mechanism for this reaction is thought to involve the 1-imino and 2-amino functionalities of guanine, reacting with the carbonyl groups of phenylglyoxal, because when the imino and amino

groups are methylated no fluorescence was detected. However, these conditions have not been tested on ss or dsDNA, only on the individual bases.

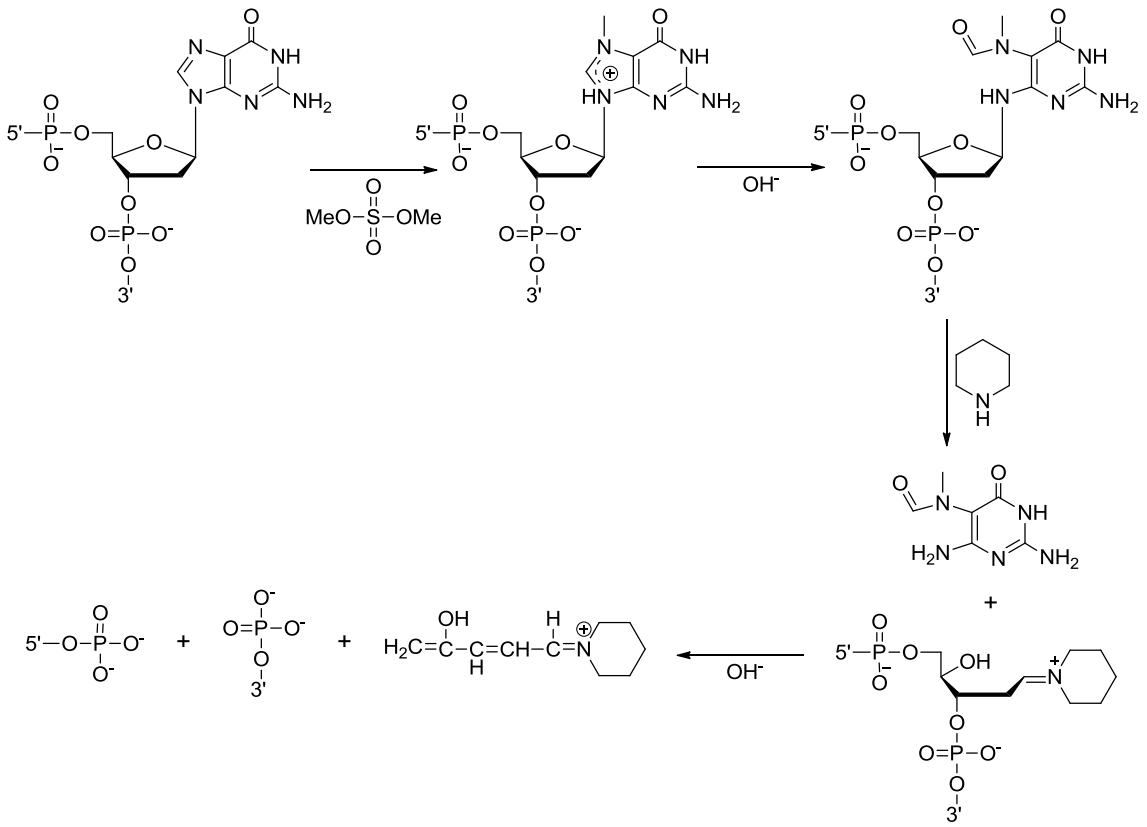


Figure 1 – 19 – *The proposed reaction mechanism for the modification of guanine by sequential treatment with dimethyl sulfate and piperidine. This figure has been adapted from Maxam et al. 1980.*[143]

Guanine is the most easily oxidised base in DNA. Its free energy is –0.15 V which is significantly lower than A (−0.05 V), T (0.1 V) and C (0.25 V).[147] This low oxidation potential makes guanine a likely target for one or multiple-electron oxidations. These oxidised products have been linked to various degenerative diseases and an increased risk of malignant cell transformation leading to cancer development.[148, 149] It is for this reason that many oxidants have been tested and the oxidative products of guanine rigorously investigated. A number of examples using several different oxidants will be discussed.[147, 150, 151]

The single electron oxidation of guanine was investigated by the treatment of dsDNA with ethidium bromide (eb) and methyl viologen (mv).[147] The eb was known to intercalate into dsDNA and promote random oxidative strand breaks, on irradiation by laser light. This reaction's selectivity was improved by the tandem employment of mv which resulted in base-specific cleavage at guanine residues. The two proposed mechanisms of reaction are shown in Figure 1 – 20.[147] In this case the products produced by oxidative damage were not investigated.

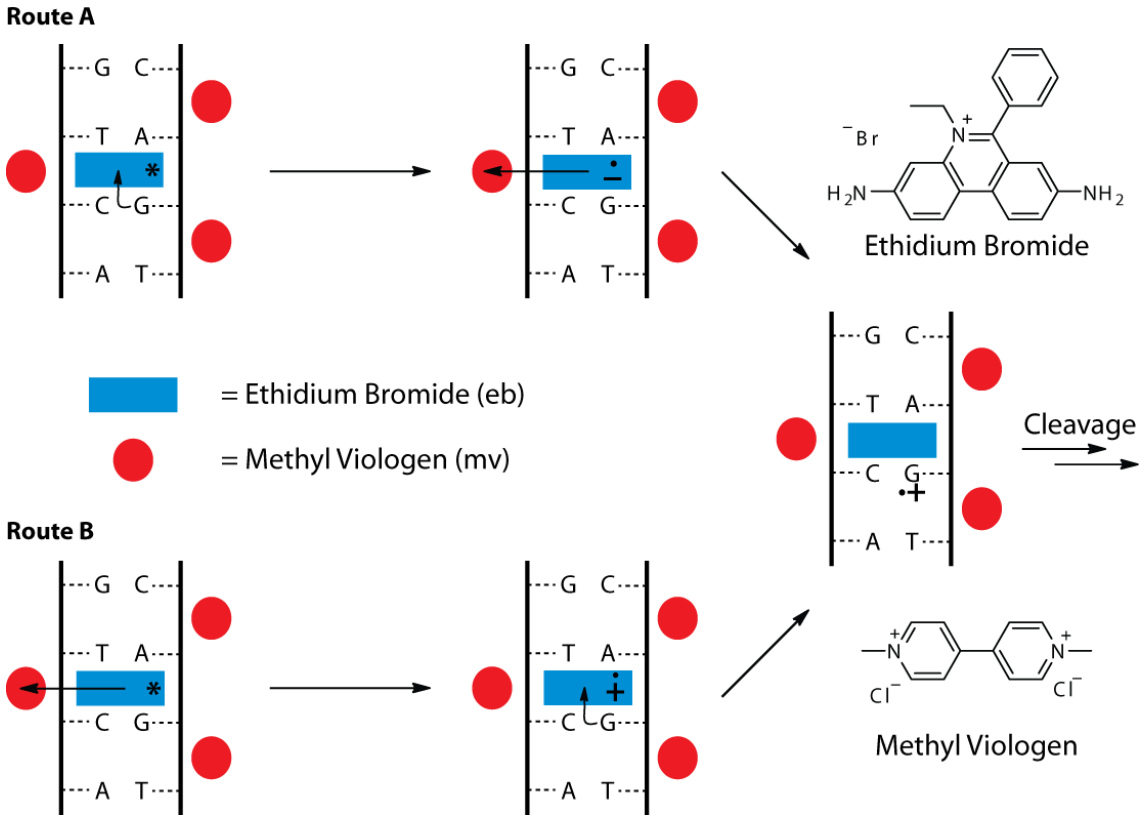


Figure 1 – 20 – *Two proposed mechanisms for the photo-induced cleavage of DNA by ethidium bromide and methyl viologen. Route A – photoexcited eb accepted an electron from guanine and this is transferred to the externally bound cosensitiser mv. Route B – the photoexcited eb initially transfers an electron to mv before accepting an electron from the adjacent guanine. This figure has been adapted from Dunn et al. 1992.*[147]

In 2000, Kochevar and co-workers utilised Mn-TMPyP/$KHSO_5$ to perform a two-electron guanine oxidation in dsDNA.[150] The reaction conditions were selective for guanine, although some residues were preferentially oxidised over others e.g. end-of-duplex G's. The main purpose of this work was to investigate possible intermediates and end-products

of the oxidation reaction. They discovered a new oxidation intermediate 5,8-dihydroxy-7,8-dihydroguanine (5,8-di-OH-dG), which further transformed to guanidinohydantoin, imidazolone and parabonic acid depending on the oxidation conditions (Figure 1 – 21).[150] All of the oxidation products were found to be alkali-labile, resulting in strand breaks if treated with piperidine.

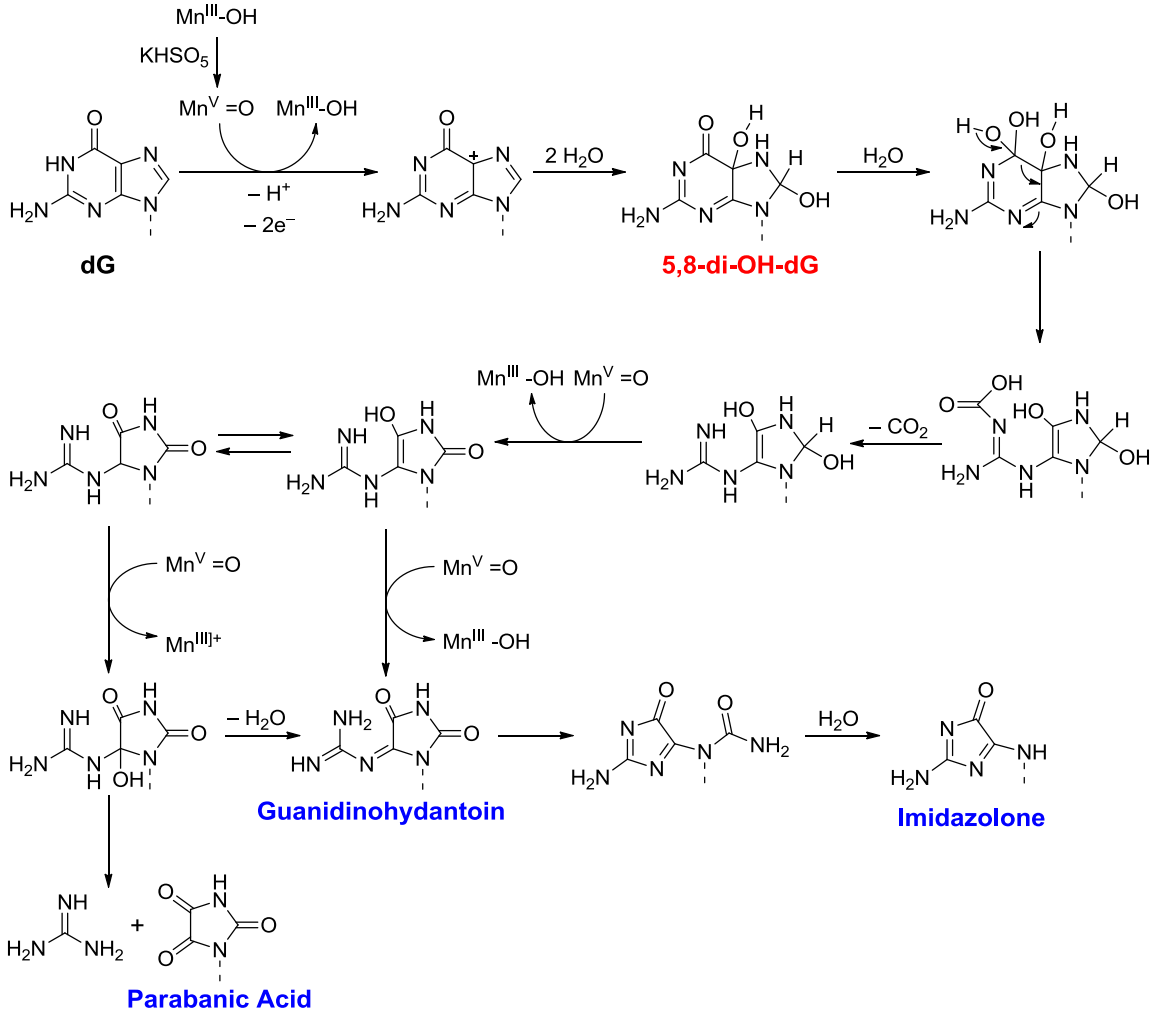


***Figure 1 – 21** – The proposed reaction mechanism for the oxidation of guanine, in dsDNA, by Mn-TMPyP/KHSO$_5$. Figure has been adapted from Vialas et al. 2000.*[150]

The oxidation intermediates and products produced from guanine depend upon the reagent used for oxidation. Joffe *et al.* utilised the biological free radical $CO_3^{\bullet-}$ as a means of four electron oxidation of guanine to form the diasteriomeric spiroiminodihydantoin (Sp) lesions, *via* the 8-oxo-G adduct (Figure 1 – 22).[151] $CO_3^{\bullet-}$ was produced by the one electron

oxidation of HCO_3^- by photochemically generated sulfate radical anions ($SO_4^{\bullet-}$). The reaction of $CO_3^{\bullet-}$ with the DNA bases was selective for guanine, however, the $SO_4^{\bullet-}$ radical anion caused non-specific DNA degradation. The reaction conditions were, therefore, optimised to keep the concentration of $SO_4^{\bullet-}$ to a minimum.[151] It was also noted that oxidation of 8-oxo-G was significantly faster than that of G. This difference in reactivity has been used by other researchers to selectively target 8-oxo-G for chemical modification.[152]

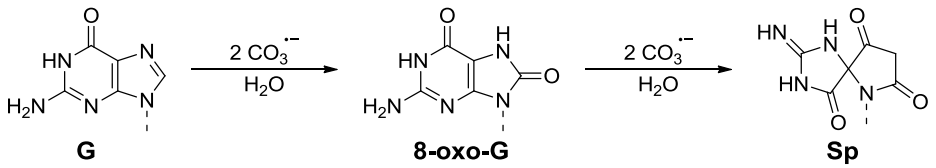


Figure 1 – 22 – *The proposed mechanism for the site-selective oxidation of guanine in DNA.*[151]

1.3.3 – Thymine

The methods developed over the years for the selective modification of thymine have mainly involved oxidation of the C5-C6 substituted double bond. Maxam and Gilbert initially employed a chemical modification using hydrazine, which reacted by the mechanism shown cleaving the base from the sugar, causing strand cleavage on treatment of piperidine (Figure 1 – 23).[1, 143] These conditions were, however, not selective for T as they also modified C. An improvement came from work by Rubin *et al.* who selectively dihydroxylated thymine using potassium permanganate.[153] Cleavage was observed at thymine residues, by treatment with piperidine, with a faint background corresponding to cleavage of C and G. This reagent was known to produce the *cis*-diol of thymine in ssDNA but not react with T's in dsDNA. This chemical modification has been utilised more recently for the detection of the thymine [2+2] photodimer, whose presence has been linked to skin cancer.[154]

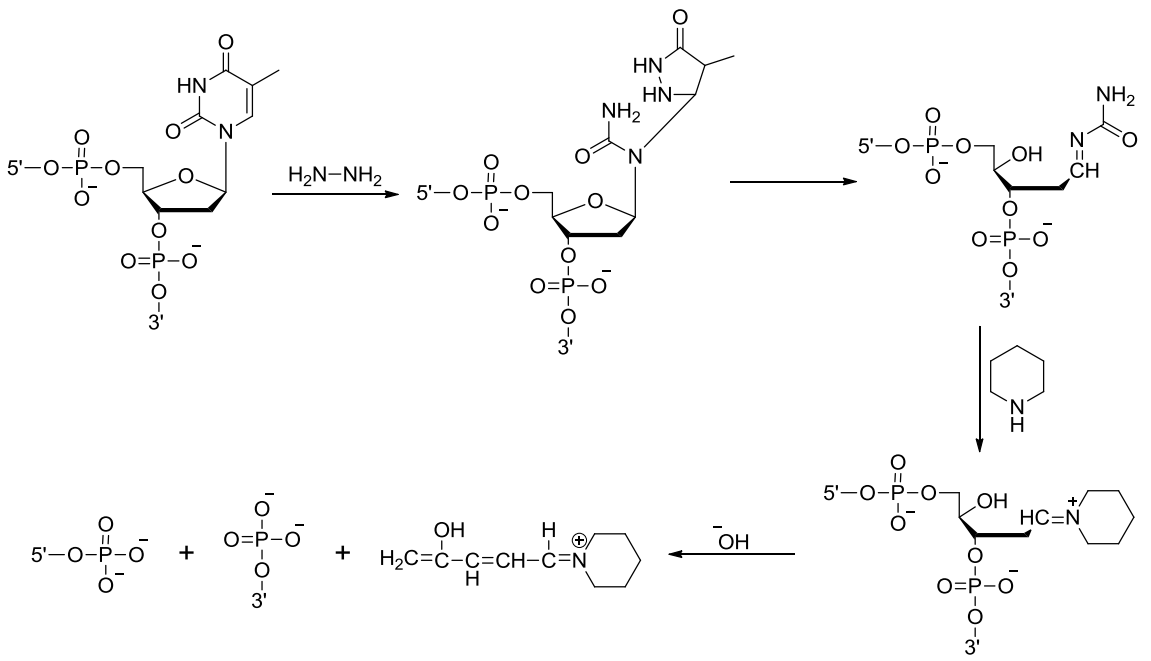


Figure 1 – 23 – *The reaction mechanism for the modification of thymine by sequential treatment with hydrazine and piperidine. This figure has been adapted from Maxam et al. 1980.*[143]

Other oxidising reagents, such as osmium sources, have also been used for the selective modification of thymine residues to form either the corresponding osmate ester[155-157] or the *cis*-diol.[158] Osmium tetroxide alone has been found to selectively dihydroxylate thymine at 23 °C, with an overall conversion of ~95%, whereas under the same conditions only 10% of cytosines were consumed and no reaction was observed with the purine nucleotides.[158] The use of nitrogen-containing ligands was then successfully employed to form more stable osmate esters, as they increased the rate of reaction with osmium by several orders of magnitude.[155] These conditions were selective for T except for partial reactivity with 5mC. To give a measure of the relative reactivity of T, 5mC and C with osmium, Okamoto *et al.* measured their half-lives upon reaction with potassium osmate and a modified bipyridine ligand. It was found that T was most reactive with a half-life of 2.2 minutes, then 5mC (11.0 minutes) and finally C (365 minutes).[159] However, thymine can be modified with complete selectivity if a bipyridine-tethered oligonucleotide is used.[156] By tethering the ligand to a short complementary strand of DNA it was possible to selectively

oxidise a single thymine that was located opposite to the tethered ligand, without modifying any other base in the strand (including other T bases). One disadvantage of this method was that the osmium reagents did not react uniformly with all thymines, as two consecutive thymines in a strand of DNA were found to oxidise more quickly than an isolated thymine.[157]

1.3.4 – Cytosine

Identification of cytosine residues for the purpose of DNA sequencing was first examined by Maxam and Gilbert in 1977.[1] They utilised a reaction of both cytosine and thymine with hydrazine to cause strand cleavage at these residues. If sodium chloride was added to the reaction solution, suppression of hydrazine's reaction with T occurred and C locations could be determined directly. However, the use of a high salt concentration had some distinct disadvantages. When the DNA was isolated by ethanol precipitation it was difficult to remove all of the salt, which resulted in anomalous electrophoretic mobility and blurriness of the sequencing gel, which made analysis of the sequence more difficult.[153] The hydrazine reagent was also known to degrade over time causing incomplete suppression of the modification of T even in high salt conditions. To overcome these issues Rubin and co-workers developed the selective modification of C using hydrazine acetate. This alternative source of hydrazine still exhibited excellent C cleavage, with only a faint T background. The reagent itself was also much more stable and more easily removed than hydrazine.[153] An additional method of cytosine modification, also developed by Rubin, was the selective reaction of hydroxylamine hydrochloride with C residues at pH 6. This reagent was known to add across the C5-C6 double bond, thus labilising the ring to internal rearrangement and strand cleavage by piperidine (Figure 1 – 24).[153] Similarly, methoxyamine has also been employed for the selective modification of cytosine bases.[160]

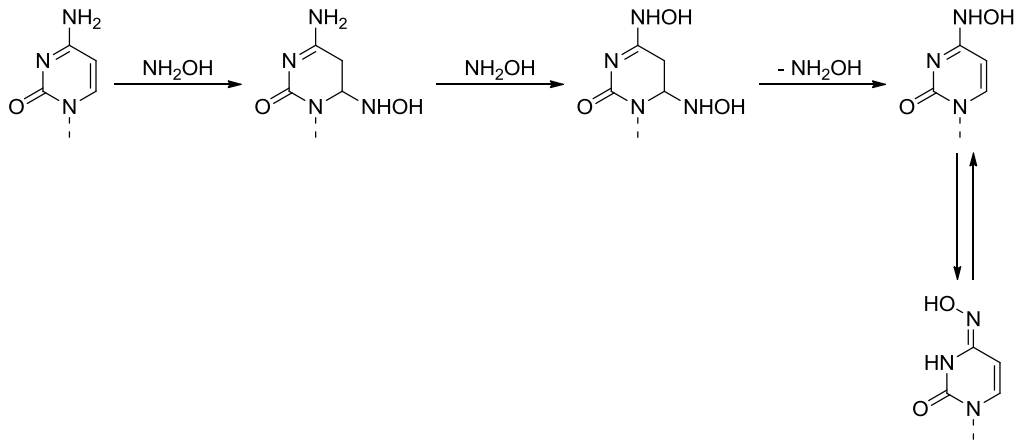


Figure 1 – 24 – *The proposed mechanism of reaction of cytosine with hydroxylamine.*

The most common method for selective modification of cytosine is by using bisulfite.[161] As previously discussed, this reaction converts cytosine residues to uracil *via* the mechanism shown in Figure 1 – 13. This reaction is also known to cause conversion of 5mC to thymine and form cytosine-5-methylenesulfonate from reaction with hmC. Furthermore, the reaction with 5mC is much slower than with C so these bases can be clearly distinguished by their difference in reactivity. The CMS adduct does not undergo deamination to form thymine, therefore, C can be distinguished from hmC by this method. It is, however, not possible to distinguish 5mC and hmC from each other by this approach.[120] Despite the fact that cytosine sometimes suffers from incomplete conversion to uracil, this is still the most widely used method for its selective chemical modification. A modification of the standard bisulfite sequencing, using semi-carbazide (1.0 M) and bisulfite (0.5 M), has also shown high levels of selectivity for cytosine modification in both DNA and RNA (Figure 1 – 25).[162, 163]

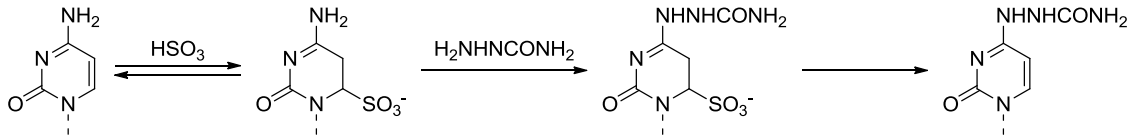


Figure 1 – 25 – *The reaction mechanism of cytosine with semicarbazide in the presence of bisulfite.*[162, 163]

1.3.5 – 5-Methylcytosine

The main types of chemical reactions which have successfully been employed to target 5-methylcytosine are oxidation reactions which dihydroxylate the C5-C6 double bond. However, with many of the conditions tested there has been competing reactivity with thymine, which is also prone to oxidation at this position (see Chapter 1, section 1.3.3). Initial attempts in 1987 to oxidise 5mC preferentially with potassium permanganate resulted in dihydroxylation of both 5mC and thymine.[164] This meant that Fritzsche and co-workers had established a chemical reaction that could successfully discriminate C from 5mC but not 5mC from T.

Okamoto *et al.* have carried out extensive investigations into the use of osmium for the selective dihydroxylation of 5mC.[165] Under the conditions developed by Okamoto for the formation of osmate esters, a chemical reaction was observed with both 5mC and T. However, a number of techniques have been optimised to overcome this selectivity issue. In 2005, Okamoto first reported that by using potassium osmate (a much less intractable source of osmium), combined with a re-oxidant potassium ferricyanide and bipyridine as a rate-accelerating ligand it was possible to form an osmate ester of 5mC.[166] When ssDNA containing a single 5mC was treated with the osmium mix mentioned previously, heated at 90 °C in piperidine (to cause strand cleavage at sites of oxidative damage) and then analysed using PAGE it was possible to observe strand cleavage at the location corresponding to 5mC (Figure 1 – 26, lane 2).[167] When a similar reaction was carried out on a strand containing a single C in the middle, negligible strand cleavage was observed. Cleavage of the DNA was completely inhibited when dsDNA was tested under the same conditions. It is thought that no reaction could have occurred in dsDNA because the π-orbital of the C5-C6 double bond is protected from attack by the osmium, by the stacking of the DNA duplex structure. However, if the strand employed a mismatched base

pair opposite the site of methylation, creating a bulged DNA duplex, then sequence-selective 5mC oxidation still occurred (Figure 1 – 26, lane 6). The rate constants for the dihydroxylation at the C and 5mC bulged DNA duplexes were calculated to be 2.51×10^{-5} and 1.11×10^{-2} s^{-1} respectively.[167] This illustrated the strong electron donating effect of the methyl group which accelerated the reaction of 5mC to give the desired selectivity.

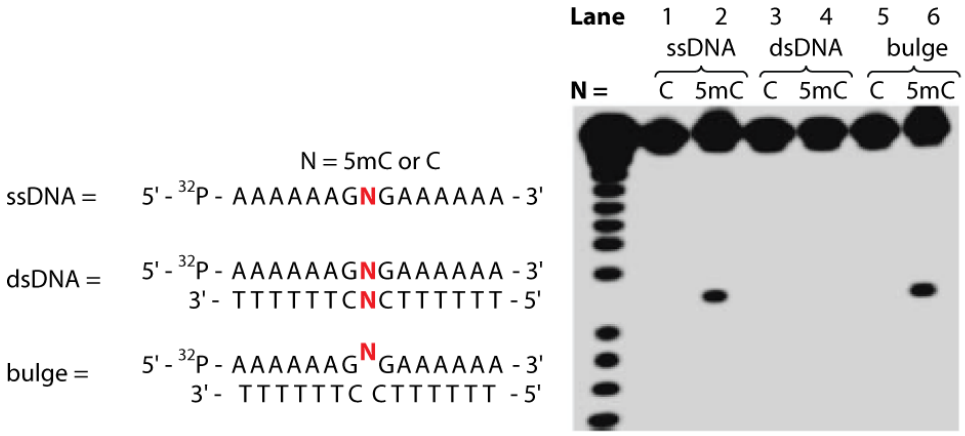


Figure 1 – 26 – *Sequences of the ^{32}P-labeled DNA that have been subjected to oxidation with osmium and then cleavage with hot piperidine. The polyacrylamide gel shows the analysis of strand cleavage. Oxidative cleavage was observed for ssDNA and bulge DNA that contained 5mC but not for dsDNA. No detectable cleavage bands were observed for C containing DNA. This figure has been adapted from Okamoto et al. 2006.*[167]

To try and improve the ease with which 5mC can be detected, Okamoto and co-workers developed methods of incorporating both fluorescent[159, 168] (Figure 1 – 27) and electrochemical[159, 169] (Figure 1 – 28) signal-sending units. These methods allowed the detection of 5mC directly on the strand, without the need for time-consuming cleavage assays, and may be further optimised for the high-throughput detection of methylation.

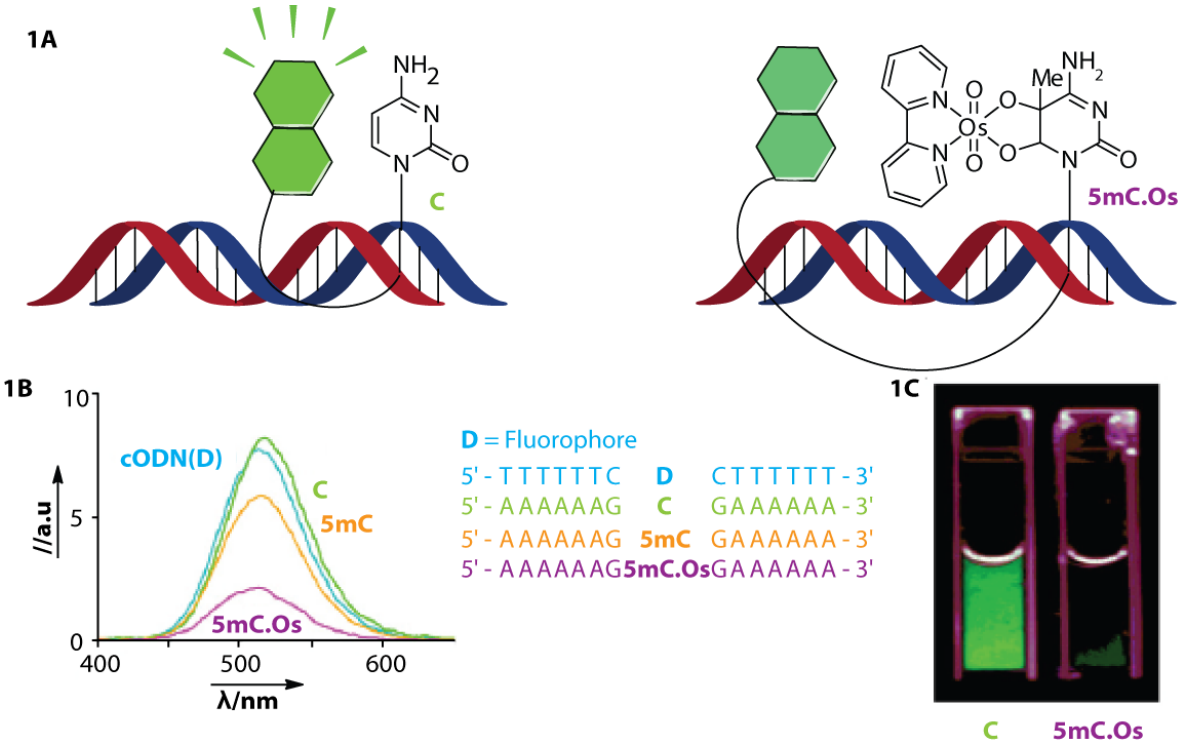


Figure 1 – 27 *– Illustration of a fluorescent signal-sending unit which can be incorporated into a complementary oligodeoxyribonucleotide strand (ODN) for the detection of 5mC. 1A) The left-hand-side highlights how cytosine does not form an osmate ester and therefore does not quench the fluorescent tag attached to the complementary strand, the right-hand-side shows how the 5mC osmate ester complex quenches the fluorescence of the attached tag. 1B) The fluorescence spectra for the ssDNA fragments upon hybridisation to the ssDNA containing the fluorescent tag. The blue line illustrates the spectra produced by the unhybridised fluorescent tag containing ssDNA (cODN(D)). The green line corresponds to cODN(D) hybridised to the DNA strand containing C where no fluorescent quenching was observed. The orange line represents when cODN(D) hybridises to the DNA strand containing 5mC, again no fluorescent quenching is detected. The purple line illustrates when cODN(D) hybridises to the DNA strand containing the 5mC osmate ester complex. Significant quenching of the fluorescence signal produced by the fluorophore in the cODN(D) tag is detected. 1C) A fluorescent image of the cODN(D) hybridised to the DNA strand containing C (left) and the cODN(D) hybridised to the DNA strand containing the 5mC osmate ester complex (right). Figure 1A-1C have been adapted from Tanaka et al. Bioorganic and Medicinal Chemistry, 2007.*[168]

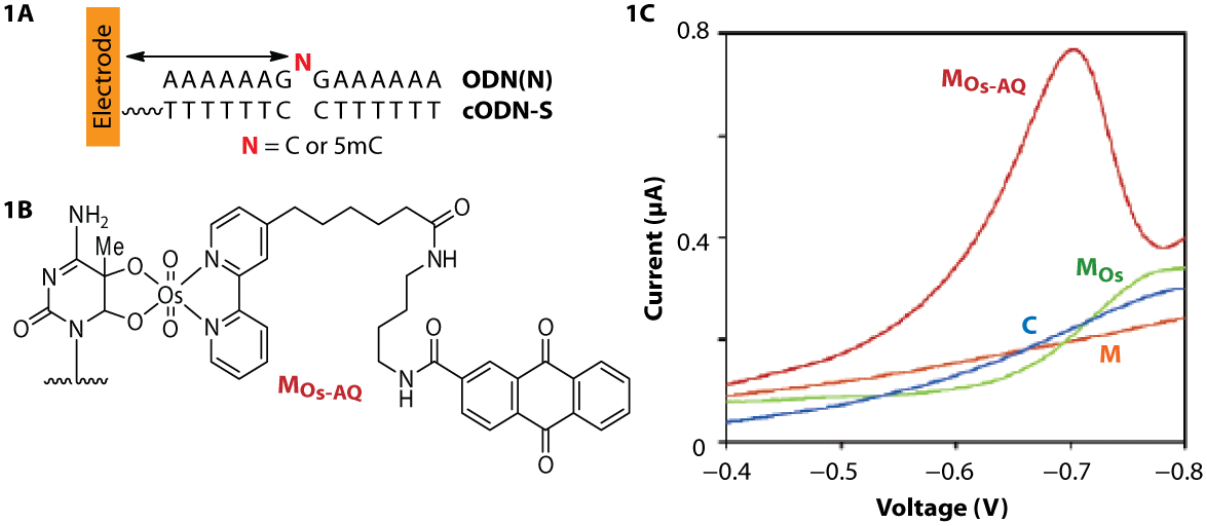


Figure 1 – 28 – *Illustration of an electrochemical signal-sending unit which can be incorporated into the bipyridine ligand for the detection of 5mC. 1A) Illustration of how a short complementary strand of DNA (cODN-S) is used to hybridise and immobilise a strand containing either C or 5mC ODN(N) onto an electrode. 1B) The structure of the osmate ester anthraquinone complex that is formed by 5mC ($M_{Os.AQ}$). 1C) The square wave voltammetry signal that is produced by the DNA strand containing C (blue line), the DNA strand containing un-modified 5mC (orange line), the DNA strand containing the osmate ester complex of 5mC without the anthraquinone derivative attached (green line), all of which do not produce an electrochemical signal. The DNA strand containing the 5mC osmate ester anthraquinone complex (red line), however, exhibits a strong electrochemical response. Figure 2A and 2B have been adapted from Tanaka et al. Journal of the American Chemical Society, 2007.*[159]

Okamoto's work up to this point had been on model strands of DNA which did not contain any competing thymines. In order to use this methodology on native DNA, which had all bases present, a number of different ways were developed to modify the rate-accelerating ligand, so that an osmate ester could only be formed at 5mC (Figure 1 – 28). One approach employed a tethered bipyridine ligand which was covalently attached to an adenine incorporated in a short oligonucleotide (known as the ICON probe).[170, 171] This tether meant the osmate ester could have only been formed at the base opposite to the modified adenosine in its complementary strand. This methodology has been used to quantify the methylation status of a tissue-specific differentially methylated region of the mouse genome in four different tissue types – kidney, liver, spleen and testis.[170] One disadvantage of this procedure was that thymine bases adjacent to the probed 5mC also exhibited

oxidation using this tethered approach.[172] Recent work by Okamoto has shown that hmC can also be oxidised to form an osmate ester using an ICON probe,[173] therefore, further work is needed before this approach can be utilised for the 100% selective modification and detection of 5mC. '

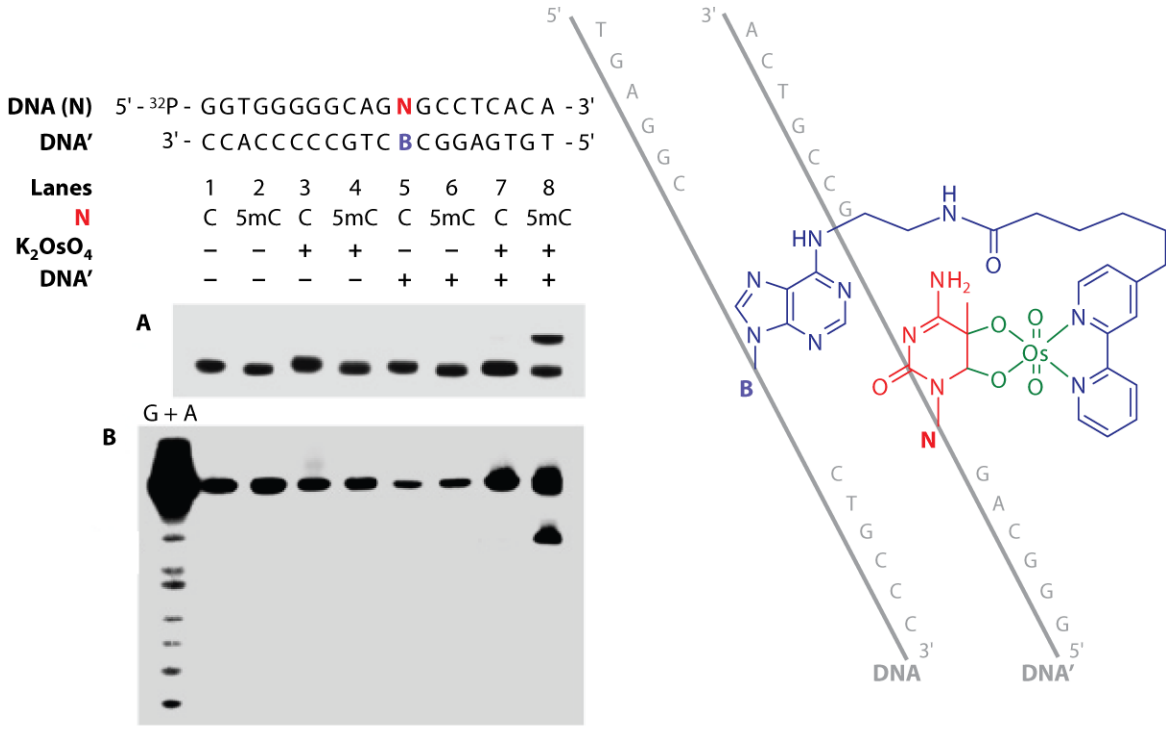


Figure 1 –28 – *Illustration of how the ICON probe (pale blue) is used for the selective detection of 5mC (red) in ssDNA, by the formation of an osmate ester inter-strand linkage (green, blue and red). A) PAGE analysis of the formation of the osmate ester inter-strand linkage. A new band appears on the gel, that has lower mobility and higher molecular weight, that corresponds to the formation of an inter-strand link between DNA' strand and the DNA (N) strand where N = 5mC. B) Illustrates the polyacrylamide gel, after treatment of the modified DNA with hot piperidine – strand cleavage indicates the formation of the inter-strand link between the modified adenine and the 5mC residue. This figure has been adapted from Tanaka et al. 2007.*[170]

In 2008, Carell and co-workers published a selective chemical reaction for the modification and detection of 5mC in DNA.[174] As 5mC is known to have a slightly lower redox potential than C and T this property was used to attempt to selectively target 5mC. Initially, a vanadium(V) species was employed to oxidise the C5-C6 double bond because the redox potential of the metal reagent could be controlled by altering cluster size and composition.[175] A short strand of DNA was reacted with a variety of vanadium sources,

treated with hot piperidine and then analysed by polyacrylamide gel electrophoresis. Cleavage was observed with 5mC but not with either T or C. However, the vanadium species also carried out a direct electron abstraction of guanine leading to oxidative cleavage at these positions in the strand. To try and prevent the direct electron abstraction of G, and increase the efficiency of the electrophilic attack of 5mC, Carell added lithium bromide to the reaction. This was observed by Emmanuvel *et al.* to accelerate the dihydroxylation reaction.[176] This was found to increase the selectivity for 5mC four-fold. After additional optimisation, conditions were found that were selective for 5mC – $NaIO_4$ (1 mM), LiBr (4 mM), Na_3PO_4 buffer (100 mM), pH 5, 40 °C 1 h. Analysis of the DNA strand, using MALDI-TOF mass spectrometry, showed a signal for the reaction product which corresponded to the bromo-hydroxylated C5-C6 bond of 5mC.[174] This method was used to identify two CpG methylation sites in a 40 base pair long fragment of the P16 promoter gene, which is a regulator of the invasive proliferation of basal cell carcinoma cells (Figure 1 – 29).[174]

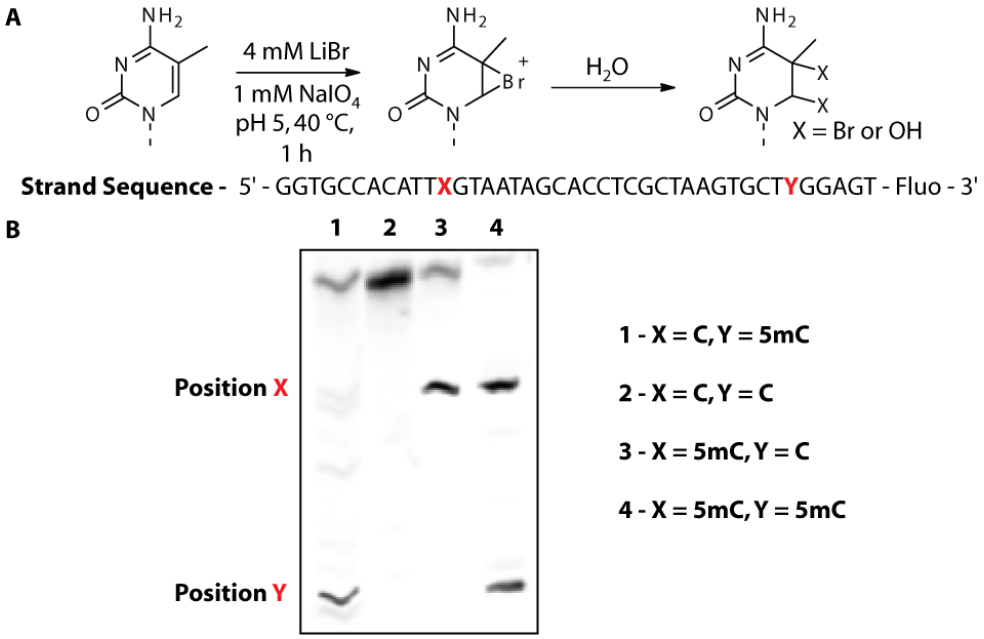


Figure 1 – 29 – *Selective modification of 5mC by sodium periodate and lithium bromide. A) The reaction intermediates for the selective modification of 5mC by sodium periodate and lithium bromide. B) PAGE analysis of a small section of the P16 promoter gene (sequence shown) exhibiting strand cleavage at two sites of DNA methylation. This figure has been adapted from Bareyt et al. 2008.*[174]

Nishimoto and co-workers have used a one-electron photooxidation protocol to selectively react and cause oxidative strand cleavage at 5mC residues.[177] A strand containing a 2-methyl-1,4-napthaquinone (NQ) chromophore was hybridised to its complementary DNA strand forming a duplex, where the NQ chromophore was opposite the methylation site being probed. This complex was irradiated with light and the photoexcitation of the NQ chromophore allowed one electron oxidation of 5mC to its radical cation intermediate, which underwent subsequent deprotonation and addition of oxygen resulting in strand cleavage at 5mC (Figure 1 – 30). Replacement of the 5mC residue with C, T or A exhibited no photooxidation. Slight photooxidation was observed with G but this was significantly reduced by the close proximity of the NQ chromophore which caused rapid charge recombination, suppressing strand cleavage. This procedure has been further modified for the detection of this chemical reaction by fluorescence.[178] This method of detection was a significant improvement on the cleavage assay as it enabled 5mC's recognition down to the sub-femtomole level.

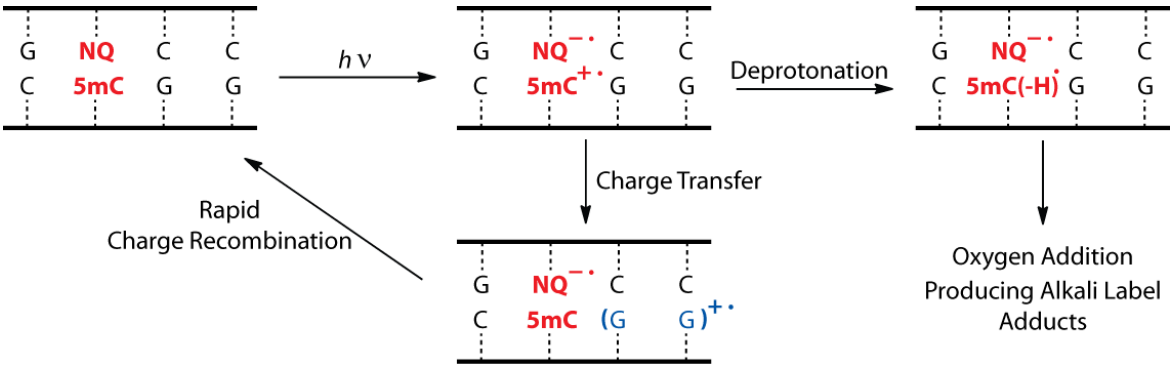


***Figure 1 – 30** – A proposed photochemical reaction mechanism for exclusive DNA strand cleavage at 5mC using an NQ chromophore. The dsDNA was irradiated with light oxidising 5mC to its radical cation intermediate which subsequently underwent deprotonation and oxidation. This produced a 5mC adduct which was susceptible to alkali strand cleavage by piperidine. When the radical cation was formed at 5mC, it was possible for it to transfer to the neighbouring guanine bases, producing G radical cations (blue). These moieties do not, however, undergo strand cleavage because the close proximity of the NQ chromophore facilitates rapid charge recombination. This figure has been adapted from Tanabe et al. 2007.*[177]

1.3.6 – 5-Hydroxymethylcytosine

There has been a great deal of interest in methods to chemical modify hmC in order to aid its detection *in vivo*. These can be split into two major strategies: 1) targeting of the product of the bisulfite modification of hmC – cytosine-5-methylenesulfonate (CMS) and 2) the use of enzymes for the incorporation of various functional groups at the free hydroxyl of hmC.

As discussed previously (see Chapter 1, section 1.2.2), when hmC is treated under the bisufite modification conditions it produces CMS. This adduct is very stable and deaminates to thymine very slowly.[129] CMS can be targeted by an anti-CMS antibody and this can be used to detect the location of hmC.[117, 122] This procedure involved initial bisulfite modification of the DNA strand, treatment with the anti-CMS antibody, PCR amplification and then sequencing. This procedure has been used for genome wide mapping of hmC[123] and for investigating hmC's link to myeloid cancers.[122]

It is possible to use an enzyme, β-glucosyltransferase, to covalently attach a glucose moiety to the free hydroxyl of hmC. At present, there are no other free hydroxyl's known to be found in DNA, therefore, this modification would be entirely selective for hmC. A number of researchers have investigated a variety of ways of detecting the attached sugar directly[179] or by modifying it in some way to make its detection easier.[123, 127] For direct detection of the sugar, when it has been attached to hmC, Robertson *et al.* used J-binding protein 1 which was coupled to magnetic beads.[179] This protein bound glycosylated hmC, allowing its detection on either a single gene, by real-time PCR, or for genome-wide analysis, by microarrays or sequencing techniques. In order to attach additional functionality, to aid the detection of hmC, a chemically modified sugar can be directly incorporated onto the free hydroxyl by β-glucosyltransferase. This was the approach employed by Song and co-workers, where a modified sugar (6-N_3-glucose) containing an

additional azide group was incorporated and then further functionalised with biotin using click chemistry.[127] Biotin was chosen in this case because it can be used for high-affinity capture and/or enrichment of hmC-containing DNA allowing for sensitive detection and deep sequencing to reveal the location of hmC residues. Furthermore, a variety of different functionalities could be incorporated using click-chemistry making this an extremely versatile approach.[127] An additional hmC-detection strategy was developed by Pastor, where an unmodified glucose moiety was attached to hmC and further derivatised when connected to the DNA strand. After attachment of the glucose moiety its vicinal hydroxyls were oxidised to aldehydes and then reacted with an aldehyde-reactive probe, which incorporated two biotin molecules for every hmC. This technique was also employed towards the genome-wide mapping of hmC. [123]

A recent publication, by Klimašauskas, demonstrated how 5-methyltransferases (C5-MTases) could be used to add exogenous aliphatic aldehydes (e.g. formaldehyde) to cytosine.[180] This prompted further investigation to see if C5-MTases could be used to incorporate additional functional groups onto other bases. It was found that when DNA duplexes, containing hmC, were treated with C5-MTases in the presence of a millimolar concentration of exogenous thiol (such as 2-mercaptoethanol, cysteamine and L-cysteine) a new product was observed.[181] ESI-MS and UV analysis showed that the thiol group had replaced the hydroxyl of hmC and was directly attached to the nucleoside. Similar reactivity was observed with selenols but less reagent was required for the modification to occur. The mechanism proposed by Klimašauskas and co-workers is detailed in Figure 1 − 31.[181] Due to the variety of co-factors which can be incorporated by the C5-MTases under mild and selective conditions, this approach could be widely used for the base-specific labelling and derivatization of DNA.

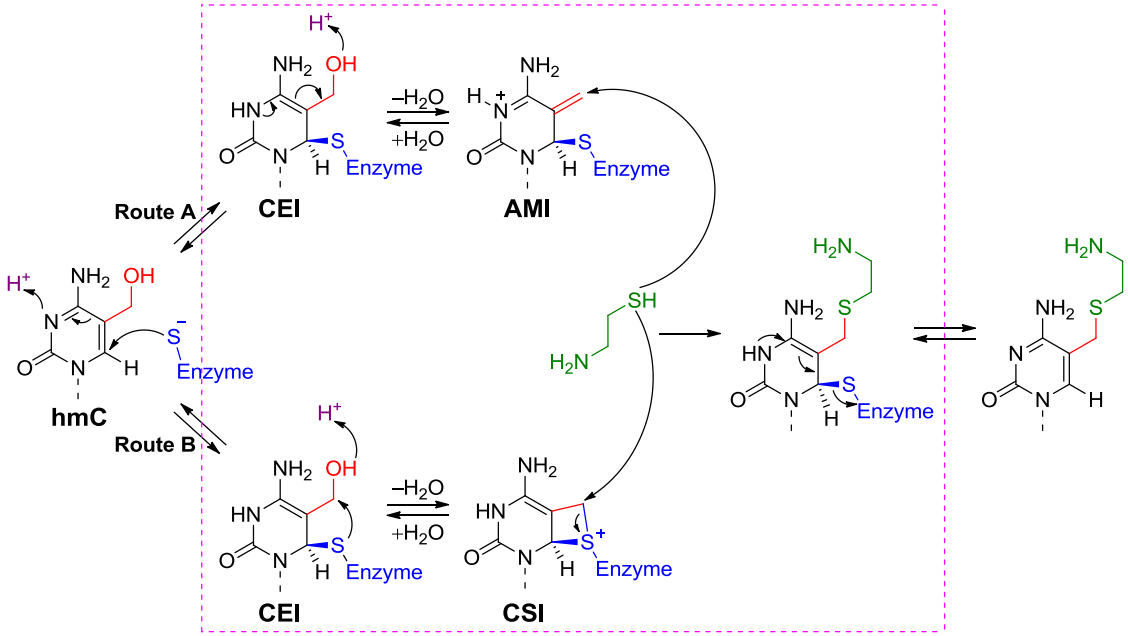


Figure 1 – 31 – *The proposed mechanism for the nucleophilic condensation of cysteamine onto hmC by C5-MTase. The reaction was thought to proceed by either route A or B. Both routes involved the initial formation of a covalent enamine intermediate (CEI), by attack of the C6 bound sulfur atom of a cysteine residue in the C5-MTase active site. Route A then proceeded by an acid-assisted dehydration of the 5-hydroxymethyl functionality giving a highly active 5-methylene intermediate (AMI). Route B employed an intramolecular attack onto the protonated 5-hydroxymethyl group, by the enzyme-borne C6-bound sulfur atom, producing the bicyclic sulfonium intermediate (CSI). Addition/attack of the cysteamine nucleophile provided the product (attack of water would have promoted the reverse reaction returning hmC). This figure has been adapted from Liutkevičiūtė et al. 2011.*[181]

The information highlighted in this introduction reviews the achievements in DNA sequencing, to identify the canonical bases (C, T, A and G) as well as the epigenetically modified bases 5mC and hmC, by both chemical and other means to date. The work portrayed in this thesis initially describes the use of nanopore techniques to identify all six bases (Chapter 2). It then moves on to discuss possible methods of selective chemical modification and detection of the modified bases (Chapters 3-6).

Chapter 2

Six-Nucleobase Discrimination by the α-Hemolysin Protein Nanopore

Chapter 2 – Six-Nucleobase Discrimination by the α-Hemolysin Protein Nanopore

Declaration

The work in this chapter is also described in the manuscript:

Wallace, E. V. B., Stoddart D., Heron, A. J., Mikhailova, E., Maglia, G., Donohoe, T. J., and Bayley, H., Identification of epigenetic DNA modifications with a protein nanopore, *Chemical Communications* **2010,** *46* (43), 8195-8197.

2.1 – Introduction

In 1999, Deamer *et al.* published a paper detailing how it was possible to distinguish RNA homopolymers (poly(A) and poly(C)) from each other by comparing the percentage channel block produced by each oligonucleotide, as it translocated through the pore.[182] Subsequently, work by Branton and co-workers demonstrated that similar discrimination was achievable between DNA homopolymers.[48] Deamer and co-workers were also able to distinguish the transition between poly(A) and poly(C) regions of a single RNA molecule of poly$A_{30}C_{70}$.[182] However, as poly(C) formed a helical structure small enough to translocate through the pore, whereas, poly(A)'s helix required unwinding before translocation was possible, this observable transition was attributed to differences in the secondary structure of poly(A) compared to poly(C) not the structure of the individual nucleobases themselves,.[49, 182] Deamer and Branton were unable to obtain single nucleotide discrimination because both DNA and RNA strands translocated too fast through the pore for accurate detection (3.3 μs/base for poly$(dA)_{100}$ and 1.2 μs/base for poly$(dC)_{100}$).[48] It is also important to consider that, owing to the current limitation of patch clamp amplifiers, it is not possible to resolve individual nucleotides travelling at these

rates because it pushes the required detector bandwidth into the MHz region which, therefore, precludes the measurement of pico-ampere steps in ion current.[25, 183]

A number of environmental factors have been investigated for their ability to slow the translocation rate and increase the residence time of DNA inside nanopores e.g. viscosity and temperature.[48, 84] Another approach for overcoming the limitations of single nucleobase discrimination has been the employment of enzymes, such as DNA polymerases[53, 54] and exonucleases, for the control of DNA translocation. Work carried out in the groups of Ghadiri and Akeson has shown that DNA polymerases, coupled with αHL, can be used to control the incorporation of single nucleoside monophosphates into growing DNA strands.[53, 54] Oxford Nanopore Technologies have also demonstrated that single nucleoside monophosphates, of C, T, A, G and the epigenetic modification 5mC, can be distinguished from each other, under asymmetric salt conditions, by using an engineered αHL mutant with a covalently attached cyclodextrin adaptamer.[63] This would mean that if a DNA exonuclease could be effectively employed, in tandem with αHL, then it would be possible to distinguish each DNA nucleobase by exonuclease-sequencing strategies. However, a great deal of optimisation would still be required before either technique could be scaled up for direct DNA sequencing.

Several groups have employed immobilisation techniques such as 5' or 3'-terminal hairpin,[50-52, 64] a short complementary dsDNA region attached to the 5' or 3' end of a ssDNA oligonucleotide[52-55] or a biotin•streptavidin complex,[52, 53, 55-61, 65] in an effort to establish whether αHL can discriminate between individual DNA nucleobases. The main advantage of immobilising DNA in this way, is that it dramatically increases the residence time of each nucleobase in the pore (e.g. >100 ms), therefore, increasing the signal to noise ratio.[50, 58, 59] Immobilisation of DNA in the αHL pore also provided a model system which

mimicked the pauses between enzymatic nucleobase additions in the two potential sequencing strategies described previously.

Bayley and co-workers have successfully utilised the biotin•streptavidin complex for immobilisation of DNA in nucleobase discrimination experiments.[59-61] Stoddart *et al.* investigated the nucleobase discriminating abilities of the E111N/K147N (NN) mutant of αHL where the Glu-111 and Lys-147 residues at the constriction site were mutated to asparagines. These mutations were made in the hope that they would result in increased current flow through the pore, when DNA was present, subsequently producing greater dispersion of residual current values. By sequentially moving a single adenine residue (in an otherwise poly(dC)$_{40}$ background) from positions 7-21 Stoddart *et al.* established that both the NN and the wild-type (WT) αHL pores have three distinct recognition points – R_1, R_2 and R_3 (Figure 2 – 1A).[59] Recognition sites R_2 and R_3 were capable of discriminating C, T A and G (in a poly(dC)$_{40}$ homopolymeric background) in the NN mutant but only R_2 was able to fully distinguish the nucleobases in the WT.[59] The NN mutant was also shown to fully discriminate the canonical nucleobases in a heteropolymeric DNA strand sequence (Figure 2 – 1B).

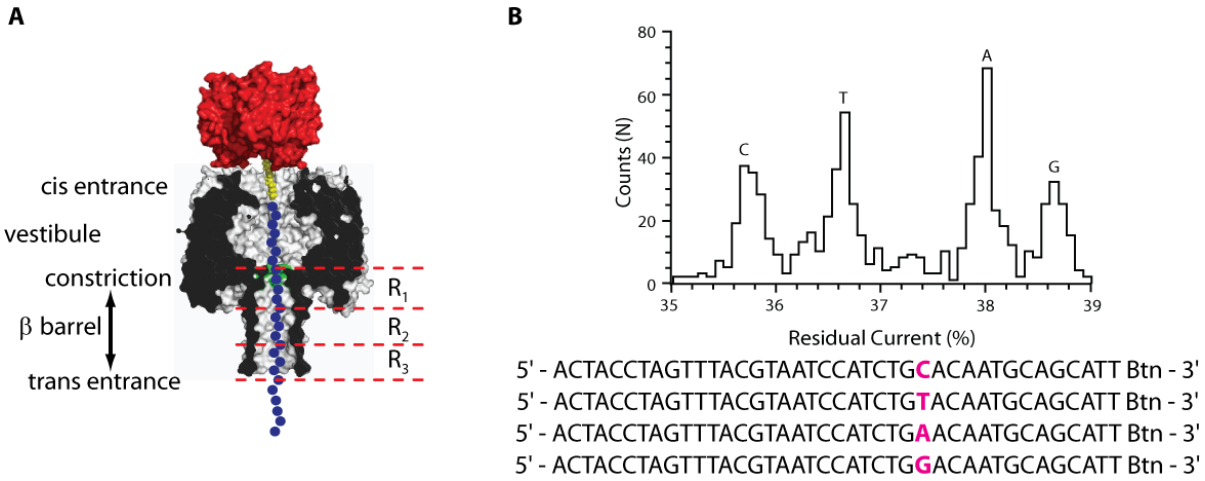


Figure 2 – 1 *– Immobilisation of ssDNA in the NN mutant of αHL illustrating nucleobase discrimination at R_2 when probed with a heteropolymeric strand sequence. A) ssDNA (blue circles) was immobilised in the αHL pore by the use of a 3' biotin (yellow circles)•streptavidin (red) complex. The location of the three recognition points R_1 (which recognises positions 8-11 relative to the 3' biotin-TEG), R_2 (which recognises positions*

13-15 relative to the 3' biotin-TEG) and R_3 (which recognises positions 16-20 relative to the 3' biotin-TEG) are indicated by red dotted lines. Figure adapted from Stoddart et al., 2010.[61] B) Histogram of the residual current levels for the E111N/K147N mutant of αHL probed with four heteropolymeric oligonucleotides that differ at only one position in their DNA sequence (pink bold). Figure adapted from Stoddart et al., 2009.[59]

To further improve αHL's nucleobase recognition it was necessary to preferentially enhance nucleobase discrimination at one recognition site and remove all others.[60] Site-directed mutagenesis has proven to be an effective strategy for altering nucleobase recognition.[59-61] The residues 113, 117 and 147 comprise the amino acids located at the central constriction of αHL. To investigate how the chemical properties of the amino acids effect recognition, 16 different E111N/K147N/M113X (NNX) mutant homoheptameric pores were probed with poly(dC)$_{40}$ oligonucleotides which had single nucleobase substitutions at recognition point R_1 (position 9 relative to the 3' biotin-TEG tag).[60] There were six different mutants (NNV, NNL, NNI, NNR, NNW and NNY) that could fully discriminate C, T, A and G. The αHL mutants which had the methionine residue at position 113 altered to a hydrophobic amino acid (valine, leucine and isoleucine), that were also capable of discriminating all four nucleobases, had bulky side-chains which could potentially provide a steric barrier to the flow of ions and interact with the DNA nucleobases by the formation of van der Waals bonds. The mutants containing aromatic amino acids at position 113 (tyrosine and tryptophan), capable of four nucleobase discrimination, also had bulky side groups but were thought to interact with the nucleobases primarily by hydrogen bonding and π-stacking interactions. Finally, it was proposed that the mutant containing a charged arginine residues at 113 interacted with DNA *via* hydrogen bonding and cation-π interactions. Moreover, by mutation of the 113 amino acid to a glycine, a residue which has no side-chain, it was possible to completely remove all recognition at R_1. Although the exact methods by which nucleobase discrimination occurred was not fully understood, this work proved that it was possible to

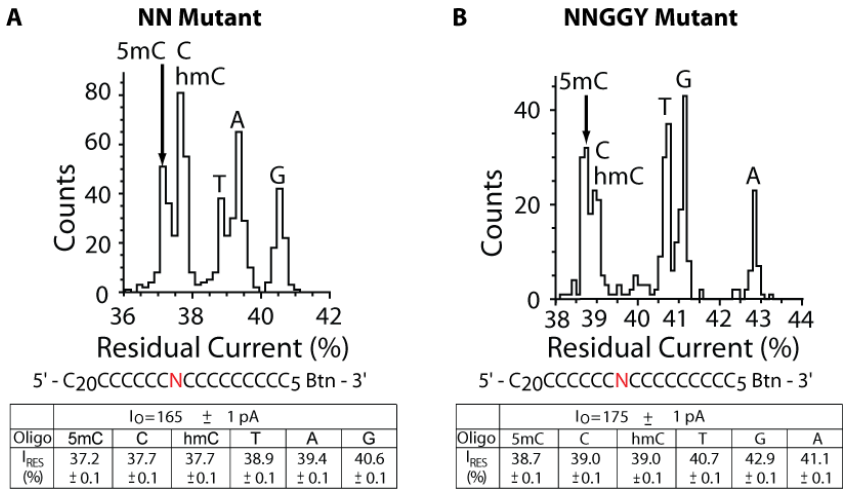
both enhance and eliminate recognition by site-directed mutagenesis.[60] The work that has so far been discussed has concentrated on discrimination of the canonical nucleobases C, T, A and G. The following sections describe how the biotin•streptavidin immobilisation approach was used to investigate whether it was possible to discriminate the epigenetic modifications 5-methylcytosoine and 5-hydroxymethylcytosine, as well as the standard nucleobases, by site-directed mutagenesis of αHL.

2.2 – Results and Discussion

2.2.1 – Probing Recognition Site R_2 with Poly(dC)$N_{(14)}$ Oligonucleotides for Six-Nucleobase Discrimination

When the recognition points of the NN mutant were first probed for four-nucleobase discrimination, by Stoddart *et al.*, R_2 produced the greatest overall spread of block levels ($\Delta I_{RES}^{OVERALL}$ = 2.7 ± 0.1%, $\Delta I_{RES}^{OVERALL}$ is defined as the difference between the two most widely dispersed I_{RES} block levels) of all three recognition sites (R_1 did not show full discrimination and R_3 $\Delta I_{RES}^{OVERALL}$ = 1.2 ± 0.0%).[59] It was for this reason that this recognition site, in the NN mutant, was initially employed for the identification of the additional epigenetically modified nucleobases (5mC and hmC). Homopolymeric ssDNA, with single-nucleobase substitutions at position 14 relative to the 3' biotin-TEG (poly(dC)$N_{(14)}$ where N = C, T, A, G, 5mC or hmC), were incubated at room temperature with streptavidin and sequentially added to the *cis* compartment (see Chapter 7, section 7.3.4). By utilising the episodic simulation mode in the pClamp software (version 10.1, Molecular Devices), the Digidata 1440A digitizer (Molecular devices) was used to apply repeated voltage steps that drove the single stranded DNA-biotin•streptavidin complexes into the pore (applied potential of +160 mV in the *trans* side relative to the *cis*) and then expelled them (−140 mV in the *trans* side relative to the *cis*). The capture of the DNA-biotin•streptavidin, by the pore, resulted in a step-wise decrease in the open pore

current (I_O) to a new block level (I_B). Another parameter that was used to describe the amount of current flowing through the pore, when blocked with DNA, was the I_{RES} which was defined as the residual current flowing through the pore expressed as a percentage of the open pore current ($I_{RES} = (I_B / I_O) \times 100$). Nucleobase discrimination by the NN mutant (protein prepared by E. Mikhailova) was investigated at +160 mV in 1 M KCl, 25 mM Tris, 100 μM EDTA, pH 8.0 (the conditions for all electrical recording experiments conducted unless otherwise stated). Under these conditions, it was possible to discriminate 5mC from C but hmC and C produced identical block levels (Figure 2 − 2). Another mutant the E111N/K147N/M113G/T115G/N123Y (NNGGY, protein prepared by E. Mikhailova), that had been designed to improve discrimination at R_2 and remove it at R_1 (initial experiments carried out by D. Stoddart, un-published results), was probed for six-nucleobase discrimination. This mutant had a larger $\Delta I_{RES}^{OVERALL}$ than the NN mutant ($\Delta I_{RES}^{OVERALL} = 4.1 \pm 0.1\%$ (n = 3) compared to the NN's $\Delta I_{RES}^{OVERALL} = 3.4 \pm 0.2\%$ (n = 4)) but still produced similar nucleobase discrimination between C and its epigenetic modifications 5mC and hmC. In the NNGGY mutant the ΔI_{RES} between C and 5mC was slightly smaller (ΔI_{RES} 5mC = −0.2 ± 0.0%) than in the NN mutant (ΔI_{RES} 5mC = −0.5 ± 0.0%), despite its greater $\Delta I_{RES}^{OVERALL}$. Therefore, full discrimination was not obtained using either the NN or the NNGGY mutants.



	I_O=165 ± 1 pA					
Oligo	5mC	C	hmC	T	A	G
I_{RES} (%)	37.2 ± 0.1	37.7 ± 0.1	37.7 ± 0.1	38.9 ± 0.1	39.4 ± 0.1	40.6 ± 0.1

	I_O=175 ± 1 pA					
Oligo	5mC	C	hmC	T	G	A
I_{RES} (%)	38.7 ± 0.1	39.0 ± 0.1	39.0 ± 0.1	40.7 ± 0.1	42.9 ± 0.1	41.1 ± 0.1

Figure 2 – 2 – *Nucleobase discrimination abilities of the NN and NNGGY pores when probed with poly(dC)$N_{(14)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 14 relative to the 3' biotin-TEG.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1	166	37.5	37.5	0.0	37.0	−0.5	38.7	1.2	39.1	1.7	40.2	2.8	3.3
2#	165	37.7	37.7	0.0	37.2	−0.5	38.9	1.2	39.4	1.7	40.6	2.9	3.4
3	169	36.7	36.7	0.0	36.2	−0.5	38.0	1.3	38.4	1.7	39.9	3.2	3.6
4	162	37.7	37.7	0.0	37.3	−0.5	39.0	1.2	39.4	1.7	40.6	2.8	3.3
Mean	166	37.4	37.4	0.0	36.9	−0.5	38.6	1.2	39.1	1.7	40.3	2.9	3.4
SD	3	0.5	0.5	0.0	0.5	0.0	0.4	0.1	0.5	0.0	0.3	0.2	0.2

Table 2 – 2A – *Residual current data for the NN mutant of αHL probed with poly(dC)$N_{(14)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 14 relative to the 3' biotin-TEG (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$_{40}$ and the poly(dC)-T$_{(14)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The $\Delta I_{RES}^{OVERALL}$ is the difference in residual current levels between the two most distant residual current levels (in this case $\Delta I_{RES}^{OVERALL}$ = I_{RES} G – I_{RES} 5mC). The # indicates the experiment which corresponds to the histogram in Figure 2 – 2A for the NN mutant.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1	171	39.1	39.1	0.0	38.9	−0.2	40.9	1.8	43.0	3.9	41.4	2.3	4.1
2#	175	39.0	39.0	0.0	38.7	−0.3	40.7	1.8	42.9	3.9	41.1	2.2	4.2
3	173	38.7	38.7	0.0	38.4	−0.3	40.5	1.8	42.4	3.7	40.9	2.3	4.0
Mean	173	38.9	38.9	0.0	38.7	−0.2	40.7	1.8	42.7	3.8	41.2	2.3	4.1
SD	2	0.2	0.2	0.0	0.2	0.0	0.2	0.0	0.3	0.1	0.2	0.1	0.1

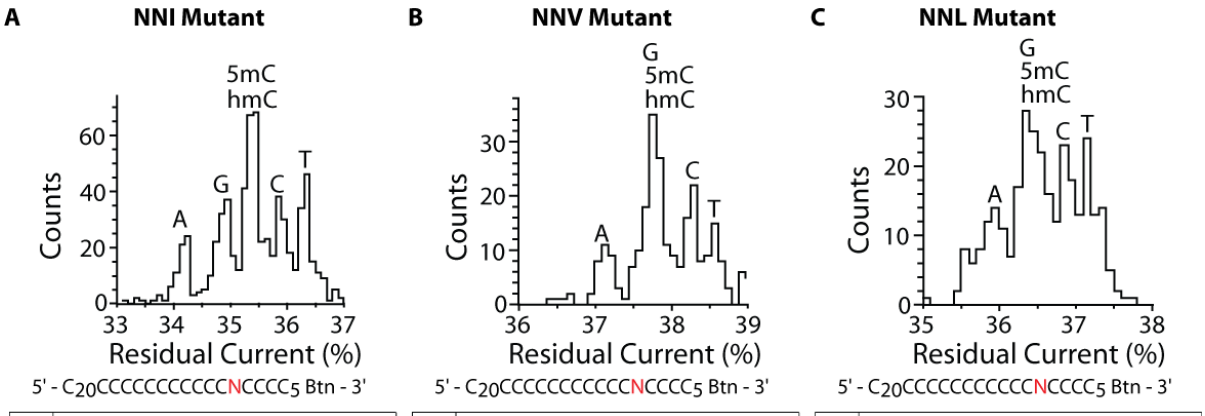
Table 2 – 2B – *Residual current data for the NNGGY mutant of αHL probed with poly(dC)$N_{(14)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 14 relative to the 3' biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for*

individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$_{40}$ and the poly(dC)-T$_{(14)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 2B for the NNGGY mutant.

2.2.2 – Probing Recognition Site R_1 with Poly(dC)$N_{(9)}$ Oligonucleotides for Six-Nucleobase Discrimination

Stoddart *et al.* had determined that out of the 16 E111N/K147N/M113X mutants investigated, only six could fully discriminate C, T, A and G.[60] The mutants that did discriminate the nucleobases had hydrophobic (NNV, NNL and NNI), basic (NNR) or aromatic (NNW and NNY) amino acids substituted at position 113 in the NNX mutants. All six mutants were probed for discrimination of the six DNA nucleobases.

The hydrophobic mutants NNV, NNL and NNI were initially tested for full discrimination of the nucleobases (NNV, NNL and NNI proteins were prepared by E. Mikhailova) (Figure 2 − 3). All three mutants produced the same order of increasing block level – A > G > hmC/5mC > C > T. However, the $\Delta I_{RES}^{OVERALL}$ for these mutants gradually decreased in the order NNI ($\Delta I_{RES}^{OVERALL}$ = 2.0 ± 0.1% (n = 4)) > NNV ($\Delta I_{RES}^{OVERALL}$ = 1.5 ± 0.1% (n = 3)) > NNL ($\Delta I_{RES}^{OVERALL}$ = 1.3 ± 0.1% (n = 4)), which resulted in poor nucleobase discrimination. NNI exhibited the best DNA nucleobase identification as only the signals from 5mC and hmC overlaid in this case. For NNL and NNV the block level associated with G was also the same as that produced by 5mC and hmC. The nucleobase discrimination of these mutants did not directly correlate with the size of the amino acid mutation made at position 113 (volume of I and L at 113 = 166.7 Å^3, whereas the volume of V at 113 = 140.0 Å^3).[60] Therefore, discrimination of the nucleobases was not purely governed by sterics, hence, the hydrophobic interactions could perhaps have played a key role in identification in these cases.



	I_O = 188 ± 2 pA					
Oligo	A	G	5mC	hmC	C	T
I_{RES} (%)	34.2 ± 0.1	34.9 ± 0.1	35.4 ± 0.1	35.4 ± 0.1	35.9 ± 0.1	36.3 ± 0.1

	I_O = 195 ± 2 pA					
Oligo	A	G	5mC	hmC	C	T
I_{RES} (%)	37.1 ± 0.1	37.8 ± 0.1	37.8 ± 0.1	37.8 ± 0.1	38.3 ± 0.1	38.6 ± 0.1

	I_O = 173 ± 1 pA					
Oligo	A	G	5mC	hmC	C	T
I_{RES} (%)	35.9 ± 0.1	36.4 ± 0.2	36.4 ± 0.2	36.4 ± 0.2	36.8 ± 0.1	37.2 ± 0.2

Figure 2 – 3 – *Nucleobase discrimination abilities of the NNI, NNV and NNL αHL pores when probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1[#]	188	35.9	35.4	−0.5	35.4	−0.5	36.3	0.4	34.2	−1.7	34.9	−1.0	2.1
2	181	32.9	32.4	−0.5	32.4	−0.5	33.3	0.4	31.3	−1.6	31.9	−1.0	2.0
3	188	35.7	35.3	−0.4	35.3	−0.4	35.9	0.2	34.0	−1.7	34.7	−1.0	1.9
4	187	36.1	35.6	−0.5	35.6	−0.5	36.4	0.3	34.3	−1.8	35.1	−1.0	2.1
Mean	186	35.2	34.7	−0.5	34.7	−0.5	35.5	0.3	33.5	−1.7	34.1	−1.0	2.0
SD	3	1.5	1.5	0.0	1.5	0.0	1.5	0.1	1.4	0.1	1.5	0.0	0.1

Table 2 – 3A – *Residual current data for the NNI mutant of αHL probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$_{40}$ and the poly(dC)-T$_{(9)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The [#] indicates the experiment which corresponds to the histogram in Figure 2 – 3A for the NNI mutant.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	189	38.8	38.3	−0.5	38.3	−0.5	39.1	0.3	37.6	−1.2	38.3	−0.5	1.5
2	191	36.2	35.7	−0.5	35.7	−0.5	36.5	0.4	35.0	−1.1	35.7	−0.5	1.5
3[#]	195	38.3	37.8	−0.5	37.8	−0.5	38.6	0.3	37.1	−1.1	37.8	−0.5	1.4
Mean	192	37.7	37.2	−0.5	37.2	−0.5	38.1	0.3	36.6	−1.2	37.2	−0.5	1.5
SD	3	1.4	1.4	0.0	1.4	0.0	1.4	0.0	1.4	0.0	1.4	0.0	0.0

Table 2 – 3B – *Residual current data for the NNV mutant of αHL probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current*

between the poly(dC)$_{40}$ and the poly(dC)-T$_{(9)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 3B for the NNV mutant.

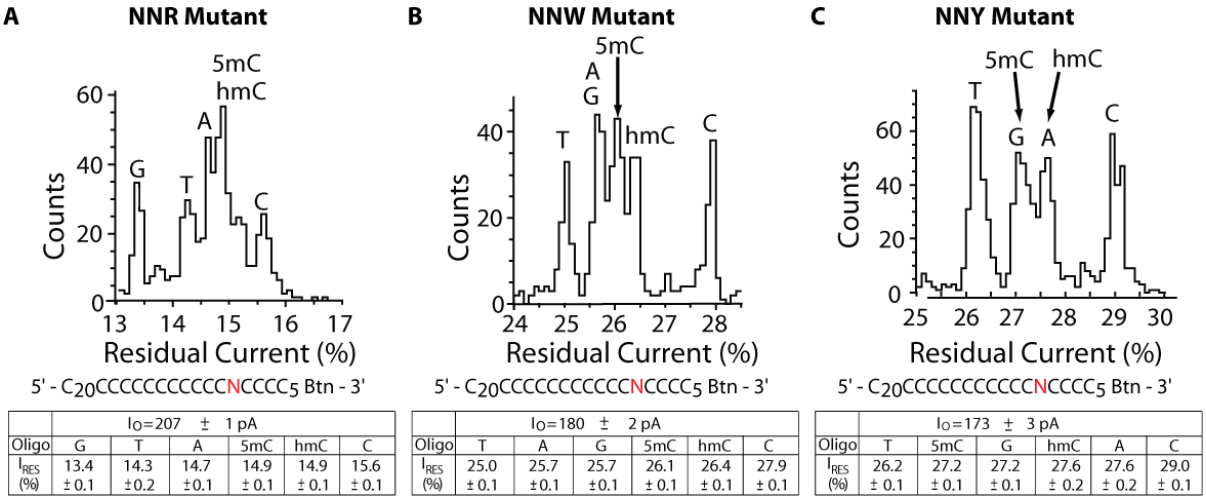
Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	172	34.5	34.0	−0.5	34.0	−0.5	34.9	0.4	33.4	−1.1	34.0	−0.5	1.5
2#	173	36.8	36.4	−0.4	36.4	−0.4	37.2	0.3	35.9	−0.9	36.4	−0.4	1.2
3	178	33.5	33.0	−0.5	33.0	−0.5	33.8	0.3	32.4	−1.1	33.0	−0.5	1.4
4	182	36.0	35.5	−0.5	35.5	−0.5	36.3	0.3	35.0	−0.9	35.5	−0.5	1.2
Mean	176	35.2	34.7	−0.5	34.7	−0.5	35.5	0.3	34.2	−1.0	34.7	−0.5	1.3
SD	5	1.5	1.5	−0.0	1.5	−0.0	1.5	0.1	1.6	−0.1	1.5	−0.0	0.1

Table 2 – 3C – *Residual current data for the NNL mutant of αHL probed with poly(dC)N$_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$_{40}$ and the poly(dC)-T$_{(9)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 3C for the NNL mutant.*

The basic (NNR) and aromatic (NNW and NNY) residue containing mutants of αHL were also probed for six-nucleobase discrimination (NNR, NNW and NNY proteins were prepared by E. Mikhailova). These amino acids can form multiple types of interaction with the DNA nucleobases e.g. hydrogen bonds and π-cation interactions with NNR and hydrogen bonding and π-stacking interactions with NNW and NNY. These additional interactions, which are generally stronger than transitory van der Waals forces, could allow improved discrimination in comparison with NNI, NNV and NNL.

When the NNR mutant was probed with poly(dC)N$_{(9)}$ oligonucleotides, the additional charged residue inside the pore caused this mutant to exhibit superior ssDNA capture, in comparison to the other mutants tested. Like the hydrophobic residues the arginine mutant was unable to differentiate the epigenetic modifications 5mC and hmC from each other but it produced a different overall order of increased current block level (G > T > A > 5mC/hmC > C) (Figure 2 – 4). The tyrosine and tryptophan mutants were,

however, both able to clearly distinguish cytosine from each of its epigenetically modified nucleobases (Figure 2 – 4). This would allow either mutant to be used in a detection assay for discrimination of 5mC, hmC and C only. However, full six-nucleobase discrimination was not obtained for NNY because G and 5mC exhibited identical block levels, as did A and hmC. The NNW mutant should have produced six distinct block levels because it had previously been shown to discriminate C, T, A and G fully and in this case 5mC and hmC's block levels did not directly overlap with any of these signals.[60] Unfortunately, when all six oligonucleotides were present in the same experiments, the block levels for G, A, 5mC and hmC all had I_{RES} values which were within ~0.3% of each other. This meant that under these conditions the block levels of A and G were observed to overlap. Probing all six R_1 mutants with poly(dC)$N_{(9)}$ oligonucleotides did not produced six distinct current signals.



	I_O=207 ± 1 pA					
Oligo	G	T	A	5mC	hmC	C
I_{RES} (%)	13.4 ± 0.1	14.3 ± 0.2	14.7 ± 0.1	14.9 ± 0.1	14.9 ± 0.1	15.6 ± 0.1

	I_O=180 ± 2 pA					
Oligo	T	A	G	5mC	hmC	C
I_{RES} (%)	25.0 ± 0.1	25.7 ± 0.1	25.7 ± 0.1	26.1 ± 0.1	26.4 ± 0.1	27.9 ± 0.1

	I_O=173 ± 3 pA					
Oligo	T	5mC	G	hmC	A	C
I_{RES} (%)	26.2 ± 0.1	27.2 ± 0.1	27.2 ± 0.1	27.6 ± 0.2	27.6 ± 0.2	29.0 ± 0.1

Figure 2 – 4 – *Nucleobase discrimination abilities of the NNR, NNW and NNY αHL pores when probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3' biotin-TEG.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	215	15.5	14.8	−0.8	14.8	−0.8	14.3	−1.2	14.8	−0.8	13.2	−2.3	2.3
2	218	15.7	14.9	−0.8	14.9	−0.8	14.4	−1.3	14.7	−1.0	13.4	−2.3	2.3
3#	207	15.6	14.9	−0.7	14.9	−0.7	14.3	−1.3	14.7	−0.9	13.4	−2.2	2.2
Mean	213	15.6	14.9	−0.8	14.9	−0.8	14.3	−1.3	14.7	−0.9	13.3	−2.3	2.3
SD	6	0.1	0.1	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0

Table 2 – 4A – *Residual current data for the NNR mutant of αHL probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$_{40}$ and the poly(dC)-$T_{(9)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 4A for the NNR mutant.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	172	28.0	26.4	−1.5	26.2	−1.8	25.1	−2.9	25.7	−2.2	25.7	−2.2	2.9
2	175	27.9	26.3	−1.6	26.0	−1.9	25.0	−2.8	25.8	−2.1	25.8	−2.1	2.8
3	179	27.5	26.0	−1.5	25.8	−1.7	24.6	−2.9	25.6	−1.9	25.6	−1.9	2.9
4#	180	27.9	26.4	−1.5	26.1	−1.9	25.0	−2.9	25.7	−2.2	25.7	−2.2	2.9
Mean	177	27.8	26.3	−1.5	26.0	−1.8	25.0	−2.9	25.7	−2.1	25.7	−2.1	2.9
SD	4	0.2	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.2	0.1	0.2	0.0

Table 2 – 4B – *Residual current data for the NNW mutant of αHL probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$_{40}$ and the poly(dC)-$T_{(9)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 4B for the NNW mutant.*

Entry	I_O (pA)	I_{RES} poly $(dC)_{40}$ (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1#	173	29.0	27.6	−1.4	27.2	−1.9	26.2	−2.8	27.6	−1.4	27.2	−1.9	2.8
2	169	31.3	29.8	−1.6	29.2	−2.1	28.4	−2.9	29.8	−1.6	29.2	−2.1	2.9
3	172	31.6	30.1	−1.4	29.5	−2.1	29.5	−2.1	30.1	−1.4	29.5	−2.1	2.1
Mean	171	30.6	29.2	−1.5	28.6	−2.0	28.0	−2.6	29.2	−1.5	28.6	−2.0	2.6
SD	2	1.4	1.4	0.1	1.3	0.1	1.7	0.5	1.4	0.1	1.3	0.1	0.5

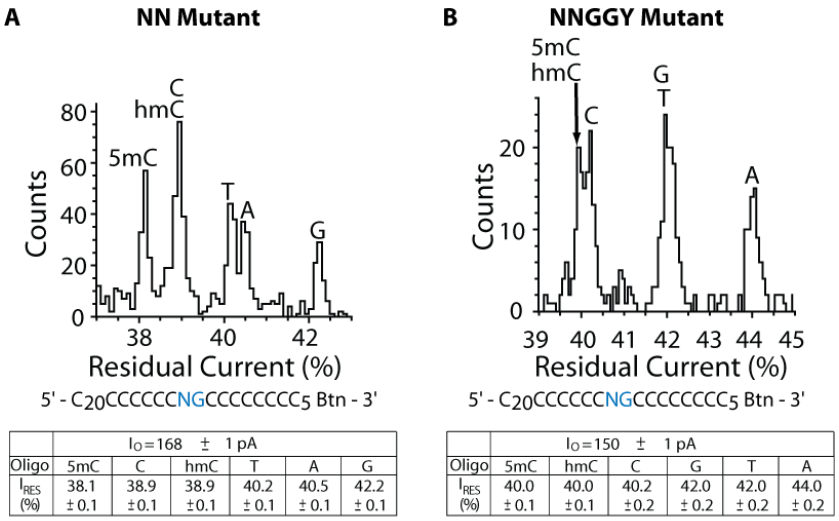
Table 2 – 4C – *Residual current data for the NNY mutant of αHL probed with poly(dC)$N_{(9)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current*

between the poly(dC)$_{40}$ and the poly(dC)-T$_{(9)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 4C for the NNY mutant.

2.2.3 – Probing Recognition Sites R_1 and R_2 with CpG Oligonucleotides for Six-Nucleobase Discrimination

Owing to the fact that both epigenetic modifications occur largely at CpG sites (see Chapter 1, section 1.2),[89, 123] additional oligonucleotide sequences that contained a guanine next to the site that was being probed (poly(dC)N$_{(9)}$G$_{(8)}$ and poly(dC)N$_{(14)}$G$_{(13)}$ where N = C, T, A, G, 5mC or hmC) were investigated, as a more relevant sequencing background.

Probing of the NN and NNGGY mutants showed some slight differences in discrimination and overall distribution of the block levels for the poly(dC)N$_{(14)}$G$_{(13)}$ oligonucleotides. In the case of the NN mutant, improved discrimination between C and 5mC was observed (ΔI_{RES} 5mC = −0.8 ± 0.1% for the poly(dC)N$_{(14)}$G$_{(13)}$ compared with ΔI_{RES} 5mC = −0.5 ± 0.0% for the poly(dC)N$_{(14)}$ see Table 2 – 2A and 2 – 5A) and a greater $\Delta I_{RES}^{OVERALL}$ was measured ($\Delta I_{RES}^{OVERALL}$ = 4.1 ± 0.1% (n = 8) for the poly(dC)N$_{(14)}$G$_{(13)}$ compared with $\Delta I_{RES}^{OVERALL}$ = 3.4 ± 0.2% (n = 4) for the poly(dC)N$_{(14)}$ see Table 2 – 2A and 2 – 5A). However, the block levels for C and hmC were still observed to directly overlap and six-nucleobase discrimination was not obtained (Figure 2 – 5A). Poor discrimination was observed with the NNGGY mutant when probed with the poly(dC)N$_{(14)}$G$_{(13)}$ oligonucleotides. hmC's residual current level no longer directly overlapped with C, it instead coincided with 5mC and the signals produced by G and T were now also identical (Figure 2 – 5B). Therefore, probing recognition point R_2, with both poly(dC)N$_{(14)}$ and poly(dC)N$_{(14)}$G$_{(13)}$ oligonucleotides, did not result in full discrimination of the six DNA nucleobases.



	I_O=168 ± 1 pA					
Oligo	5mC	C	hmC	T	A	G
I_{RES} (%)	38.1 ± 0.1	38.9 ± 0.1	38.9 ± 0.1	40.2 ± 0.1	40.5 ± 0.1	42.2 ± 0.1

	I_O=150 ± 1 pA					
Oligo	5mC	hmC	C	G	T	A
I_{RES} (%)	40.0 ± 0.1	40.0 ± 0.1	40.2 ± 0.2	42.0 ± 0.2	42.0 ± 0.2	44.0 ± 0.2

Figure 2 – 5 – *Nucleobase discrimination abilities of the NN and NNGGY pores when probed with poly(dC)$N_{(14)}G_{(13)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 14 relative to the 3' biotin-TEG.*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	167	39.0	39.0	0.0	38.1	−0.8	40.1	1.1	40.5	1.5	42.2	3.2	4.1
2	159	38.3	38.3	0.0	37.6	−0.7	39.6	1.3	39.9	1.6	41.6	3.3	4.0
3	161	39.1	39.1	0.0	38.2	−0.9	40.2	1.1	40.5	1.4	42.3	3.3	4.1
4#	168	38.9	38.9	0.0	38.1	−0.8	40.2	1.2	40.5	1.6	42.2	3.3	4.1
5	162	39.2	39.2	0.0	38.4	−0.8	40.3	1.2	40.8	1.6	42.6	3.4	4.1
6	161	39.3	39.3	0.0	38.4	−0.9	40.4	1.1	40.8	1.5	42.6	3.3	4.2
7	164	38.9	38.9	0.0	38.0	−1.0	40.2	1.2	40.5	1.6	42.2	3.2	4.2
8	171	38.5	38.5	0.0	37.6	−0.8	39.6	1.2	40.0	1.5	41.7	3.2	4.0
Mean	164	38.9	38.9	0.0	38.1	−0.8	40.1	1.2	40.4	1.5	42.2	3.3	4.1
SD	4	0.3	0.3	0.0	0.3	0.1	0.3	0.1	0.3	0.1	0.4	0.1	0.1

Table 2 – 5A – *Residual current data for the NN mutant of αHL probed with poly(dC)$N_{(14)}G_{(13)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 14 relative to the 3' biotin-TEG (n = 8). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(14)}G_{(13)}$ and the poly(dC)$T_{(14)}G_{(13)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 5A for the NN mutant.*

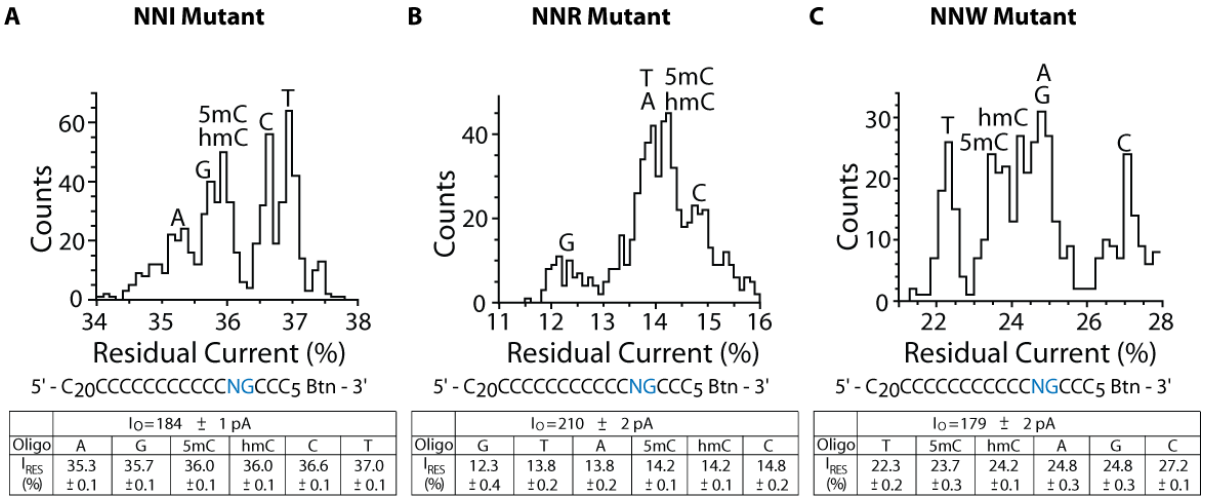
Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	171	39.1	38.7	−0.3	38.7	−0.3	40.7	1.6	42.9	3.9	40.7	1.6	4.2
2	155	41.3	41.0	−0.3	41.0	−0.3	43.2	1.9	45.3	4.0	43.2	1.9	4.3
3#	150	40.2	40.0	−0.2	40.0	−0.2	42.0	1.8	44.0	3.8	42.0	1.8	4.1
Mean	159	40.2	39.9	−0.3	39.9	−0.3	42.0	1.8	44.1	3.9	42.0	1.8	4.2
SD	11	1.1	1.1	0.1	1.1	0.1	1.2	0.1	1.2	0.1	1.2	0.1	0.1

Table 2 – 5B – *Residual current data for the NNGGY mutant of αHL probed with poly(dC)$N_{(14)}G_{(13)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G,*

5mC or hmC) at position 14 relative to the 3' biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(14)}G_{(13)}$ and the poly(dC)$T_{(14)}G_{(13)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 5B for the NNGGY mutant.

As NNI, NNV and NNL were previously observed to have produced similar discrimination patterns, when probed with poly(dC)$N_{(9)}$ oligonucleotides, only NNI, which exhibited the greatest nucleobase identification ability, was probed with the poly(dC)$N_{(9)}G_{(8)}$ ssDNA sequences. The $\Delta I_{RES}{}^{OVERALL}$, when probed with poly(dC)$N_{(9)}G_{(8)}$, was smaller than for the poly(dC)$N_{(9)}$ strands ($\Delta I_{RES}{}^{OVERALL}$ = 1.6 ± 0.1% (n = 3) for the poly(dC)$N_{(9)}G_{(8)}$ compared with $\Delta I_{RES}{}^{OVERALL}$ = 2.0 ± 0.1% (n = 4) for the poly(dC)$N_{(9)}$ see tables 2 – 3A and 2 – 6A). As the overall signal distribution was smaller, this resulted in poorer nucleobase discrimination because the block levels, which were in the same order of increasing block (A > G > 5mC/hmC > C > T), were all much closer together (Figure 2 – 6A). Inferior nucleobase discrimination was also observed when NNR was probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides (Figure 2 – 6B). A similar range of block levels was noted, when compared to the poly(dC)$N_{(9)}$ signals, however, it was not possible to distinguish T from A or 5mC from hmC with this αHL mutant. However, a significant improvement in nucleobase identification was noted when the NNW mutant was probed with the poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides (Figure 2 – 6C). The $\Delta I_{RES}{}^{OVERALL}$ had increased by over 1.5 times when compared to the block level range produced by poly(dC)$N_{(9)}$ ($\Delta I_{RES}{}^{OVERALL}$ = 4.9 ± 0.1% (n = 3) for the poly(dC)$N_{(9)}G_{(8)}$ compared with $\Delta I_{RES}{}^{OVERALL}$ = 2.9 ± 0.0% (n = 4) for the poly(dC)$N_{(9)}$). This in turn resulted in improved discrimination of the two epigenetic modifications whose block levels were quite close when probed with poly(dC)$N_{(9)}$ containing ssDNA ($\Delta I_{RES}{}^{(poly(dC)hmC(9)G(8) - poly(dC)5mC(9)G(8))}$ = 0.8 ± 0.2% for the poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides compared with $\Delta I_{RES}{}^{(poly(dC)hmC(9) - poly(dC)5mC(9))}$ = 0.3 ± 0.1% for the

poly(dC)$N_{(9)}$ see table 2 – 4B and 2 – 6C). Despite the dramatic increase in $\Delta I_{RES}^{OVERALL}$ the block levels produced by A and G still directly overlaid with each other.



	I_O=184 ± 1 pA					
Oligo	A	G	5mC	hmC	C	T
I_{RES} (%)	35.3 ± 0.1	35.7 ± 0.1	36.0 ± 0.1	36.0 ± 0.1	36.6 ± 0.1	37.0 ± 0.1

	I_O=210 ± 2 pA					
Oligo	G	T	A	5mC	hmC	C
I_{RES} (%)	12.3 ± 0.4	13.8 ± 0.2	13.8 ± 0.2	14.2 ± 0.1	14.2 ± 0.1	14.8 ± 0.2

	I_O=179 ± 2 pA					
Oligo	T	5mC	hmC	A	G	C
I_{RES} (%)	22.3 ± 0.2	23.7 ± 0.3	24.2 ± 0.1	24.8 ± 0.3	24.8 ± 0.3	27.2 ± 0.1

Figure 2 – 6 – *Nucleobase discrimination abilities of the NNI, NNR and NNW αHL pores when probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3' biotin-TEG.*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1#	184	36.6	36.0	−0.7	36.0	−0.7	37.0	0.3	35.3	−1.3	35.7	−0.9	1.7
2	186	36.4	35.7	−0.8	35.7	−0.8	36.7	0.3	35.2	−1.3	35.5	−1.0	1.6
3	186	36.0	35.3	−0.7	35.3	−0.7	36.3	0.3	34.7	−1.3	35.0	−1.0	1.7
Mean	185	36.4	35.6	−0.7	35.6	−0.7	36.7	0.3	35.0	−1.3	35.4	−1.0	1.6
SD	2	0.3	0.3	0.1	0.3	0.1	0.3	0.0	0.3	0.0	0.4	0.0	0.1

Table 2 – 6A – *Residual current data for the NNI mutant of αHL probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3' biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(9)}G_{(8)}$ and the poly(dC)$T_{(9)}G_{(8)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 6A for the NNI mutant.*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1#	210	14.8	14.2	−0.6	14.2	−0.6	13.8	−1.0	13.8	−1.0	12.3	−2.5	2.5
2	213	14.8	14.2	−0.6	14.2	−0.6	13.8	−1.0	13.8	−1.0	12.5	−2.3	2.3
3	215	13.8	13.4	−0.4	13.4	−0.4	12.9	−0.9	12.9	−0.9	11.3	−2.5	2.5
Mean	212	14.5	13.9	−0.5	13.9	−0.5	13.5	−0.9	13.5	−0.9	12.0	−2.4	2.4
SD	2	0.6	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.6	0.1	0.1

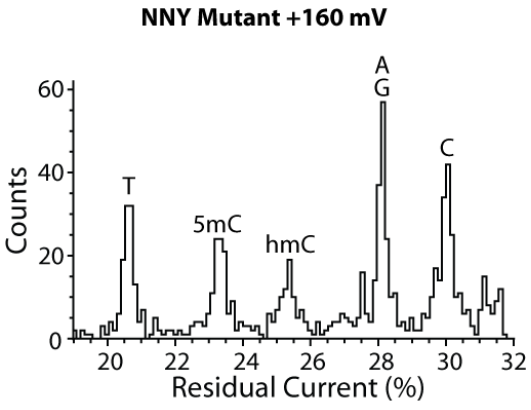
Table 2 – 6B – *Residual current data for the NNR mutant of αHL probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3' biotin-TEG (n = 3). The I_O and I_{RES} given for*

each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(9)}G_{(8)}$ and the poly(dC)$T_{(9)}G_{(8)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 6B for the NNR mutant.

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1	170	26.6	23.8	−2.7	23.0	−3.6	21.6	−4.9	24.2	−2.4	24.2	−2.4	4.9
2#	179	27.2	24.2	−2.9	23.7	−3.5	22.3	−4.9	24.8	−2.4	24.8	−2.4	4.9
3	177	27.9	25.1	−2.8	24.0	−3.8	22.9	−5.0	25.6	−2.3	25.6	−2.3	5.0
Mean	175	27.2	24.4	−2.8	23.5	−3.6	22.3	−4.9	24.9	−2.3	24.9	−2.3	4.9
SD	5	0.6	0.6	0.1	0.5	0.2	0.6	0.0	0.7	0.0	0.7	0.0	0.0

Table 2 – 6C – *Residual current data for the NNW mutant of αHL probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(9)}G_{(8)}$ and the poly(dC)$T_{(9)}G_{(8)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The # indicates the experiment which corresponds to the histogram in Figure 2 – 6C for the NNW mutant.*

Finally, the NNY mutant was probed for six-nucleobase discrimination with the poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides. This mutant showed a notable improvement as the $\Delta I_{RES}^{OVERALL}$ ($\Delta I_{RES}^{OVERALL}$ = 9.6 ± 0.4% (n = 3) for the poly(dC)$N_{(9)}G_{(8)}$ compared with $\Delta I_{RES}^{OVERALL}$ = 2.6 ± 0.5% (n = 3) for poly(dC)$N_{(9)}$ see tables 2 – 3C and 2 – 7) had increased by over 3.5 times and the signals for 5mC and hmC no longer produced identical block levels with G and A respectively. The identification of C from its epigenetically modified nucleobases was easily observed, owing to the very large block level range (ΔI_{RES} 5mC = −6.8 ± 0.2% and ΔI_{RES} hmC = −4.8 ± 0.2% see table 2 – 7). Similarly to the NNW mutant, the NNY still produced block levels for A and G that were identical. Therefore, all of the mutants tested at R_1 did not facilitate six-nucleobase discrimination upon probing with both poly(dC)$N_{(9)}G_{(8)}$ and poly(dC)$N_{(9)}$ oligonucleotides at an applied potential of +160 mV.



	I_O = 160 ± 3 pA					
Oligo	T	5mC	hmC	A	G	C
I_{RES} (%)	20.3 ± 0.1	22.9 ± 0.1	24.9 ± 0.2	27.7 ± 0.1	27.7 ± 0.1	29.5 ± 0.1

Figure 2 – 7 – *Nucleobase discriminating ability of the NNY αHL pore when probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG at an applied potential of +160 mV.*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1[#]	160	29.5	24.9	−4.6	22.9	−6.6	20.3	−9.2	27.7	−1.9	27.7	−1.9	9.2
2	169	31.8	27.0	−4.9	25.0	−6.9	22.1	−9.7	29.9	−1.9	29.9	−1.9	9.7
3	163	31.7	26.9	−4.8	24.8	−6.8	21.7	−9.9	29.7	−1.9	29.7	−1.9	9.9
Mean	164	31.0	26.2	−4.8	24.2	−6.8	21.4	−9.6	29.1	−1.9	29.1	−1.9	9.6
SD	4	1.3	1.2	0.1	1.1	0.2	1.0	0.4	1.3	0.0	1.3	0.0	0.4

Table 2 – 7 – *Residual current data for the NNY mutant of αHL probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG, at an applied potential of +160 mV (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(9)}G_{(8)}$ and the poly(dC)$T_{(9)}G_{(8)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The [#] indicates the experiment which corresponds to the histogram in Figure 2 – 7 for the NNY mutant.*

2.2.4 – Optimisation of Experimental Conditions for Six-Nucleobase Discrimination by the αHL Mutant E111N/K147N/M113Y Upon Probing with CpG Oligonucleotides

The $\Delta I_{RES}^{OVERALL}$ for the E111N/K147N/M113Y was so large, it was proposed that if the experimental conditions were optimised (+160 mV in 1 M KCl, 25 mM Tris, 100 μM EDTA, pH 8.0), it might be possible to separate the signals for A and G and obtain six-nucleobase discrimination. Previous work in the group had shown that by altering the physical conditions, such as: pH, salt concentration, salt type and applied potential it was

possible to modify nucleobase discrimination (unpublished results carried out by L. Franceschini, Z. Epsek, D. Stoddart, A. Heron and G. Maglia). In an attempt to enhance the signal separation between A and G, applied potentials between +100 mV and +200 mV were examined (Figure 2 – 8A). At applied potentials below +130 mV it was possible to discriminate the residual current signals produced by A and G. Therefore, an applied potential of +110 mV was selected for further investigation as, discrimination was improved sufficiently to distinguish the two nucleobases (Figure 2 – 8).

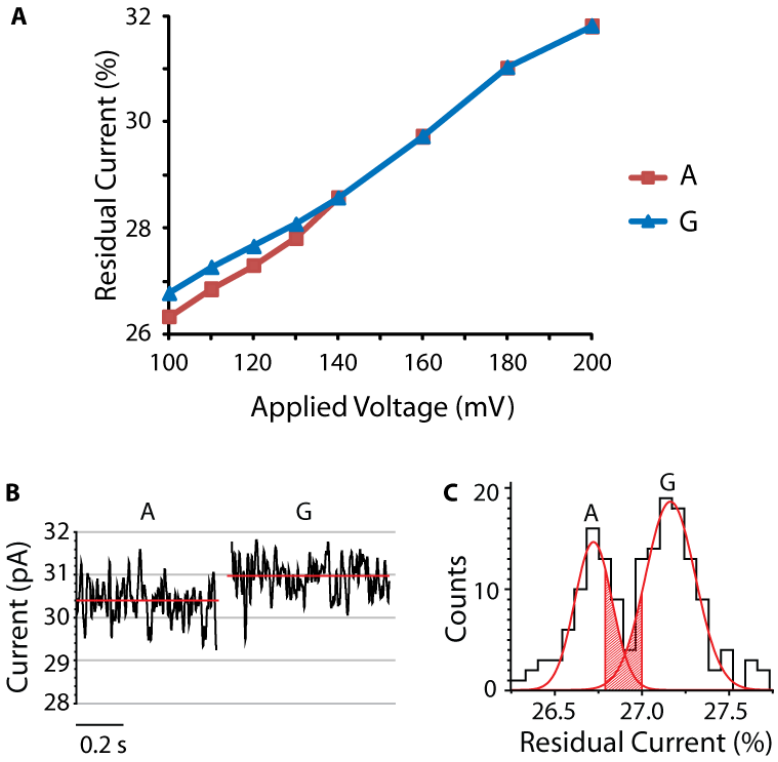
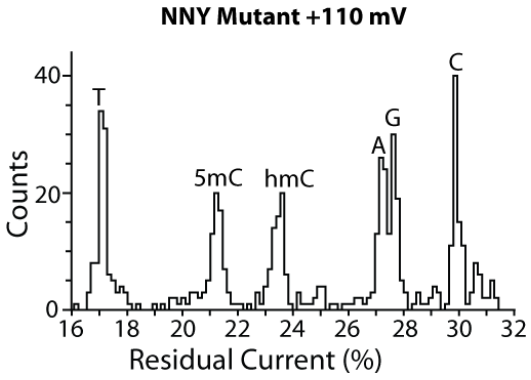


Figure 2 – 8 – *The discrimination of the poly(dC)$A_{(9)}G_{(8)}$ and poly(dC)$G_{(9)}G_{(8)}$ signals in the NNY mutant. A) Typical plot of the residual current values for poly(dC)$A_{(9)}G_{(8)}$ and poly(dC)$G_{(9)}G_{(8)}$ over the applied voltage range +100-200 mV for a single experiment (n = 1). Below +140 mV it was possible to distinguish the two oligonucleotides from each other. B) Typical current levels for the NNY mutant when blocked with poly(dC)$A_{(9)}G_{(8)}$ and poly(dC)$G_{(9)}G_{(8)}$ at an applied potential of +110 mV (filtered at 50 Hz for display purposes). C) Event histogram displaying only the residual current levels resulting from poly(dC)$A_{(9)}G_{(8)}$ and poly(dC)$G_{(9)}G_{(8)}$ levels. The shaded region corresponds to the region where nucleobases cannot be called with greater than 95% confidence (~17% of the total events). Figure 2 – 8 section C was produced and analysed by A. Heron. This figure has been adapted from Wallace et al. 2010.*[184]

At an applied potential of +110 mV six distinct residual current levels were observed (Figure 2 – 9), over an even greater range ($\Delta I_{RES}^{OVERALL}$) than was observed at +160 mV

($\Delta I_{RES}^{OVERALL}$ = 9.6 ± 0.4% (n = 3) at +160 mV increased to $\Delta I_{RES}^{OVERALL}$ = 12.1 ± 0.5% (n = 4) at +110 mV). The ΔI_{RES} values for hmC and 5mC were also even greater at this applied potential than the previously measured values at +160 mV (ΔI_{RES} hmC = −6.2 ± 0.2% and ΔI_{RES} 5mC = −8.4 ± 0.3%, Table 2 – 9). By optimisation of the experimental conditions, it was possible to distinguish the epigenetic modifications 5mC and hmC from the other four canonical nucleobases C, T, A and G.



	I_O=113 ± 1 pA					
Oligo	T	5mC	hmC	A	G	C
I_{RES} (%)	17.1 ±0.2	21.3 ±0.2	23.6 ±0.2	27.4 ±0.1	27.8 ±0.1	30.0 ±0.1

Figure 2 – 9 – *Nucleobase discriminating ability of the NNY αHL pore when probed with poly(dC)$G_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG at an applied potential of +110 mV.*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	$\Delta I_{RES}^{OVERALL}$ %
1[#]	113	30.0	23.6	−6.4	21.3	−8.7	17.1	−12.9	27.4	−2.6	27.8	−2.2	12.9
2	121	29.4	23.4	−6.0	21.3	−8.1	17.3	−12.1	26.8	−2.6	27.3	−2.1	12.1
3	113	29.7	23.5	−6.2	21.4	−8.3	17.8	−11.9	27.0	−2.7	27.5	−2.2	11.9
4	114	29.4	23.1	−6.3	20.8	−8.6	16.8	−12.6	26.7	−2.7	27.1	−2.3	12.6
Mean	115	29.6	23.4	−6.2	21.2	−8.4	17.3	−12.4	27.0	−2.7	27.4	−2.2	12.4
SD	4	0.3	0.2	0.2	0.3	0.3	0.4	0.5	0.3	0.1	0.3	0.1	0.5

Table 2 – 9 – *Residual current data for the NNY mutant of αHL probed with poly(dC)$N_{(9)}G_{(8)}$ oligonucleotides with a single substituted nucleobase (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG, at an applied potential of +110 mV (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dC)$C_{(9)}G_{(8)}$ and the poly(dC)$T_{(9)}G_{(8)}$ levels (ΔI_{RES} T = I_{RES} T − I_{RES} C). The [#] indicates the experiment which corresponds to the histogram in Figure 2 – 9 for the NNY mutant.*

2.3 – Conclusions

By investigating a number of αHL mutants, which previously exhibited four-nucleobase discrimination at either recognition site R_1 or R_2,[59, 60] it was possible to identify the NNY mutant which had improved discrimination when probed with the epigenetically relevant sequence poly(dC)$N_{(9)}G_{(8)}$ (where N = C, T, A, G, 5mC or hmC). This mutant had a large overall block level range and could distinguish cytosine from its epigenetically modified nucleobases. Optimisation of the experimental conditions, by applying a lower potential of +110 mV, resulted in six-nucleobase discrimination by probing the recognition site R_1 of the NNY mutant pore.

There are a number of factors which seem to contribute to overall nucleobase discrimination, these include the amino acid residues present at the recognition sites of the αHL mutant being probed, the sequence of the probing oligonucleotide and the experimental conditions. The molecular reasoning behind the observed changes in nucleobase discrimination is, however, not well understood. Clearly, mutation of the amino acids that are located at the constriction has a great effect on nucleobase discrimination at this recognition site, as it was observed that R_1 could be enhanced (e.g. in the case of NNY) or removed completely (e.g. NNG mutant).[60] These differences in discrimination have also been mirrored in the experiments documented here where by varying the side chain of the amino acid located at 113, from hydrophobic NNI to aromatic NNY, a wide change in discrimination of the epigenetically modified nucleobases was also observed.

Work carried out by Okamoto and co-workers, has shown that a chemically designed zinc finger peptide was able to distinguish hmC from 5mC.[134] The selectivity for the two nucleobases was explained on the basis of a CH-π interaction between a phosphotyrosine residue, on the zinc finger peptide, and the epigenetic modifications. A strong CH-π

interaction was observed between 5mC and the phosphotyrosine amino acid, whereas, hmC's hydroxymethyl group sterically hindered the approach of the peptide making the CH-π interaction significantly weaker. This difference in binding strength could account for the 5mC containing oligonucleotide producing a lower residual current signal, than that which contained the hmC nucleobase, when probing the NNY mutant. A similar CH-π interaction may have resulted between 5mC and the tyrosine amino acid at position 113 in the NNY αHL mutant pore. The greater current block produced for 5mC could be rationalised by the increased interaction between this nucleobase and the pore resulting in reduced current flow through the constriction of αHL.[134] However, further experiments would need to be performed to confirm if this is indeed the case.

Stoddart *et al.* observed that R_1 and R_2 showed maximum recognition at positions 9 and 14 relative to the 3' biotin-TEG, however, these recognition sites were broad and therefore could recognise nucleobases either side of this position (Figure 2 – 10).[59] As the additional guanine nucleobase present in the CpG oligonucleotide was within the recognition region, this could account for why changing the nucleobase at position 8 and 13, respectively, to a guanine had a dramatic effect on nucleobase discrimination. However, the exact interactions or conformational changes that the single additional guanine promoted in the ssDNA oligonucleotides, that caused such dramatic alterations in the $\Delta I_{RES}^{OVERALL}$ and single nucleobase discrimination, is not clearly understood.

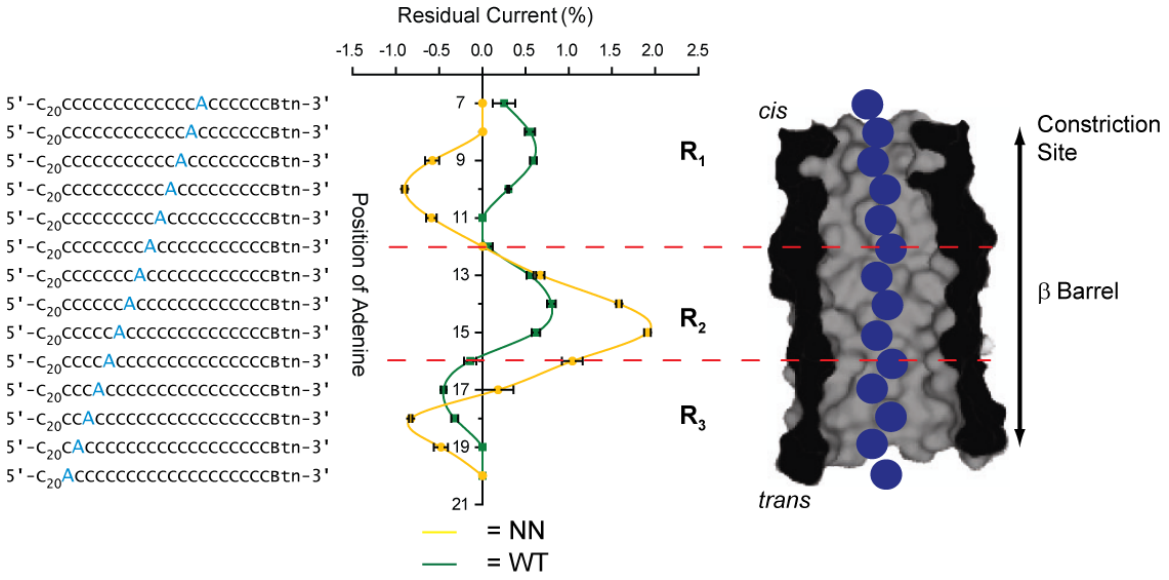


Figure 2 – 10 *– The location of the different recognition sites within the αHL pore. The central graph plots the difference in residual current (ΔI_{RES}) between poly(dC)$_{40}$ and poly(dC) with a single adenine nucleobase (the sequence of each oligonucleotide is shown on the left) at various different positions in the strand, for both the WT and the NN pores. The sites of discrimination are labelled R_1, R_2 and R_3 and their probable location within the αHL barrel is indicated on the right. Figure adapted from Stoddart et al, 2009.*[59]

As the applied potential was lowered from +160 mV to +110 mV, it was possible to enhance discrimination of the six DNA nucleobases. This initially seemed counter intuitive because at higher applied potentials discrimination can improve owing to an increase in the overall current flowing through the pore resulting in improved signal to noise ratios.[185] However, at lower applied potentials the DNA will experience a weaker overall force pulling on the immobilised ssDNA-biotin•streptavidin complex, allowing an increased interaction with the αHL pore.[65, 186, 187] Therefore at a lower applied potential the DNA may be able to adopt a different base-specific conformation, as it is in a slightly more relaxed state, and this alternative conformation may have resulted in improved discrimination. Further work is required before it will be possible to fully distinguish the mechanisms behind nucleobase discrimination in biological nanopores.

Despite not fully understanding the manner by which αHL discriminates the DNA nucleobases, this work demonstrated that by site-directed mutagenesis and experimental optimisation it was indeed possible to use nanopore technology to detect all six

nucleobases. This would mean that if this technology can be integrated into a practicable nanopore sequencing device, it could potentially be used to study epigenetic modifications. This would be of great benefit as both 5mC and hmC have been linked with a number of diseases and types of cancer.[89, 96, 122] It has also been suggested that both nucleobases could be used as potential biomarkers in the early diagnosis of cancer,[89, 122] therefore if a technique can be developed that can accurately distinguish cytosine from its epigenetically modified adducts then it could contribute to future fundamental medical diagnostic studies.

Chapter 3

Screening Chemical Reactions for the Selective Modification of DNA

Chapter 3 – Screening Chemical Reactions for the Selective Modification of DNA

3.1 – Introduction

The impetus behind developing chemical reactions which are selective for only one DNA nucleobase is generally to aid detection of that base for DNA sequencing. A large number of different types of reaction have been developed for this purpose (see Chapter 1, section 1.3). Guanine and adenine have both been targeted for the selective modification by methylating agents.[1, 143] The C5-C6 substituted double bond of thymine and 5-methylcytosine has been extensively investigated for modification by various oxidising agents.[153, 156, 158, 159, 174] Cytosine has been selectively modified by the bisulfite reaction, which is the main method to date for determining cytosine from its epigenetically modified base 5mC.[109] Since the discovery of 5-hydroxymethylcytosine in mammals in 2009, it has also been targeted for selective modification, mostly by enzymes which target the free hydroxymethyl group at the C5 position.[123, 127, 181] Despite the documentation of many different DNA modification methods in the literature, more selective methods that occur under milder conditions, resulting in less DNA damage, are still required.

There are a number of factors which need to be considered when designing a selective chemical reaction for one of the DNA bases. Firstly, it is necessary to consider which of the six bases (C, T, A, G, 5mC or hmC) is to be targeted by the reaction and, in particular, what functional moiety present in that base one wishes the reaction to chemically modify. For example a reaction which would target the free hydroxyl of hmC could prove to be highly selective because there are no other free hydroxyls present in DNA. On the other hand a modification which could target free NH_2 is less likely to be selective owing to the fact that C, A, G, 5mC and hmC all contain this reactive functional group (Figure 3 – 1).

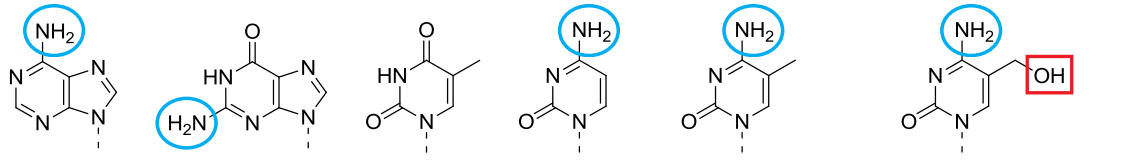


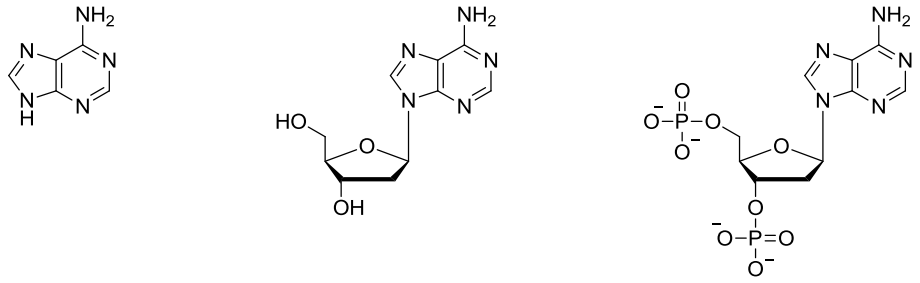
Figure 3 – 1 *– The DNA bases adenine, guanine, thymine, cytosine, 5-methylcytosine and 5-hydroxymethylcytosine. Blue circles highlight the number of free NH_2 present in the DNA bases, whereas, the single free hydroxyl is highlighted by a red square.*

Once the base to be modified has been chosen, the conditions employed to carry out the reaction must be designed. ssDNA is a relatively stable entity, particularly in comparison to other biological molecules such as proteins, however it does impose certain constrictions on the conditions that can be employed to modify it. As DNA requires additional additives to increase its solubility in organic solvents,[188] the reaction conditions utilised would require water as a co-solvent to solubilise the DNA. The presence of water in the reaction solution would, therefore, preclude the use of any reagents which decompose upon exposure to water e.g. such as *tert*-butyllithium. DNA is prone to depurination of adenine and guanine residues and hydrolytic deamination of cytosine and 5-methylcytosine residues.[189] The rates of these processes are affected by temperature and pH. Therefore, chemical reactions employed for the selective modification of DNA should be designed to react with DNA at temperatures below 40 °C and within a pH range of approximately 5-9, in order to prevent these undesired modifications. This chapter details several different types of chemical reaction that were screened in an attempt to identify a chemical reagent that was capable of selectively modifying one of the DNA bases in the presence of all the others. Those which were found to be selective were further optimised for modification of individual bases within ssDNA.

3.2 – Results and Discussion

3.2.1 – Identification of an Effective Model System for the Individual DNA Bases

With the aim of screening a variety of different reaction conditions for the selective chemical modification of DNA, it was necessary to develop a system that would provide an accurate model for the DNA bases, as well as facilitating easy and efficient purification of the reaction products. Several options were available, these included individual protected/unprotected analogues of the 1) nitrogenous nucleobases, 2) deoxynucleosides or 3) deoxynucleotides 3',5'-bisphosphates (Figure 3 – 2). Unprotected analogues of the nitrogenous nucleobases were deemed a poor model because they contained a free NH moiety, that is not present in ssDNA and could augment the chemical reactivity of the bases. The nucleotides were considered unsuitable owing to the difficulties involved with purification of these molecules e.g. the pendant phosphate groups would cause these molecules to enter the aqueous phase in chemical purification and would require reverse phase HPLC rather than standard flash chromatography. The unprotected nucleosides were deemed another poor model system because they contained two free hydroxyls that are protected with phosphate groups in ssDNA. The presence of these functional units could again affect reaction selectivity and efficiency. However, if the free hydroxyls were protected with a relatively stable and functionally inert group, such as a silicon protecting-group, then this should provide a system that has similar reactivity to the DNA bases, when present in ssDNA, and decrease the polarity of the bases aiding purification by flash chromatography.



Adenine
Nitrogenous Nucleobase

Deoxyadenosine
Deoxynucleoside

Deoxyadenosine 3',5'-bisphosphate
Deoxynucleoside 3',5'-bisphosphate

Figure 3 – 2 – *The chemical structure of the adenine nitrogenous nucleobase, deoxynucleoside and deoxynucleotide.*

tert-Butyldimethylsilyl (TBS) groups have previously been used to selectively protect the free alcohols of DNA bases for the synthesis of a variety of different DNA analogues, in the presence of both primary and secondary amine functionalities.[190-192] Accordingly, this protecting group was selected for protection of the deoxyribose sugar. *tert*-Butyldimethylsilyl chloride was used to protect all five DNA deoxynucleosides (**1**, **3**, **5**, **7** and **9**), because *tert*-butyldimethylsilyl triflate was observed to be too reactive a source resulting in decomposition of the deoxynucleosides (Figure and Table 3 – 3). The TBS protection of hmC was not investigated as the deoxynucleoside analogue was not commercially available. The syntheses of all five *bis*-TBS protected deoxynucleosides were established in high yields. It was evident in all cases (**2, 4, 6, 8** and **10**) that the TBS protection had occurred at both hydroxyls, because the ^{1}H NMR spectrum exhibited integrals of 12H and 18H respectively for the characteristic methyls of the TBS moiety. This provided a model system upon which various selective chemical reactions could be tested.

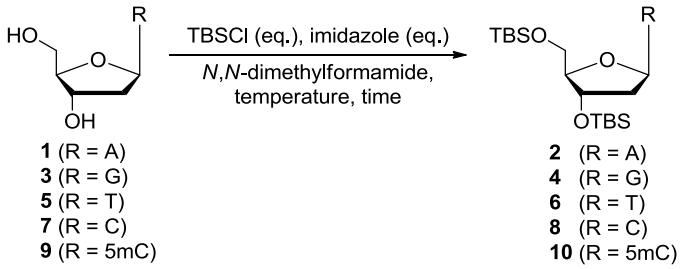


Figure 3 – 3 – *General reaction scheme for the bis-TBS protection of the deoxynucleosides by tert-butyldimethylsilyl chloride.*

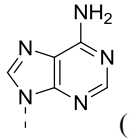
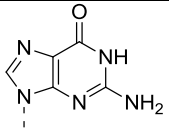
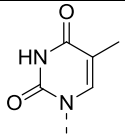
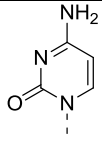
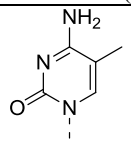
Entry	R	Conditions	Yield (%)
1	 (**1**)	TBSCl (4.0 eq.), imidazole (7.0 eq.), rt, 24 h [191]	89% of **2**
2	 (**3**)	TBSCl (4.0 eq.), imidazole (7.0 eq.), rt, 24 h	88% of **4**
3	 (**5**)	TBSCl (2.2 eq.), imidazole (4.0 eq.), 50 °C, 6 h [190]	94% of **6**
4	 (**7**)	TBSCl (2.2 eq.), imidazole (4.0 eq.), 50 °C, 8 h	95% of **8**
5	 (**9**)	TBSCl (4.0 eq.), imidazole (7.0 eq.), rt, 4 d	90% of **10**

Table 3 – 3 – *Conditions for bis-TBS protection of the deoxynucleosides.*

3.2.2 – Selective Modification of the *bis*-TBS protected Deoxynucleosides by Bromination

Selective halogenation has been successfully used as a tool for the introduction of a functional handle into the DNA bases. This has enabled them to be further derivatised by various chemical means e.g. palladium couplings,[193-196] nitration reactions,[197] thiol couplings,[198] S_N2 reactions for the attachment of various aromatic rings[199] and oxidation to alcohols.[190] Although some of the reaction conditions reported were suitable for testing on the DNA model system,[193, 197, 198] many were carried out under anhydrous conditions in

organic solvents.[190, 194-196, 199] Halogenation reactions have been successfully employed to react at a number of positions on the DNA bases indicated below (Figure 3 – 4).

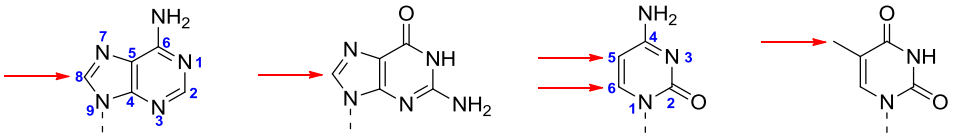


Figure 3 – 4 – *The positions at which the DNA bases have previously been observed to have been halogenated by various chemical methods. For clarity, the IUPAC numbering system for the purine and pyrimidine bases has been indicated for both adenine and cytosine (blue numbers).*

It was proposed that despite the high level of reactivity that has been observed for the DNA bases, with a variety of brominating agents, it could be possible to selectively react one of the bases in the presence of all the others. Any selective modifications could then be further optimised for the use with ssDNA. A number of conditions were investigated for the selective modification of the bis-TBS protected deoxynucleosides with brominating reagents. The first of which employed bromine itself as the halogenating agent (Figure and Table 3 – 5).[193] Compound **2** was observed to undergo bromination at the 8-position to form **11** in 43% yield. The structure of **11** was confirmed by analysis of the ^{1}H and ^{13}C NMR spectra which showed the disappearance of the singlet at 8.35 ppm (H8) in the ^{1}H NMR spectrum of **2** and a shift in the corresponding carbon signal from 152.5 to 139.0 ppm. The MS spectrum also showed the characteristic isotope splitting pattern for mono-bromination of compound **2**. Guanine (**4**) reacted, in a similar way to that of adenine, resulting in bromine substitution at the 8-position, but in a lower 21% yield (**12**). Modification of **6** resulted in the production of a single diastereomer of the bromohydrin of thymine (**13**) in a modest 37% yield (this structure was confirmed by X-ray crystallography, see Figure 3 – 6). The low yield was attributed to decomposition of the starting material under these conditions. The bromo-hydroxylation of the C5-C6 double bond resulted in the shifting of the characteristic singlet C6 alkene signal, at 7.40 ppm,

upfield to a doublet (owing to coupling with the appended hydroxyl) at 5.13 ppm in the ^{1}H NMR. The modification of cytosine (**8**) produced the desired mono-brominated analogue **14** in 46% yield. The substitution of the bromine at the C5 position was confirmed by the disappearance of the alkene peak (corresponding to the H5 proton of **8**) and the replacement of the H6 doublet proton signal (at 7.82 ppm) with a singlet at 8.08 ppm. Finally, 5-methylcytosine (**10**) was not observed to undergo any reaction with bromine under these conditions and only starting material was recovered. Bromine was observed to react to a certain extent with A, G, T and C, however, 5mC was not modified. This low degree of selectivity was not unsurprising, however, as bromine is an extremely reactive electrophile.

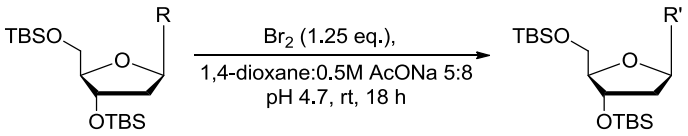


Figure 3 – 5 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) by bromine.*

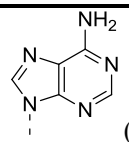
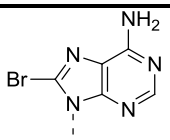
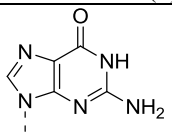
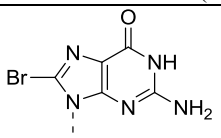
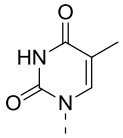
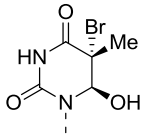
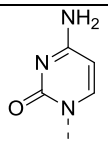
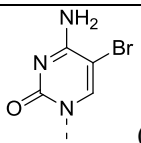
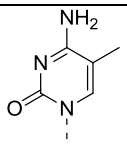
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	 (**11**)	43% **11** 54% RSM
2	 (**4**)	 (**12**)	21% **12** 67% RSM
3	 (**6**)	 (**13**)	37% **13**
4	 (**8**)	 (**14**)	46% **14** 39% RSM
5	 (**10**)	-	72% RSM

Table 3 – 5 *– The yields for the reaction of bromine with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) (RSM refers to recovered starting material).*

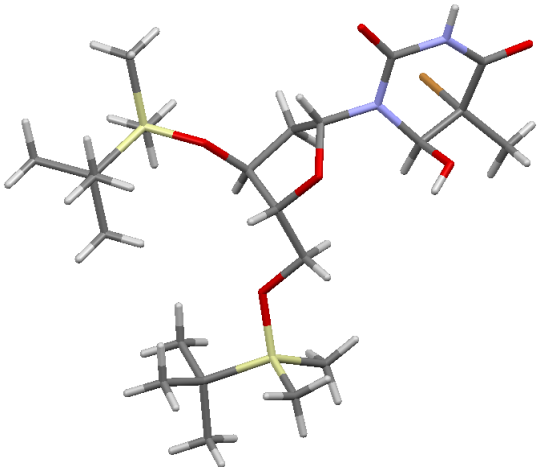


Figure 3 – 6 *– Crystal structure of the major diastereomer of the bromo-hydrin of thymine (**13**). It shows that bromination has occurred from the top-face producing the (5R,6R) distereomer. See Chapter 9, appendix 9.1 for additional crystallographic data. This crystal structure was run and characterised by D. Baker.*

Another commonly utilised brominating reagent, *N*-bromosuccinimide (NBS, **15**),[197] was examined to see if it could facilitate selective modification of one of the five *bis*-TBS protected deoxynucleosides (Figure and Table 3 – 7). This reagent was slightly more

selective than bromine as it did not brominate the 8-position of adenine, with only starting material being recovered. It was still observed to modify guanine, although the yield of **12** was lower than that formed by reaction with bromine (21% with bromine compared with 12% with NBS). NBS was found to promote bromination at the C5-C6 bond of thymine, from either face producing diastereomers **13** and **16** in 37% and 27% yield respectively. The difference between the two diastereomers is most notable at the junction between the deoxyribose sugar and the nucleobase, where the ^{1}H and ^{13}C NMR spectra showed signals at 6.51 ppm and 84.8 ppm for **13** and 5.84 ppm and 87.1 ppm for **16**. The reaction of NBS with **8** resulted in a low yield of desired product **14**, with no recoverable unmodified starting material; this was due to decomposition of the starting material. Again no modification of 5mC (**10**) was noted and only significant decomposition was observed upon reaction of NBS with this nucleobase. NBS was slightly more selective than bromine, although the yields of brominated products were much lower and decomposition of the protected deoxynucleosides was more prevalent with this reagent.

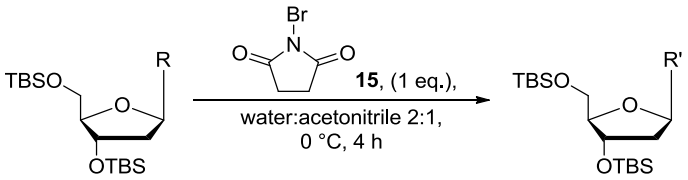


Figure 3 – 7 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2, 4, 6, 8** and **10**) by NBS (**15**).*

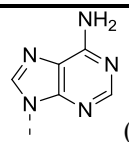
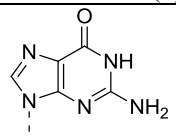
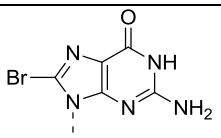
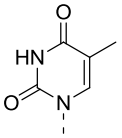
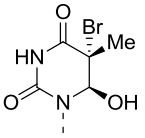
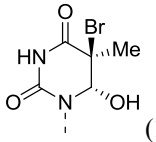
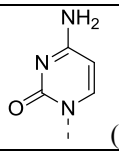
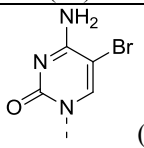
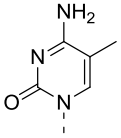
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	-	88% RSM
2	 (**4**)	 (**12**)	12% of **12** 68% RSM
3	 (**6**)	 (**13**)  (**16**)	64% of **13** and **16** (d.r = 58:42)
4	 (**8**)	 (**14**)	25% of **14**
5	 (**10**)	-	43% RSM

Table 3 – 7 – *The yields for the reaction of NBS (**15**) with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) (RSM refers to recovered starting material).*

2,4,4,6-Tetrabromocyclohexa-2,5-dienone (TBCD, **17**) has been employed in the Donohoe group for the selective bromination of the pyrrolidinone ring of the natural product microsclerodermin F (work carried out by J. Kershaw and C. Winter).[200] It was noted that TBCD was a less reactive and more selective brominating source than bromine and NBS. This brominating agent was also examined for selective modification of the *bis*-TBS protected deoxynucleosides (Figure and Table 3 – 8). TBCD was indeed observed to be a much milder source of bromine because it resulted in significantly less decomposition of the starting material, as the yields of recovered unmodified A, C and 5mC were much higher than that previously observed with NBS. However, the selectivity of the reaction showed little change from that observed by NBS, as the same three bases C, T and G reacted. This reagent, like bromine, was also observed to produce a single diastereomer of the bromohydrin **13**.

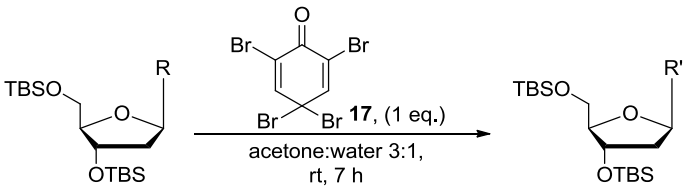


Figure 3 – 8 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2, 4, 6, 8** and **10**) by TBCD (**17**).*

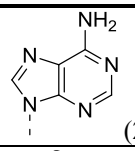
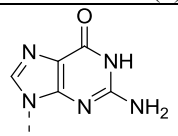
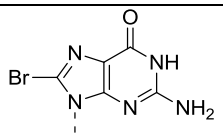
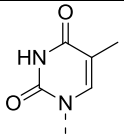
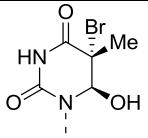
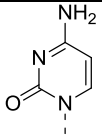
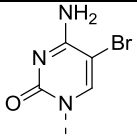
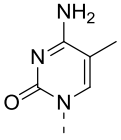
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	-	100% RSM
2	 (**4**)	 (**12**)	11% of **12** 40% RSM
3	 (**6**)	 (**13**)	19% of **13** 75% RSM
4	 (**8**)	 (**14**)	19% of **14** 75% RSM
5	 (**10**)	-	93% RSM

Table 3 – 8 – *The yields for the reaction of TBCD (**17**) with the bis-TBS protected deoxynucleosides (**2, 4, 6, 8** and **10**) (RSM refers to recovered starting material).*

The brominating agent 1,3-dibromo-5,5-dimethylimidazolidine-2,4-dione (**18**) was used by Magnus *et al.* for the formation of the bromohydrin of an isolated double bond, in an intermediate in the synthesis of codeine.[201] The reaction promoted the formation of a bromonium ion, which was subsequently opened under the aqueous conditions. This bromine source was tested in the hope that it would selectively produce the bromohydrin diastereomers of T (**13** and **16**) (Figure and Table 3 – 9). As predicted, this reagent formed the bromohydrins of T (**13** and **16**) in the highest combined yield observed so far (53% of **13** and 32% of **16**). However, it also reacted to a certain extent with the *bis*-TBS protected analogues of A, C and G. In the case of G, these conditions not only promoted bromination

but they resulted in the removal of the TBS protecting group on the primary hydroxyl of the sugar, affording **19** in a yield of 37%. Interestingly, this reagent was reactive enough to also form the bromohydrins of 5mC (**20** and **21**) in quantitative yield (56% and 44%). The two bromohydrin distereomers of 5mC were identified by the characteristic upfield shift of the proton at C6 from 7.52 ppm in **10** to 5.40 ppm and 5.03 ppm in the ^{1}H spectra of the respective bromohydrins (the exact stereochemistry was not defined for **20** and **21** because the diastereomers were not crystalline). Despite this reagent's increased propensity to form the bromohydrins of T and 5mC, the reagent was unselective and was not investigated further.

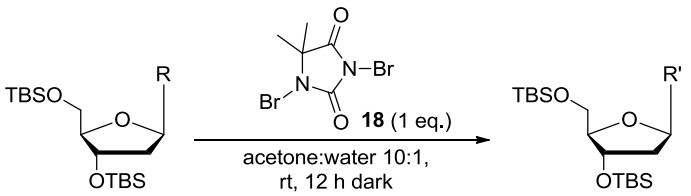


Figure 3 – 9 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) by **18**.*

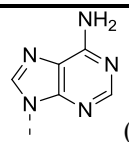
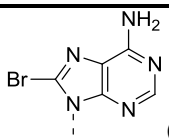
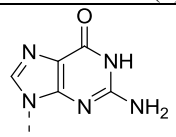
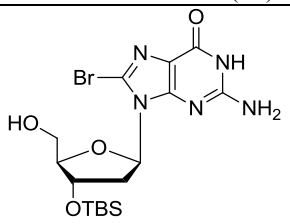
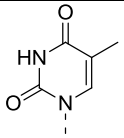
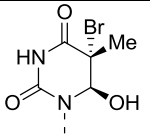
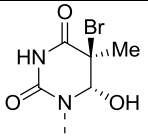
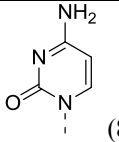
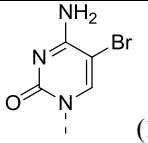
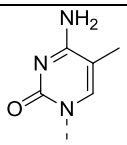
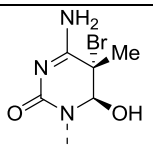
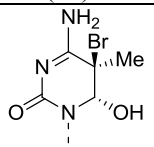
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	 (**11**)	46% of **11** 54% RSM
2	 (**4**)	 (**19**)	37% of **19**
3	 (**6**)	 (**13**)  (**16**)	85% of **13** and **16** (d.r. = 62:38)
4	 (**8**)	 (**14**)	17% of **14**
5	 (**10**)	 (**20**)  (**21**)	100% of **20** and **21** (d.r. = 56:44)

Table 3 – 9 – *The yields for the reaction of* ***18*** *with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) (RSM refers to recovered starting material).*

Conditions for the oxidative bromination of unsaturated C-C bonds, mediated by Selectfluor® (**22**) and potassium bromide, were developed by Shreeve and co-workers.[202] By employing these conditions, they were able to brominate an alkene that was conjugated to a carbonyl group. This is a similar system to that of the double bond of thymine, except that T is substituted at the five position. These conditions were tested on the five *bis*-TBS protected nucleosides in the hope that they may show selectivity towards thymine (Figure and Table 3 – 10). These conditions provided a reasonably high yield of the bromohydrins of T (combined yield of **13** and **16** was 67%). However, the reaction selectivity was poor as the conditions used promoted a reaction with A, G and C as well as causing complete decomposition of 5mC (**10**). This particular reagent produced a 20% yield of the brominated guanine base **12**. However, the reaction conditions also resulted in the removal of the TBS protecting group, resulting in the formation of the brominated base **19**. If the

reaction conditions promoted the removal of the TBS groups then the unprotected bases would be soluble in the aqueous work-up conditions and would not be recovered, which could account for the decomposition of 5mC and lack of recovered starting material for the other bases. Due to the lack of selectivity and high level of decomposition these conditions were not investigated further.

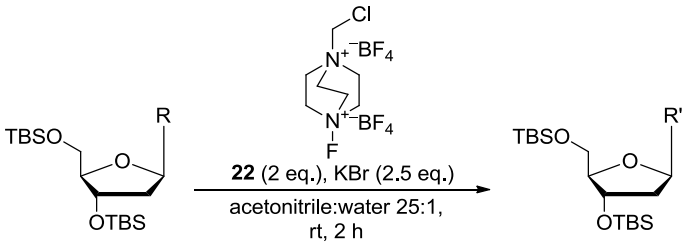


Figure 3 – 10 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) by Selectfluor® (**22**) and potassium bromide.*

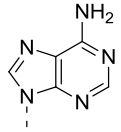
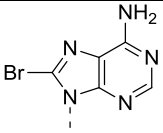
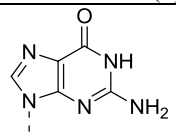
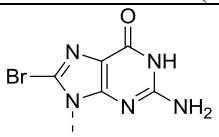
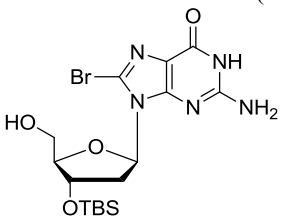
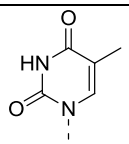
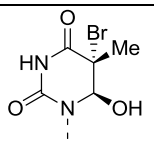
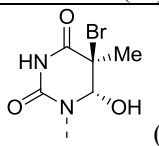
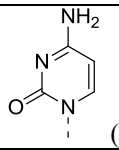
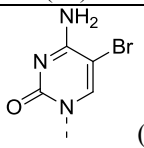
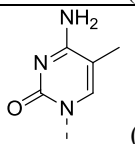
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	 (**11**)	43% of **11** 25% RSM
2	 (**4**)	 (**12**)  (**19**)	20% of **12** 36% of **19**
3	 (**6**)	 (**13**)  (**16**)	67% of **13** and **16** (d.r. = 52:48)
4	 (**8**)	 (**14**)	57% of **14**
5	 (**10**)	-	Decomp.

Table 3 – 10 – *The yields for the reaction of Selectfluor (**21**) and potassium bromide with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) (RSM refers to recovered starting material).*

Another source of bromine that was examined for the selective modification of the *bis*-TBS protected deoxynucleosides was tetrabutylammonium tribromide (TBATB). This reagent was investigated under two different sets of conditions one of which was carried out at room temperature, the other at 0 °C.[203, 204] The reaction at room temperature in a tetrahydrofuran:water (1:1) mix was tested first (Figure and Table 3 – 11).[203] These conditions were found to be highly reactive towards the pyrimidine bases, as high levels of decomposition were noted under these conditions and only a small amount of **13**, **14** and **16** were detected. It was, however, observed to be less reactive towards the purine bases producing only a trace of **11** and a low yield of **12**. As these conditions were highly reactive, lower temperature conditions which also employed a shorter reaction time were investigated (Figure and Table 3 – 12).[204] The milder brominating conditions were found to be the most selective tested. High levels of unmodified G, A and 5mC were recovered and no decomposition was observed for C and T. A 20% yield of the single diastereomer **13** was produced by the reaction of T with TBATB at 0 °C in *N*,*N*-dimethylformamide:water (7:3), with a high yield of unreacted **6** also recovered. A similar yield and level of unmodified cytosine was also observed (19% of **14** and 64% RSM). The lower temperature and shorter reaction time made a significant difference to the reaction selectivity when using tetrabutylammonium tribromide. However, these conditions, despite being the most selective tested, were not further optimised because a reaction which promoted complete base selectivity was required.

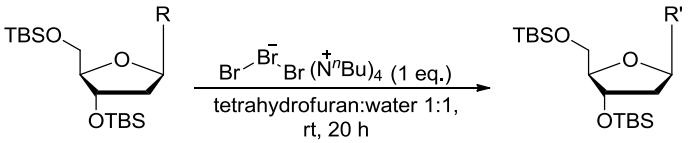


Figure 3 – 11 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) by TBATB at room temperature.*

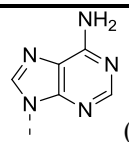
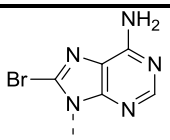
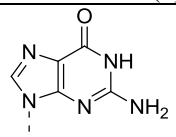
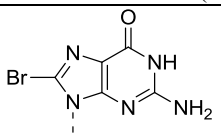
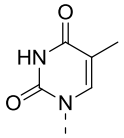
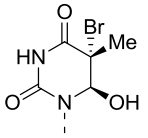
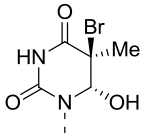
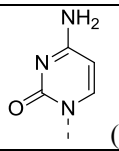
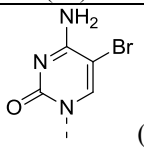
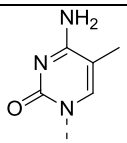
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	 (**11**)	1% of **11** 78% RSM
2	 (**4**)	 (**12**)	17% of **12** 48% RSM
3	 (**6**)	 (**13**)  (**16**)	11% of **13** and **16** (d.r. = 79:21) 8% RSM
4	 (**8**)	 (**14**)	20% of **14**
5	 (**10**)	-	Decomp.

Table 3 – 11 – *The yields for the reaction of TBATB with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) (RSM refers to recovered starting material).*

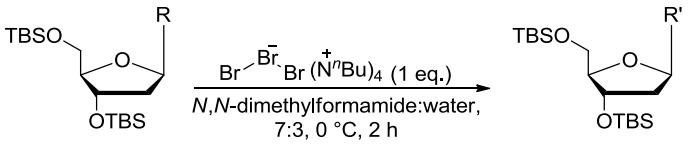


Figure 3 – 12 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) with TBATB at 0 °C.*

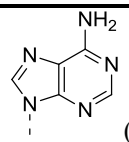
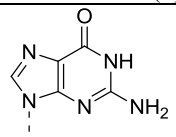
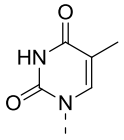
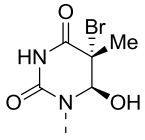
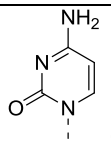
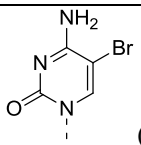
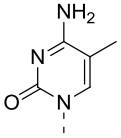
Entry	R	R'	Yield R:R' (%)
1	 (**2**)	-	100% RSM
2	 (**4**)	-	95% RSM
3	 (**6**)	 (**13**)	20% of **13** 73% RSM
4	 (**8**)	 (**14**)	19% of **14** 64% RSM
5	 (**10**)	-	85% RSM

Table 3 – 12 *– The yields for the reaction of TBATB with the bis-TBS protected deoxynucleosides (**2, 4, 6, 8** and **10**) (RSM refers to recovered starting material).*

In summary, seven different sets of conditions, and a variety of different bromine sources were tested for the selective modification of DNA. All brominating agents were found to produce a variety of different products upon reaction with the bases, however, all of those tested also reacted with at least two different bases. Therefore, no further optimisation was carried out for selective halogenation of the DNA bases.

3.2.3 – Selective Modification of the *bis*-TBS Protected Deoxynucleosides by the Palladium-Catalysed Heck Reaction

Palladium-catalysed cross-coupling reactions provide great versatility in organic synthesis. Various palladium reagents have been used to couple functionally derivatised DNA[193, 195, 197] as well as to form metal complexes with the bases themselves.[205-207] It was proposed that palladium could be used to couple additional aromatic functional groups directly onto individual DNA bases. If this is possible it would prove an extremely versatile method of directly derivatising DNA strands. Although there are many types of

palladium-catalysed cross-coupling reactions, the most useful for this system would be the Heck reaction. This is due to the fact that it could be used directly with the DNA bases rather than requiring the addition of another functional unit e.g. a boronic ester for Suzuki cross-couplings. The standard conditions employed for Heck reactions, generally involve heating at high temperature (~80 °C) in organic solvent under basic conditions. Although the Heck reaction has been carried out under aqueous conditions, the high temperature and basic conditions employed would not be tolerated by DNA.[208-211]

In 2009, Davis *et al.* published conditions for the aqueous Suzuki-Miyaura Cross coupling reaction of aryl boronic acids to a model protein.[212] They had developed an aqueous soluble palladium catalyst (**23**) that could successfully couple amino acids to aromatic boronic acids in high yields at 37 °C in phosphate buffer in the presence of oxygen. These conditions appeared ideal for the coupling of DNA to aryl groups. It was proposed that by optimisation of the reaction conditions, it could be possible to use the newly developed catalyst to promote Heck reactions with the *bis*-TBS protected deoxynucleosides. It was thought that the reaction was most likely to affect coupling with cytosine as its C5-C6 double bond was least sterically hindered. To test the activity of the palladium catalyst **23**, iodo-tyrosine analogue **24** was successfully coupled to phenyl boronic acid producing **26** in 70% yield (literature value was 94% yield, Figure 3 – 13).[212]

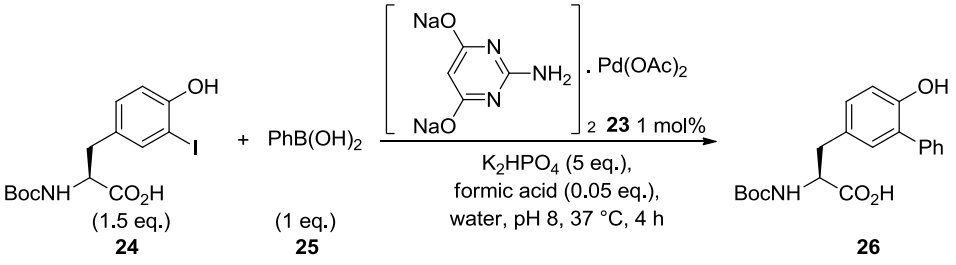


Figure 3 – 13 – *The reaction of tyrosine analogue (**24**) with phenyl boronic acid (**25**), catalysed by an aqueous soluble palladium catalyst (**23**).*[212]

The reaction conditions were then investigated for the selective palladium catalysed Heck reaction of cytosine (**8**) with boc-3-iodo-L-tyrosine (**24**) (Figure and Table 3 – 14). In all

cases the reaction was left for 24 h to allow sufficient time for the Heck reaction to proceed. Initially, a 1 mol% catalyst loading, employed by Davis and co-workers,[212] was used to catalyse the reaction, however, under these conditions no reaction was observed. High levels of **8** were recovered suggesting that the reaction was not resulting in decomposition of the starting materials. The catalyst loading was increased ten-fold in an attempt to promote the reaction, however, this was unsuccessful and again high levels of **8** were recovered. It was hypothesised that because of the TBS protecting groups, employed to decrease the polarity of the bases, **8** was not sufficiently soluble in the reaction mixture. Therefore a 1:1 mixture of water and acetonitrile was tested, which again resulted in the isolation of **8** only. These conditions (Table 3 – 14, entry 3) were also tested on the other *bis*-TBS protected bases, to see if the Heck reaction could be promoted for T, G, A or 5mC (Figure 3 – 15). However, only unreacted starting material was recovered in all cases. After optimisation of the conditions, originally developed for the Suzuki-Miyaura palladium-catalysed cross-coupling, it was established that these were not compatible for the implementation of the Heck reaction. Owing to the lack of other DNA compatible reaction conditions for promotion of the Heck reaction,[211] no further investigations were carried out on palladium couplings of the DNA bases.

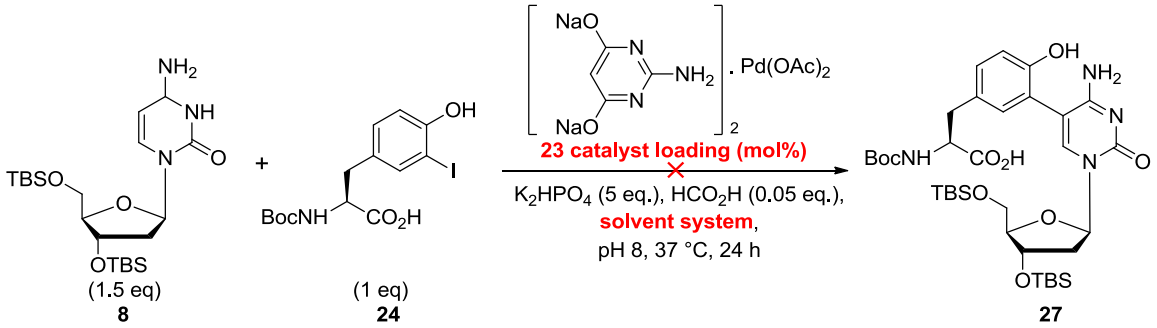


***Figure 3 – 14** – The conditions that were employed in Table 3 – 14 for the palladium-catalysed Heck reaction of **8** with **24**. The catalyst loading and solvent system used were investigated for their effect on the yield of **27**.*

Entry	23 Catalyst Loading (mol %)	Solvent System	Yield 8:27 (%)
1	1	100% water	95% RSM
2	10	100% water	98% RSM
3	10	1:1 water:acetonitrile	81% RSM

Table 3 – 14 – *Conditions tested for the optimisation of the palladium-catalysed Heck reaction of* ***8*** *and* ***24*** *upon variation of the ratio of the catalyst loading and the solvent system (RSM refers to recovered starting material).*

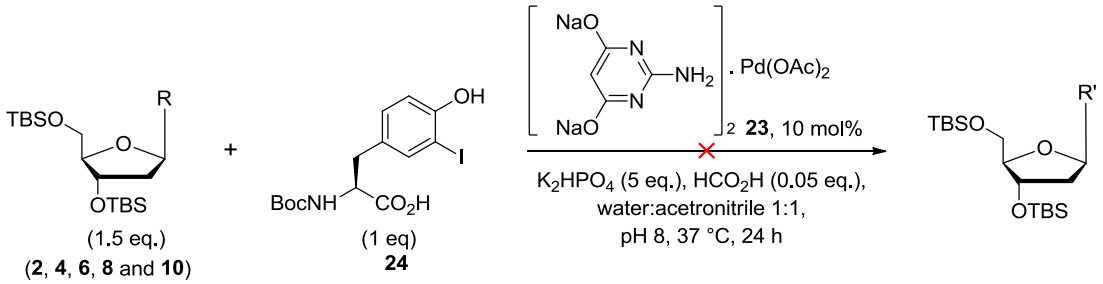


Figure 3 – 15 – *Reaction conditions for the palladium catalysed Heck reaction of the bis-TBS protected deoxynucleosides (****2****,* ***4****,* ***6****,* ***8*** *and* ***10****) with* ***24****. No reaction was observed with any of the bases under the conditions shown and >80% RSM was obtained in all cases.*

3.2.4 – Selective Modification of the *bis*-TBS Protected Deoxynucleosides by Diels-Alder Reactions

Owing to a desire to make chemical reactions more environmentally friendly, there has been an increased interest in those which can occur in pure water (the ultimate 'green solvent').[208] One such reaction which has been widely and successfully employed in an aqueous environment is the Diels-Alder reaction – one of the most synthetically useful reactions in organic chemistry for the synthesis of highly complex six-membered rings.[213] In 1980, work by Breslow and co-workers highlighted that certain Diels-Alder reactions could occur in water with drastically improved reaction rates, in comparison to those observed with organic solvents.[214] It was clear from work subsequently carried out by Griesbeck, that the reaction conversion and *endo/exo* selectivity could be improved and tailored when water was used as the solvent at various different overall reaction concentrations (Figure and Table 3 – 16).[215] It was proposed that it might be possible to use this reaction, which favours implementation in an aqueous environment, for the

selective modification of one of the DNA bases. There was no literature precedent indicating that Diels-Alder reactions have been carried out directly on any of the unmodified DNA bases themselves, however, Diels-Alder reactions have occurred in the presence of DNA[216-218] and have even been catalysed by it.[219] It was hoped, that owing to the Diels-Alder reaction's sensitivity to sterics, cytosine could be selectively modified under these conditions.

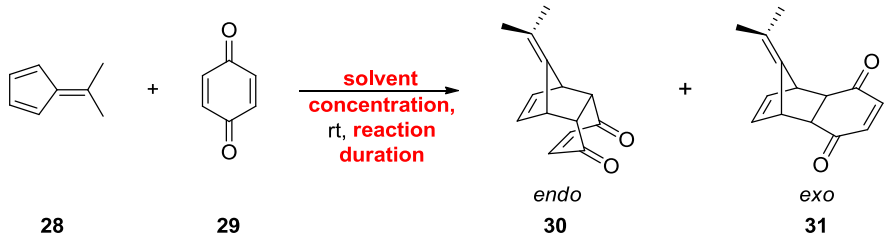


***Figure 3 – 16** – The two possible products from the Diels-Alder reaction of **28** with **29**. The solvent and concentration of a 1:1 mixture of **28** and **29** were investigated for their effect on the yield of **30** and **31**. Figure adapted from Griesbeck 1988.*[215]

Entry	Solvent	Concentration of a 1:1 mixture of 28 and 29 (M)	Reaction duration	% Conversion	*Endo:Exo*
1	carbon tetrachloride	0.15	14 d	40	55:45
2	methanol	0.15	10 d	95	44:56
3	water	0.15	11 h	100	65.35
4	water	0.0010	24 h	-	14:86
5	water	1.60	8 h	-	88:12

***Table 3 – 16** – Illustration of how the % conversion and endo vs. exo selectivity was affected by the type of solvent used and the concentration of a 1:1 mixture of **28** and **29**. Table adapted from Griesbeck, 1988.*[215]

Owing to its high reactivity in Diels-Alder reactions, cyclopentadiene was initially investigated for reaction with the *bis*-TBS protected deoxynucleosides.[220] The Diels-Alder reaction of cyclopentadiene (**32**) was initially carried out on two test substrate dienophiles (**33** and **34**) (Figure and Table 3 – 16), before advancing onto the *bis*-TBS protected deoxynucleoside system. Dienophiles **33** and **34** were selected as test substrates because they are less volatile versions of substrates which have previously been shown to successfully undergo Diels-Alder reactions in water (e.g. methyl acrylate and dimethyl

maleate).[220] The reaction of **32** with **33** was high yielding in both ethanol and water (Figure 3 – 17, entries 1 and 3). However, the reaction of **32** with **34** clearly illustrated the rate enhancement that can be experienced when the Diels-Alder reaction is performed in water in comparison to organic solvents (Figure 3 – 17, entries 2 and 4). Due to the low molecular weight of cyclopentadiene and the small scale on which the DNA modification reactions were performed, it was proposed that an excess of cyclopentadiene could be employed in these reactions. Therefore, **32** was reacted with **33** in a ten-fold excess to establish whether the reaction yield would be retarded by dimerisation of the excess **32**, pleasingly this was not the case (see Table 3 – 17, entries 3 and 5).

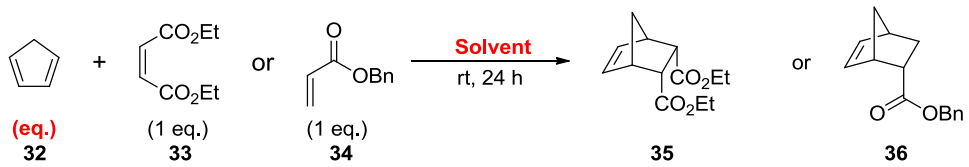


***Figure 3 – 17** – The Diels-Alder reaction of **32** with **33** and **34**. The equivalents of **32** employed and the solvent system used were investigated for their effect on the yield of **35** and **36**.*

Entry	Solvent	Dienophile	Diene (eq.)	Yield (%)
1	ethanol	**33**	1	72
2	ethanol	**34**	1	30
3	water	**33**	1	73
4	water	**34**	1	66
5	water	**33**	10	70

***Table 3 – 17** – The reaction conditions and yields for the Diels-Alder reaction of **32** with **33** and **34**, upon variation of the solvent, dienophile and equivalents of diene.*

The reaction of cyclopentadiene (**32**) with the *bis*-TBS protected deoxynucleosides was first investigated in pure water at room temperature (Figure 3 – 18), and under these conditions no reaction was observed with any of the DNA bases.[220] To investigate if the lack of reactivity was owing to an issue with solubility, the reaction was repeated with **2**, **4**, **6**, **8** and **10** in ethanol (a solvent in which the bases were fully soluble) and on the unprotected deoxynucleosides of A (**1**), G (**3**), T (**5**), C (**7**) and 5mC (**9**) in water. Both sets of conditions resulted in the recovery of unmodified starting material in high yields. In an

effort to encourage a Diels-Alder reaction to occur with DNA, several different Lewis acids were employed in both catalytic and stoichiometric amounts. The catalysts investigated included silver trifilate,[221] fluoroboric acid (with and without SDS as an additional surfactant)[222] and copper(II) nitrate.[223] Despite the implementation of catalysts known to improve Diels-Alder reactions, no reaction was observed with any of the conditions investigated.

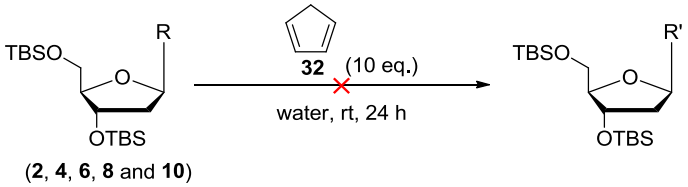


Figure 3 – 18 – *Reaction conditions for the Diels-Alder reaction of cyclopentadiene (**32**) with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**). No reaction was observed with any of the bases under the conditions shown and >90% RSM was obtained in all cases.*

Another diene which has previously shown high reactivity for Diels-Alder reactions in water is Danishefsky's diene (**37**).[221, 222] This diene was also investigated for the selective modification of one of the DNA bases by a Diels-Alder reaction. In this instance, only a three-fold excess of Danishefsky's diene was employed for the reaction of **37** with the *bis*-TBS protected deoxynucleosides, owing to its higher molecular weight (Figure 3 – 19). When **2**, **4**, **6**, **8** and **10** were treated with three equivalents of **37** at room temperature in water, no reaction was observed with any of the bases with this alternative diene source.[220] The reaction of **37** with **2**, **4**, **6**, **8** and **10** was again repeated in ethanol and in water with stoichiometric amounts of the Lewis acid catalysts silver trifilate,[221] fluoroboric acid[222] and copper(II) nitrate.[223] Without exceptions, no chemical modification of any of the DNA bases was observed under these different conditions. Owing to lack of reactivity, no further investigations were carried out to promote Diels-Alder reactions of the *bis*-TBS protected deoxynucleosides.

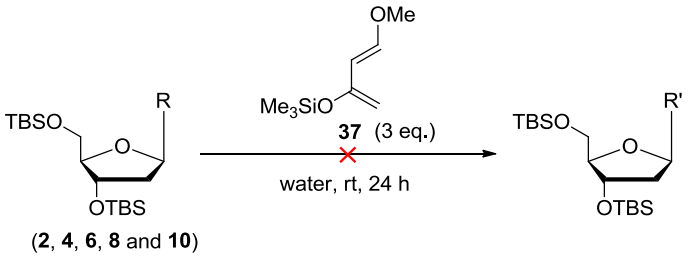


Figure 3 – 19 – *Reaction conditions for the Diels-Alder reaction of Danishefsky's diene (**37**) with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**). No reaction was observed with any of the bases under the conditions shown and >70% RSM was obtained in all cases.*

3.2.5 – Selective Modification of the *bis*-TBS Protected Deoxynucleosides by a Thiol-Michael Addition

The 1,4-addition of thiols to conjugated alkenes has attracted considerable interest for a number of reasons. The olefinic double bond of the α,β-unsaturated carbonyl species can be protected as its β-sulfido complex and can be easily regenerated by either copper(I)-induced[224] or oxidative induced elimination.[225] Organo-sulfur compounds are also extremely useful because they themselves have multiple applications in the synthesis of biologically active molecules e.g. calcium antagonist diltiazem.[226, 227] It was hypothesised that this particular method of reaction could be implemented for the selective modification of thymine in ssDNA (examples of cyclic alkenes that have undergone this transformation are shown in Figure 3 – 20).

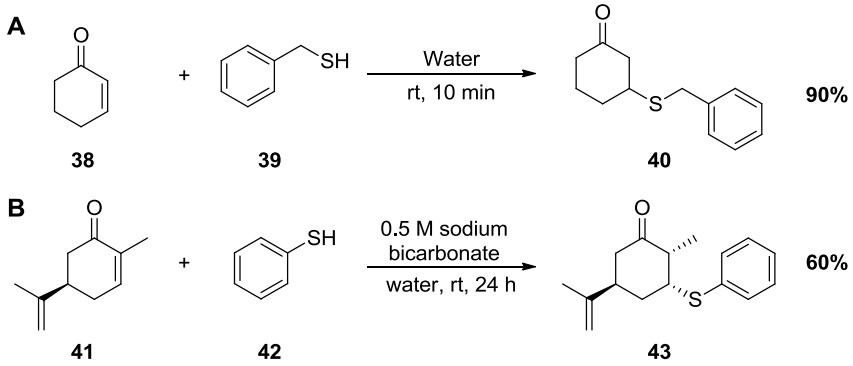


Figure 3 – 20 – *Two examples from the literature showing successful thiol-Michael additions. A) This example was from Khatik et al. 2006.[228] B) This example was from Rodrigues de Almeida et al. 2008.[229]*

Initially, the reaction between the thiol, benzyl mercaptan (**39**), and the bis-TBS protected deoxynucleosides was investigated. This thiol was chosen because there were several examples in the literature of it undergoing this type of transformation.[228, 230, 231] **39** was stirred at 35 °C for 24 h with each of the protected nucleosides **2**, **4**, **6**, **8** and **10** (Figure 3 − 21).[228] For all five bases tested, only high levels of starting material were recovered (indicating that decomposition did not occur under the reaction conditions). Several Lewis acids were investigated (on bases **2**, **4**, **6**, **8** and **10**) to see if they could promote the reaction between benzyl mercaptan and DNA; these included borax,[226] manganese(II) chloride[232] and copper tetrafluoroborate.[230] Unfortunately, only unmodified *bis*-TBS protected deoxynucleosides were recovered under these conditions. Owing to the fact that these reactions can be catalysed by both acid and base, conditions which employed sodium bicarbonate were also investigated, however, these to proved unreactive towards DNA.[229] Work by Baishya *et al.* showed that the thiol-Michael addition could also take place in an ionic liquid.[231] These conditions were tested, in the anticipation that the ionic liquid would be able to solubilise both organic reagents and enhance the rate of reaction and selectivity in this case. Unfortunately, no reaction was observed in this reaction medium either. Under all the conditions tested no reaction was observed for any of the bases with benzyl mercaptan (**39**).

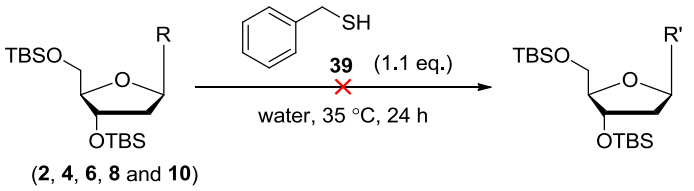


***Figure 3 – 21** – Reaction conditions for Thiol-Michael addition of benzyl mercaptan (**39**) with the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**). No reaction was observed with any of the bases under the conditions shown and >70% RSM was obtained in all cases.*

An additional thiol, that has also been successfully employed in the thiol-Michael addition, was also investigated – thiophenol (**42**).[226, 228-232] The previous Lewis acid catalysed

conditions (utilising borax,[226] manganese(II) chloride[232] and copper tetrafluoroborate[230]) were examined for the implementation of the thiol-Michael addition reaction of thiophenol (**42**) with the protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**). Again, in all cases, no reaction was observed with any of the DNA bases. Owing to the complete lack of reactivity of the system to thiol-Michael additions, investigations into these reactions ceased.

3.2.6 – Selective Modification of the *bis*-TBS Protected Deoxynucleosides by an Osmium-Catalysed Dihydroxylation Reaction

There are many examples in the literature of modification of the C5-C6 bond, of thymine and 5-methylcytosine, to form osmate esters.[155, 156, 158, 159, 233] Experiments by Okamoto and co-workers have also shown that T was more susceptible to oxidation than 5mC.[159] Work by Barvian *et al.* also illustrated that it was possible to use various osmium oxidation conditions to dihydroxylate thymine, however, the reactivity of these conditions for other bases was not examined.[192] Therefore, it was hypothesised that it might be possible to selectively dihydroxylate T in the presence of all the other bases. If this could be achieved, it may be possible to condense an osmate ester onto the diol at a later stage to produce an even bulkier modification.

A number of dihydroxylation reactions were examined for the selective modification of the pyrimidine bases (**6**, **8**, **and 10**). The Upjohn dihydroxylation conditions, initially tested for modification of T by Barvian were investigated for modification of the pyrimidine bases (Figure and Table 3 – 22).[192] A mixture of the two diastereomers of the diol of thymine were formed in an 84% yield in a ratio of 81:19. The major diastereomer reported by Barvian was the (5*S*,6*R*) diastereomer, producing a signal at 5.22 for the C6 proton. This was used to reference and integrate the signal obtained for this proton in the ^{1}H NMR spectrum of **44** (the minor diastereomer **45** shows a signal at 4.95 ppm for this proton).

These conditions were found to be very reactive, as a high level of decomposition was observed for **8** (only 50% yield of **8** was recovered). Moreover, no dihydroxylation of this base was observed, this was likely owing to the lower reactivity of *cis* alkenes verses tri-substituted alkenes towards dihydroxylation.[234] Upon treatment of **10** under the same reaction conditions, it was noted that the C5-C6 bond of 5mC was susceptible to dihydroxylation. However, the diol of 5mC was unstable under the reaction conditions and spontaneously deaminated to give 51% yield of the two diastereomers of the thymine diol (ratio of **44**:**45** = 84:16). The instability of the 5mC diol was not unsurprising, as there are several examples in the literature of 5mC-diols deaminating to T-diol under oxidising conditions.[235, 236] Owing to lack of selectivity, these conditions were not investigated further.

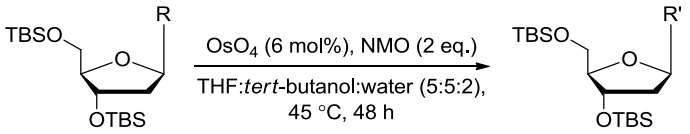


Figure 3 – 22 – *General reaction scheme for the Upjohn dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**).*

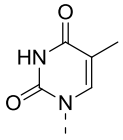
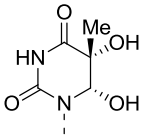
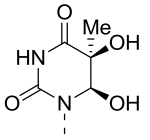
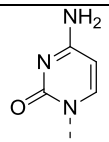
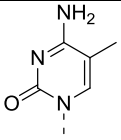
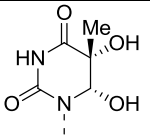
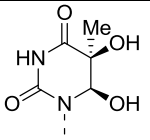
Entry	R	R'	Yield R:R' (%)
1	 (**6**)	 (**44**)  (**45**)	84% of **44** and **45** (d.r. 81:19)
2	 (**8**)	-	50% RSM
3	 (**10**)	 (**44**)  (**45**)	51% of **44** and **45** (d.r. 84:16) 29% RSM

Table 3 – 22 – *The yields for the Upjohn dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**) (RSM refers to recovered starting material).*

In 1992, Sharpless and co-workers developed the PHAL ligands for the asymmetric dihydroxylation (AD) of substituted alkenes.[237] Depending on the ligand used

($(DHQ)_2PHAL$ or $(DHQD)_2PHAL$), it was possible to preferentially dihydroxylate one face of an alkene producing a single enantiomer.[238] The conditions optimised by Sharpless and co-workers for the asymmetric dihydroxylation of tri-substituted alkenes were investigated for the selective modification of the *bis*-TBS protected deoxynucleosides.[238] Initially, the use of the $(DHQ)_2PHAL$ was investigated (Figure 3 – 23). This ligand should have promoted the dihydroxylation of thymine from the more hindered pro-(5*S*,6*R*) face (mismatched). Unfortunately due to the $(DHQ)_2PHAL$ ligand promoting the formation of the disfavoured diastereomer, the reaction conditions resulted in complete decomposition of **6**. These conditions did not promote any reaction with 5mC or C (82% and 60% RSM was obtained respectively). The $(DHQD)_2PHAL$ ligand was then examined for the formation of the favoured diastereomer of T (5*R*,6*S,* **44**) (matched). Pleasingly, it produced **44** in 59% yield with complete diastereoselectivity. These conditions were not found to modify **8** but did cause a very low conversion of 5mC to the diol of T (8% yield with 98:2 ratio of **44**:**45**, Figure and Table 3 – 24). To try and improve the rate of reaction, the solid pre-made AD-mix β was also tested because inhomogeneities in the reaction mixture, brought about by the separate addition of all the individual components, have previously resulted in reaction rate variations.[238] The pre-made AD-mix β was tested and showed comparable results to when each reagent was added separately. These conditions were much more selective than those previously tested and had potential for further optimisation.

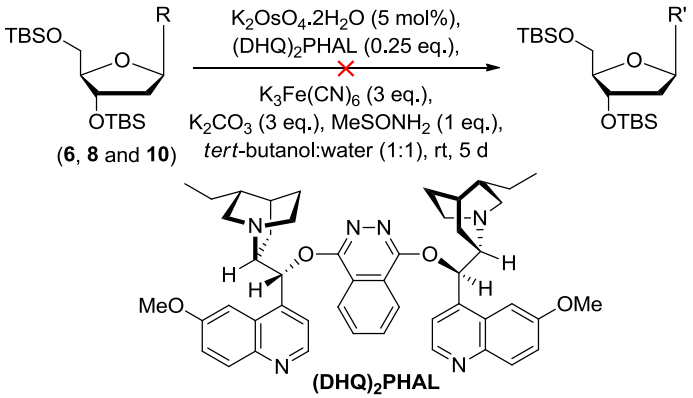


Figure 3 – 23 – *General reaction scheme for the asymmetric dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**) (employing the (DHQ)$_2$PHAL ligand).*

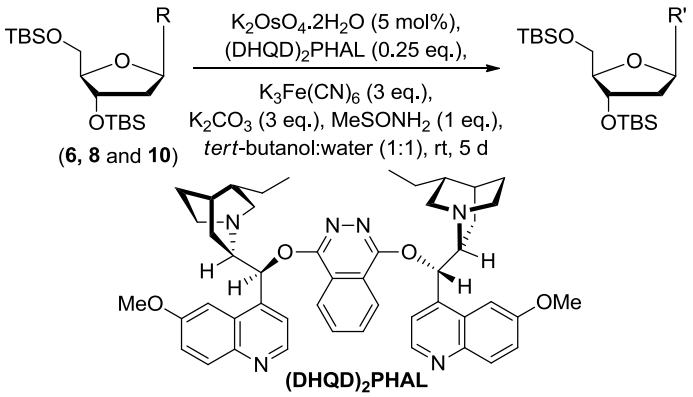


Figure 3 – 24 – *General reaction scheme for the asymmetric dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**) (employing the (DHQD)$_2$PHAL ligand).*

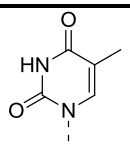
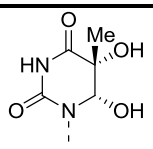
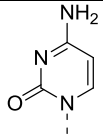
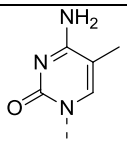
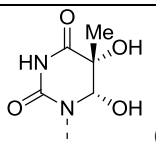
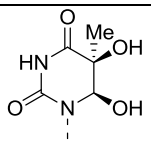
Entry	R	R'	Yield R:R' (%)
1	 (**6**)	 (**44**)	59% of **44**
2	 (**8**)	-	89% RSM
3	 (**10**)	 (**44**)  (**45**)	8% of **44** and **45** (d.r. 98:2) 58% RSM

Figure 3 – 24 – *The yields for the asymmetric dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**) (employing the (DHQD)$_2$PHAL ligand) (RSM refers to recovered starting material).*

Another set of dihydroxylation conditions, also developed by Sharpless, which employed the additive citric acid were tested.[239] The use of citric acid in the dihydroxylation of

olefins has three distinct advantages; 1) it maintains the pH of the reaction in its optimum pH range (4-6) because it neutralises 4-methymorpholine (the bi-product produced from the co-oxidant 4-methylmorpholine-*N*-oxide (NMO)); 2) it acts as a ligand for osmium stabilising the osmium(VI) form, therefore, preventing disproportionation to Os(VIII) and Os(IV) (an insoluble form of osmium that precipitates out of the reaction solution, therefore, reducing the amount of available osmium to catalyse the reaction); 3) it results in a very clean reaction, thus, making product isolation easier. This final set of conditions produced an 89% yield of a 90:10 mixture of **44** and **45** in only 36 h. As expected, the diastereoselectivity was similar to that observed for the Upjohn conditions (d.r. 81:19). Neither **8** nor **10** were observed to react under these conditions. For this particular application, it was not necessary to produce a single diastereomer of T, therefore, the citric acid dihydroxylation conditions (Figure and Table 3 – 25) were considered to be the best and most selective conditions examined. In order for these conditions to be tested on ssDNA it was necessary for them to occur under buffered conditions, therefore, the reaction conditions were selected for further optimisation.

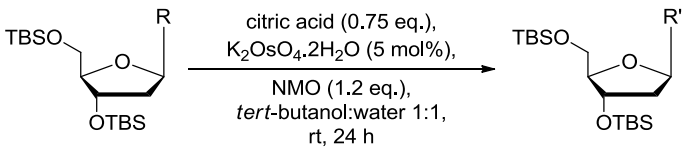


Figure 3 – 25 – *General reaction scheme for the citric acid dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**).*

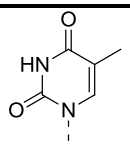
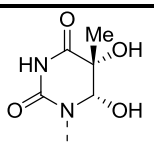
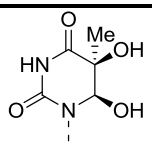
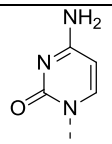
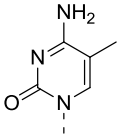
Entry	R	R'	Yield R:R' (%)
1	 (**6**)	 (**44**)  (**45**)	89% of **44** and **45** (d.r. 90:10)
2	 (**8**)	-	89% RSM
3	 (**10**)	-	96% RSM

Table 3 – 25 – *The yields for the citric acid dihydroxylation of the bis-TBS protected deoxynucleosides (**6**, **8** and **10**) (RSM refers to recovered starting material).*

3.2.7 – Optimisation of the Citric Acid Dihydroxylation of Thymine

It was necessary to further optimise the reaction conditions for the citric acid dihydroxylation owing to the fact that the pH of the reaction solution was ~4. This pH is likely to be too acidic for the modification of ssDNA as the conditions may have resulted in unwanted side reactions e.g. deamination and depurination.[140, 189] Initially, the reaction duration was reduced to 24 h and repeated with thymine (the conditions and percentage conversion of **6** to **44** and **45** that were observed are shown in Figure 3 – 26). The yields quoted for the formation of the diol of thymine are denoted as percentage conversions because it was not possible to separate **6** from **44** and **45** using flash chromatography (the % conversion was established by determining the relative integrals in the ^{1}H NMR of the C6 proton in **6**, **44** and **45**).

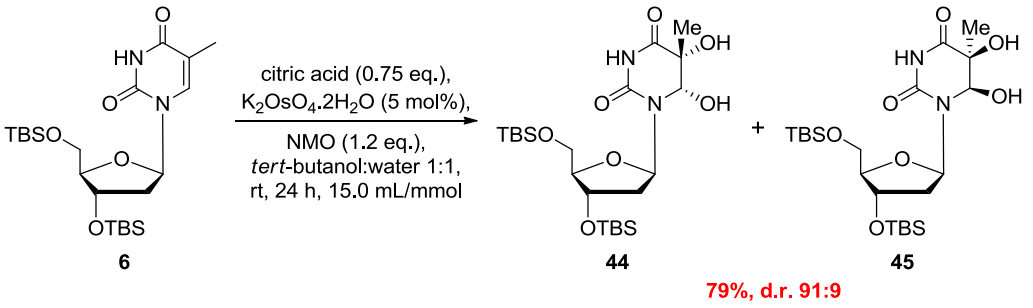


Figure 3 – 26 – *The standard un-optimised conditions for the citric acid dihydroxylation of **6**.*

Before attempting to buffer the reaction, investigations were made into how the co-solvent, overall concentration and the ratio of water to co-solvent affected the percentage conversion to the diols **44** and **45**. Sharpless and co-workers had observed that the dihydroxylation reaction, carried out in the presence of citric acid, could be performed using either *tert*-butanol or acetonitrile as a co-solvent with water.[239] When acetonitrile was employed in the conditions described in Figure 3 – 26 the yield decreased to 58% (d.r. = 81:19 (**44**:**45**)). As a result, all future experiments were carried out with *tert*-butanol as the preferred organic solvent.

The effect of concentration of the overall reaction mixture was then investigated. A significant improvement in reaction conversion was observed at an overall concentration of 30 mL/mmol (93% conversion (d.r. = 94:6)), however, at 60 mL/mmol the reaction conversion was retarded in comparison to the original system (61% conversion (d.r. = >99:1)). Concentrations lower than 15 mL/mmol were not investigated, owing to lack of solubility of the reagents. Finally, the effect of the ratio of *tert*-butanol to water was probed (see Figure and Table 3 – 27). A marked enhancement in % conversion was obtained at a 3:2 ratio of *tert*-butanol:water (Figure and Table 3 – 27, entry 2). The ratio of *tert*-butanol:water and the overall reaction concentration were investigated further, once optimised buffer conditions were obtained.

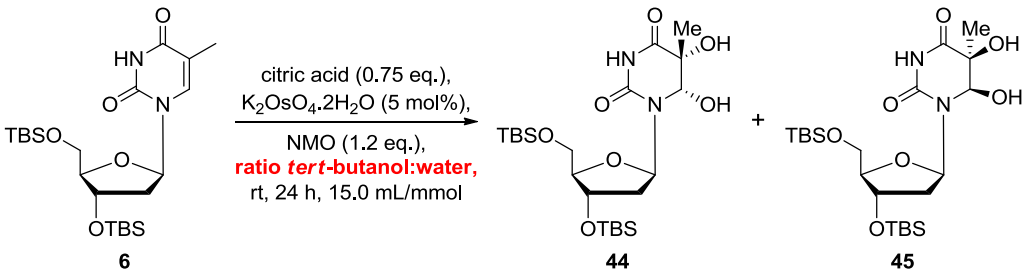


***Figure 3 – 27** – The conditions that were employed in Table 3 – 27 for the citric acid dihydroxylation of **6**. The ratio of tert-butanol:water was investigated for its effect on the % conversion of **6** to **44** and **45**.*

Entry	Ratio *tert*-butanol:water	% Conversion	d.r. (44:45)
1	4:1	77	98:2
2	3:2	100	84:16
3	1:1	79	91:9
4	2:3	70	98:2
5	1:4	68	89:11

Table 3 – 27 – *The calculated % conversions and d.r.'s for the dihydroxylation reaction (Figure 3 – 27), upon variation of the ratio of tert-butanol:water.*

The main concern with the original conditions (shown in Figure 3 – 26) was that the reaction pH was too low for selective modification of ssDNA. Therefore, several buffers at two different pH's (5 and 6) were investigated for the conversion of **6** to **44** and **45** (Figure and Table 3 – 28). The table includes the buffering agent and the solution used to alter the pH to the desired value (e.g. pH 5 or 6). The use of citric acid/sodium hydroxide and sodium citrate/hydrochloric acid had a significant effect on the % conversion reducing it by almost half when compared to the original conditions (Figure 3 – 26). However, the use of potassium citrate/hydrochloric acid showed a marked improvement when compared to sodium citrate/hydrochloric acid (52 and 60% conversion at pH's 5 and 6 for potassium citrate/HCl when compared to 26 and 39% conversion for sodium citrate/hydrochloric acid). The greatest % conversion observed for the buffered reactions was when potassium citrate/citric acid buffer was used. This could be due to the fact that no new counter ions had been added to the system, which could have reduced the reaction efficiency (i.e. potassium was already present owing to the potassium osmate and the citrate anion was present from the citric acid). The concentration of potassium citrate/citric acid was chosen so that when the buffer was added to the reaction mixture there would be 0.75 eq. of citrate anion present in the solution. The pH 5 and 6 potassium citrate/citric acid buffers were used for all subsequent experimental optimisation.

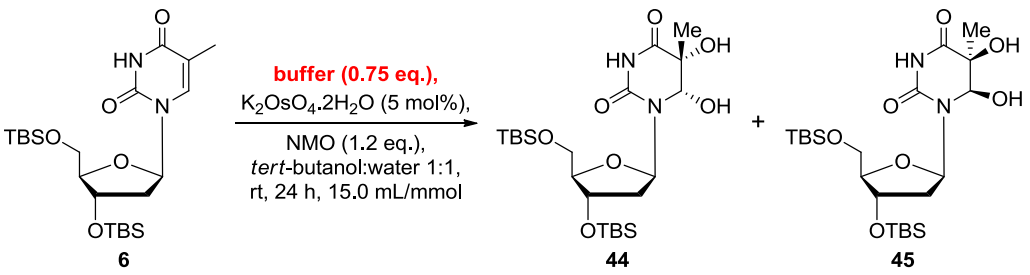


Figure 3 – 28 – *The conditions that were employed in Table 3 – 28 for the citric acid dihydroxylation of* ***6****. The buffer, that was used for stabilisation of the reaction pH, was investigated for its effect on the % conversion of* ***6*** *to* ***44*** *and* ***45****.*

Entry	Buffering Reagent	Solution used to alter Buffer pH	pH	% Conversion	d.r. (44:45)
1	99 mM citric acid	1 M sodium hydroxide	5	40	100:0
2	99 mM citric acid	1 M sodium hydroxide	6	39	100:0
3	99 mM sodium citrate	1 M hydrochloric acid	5	26	100:0
4	99 mM sodium citrate	1 M hydrochloric acid	6	39	100:0
5	99 mM potassium citrate	1 M hydrochloric acid	5	52	100:0
6	99 mM potassium citrate	1 M hydrochloric acid	6	60	99:1
7	99 mM potassium citrate	99 mM citric acid	5	72	89:11
8	99 mM potassium citrate	99 mM citric acid	6	87	100:0

Table 3 – 28 – *The calculated % conversions and d.r.'s for the dihydroxylation reaction (Figure 3 – 28), upon variation of the buffer.*

Several more aspects of the dihydroxylation required additional investigation to determine whether the reaction conversion could be further improved from 87% conversion at pH 6 in potassium citrate/citric acid buffer (Table 3 – 28, entry 8). The mol% of potassium osmate was increased from 5 to 10, 15 and 25 mol%, at both pH 5 and 6, however, this did not result in a marked increase in % conversion (see Chapter 9, appendix 9.2 for yields and d.r. values). As Sharpless had previously shown that this could be beneficial for increasing the rate of some slower dihydroxylations,[239] the amount of citrate anion present in the reaction was doubled to 1.5 eq. However, in this system no significant yield increase was observed. Finally, the previously observed improvements in reaction conversion, that were

detected by altering both the overall concentration and the ratio of *tert*-butanol:water, were investigated in the potassium citrate/citric acid buffered system (Figure and Table 3 – 29). It was found that although these alterations improved the reaction conversion when water was used as a co-solvent, when potassium citrate/citric acid buffer was used this enhancement in % conversion was not observed.

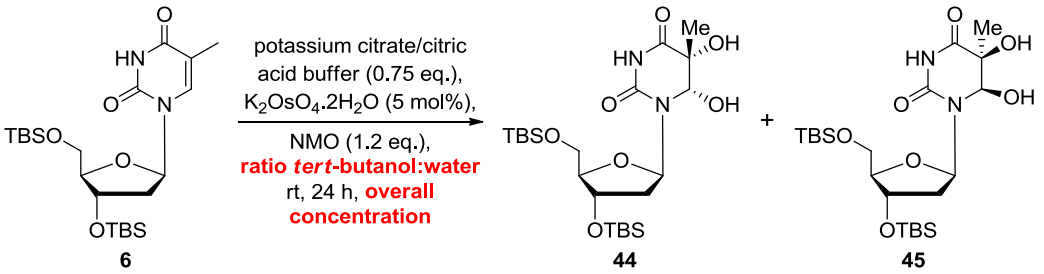


Figure 3 – 29 – *The conditions that were employed in Table 3 – 29 for the citric acid dihydroxylation of* ***6****. The ratio of tert-butanol:water and the overall concentration were investigated for their effect on the % conversion of* ***6*** *to* ***44*** *and* ***45****.*

Entry	pH	Overall Concentration	Ratio *tert*-butanol:water	Yield %	d.r. (44:45)
1	5	15 mL/mmol base	1:1	72	89:11
2	6	15 mL/mmol base	1:1	87	100:0
3	5	30 mL/mmol base	1:1	44	97:3
4	6	30 mL/mmol base	1:1	28	98:2
5	5	15 mL/mmol base	3:2	53	99:1
6	6	15 mL/mmol base	3:2	39	99:1
7	5	30 mL/mmol base	3:2	37	100:0
8	6	30 mL/mmol base	3:2	42	100:0

Table 3 – 29 – *The calculated % conversions and d.r.'s for the dihydroxylation reaction (Figure 3 – 29), upon variation of the overall concentration and solvent ratio.*

After optimisation, the conditions which led to the highest conversion of thymine to its corresponding distereomeric diol was 5 mol% potassium osmate, 1.2 eq. NMO, potassium citrate/citric acid (0.75 eq.) buffer, pH 6, *tert*-butanol:water (1:1), overall concentration of 15 mL/mmol, at room temperature for 24 h (Table 3 – 29, entry 2). These conditions were

tested on the other bases A, C, G and 5mC, to see if selectivity was maintained under the buffered conditions (Figure 3 – 30). Pleasingly, no modification of the *bis*-TBS protected deoxynucleosides was observed under these reaction conditions. The next step, was to test the optimised and selective conditions on ssDNA, to see if the selectivity for thymine was maintained in this system (Chapter 4).

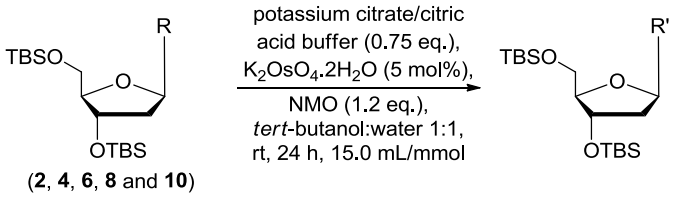


Figure 3 – 30 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **6**, **8** and **10**) by the optimised buffered dihydroxylation conditions. The conditions shown resulted in a 87% conversion to the diols **44** and **45**, for the modification of thymine (**6**), and >80% RSM was obtained for A, C, G and 5mC.*

3.2.8 – Investigations into the Formation of Osmate Esters Directly from the Diol of Thymine

It has been shown previously by others that it was possible to selectively form osmate esters of both thymine and 5-methylcytosine.[155-159] Although both bases reacted to form these osmium-containing metal complexes, thymine had been observed to be the most reactive of the two bases.[159] It was proposed that the diols of thymine (**44** and **45**, formed by the optimised citric acid dihydroxylation conditions) could be further modified by the subsequent formation of an osmate ester. This would enable further modification of thymine, into a bulkier and possibly more easily identified base, as the increased size of the modified complex could result in a greater current block when immobilised in the αHL pore during electrical recording experiments.

Now that conditions have been successfully developed for the selective dihydroxylation of T (Chapter 3, sections 3.2.6 and 3.2.7), it was necessary to establish if it was possible to form and isolate stable osmate esters of thymine and 5-methylcytosine (the two bases

which have been previously shown to react to form osmate esters).[155, 156, 158, 159] If the TMEDA (*N*,*N*,*N*',*N*'-tetramethylethylenediamine) osmate esters of thymine were found to be stable then the direct condensation onto the diols of T (**44** and **45**) could be investigated.

Donohoe and co-workers have illustrated that it was possible to isolate osmate esters of diols derived from cyclic allylic alcohols and trichloroacetamides.[240] This procedure was used to attempt the isolation of the *bis*-TBS protected deoxynucleoside TMEDA osmate esters of T and 5mC (Figure 3 – 31). The osmate esters **46, 47** and **48** were formed in almost quantitative yields (and in the case for 5mC with complete diastereoselectivity). The TMEDA osmate esters were easily identified by the characteristic four individual singlet peaks in the ^{1}H NMR (all of which integrated to 3 protons) which corresponded to the four methyl groups of the TMEDA when bound to the osmate ester complex. A distinctive osmium isotope splitting pattern was also produced in the MS spectrum. Diastereomer **46** was assigned as the major product by analogy to the facial selectivity observed by osmium complexes for the dihydroxylation of thymine. As compound **48** was not crystalline, it was not possible to determine the stereochemistry of the osmate ester.

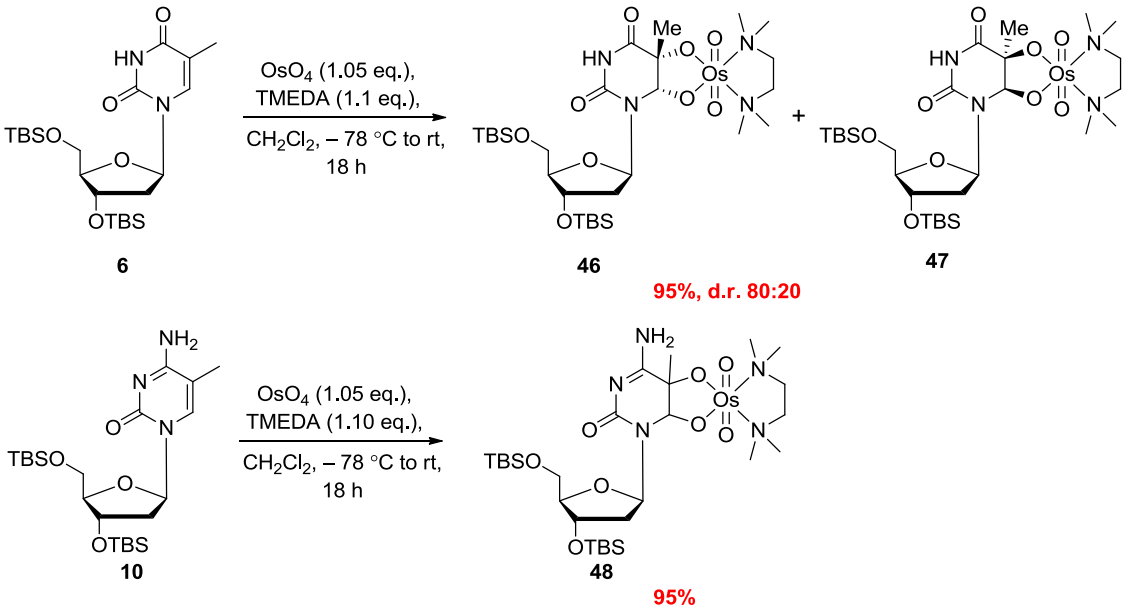


***Figure 3 – 31** – The conditions employed for the formation of the bis-TBS protected deoxynucleoside TMEDA osmate esters of T and 5mC (**46**, **47** and **48**).*

With osmate esters **46, 47** and **48** shown to be stable to column chromatography, it was necessary to investigate a new system for the condensation of an osmate ester directly onto the diols of thymine (**44** and **45**) under aqueous conditions. Unlike the previous conditions, these needed to be carried out with water as a co-solvent in the temperature range 0-40 °C and pH range 5-9. In 2008, Donohoe and co-workers developed conditions for the condensation of a TMEDA osmate ester onto an *N*-tosyl amino alcohol at pH 7.[241] These conditions were tested on an inseparable mixture of **44** and **45** and, after 24 h, 96% of the TMEDA osmate esters **46** and **47** were produced with a d.r. of 94:6 (Figure 3 – 32). By comparison of the ^{1}H and ^{13}C NMRs of the major diastereomer, produced from the condensation of osmium onto **44**, it was possible to confirm that **46** was the major disatereomer produced by direct osmate ester formation onto the alkene of **6**. The scope of the reaction for the employment of a variety of different ligands was further investigated because it was proposed that bulkier amines could facilitate better base discrimination when investigated by electrical recording experiments. To this end, two bulkier amine ligands were tested for the condensation of an osmate ester onto **44** and **45** – 2,2'-bipyridine (BIPY) and 2,10-phenanthroline (PHEN) (Figure 3 – 32). An almost quantitative yield of the BIPY osmate esters **49** and **50** was obtained (96% yield, d.r. of 95:5) but the reduced reactivity of the PHEN ligand meant only a 51% yield was obtained of these osmate esters (**51** and **52**, d.r. of 93:7). Both types of osmate ester exhibited the distinctive osmium isotope splitting pattern in the MS spectrum. In order to investigate the rate with which complexation occurred, reactions to form the TMEDA and BIPY osmate esters were repeated for only 2 h. The TMEDA osmate ester was isolated in 84% yield (d.r. 93:7) and the BIPY osmate ester in 54% yield (d.r. 94:6), indicating that the ligand acceleration effect was greater for TMEDA than BIPY.[240] A two-step, one-pot procedure was investigated for the selective modification of thymine (**6**) by dihydroxylation and subsequent osmate ester condensation. However, only 18% of the desired TMEDA osmate

esters (**46** and **47**, d.r. 80:20) and 80% of the diols of thymine (**44** and **45,** d.r. 75:25) were detected. Further investigation showed that the presence of both NMO and citric acid (which are necessary for step 1) inhibited the formation of the osmate esters **46** and **47**. Therefore, as a one-pot procedure was not viable, it was necessary to investigate methods for buffering the reaction to form the osmate esters of thymine *via* a two-step procedure.

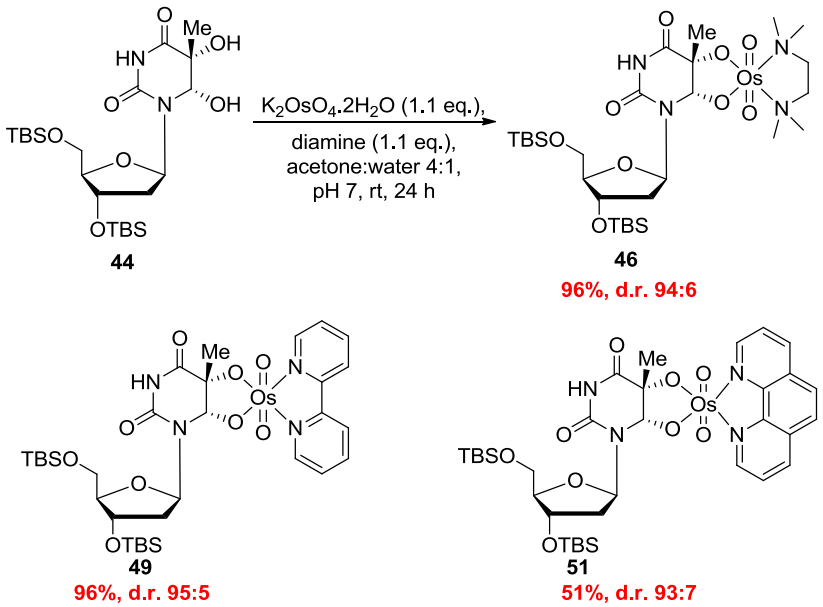


Figure 3 – 32 – *The aqueous conditions employed for the condensation of the TMEDA, BIPY and PHEN osmate esters onto the two diastereomers of the diols of thymine (**44** and **45**). In all cases, for simplicity, only the chemical structures of the major diastereomers were depicted (the stereochemistry of the osmate ester products was inferred from the major diastereomer of the diol **44**).*

Although the condensation of an osmate ester onto the diol of thymine occurs at pH 7, a solution of potassium osmate was initially added to the reaction mixture and then the pH altered to pH 7 by the addition of 1 M hydrochloric acid. Once the addition of the osmium reagent was complete, the initial pH of the solution was around 11 which was too high for experimentation on ssDNA. In order to maintain the pH of the reaction at pH 7 from the start, 10 mM phosphate buffer was utilised as the co-solvent with acetone (Figure and Table 3 − 33). The substitution of water with 10 mM potassium phosphate buffer resulted in a dramatic drop in yield of the TMEDA osmate esters (**46** and **47**). Increasing the reaction time and altering the co-solvent ratio failed to improve the yield of the desired

product. Therefore, no further optimisation was carried out on this system owing to poor conversion to the desired product under the buffered conditions.

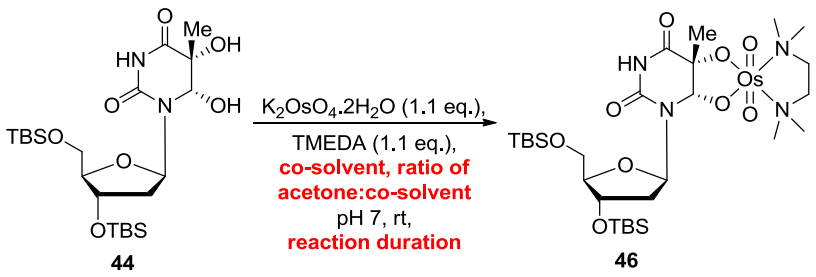


Figure 3 – 33 – *The conditions that were employed in Table 3 – 33 for the condensation of a TMEDA osmate ester onto the diols of thymine (**44** and **45**). The co-solvent, ratio of acetone:co-solvent and reaction duration were investigated for their effect on the yield of **46** and **47**. For simplicity, only the chemical structures of the major diastereomers (**44** and **46**) were depicted.*

Entry	Co-solvent	Ratio of acetone:co-solvent	Reaction duration (h)	Yield (%)	d.r. of 46:47
1	water	1:1	2	84	93:7
2	water	4:1	2	54	>99:1
3	10 mM potassium phosphate	1:1	2	6 of **46** 89 **of 44**	>99:1
4	10 mM potassium phosphate	4:1	2	10 of **46** 89 of **44**	>99:1
5	10 mM potassium phosphate	1:1	24	8 of **46** 90 of **44**	>99:1

Table 3 – 33 – *The calculated yields and d.r.s for the condensation of a TMEDA osmate ester onto the diols of thymine (shown in Figure 3 – 33) upon variation of the co-solvent, ratio of acetone:co-solvent and reaction duration.*

It was hypothesised that another source of osmium(VI) could be used for the condensation of an osmate ester onto the diol of thymine. Sharpless and co-workers observed that when osmium(VIII) tetroxide coordinated to quinuclidine it formed a bright orange complex (**53**), which upon reaction with cyclohexene turned bright green (**54**) indicating the formation of the osmate ester and the reduction of osmium(VIII) to osmium(VI).[242] This cyclohexene osmate ester complex could be employed as a source of osmium(VI) (**54**) for the condensation of an osmate ester onto the thymine diols (**44** and **45**). Under aqueous conditions it would be expected that the osmium complex would be in equilibrium, between being bound to either the cyclohexene diol or the thymine diol, because

quinuclidine is a mono-dentate ligand that does not form stable osmate esters. It was proposed that the osmium complex would be more stable when attached to thymine because it has a more substituted electron rich system (as it contains a methyl group at C5)[234] and the carbonyl oxygen (at C4) might be able to provide additional stabilisation through weak coordination to the osmium. If the thymine osmate ester was the preferred complex, then if an additional one equivalent of TMEDA was added to the reaction it would displace quinuclidine, thereby trapping out the desired stable osmate ester of T. Therefore, this system could provide an alternative source of osmium(VI) which would be easily monitored by the distinct colour changes of the various osmium complexes (Figure 3 – 34).[241-243]

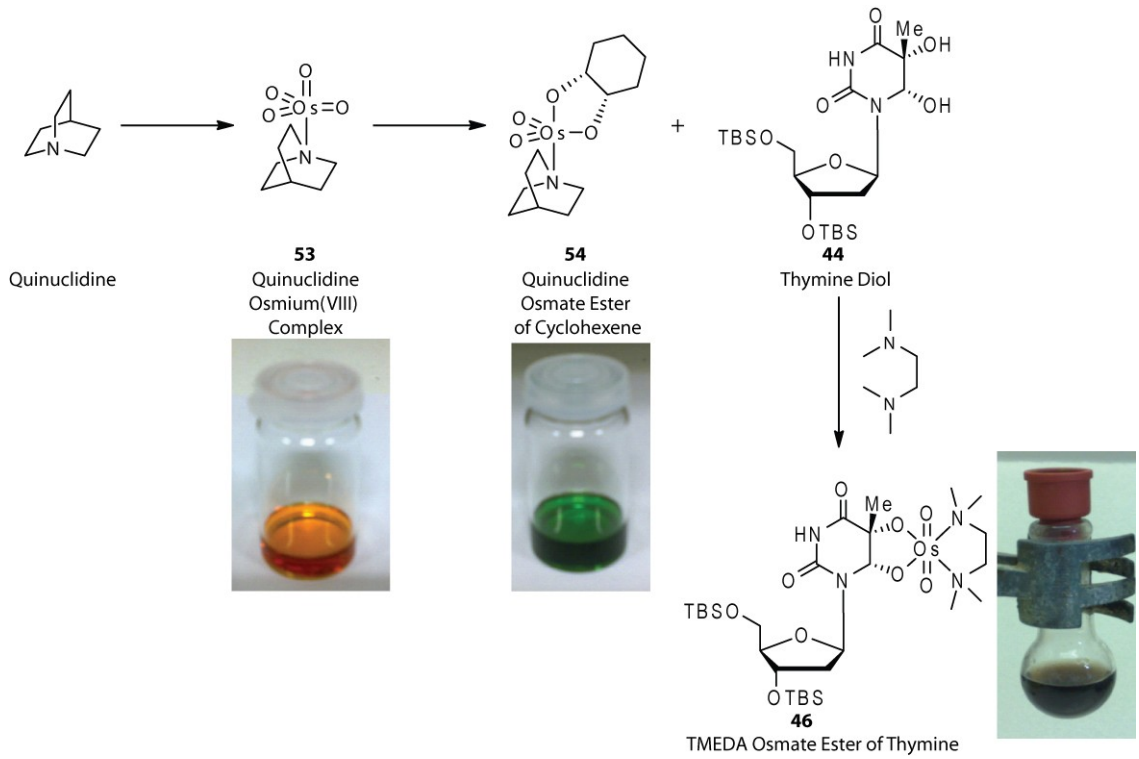


Figure 3 – 34 – *Proposed mechanism for condensation of an osmate ester onto the diols of thymine using* ***54*** *(only the major diastereomers* ***44*** *and* ***46*** *were depicted for simplicity). This scheme illustrates the characteristic colour changes that would be observed for the formation of the various osmium complexes.*

The hypothesised method for the condensation of an osmate ester onto thymine was initially investigated in both organic and aqueous solvent systems (Figure and

Table 3 – 35). In both dichloromethane and acetone:water (4:1), no osmate ester of thymine was observed when additional quinuclidine was added to the reaction solution (Table 3 – 35, entries 1 and 3). This highlighted the fact that quinuclidine was not able to form stable and isolatable osmate esters of thymidine. Pleasingly, when additional TMEDA was added to the reaction, it was possible to isolate high yields of the desired osmate ester in both organic and aqueous solvent systems (Table 3 – 35, entries 2, 4 and 5). This proved in principle that it was possible to condense an osmate ester onto the diol of thymine using complex **54** as the source of osmium(VI).

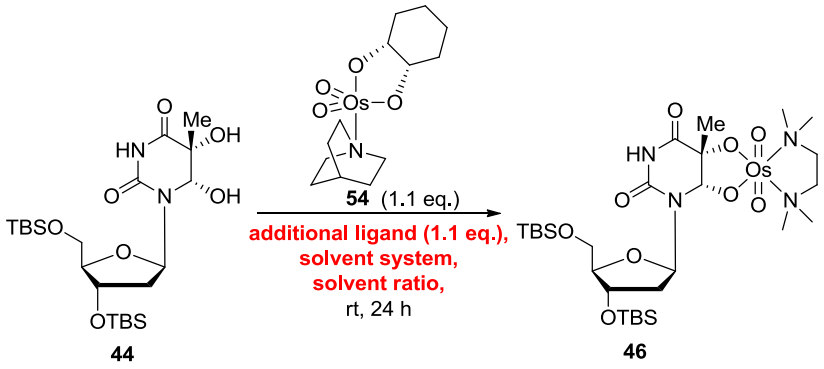


Figure 3 – 35 – *The conditions that were employed in Table 3 – 35 for the formation of a TMEDA osmate ester from the diols of thymine (**44** and **45**), using **54** as a source of osmium(VI). The ligand, solvent system and the ratio of the solvent system were investigated for their effect on the yield of **46** and **47**. For simplicity, only the chemical structures of the major diastereomers (**44** and **46**) were depicted.*

Entry	Solvent System	Solvent Ratio	Additional Ligand	Yield	d.r. 46:47
1	dichloromethane	1	quinuclidine	100% RSM	-
2	dichloromethane	1	TMEDA	88%	92:8
3	acetone:water	4:1	quinuclidine	100% RSM	-
4	acetone:water	4:1	TMEDA	67%	94:6
5	acetone:water	1:1	TMEDA	60%	92:8

Table 3 – 35 – *The calculated yields and d.r.'s for the formation of a TMEDA osmate ester from the diols of thymine (shown in Figure 3 – 35) using 54 as a source of osmium(VI). The ligand, solvent system and the ratio of the solvent system were investigated for their effect on the yield of **46** and **47** (RSM refers to recovered starting material).*

In order for the new osmium source to be suitable for reaction with ssDNA it was necessary to buffer the reaction, as the pH of the unbuffered reaction solution was recorded as ~11. Potassium phosphate (50 mM) at pH 7 (at a ratio of both 1:1 and 4:1

acetone:buffer) was tested as the buffering agent for the formation of the TMEDA osmate esters (**46** and **47**) from their corresponding diols (**44** and **45**). These conditions produced moderate yields of the desired product (Figure and Table 3 – 36, entries 1 and 2), although, these were significantly higher than those previously observed with the conditions developed by Donohoe and co-workers (Figure and Table 3 – 33).[241] In an effort to further increase the yield of **46** and **47** 2.2 equivalents of TMEDA was added to the reaction mixture, however, this reduced the yield of product by 20%. Finally, increasing the temperature of the reaction was probed for its effect on the yield of the desired osmate ester. Pleasingly, when the reaction was heated to 40 °C an increase in overall yield was observed at both solvent ratios tested (Table 3 – 36, entries 4 and 5). Therefore, the newly developed osmium(VI) source **54** was able to condense an osmate ester onto the diols of thymine (**44** and **45**) under pH 7 buffered conditions at 40 °C.

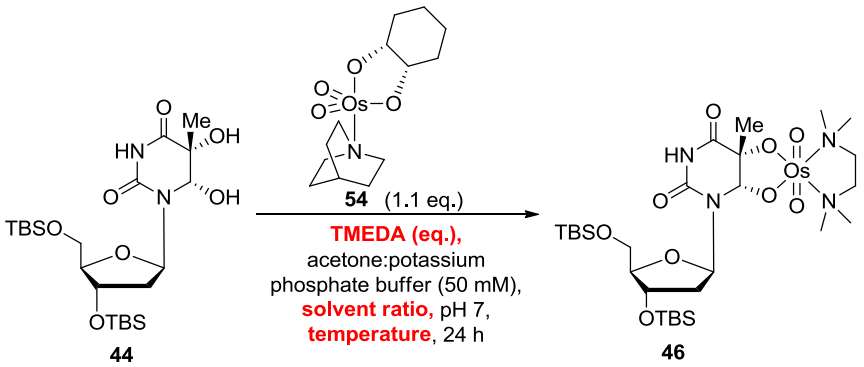


Figure 3 – 36 *– The buffered conditions that were employed in Table 3 – 36 for the formation of a TMEDA osmate ester from the diols of thymine (**44** and **45**), using **54** as a source of osmium(VI). The number of equivalents of ligand, the ratio of the solvent system and the temperature were investigated for their effect on the yield of **46** and **47**. For simplicity, only the chemical structures of the major diastereomers (**44** and **46**) were depicted.*

Entry	Solvent Ratio	Temperature (°C)	TMEDA (eq.)	Yield (%)	d.r. 46:47
1	4:1	25	1.1	63	92:8
2	1:1	25	1.1	58	94:6
3	4:1	25	2.2	40	95:5
4	4:1	40	1.1	77	93:7
5	1:1	40	1.1	66	93:7

Table 3 – 36 – *The calculated yields and d.r.'s for the formation of TMEDA osmate esters from the diols of thymine (shown in Figure 3 – 36) upon variation of the ratio of the solvent system, temperature and equivalents of ligand.*

Finally, it was necessary to test if the reaction of osmium complex **54** was selective for the modification of thymine diols. The optimised conditions (Table 3 – 36, entry 4) were tested on the *bis*-TBS deoxynucleosides of C, A, G and 5mC and no reaction was detected with any of these nucleobases (Figure 3 – 37). Therefore, the two-step selective modification of thymine to its TMEDA osmate ester had been successfully optimised for use under buffered conditions.

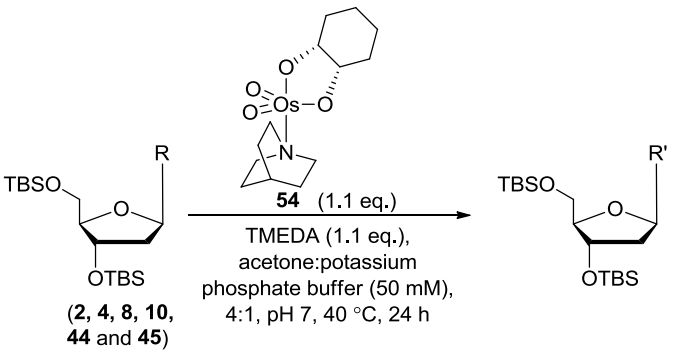


Figure 3 – 37 – *General reaction scheme for the modification of the bis-TBS protected deoxynucleosides (**2**, **4**, **8**, **10**, **44** and **45**) by the optimised TMEDA osmate ester condensation conditions. The conditions shown resulted in a 77% yield of the osmate esters **46** and **47**, by condensation onto the diols (**44** and **45**), and >75% RSM was obtained for A, C, G and 5mC.*

3.3 – Conclusions

A model system has been successfully developed for the screening of the reactivity of the canonical DNA bases, C, T, A and G, as well as the epigenetic modification 5mC. A number of different types of chemical reaction have been investigated for the selective modification of the deoxynucleosides. Halogenation conditions proved too reactive, and in

all of the conditions tested, at least two of the bases were found to be modified. Furthermore, many of the conditions examined were found to lead to significant amounts of decomposition of the starting materials (**2**, **4**, **6**, **8** and **10**). Several other types of reaction were also investigated, and were found to show complete lack of reactivity with this model system. Finally, the osmium-catalysed dihydroxylation reactions were probed and the citric acid dihydroxylation was observed to be selective for thymine. Despite the high degree of selectivity, the low pH of the reaction was not suitable for use on ssDNA. Therefore, these conditions were extensively optimised and new buffered conditions were found that resulted in 87% conversion to the diol of thymine (**44**) whilst still maintaining the reaction selectivity

It was hypothesised that an osmate ester could be directly condensed onto the diol of thymine. This would result in further modification of the thymine base producing a bulky osmate ester which could potentially be more easily identified by electrical recording experimental techniques, as the increased size of the modified complex could result in a greater current block when immobilised in the αHL pore. Conditions that were initially developed by Donohoe and co-workers, for the condensation of an osmate ester onto an *N*-tosyl amino alcohol, were unsuccessful for the formation of TMEDA osmate esters of the thymine diol under buffered conditions because only ~10% yield of the desired osmate ester was isolated.[241] However, employment of the osmium(VI) source **54**, under buffered conditions, allowed the formation of the desired TMEDA osmate ester in 77% yield without modification of any of the other bases. As the two-step process for the selective dihydroxylation and subsequent osmate ester formation of thymine had been optimised, it was necessary to investigate the modification reactions by electrical recording experiments (Chapter 4).

In conclusion, by utilising an appropriate model system (the *bis*-TBS protected deoxynucleosides) it was possible to investigate a number of chemical reactions for the selective modification of the DNA bases. Those conditions which were found to be selective (in this case selective dihydroxylation and osmate ester formation of thymine) were then optimised for use with ssDNA.

Chapter 4

Chemical Modification of the DNA Bases

Monitored by Electrical Recording Experiments

Chapter 4 – Chemical Modification of the DNA Bases Monitored by Electrical Recording Experiments

4.1 – Introduction

There has been a great deal of interest in chemical reactions which can selectively modify a single base in a strand of DNA. By reacting one of the bases, it was hoped that the base could be identified more easily for the purposes of DNA sequencing.[1] For example the selective oxidation of thymine, using potassium permanganate, formed the diol of T which was susceptible to base-catalysed strand cleavage. Therefore, the position of the thymine base in a strand of DNA was determined by running the reaction cleaved products on a polyacrylamide gel, where the cleavage band corresponded to the location of the thymine base.[164] The work included in this chapter discusses two different oxidation reactions which have been optimised for modification of one of the DNA base: 1) Carell's selective oxidation of the C5-C6 bond of 5mC and 2) the citric acid dihydroxylation which was optimised (Chapter 3) for the selective modification of thymine.

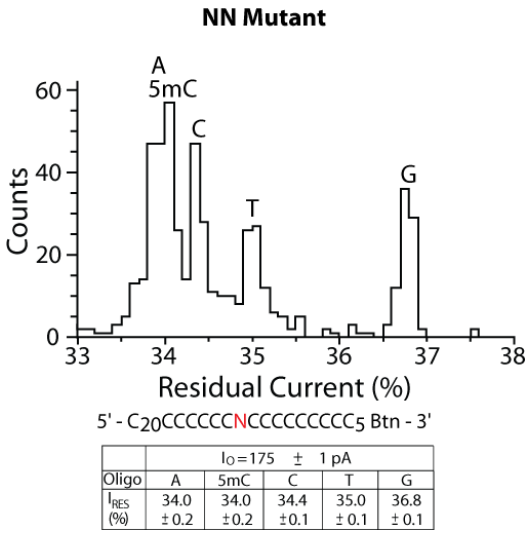
4.2 – Results and Discussion

4.2.1 – Discrimination of Unmodified Poly(dA)$N_{(14)}$ Oligonucleotides by the NN αHL Mutant

A variety of αHL mutants have been shown to discriminate C, T, A and G, when probed with homopolymeric ssDNA containing a single altered base at positions 9 (R_1), 14 (R_2) or 18 (R_3).[59-61] The E111N/K147N/M113Y has also been shown to discriminate the epigenetic modifications (5mC and hmC) as well as the canonical bases.[184] It was hypothesised that αHL mutants could be used to detect the reaction conversion and the product(s) of a selective chemical reaction of one of the DNA bases. To this end, two

oxidation reactions which had been developed for selective modification of either 5mC (Carell's oxidation) or T (citric acid dihydroxylation) were tested on ssDNA.

Poly(dA)$N_{(14)}$ oligonucleotides (where N = A, C, T, G or 5mC, located at position 14 relative to the 3' biotin TEG in 40nt long ssDNA) were utilised for the investigation of the chemical modification of the DNA bases because the oxidation conditions tested were not observed to react with adenine (the base comprising the majority of the poly(dA)$N_{(14)}$ oligonucleotide sequence, Chapter 3).[174] The NN mutant of αHL (protein prepared by E. Mikhailova) was initially probed with unmodified poly(dA)$N_{(14)}$ (where N = A, C, T, G or 5mC), to determine the base discrimination properties of this mutant (the experimental conditions used, unless otherwise stated, were 1 M KCl, 25 mM Tris, 100 μM EDTA, pH 8.0 at +160 mV, Figure and Table 4 – 1). The NN mutant was able to fully discriminate C, T, A and G but 5mC and A produced identical block levels. A different order of increasing block level and a slightly smaller $\Delta I_{RES}^{OVERALL}$ was also noted, upon probing with poly(dA)$N_{(14)}$ oligonucleotides, when compared with poly(dC)$N_{(14)}$ oligonucleotides (5mC > C/hmC > T > A > G and a $\Delta I_{RES}^{OVERALL} = 3.4 \pm 0.2\%$ (n = 4) for poly(dC)$N_{(14)}$ (Figure and Table 2 – 2A) and 5mC/A > C > T > G and a $\Delta I_{RES}^{OVERALL} = 2.8 \pm 0.0\%$ (n = 4) for poly(dA)$N_{(14)}$ (Figure and Table 4 – 1)). Poly(dA)$N_{(14)}$ oligonucleotides were also observed to block NN to a greater extent than the poly(dC). Therefore, under the electrical recording experimental conditions which were utilised for investigation of the chemical modification of the DNA bases (Chapters 4.2.2 – 4.2.3), the NN mutant was not capable of identifying a distinct block level for each of the five DNA bases tested.



	I_O = 175 ± 1 pA				
Oligo	A	5mC	C	T	G
I_{RES} (%)	34.0 ± 0.2	34.0 ± 0.2	34.4 ± 0.1	35.0 ± 0.1	36.8 ± 0.1

Figure 4 – 1 – *Base discrimination abilities of the NN pore when probed with poly(dA)$N_{(14)}$ oligonucleotides with a single substituted base (N = C, T, A, G or 5mC) at position 14 relative to the 3' biotin-TEG.*

Entry	I_O (pA)	I_{RES} poly(dA)$_{40}$ (pA)	I_{RES} C (pA)	ΔI_{RES} C %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} T (pA)	ΔI_{RES} T %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1[#]	175	34.0	34.4	0.4	34.0	0.0	35.0	1.0	36.8	2.8	2.8
2	172	34.5	34.9	0.4	34.5	0.0	35.6	1.0	37.3	2.8	2.8
3	171	35.1	35.5	0.4	35.1	0.0	36.2	1.1	38.0	2.9	2.9
4	173	34.8	35.2	0.3	34.8	0.0	35.9	1.0	37.7	2.8	2.8
Mean	173	34.6	35.0	0.4	34.6	0.0	35.7	1.1	37.4	2.8	2.8
SD	2	0.5	0.5	0.0	0.5	0.0	0.5	0.0	0.5	0.0	0.0

Table 4 – 1 – *Residual current data for the NN mutant of αHL probed with poly(dA)$N_{(14)}$ oligonucleotides with a single substituted base (N = C, T, A, G or 5mC) at position 14 relative to the 3' biotin-TEG (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} T, for example, is the difference in residual current between the poly(dA)$_{40}$ and the poly(dA)-$T_{(14)}$ levels (ΔI_{RES} T = I_{RES} T – I_{RES} A). The [#] indicates the experiment which corresponds to the histogram in Figure 4 – 1 for the NN mutant.*

4.2.2 – Electrical Recording Studies of Carell's Selective Oxidation of 5mC

Thomas Carell and co-workers investigated the differing redox-potentials of 5mC, C and T, in order to selectively oxidise the C5-C6 bond of 5mC in the presence of all other bases.[174] The conditions that were initially developed employed vanadium(V)-containing clusters to oxidise 5mC and subsequent piperidine cleavage to observe the site of reaction. However, under these conditions, cleavage was also observed at guanine bases because the vanadium species promoted direct electron abstraction producing oxidatively damaged

DNA lesions at these positions. Experimental optimisation produced conditions which were completely selective for 5mC ($NaIO_4$/LiBr, phosphate buffer, pH 5 40 °C, Figure 4 − 2). In order to determine the possible product of the oxidation reaction, Carell and co-workers analysed DNA, treated under the oxidation conditions, by mass spectrometry (MS). An increase in molecular weight of *m/z* 96 was observed, corresponding to the formation of the bromohydrin of 5mC.

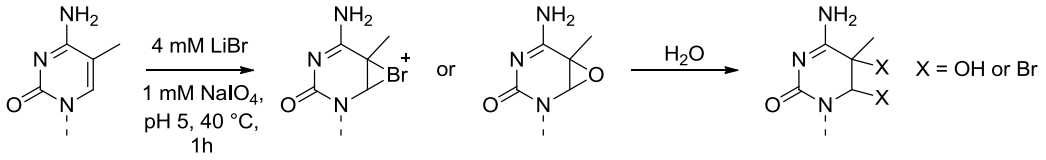


Figure 4 – 2 – *Proposed reaction intermediates for the selective chemical modification of 5mC. Figure adapted from Bareyt et al. 2008.*[174]

It was hypothesised that Carrel's conditions for the selective modification of 5mC, using sodium periodate and lithium bromide, could also be investigated by electrical recording experiments.[174] Carell and co-workers monitored the reaction selectivity by observing cleavage bands that were produced at the location of the oxidatively damaged DNA residues. Polyacrylamide gel analysis showed relatively weak cleavage bands, in comparison to the unmodified strand, indicating low reaction conversion for the selective modification of 5mC. However, Carell and co-workers observed a loss of base selectivity upon increasing both the reaction duration and temperature.[174] Therefore, to maintain selectivity the conditions that were tested for the modification of poly(dA)$N_{(14)}$ (where N = C, T, A, G or 5mC) oligonucleotides were – 1mM sodium periodate, 4 mM lithium bromide, 100 mM sodium phosphate buffer, pH 5.0 for 1 h at 40 °C.[174] Under these conditions it would be expected that only poly(dA)5m$C_{(14)}$ would be observed to react by electrical recording experiments.

The poly$(dA)_{40}$ oligonucleotide was initially tested for reactivity under these conditions, because it was necessary to confirm, that the nucleobase which comprised the majority of

the poly(dA)$N_{(14)}$ background DNA sequence, was not modified under the oxidation conditions being investigated. For the electrical recording experiments, modified poly(dA)$_{40}$ (A-CR) (that had been reacted by the conditions described above, purified and then pre-incubated with excess streptavidin at room temperature for at least 5 min) was added to the *cis* compartment and a one second pulling and releasing voltage sweep (see Chapter 7, section 7.3.5) was applied (for a minimum of 200 sweeps), to sample the various reacted oligonucleotide species. In the case of poly(dA)$_{40}$, when the modified DNA (A-CR) was added to the *cis* compartment, a single residual current level was observed in the residual current histogram (Figure 4 – 3A left). As only one residual current level was produced, this indicated that either poly(dA)$_{40}$ had completely reacted to produce a new product or no reaction had occurred. To determine which scenario was correct, a sample of unreacted stock solution of poly(dA)$_{40}$ (A) was then added to the *cis* compartment and measured (for a minimum of 200 sweeps, Figure 4 – 3A, right). This resulted in a histogram containing a single residual current population that had an increased count number (i.e. more DNA block events had occurred, in the same length of time, implying the number of ssDNA biotin•streptavidin complexes resulting in that particular block level had increased). As no new population had been observed in the histogram, and the population of the previously measured block level had increased in count number, it was assumed that adenine had not reacted under these conditions and no chemical modification had occurred. However, to confirm that this was indeed the case, a second control experiment was performed in which the unmodified poly(dA)$_{40}$ (A) was added to the *cis* compartment first to verify that this oligonucleotide produced a single block level (Figure 4 – 3B, left). Subsequent addition of A-CR again produced a single block level with increased count number. Therefore, under the reaction conditions investigated, no modification of adenine (the base which comprises the majority of the poly(dA)$N_{(14)}$ (where N = C, T, A, G or 5mC) oligonucleotide sequences) was observed,

which was supported by previous data in the literature on the base selectivity of Carell's oxidation conditions.[174] Therefore, as adenine is not modified under the reaction conditions, it will provide a suitable unreactive background oligonucleotide sequence upon which selective oxidation of 5mC, by Carell's oxidation conditions,[174] can be investigated.

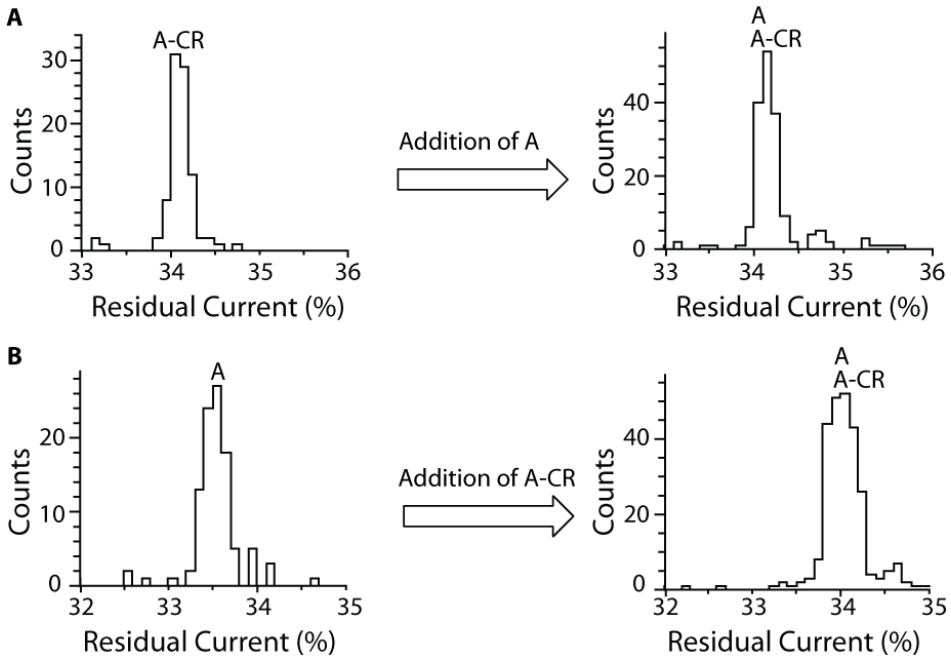


Figure 4 – 3 – *The NN pore probed with reacted and unreacted poly(dA)$_{40}$. A) The histogram on the left shows only the presence of A-CR in the cis compartment. The histogram on the right corresponds to when A was added to the cis compartment and a single residual current level with increased count number was observed when both modified (A-CR) and unmodified oligonucleotides (A) were present in the cis compartment. B) The control experiment on the left shows only the presence of A in the cis compartment – a single block level is observed in the histogram. The histogram on the right corresponds to when A-CR was added to the cis compartment and again a single residual current level with increased count number was detected, when both the modified and unmodified oligonucleotides were present in the cis compartment. This suggests that no reaction had occurred.*

Entry	I_O (pA)	I_{RES} A (pA)	I_{RES} A-CR (pA)	ΔI_{RES} (A-CR – A) %
1a[#]	156	-	34.10	-
1b[#]	158	34.15	34.15	0.00
2a[#]	166	33.52	-	-
2b[#]	164	34.02	34.02	0.00
3a	181	-	36.66	-
3b	184	36.57	36.57	0.00
Mean	168	34.57	35.10	0.00
SD	12	1.37	1.38	0.00

Table 4 – 3 – *Residual current data for the NN pore probed with reacted poly(dA)$_{40}$ (A-CR) and un-reacted poly(dA)$_{40}$ (A) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case A-CR) and 'b' denotes when both oligonucleotides are present in the cis compartment (A and A-CR). The [#] indicates the experiments which correspond to the histograms in Figure 4 – 3 (A corresponds to 1a and 1b and B corresponds to 2a and 2b).*

Carell and co-workers have previously observed direct one-electron abstraction with guanine, under some of the conditions employed for oxidation of the C5-C6 bond of 5mC.[174] The conditions tested in this system, however, have been optimised by Carell and co-workers to minimise this unwanted side reaction.[174] Poly(dA)G$_{(14)}$ was investigated for modification by the sodium periodate/lithium bromide reaction mixture and no reaction was detected with this nucleobase (e.g. a single block level, with increased count number, was detected when G and G-CR were both present in the *cis* compartment, Figure 4 – 4A, right). The control experiment, where G was added to the *cis* compartment first and G-CR second, also resulted in the observation of a single block level when both modified and unmodified poly(dA)G$_{(14)}$ samples were present (Figure 4 – 4B). Therefore, the direct one-electron abstraction had been successfully suppressed in the modification conditions tested (Figure and Table 4 – 4).

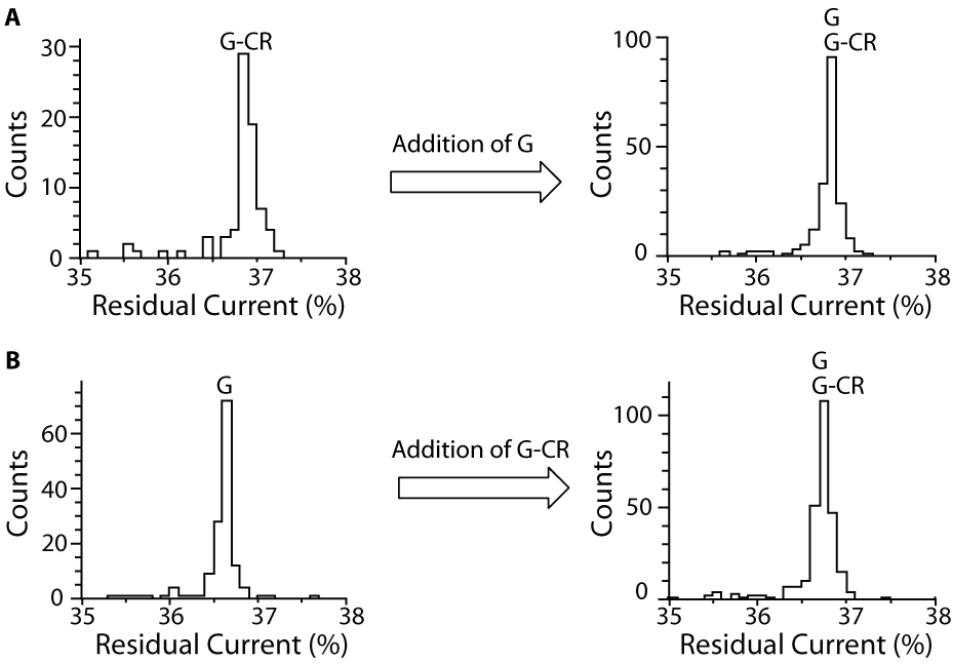


Figure 4 – 4 – *The NN pore probed with reacted poly(dA)$G_{(14)}$ (G-CR) and un-reacted poly(dA)$G_{(14)}$ (G). A) The histogram on the left shows only the presence of G-CR in the cis compartment. The histogram on the right corresponds to when G was added to the cis compartment and a single residual current level with increased count number was observed when both modified (G-CR) and unmodified oligonucleotides (G) were present in the cis compartment. B) The control experiment on the left shows only the presence of G in the cis compartment – a single block level is observed in the histogram. The histogram on the right corresponds to when G-CR was added to the cis compartment and again a single residual current level with increased count number was detected, when both the modified and unmodified oligonucleotides were present in the cis compartment. This suggests that no reaction had occurred.*

Entry	I_O (pA)	I_{RES} G (pA)	I_{RES} G-CR (pA)	ΔI_{RES} (G-CR – G) %
1a#	163	36.63	-	-
1b#	162	36.75	36.75	0.00
2b	163	36.84	36.84	0.00
3a#	160	-	36.88	-
3b#	160	36.84	36.84	0.00
Mean	162	36.77	36.83	0.00
SD	2	0.10	0.06	0.00

Table 4 – 4 – *Residual current data for the NN pore probed with reacted poly(dA)$G_{(14)}$ (G-CR) and un-reacted poly(dA)$G_{(14)}$ (G) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case G) and 'b' denotes when both oligonucleotides are present in the cis compartment (G and G-CR). The # indicates the experiments which correspond to the histograms in Figure 4 – 4 (A corresponds to 3a and 3b and B corresponds to 1a and 1b).*

The pyrimidine bases T and C were also investigated for chemical oxidation by Carell's method. Modified poly(dA)$T_{(14)}$ (T-CR) produced two distinguishable current signals, after treatment with sodium periodate and lithium bromide (Figure 4 – 5A, left). The two signals produced could have corresponded to either 1) incomplete modification of thymine to a single product, with unmodified poly(dA)$T_{(14)}$ still present in the sample (e.g. a 3:1 or 1:3 mixture of modified:unmodified poly(dA)$T_{(14)}$ (n = 1)) or 2) thymine had been completely modified to two different products in a 3:1 ratio. Upon addition of unmodified poly(dA)$T_{(14)}$ it was confirmed that incomplete modification of T had occurred, with the modified product species producing a greater current block, when immobilised in the NN pore (Figure 4 – 5A, right). The control experiment (Figure 4 – 5B) confirmed that the unmodified strand produced a single block level when added to the *cis* compartment and that two distinct block levels were observed upon subsequent addition of modified poly(dA)$T_{(14)}$ (T-CR). The electrical recording data indicated that the $\Delta I_{RES}^{(T\text{-}CR\ -\ T)} = -0.41 \pm 0.08\%$ (n = 3) (Table 4 – 5A,B). Poly(dA)$C_{(14)}$ was also found to react with the sodium periodate/lithium bromide mixture. However, cytosine was completely converted to the new product within the 1 h reaction time-period because C-CR alone produced a single block level (Figure 4 – 5C, left) and addition of unmodified poly(dA)$C_{(14)}$ (C) resulted in two distinct residual current signals (Figure 4 – 5C, right). The control experiment corroborated these findings (Figure 4 – 5D). The $\Delta I_{RES}^{(C\text{-}CR\ -\ C)}$ observed was less than that measured for T ($\Delta I_{RES}^{(C\text{-}CR\ -\ C)} = -0.29 \pm 0.03\%$ (n = 4), Table 4 – 5C,D). The observed reactivity was not predicted as Carell's oxidation conditions were selective for 5mC.[174] However, some T bases were observed, by Carell and co-workers, to react and form cleavage bands, which was explained by hypothesising that the most reactive sites e.g. 5mC had been shielded by DNA secondary structure.[174] Therefore, as the poly(dA)$T_{(14)}$ and poly(dA)$C_{(14)}$ strands did not contain any reactive 5mC's, the reagents could have modified the next most reactive bases – C and T.

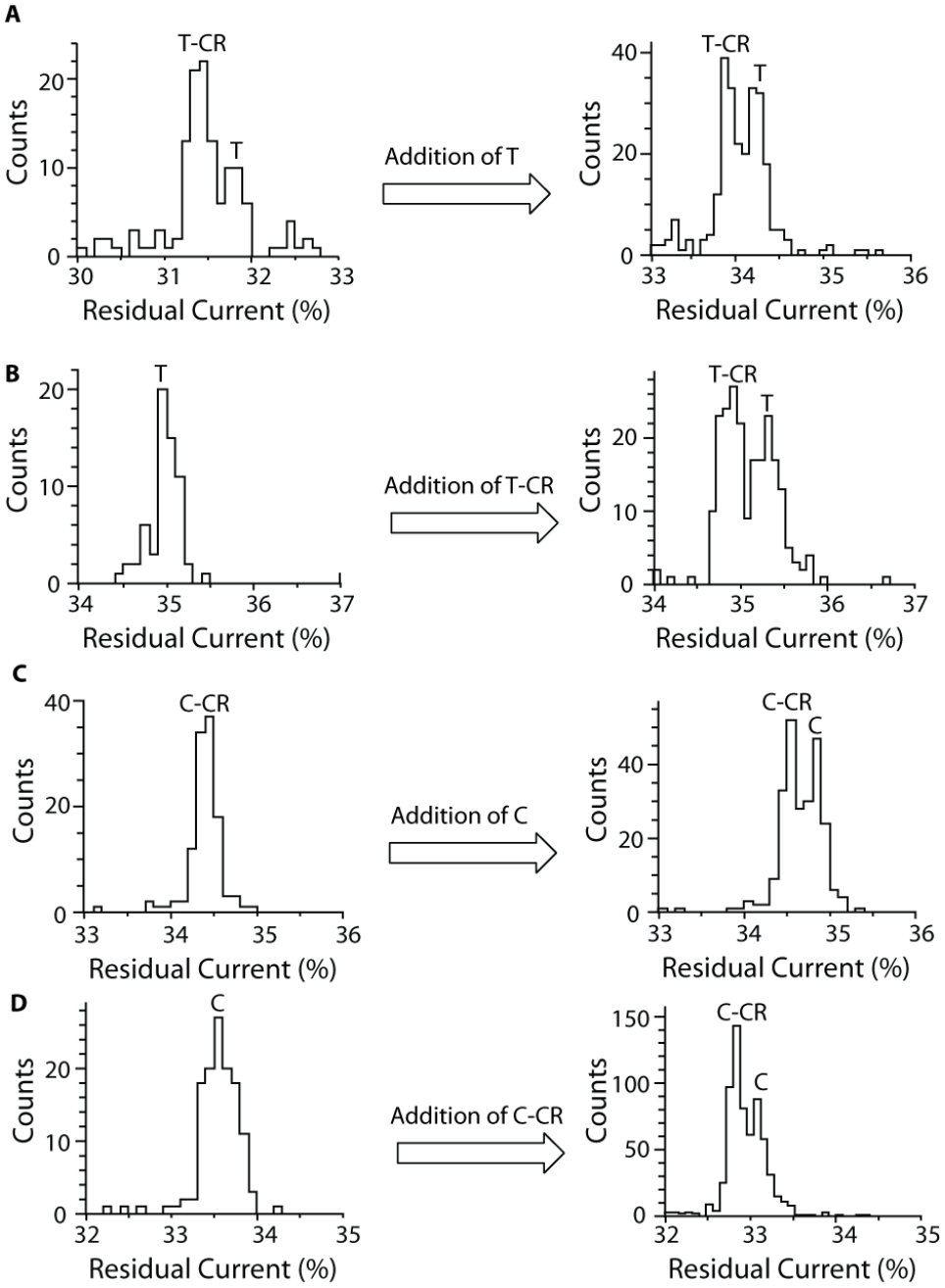


Figure 4 – 5 – *The NN pore probed with reacted and unreacted poly(dA)$T_{(14)}$ (A and B) and poly(dA)$C_{(14)}$ (C and D). A) The histogram on the left shows only the presence of T-CR – this suggested that T had undergone incomplete conversion to the product, with the modified base producing a lower residual current. The histogram on the right corresponds to when T was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – one of the residual current levels increased its count number upon addition of T indicating this peak corresponded to the un-modified poly(dA)$T_{(14)}$ (these histograms correspond to different experiments). B) The control experiment on the left shows only the presence of T in the cis compartment – a single block level was observed. The histogram on the right corresponds to when T-CR was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – the appearance of a new lower residual current block level corresponded to the product species (T-CR). C) The histogram on the left shows only the presence of C-CR – this suggested that either C had undergone complete conversion to the product or no reaction had occurred at all. The histogram on*

the right corresponds to when C was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – a new residual current level was observed upon addition of C indicating this peak corresponded to the unmodified poly(dA)$C_{(14)}$, with the modified base (C-CR) producing a lower residual current. Therefore, under these conditions complete modification of C was observe. D) The control experiment on the left shows only the presence of C in the cis compartment – a single block level was observed. The histogram on the right corresponds to when C-CR was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new lower residual current block level corresponded to product species (C-CR) (these histograms correspond to different experiments).

Entry	I_O (pA)	I_{RES} T (pA)	I_{RES} T-CR (pA)	ΔI_{RES} (T-CR – T) %
1a#	163	31.88	31.40	−0.47
2b#	169	34.23	33.92	−0.32
3a	168	35.35	35.02	−0.34
3b	170	35.37	34.92	−0.45
4a#	165	35.01	-	-
4b#	166	35.34	34.86	−0.48
Mean	167	34.53	34.02	−0.41
SD	3	1.37	1.53	0.08

Table 4 – 5A,B – *Residual current data for the NN pore probed with reacted poly(dA)$T_{(14)}$ (T-CR) and un-reacted poly(dA)$T_{(14)}$ (T) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 3a and 3b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case T-CR) and 'b' denotes when both oligonucleotides are present in the cis compartment (T and T-CR). The # indicates the experiments which correspond to the histograms in Figure 4 – 5 (A corresponds to 1a and 2b and B corresponds to 4a and 4b).*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} C-CR (pA)	ΔI_{RES} (C-CR – C) %
1a#	163	33.57	-	-
2b#	166	35.09	34.83	−0.26
3a#	164	-	34.41	-
3b#	165	34.85	34.54	−0.31
4b	165	34.82	34.55	−0.27
5a	163	-	34.96	-
5b	164	35.31	34.99	−0.33
Mean	164	34.73	34.71	−0.29
SD	1	0.68	0.24	0.03

Table 4 – 5C,D – *Residual current data for the NN pore probed with reacted poly(dA)$C_{(14)}$ (C-CR) and un-reacted poly(dA)$C_{(14)}$ (C) (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 3a and 3b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case C-CR) and 'b' denotes when both oligonucleotides are*

present in the cis compartment (C and C-CR). The # indicates the experiments which correspond to the histograms in Figure 4 – 5 (C corresponds to 3a and 3b and D corresponds to 1a and 2b).

Finally poly(dA)5mC$_{(14)}$ was investigated for its reactivity under the oxidative sodium periodate/lithium bromide conditions. When the modified (poly(dA)5mC$_{(14)}$ ssDNA (5mC-CR) was initially added to the *cis* compartment a number of distinguishable block levels were observed (Figure 4 – 6A, left). This suggested that a number of different products had been detected by the NN mutant of αHL. The mechanism that was proposed by Carell *and* co-workers showed that a number of possible intermediates and products could have been produced upon oxidation of 5mC (Figure 4 – 7).[174] It was possible that upon reaction with sodium periodate/lithium bromide 5mC had reacted and produced a number of these intermediates or end products which were stable and detectable under the electrical recording experimental conditions. However, upon addition of unmodified poly(dA)5mC$_{(14)}$ the residual current signal that resulted was a single block level over a very wide range (~2%, Figure 4 – 6A, right). This was because the individual signals, produced by the multiple products of 5mC, were broad and located relatively close together. Therefore, when additional unreacted ssDNA was added, and its signal appeared in close proximity to the signals corresponding to 5mC-CR (Figure and Table 4 – 6A), one broad residual current population was measured. The control experiment showed a single residual current level upon addition of un-modified poly(dA)5mC$_{(14)}$ (5mC) (Figure 4 − 6B, left). Subsequent addition of the modified oligonucleotide resulted in a broadening on the signal, which was indicative of the multiple products observed for reaction with 5mC under Carell's oxidation conditions. Owing to the resultant multiple signal overlap, it was not possible to calculate a value for $\Delta I_{RES}^{(5mC\text{-}CR\ -\ 5mC)}$. Therefore, by probing the NN mutant of αHL it was possible to detect multiple products from the reaction of 5mC with lithium bromide/sodium periodate.

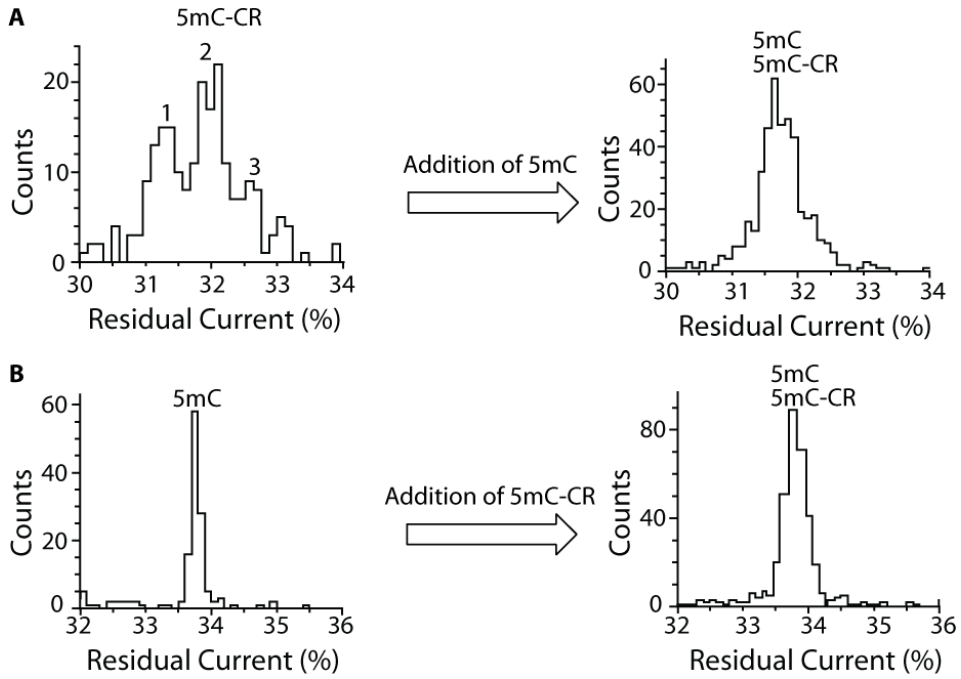


Figure 4 – 6 – *The NN pore probed with reacted poly(dA)5mC$_{(14)}$ (5mC-CR) and un-reacted poly(dA)5mC$_{(14)}$ (5mC). A) The histogram on the left shows only the presence of 5mC-CR in the cis compartment – this suggested that multiple products were produced as there was more than one residual current level. The histogram on the right corresponds to when 5mC was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, – one broad residual current level was observed. B) The histogram on the left shows only the presence of 5mC in the cis compartment – one narrow signal was produced. The histogram on the right corresponds to when 5mC-CR was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – one broad residual current level was observed.*

Entry	I_O (pA)	I_{RES} 5mC (pA)	I_{RES} 5mC-CR Major (2) (pA)	I_{RES} 5mC-CR Minor (1) (pA)	I_{RES} 5mC-CR Minor (3) (pA)
1a$^{\#}$	161	-	32.00	31.27	32.58
1b$^{\#}$	170	31.73	31.73	-	-
2a$^{\#}$	176	33.76	-	-	-
2b$^{\#}$	176	33.78	33.78	-	-
3a	172	-	34.32	33.49	35.07
3b	176	34.42	34.42	-	-
Mean	172	33.42	33.25		
SD	6	1.17	1.29		

Table 4 – 6 – *Residual current data for the NN pore probed with reacted poly(dA)5mC$_{(14)}$ (5mC-CR) and un-reacted poly(dA)5mC$_{(14)}$ (5mC) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case 5mC-CR) and 'b' denotes when both oligonucleotides are present in the cis compartment (5mC and 5mC-CR). The $^{\#}$ indicates the experiments which correspond to the histograms in Figure 4 – 6 (A corresponds to 1a and 1b and B corresponds to 2a and 2b).*

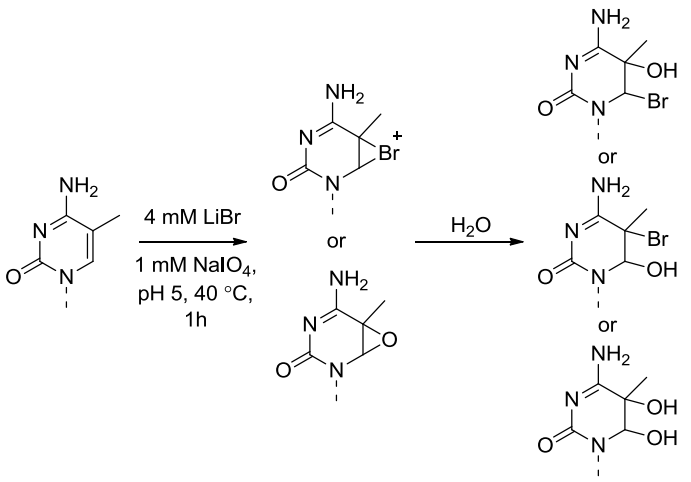


Figure 4 – 7 – *Proposed reaction intermediates for the selective chemical modification of 5mC. Figure adapted from Bareyt et al. 2008.*[174]

4.2.3 – Investigation of the Optimised Dihydroxylation Conditions by Electrical Recording Experiments

It was proposed that the newly developed conditions for the selective dihydroxylation of thymine could be investigated by electrical recording experiments. As an intermediate system between the single *bis*-TBS protected deoxynucleosides and 40 nucleotide ssDNA, the optimised dihydroxylation conditions were investigated by ^{1}H NMR spectroscopy on a 10 nucleotide ssDNA fragment (sequence 5' – GGAATGGCAA – 3'). Signals for the H5 and H6 alkene protons of cytosine and the CH_3 and the H6 protons of thymine were easily identified in the ^{1}H NMR spectrum of the 10 nucleotide ssDNA (see Chapter 9, appendix 9.3 for the ^{1}H NMR spectrum). It was also possible to distinguish the H1 proton of the deoxyribose sugar for each of the ten bases in the strand. The ssDNA fragment was subjected to the optimised dihydroxylation conditions, however, upon purification of the ssDNA, by a biological spin column, significant broadening of the ^{1}H NMR spectrum of the modified product was observed. This was likely due to introduction of paramagnetic material, such as copper(II) salts, into the sample during the purification step. Unfortunately, no accurate information relating to the reactivity and selectivity of the DNA bases in the short strand of DNA could, therefore, be determined.

Investigation of the selective chemical modification of ssDNA, by the optimised dihydroxylation conditions, was performed on the poly(dA)$N_{(14)}$ (where N = C, T, A, G or 5mC) oligonucleotide system. This particular system was selected because the *bis*-TBS protected deoxyadenosine base had not been observed to react under the dihydroxylation conditions (Chapter 3). Due to the significant difference in mass between the *bis*-TBS protected deoxynucleosides and the 40 nucleotide ssDNA, it was not possible to accurately replicate the concentration of the reaction solution of the deoxynucleosides with the larger ssDNA (66.4 mM for the deoxynucleosides, whereas the ssDNA stock solution was only 100 μM). An excess of reagents (1000 eq.) was initially employed for the selective modification of the poly(dA)$N_{(14)}$ oligonucleotides, in order to compensate for the low concentration of ssDNA and ensure that modification of the ssDNA was observed.

In order to confirm that no reaction was observed with adenine (the nucleobase that comprises the majority of the poly(dA)$N_{(14)}$ oligonucleotide background DNA sequence) the poly$(dA)_{40}$ oligonucleotide was initially investigated for modification under the optimised dihydroxylation conditions (1000 eq.). No chemical reaction was detected for the poly$(dA)_{40}$ oligonucleotide under these conditions, as the modified oligonucleotide sample (A-D) produced a single block level, which increased in count number upon addition of unmodified (A) (Figure 4 – 8A). The control experiment, where unmodified poly$(dA)_{40}$ (A) was added to the *cis* compartment first and modified poly$(dA)_{40}$ (A-D) second, was investigated, in order to confirm that A produced a single block level (Figure 4 – 8B). Under the reaction conditions tested, no modification of adenine was observed, which supports the previous experimental results obtained for the *bis*-TBS protected model system (subsection 3.2.6). Therefore, as adenine is not modified under the reaction conditions, it will provide an unreactive background oligonucleotide sequence

upon which selective oxidation of T, by the citric acid dihydroxylation conditions, can be investigated.

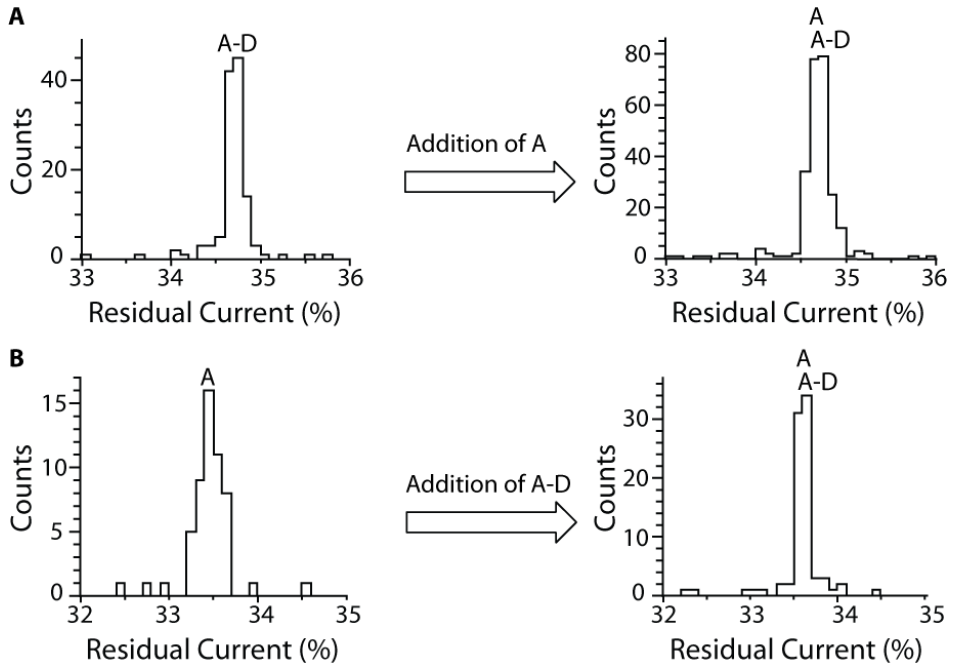


Figure 4 – 8 – *The NN pore probed with reacted poly(dA)$_{40}$ (A-D) and un-reacted poly(dA)$_{40}$ (A). A) The histogram on the left shows only the presence of A-D in the cis compartment. The histogram on the right corresponds to when A was added to the cis compartment and a single residual current level with increased count number was observed when both modified (A-D) and unmodified oligonucleotides (A) were present in the cis compartment. B) The control experiment on the left shows only the presence of A in the cis compartment – a single block level is observed in the histogram. The histogram on the right corresponds to when A-D was added to the cis compartment and again a single residual current level with increased count number was detected, when both the modified and unmodified oligonucleotides were present in the cis compartment. This suggests that no reaction had occurred.*

Entry	I_O (pA)	I_{RES} A (pA)	I_{RES} A-D (pA)	ΔI_{RES} (A-D – A) %
1a#	165	-	34.71	-
1b#	166	34.70	34.70	0.00
2a#	165	33.47	-	-
2b#	167	33.60	33.60	0.00
3a	168	-	34.15	-
3b	170	34.20	34.20	0.00
Mean	167	33.99	34.27	0.00
SD	2	0.57	0.46	0.00

Table 4 – 8 – *Residual current data for the NN pore probed with reacted poly(dA)$_{40}$ (A-D) and un-reacted poly(dA)$_{40}$ (A) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in*

this case A-D) and 'b' denotes when both oligonucleotides are present in the cis compartment (A and A-D). The # indicates the experiments which correspond to the histograms in Figure 4 – 8 (A corresponds to 1a and 1b and B corresponds to 2a and 2b).

These reaction conditions were also investigated for the modification of the other purine base guanine. The modified poly(dA)$G_{(14)}$ oligonucleotide (G-D) produced a single residual current level (Figure 4 – 9A, left). Upon addition of unmodified poly(dA)$G_{(14)}$ (G) a single block level with an increased count number was observed (Figure 4 – 9A, right). The control experiment, where G was added to the *cis* compartment first and G-D second, also observed a single residual current level when both reacted and unreacted poly(dA)$G_{(14)}$ were present (Figure 4 – 9B). Therefore, as was previously observed with the *bis*-TBS protected deoxynucleoside system, the optimised dihydroxylation conditions did not cause modification of either of the purine bases.

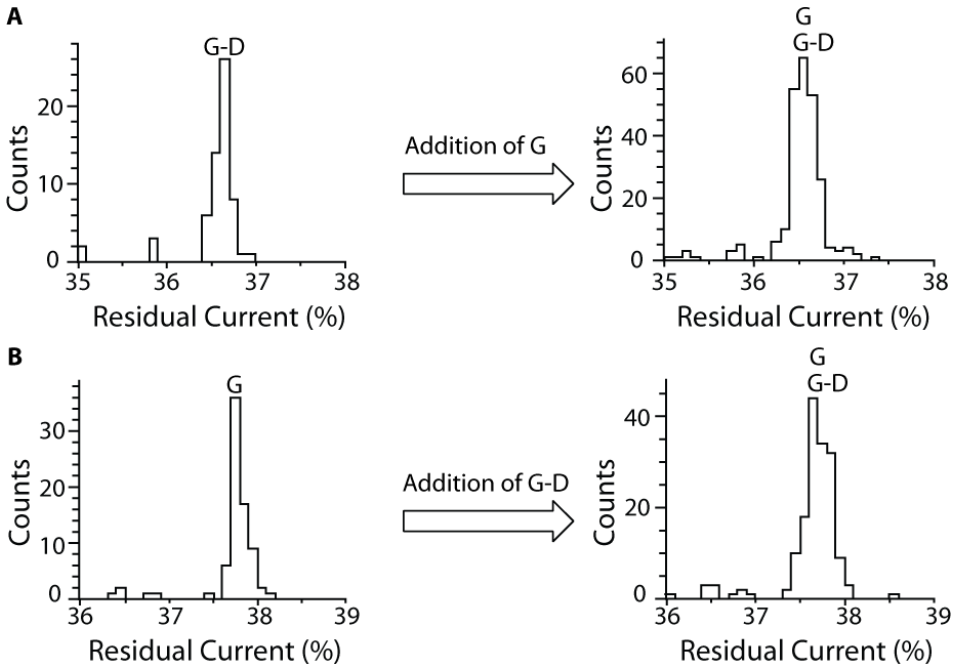


Figure 4 – 9 – *The NN pore probed with reacted poly(dA)$G_{(14)}$ (G-D) and un-reacted poly(dA)$G_{(14)}$ (G). A) The histogram on the left shows only the presence of G-D in the cis compartment. The histogram on the right corresponds to when G was added to the cis compartment and a single residual current level with increased count number was observed when both modified (G-D) and unmodified oligonucleotides (G) were present in the cis compartment. B) The control experiment on the left shows only the presence of G in the cis compartment – a single block level is observed in the histogram. The histogram on the right corresponds to when G-D was added to the cis compartment and again a single residual current level with increased count number was detected, when both the modified*

and unmodified oligonucleotides were present in the cis compartment. This suggests that no reaction had occurred.

Entry	I_O (pA)	I_{RES} G (pA)	I_{RES} G-D (pA)	ΔI_{RES} (G-D – G) %
1a[#]	169	-	36.50	-
1b[#]	170	36.56	36.56	0.00
2b	171	33.14	33.14	0.00
3a[#]	160	37.77	-	-
3b[#]	163	37.71	37.71	0.00
Mean	167	36.30	35.91	0.00
SD	5	2.18	1.57	0.00

Table 4 – 9 – *Residual current data for the NN pore probed with reacted poly(dA)$G_{(14)}$ (G-D) and un-reacted poly(dA)$G_{(14)}$ (G) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case G-D) and 'b' denotes when both oligonucleotides are present in the cis compartment (G and G-D). The [#] indicates the experiments which correspond to the histograms in Figure 4 – 9 (A corresponds to 1a and 1b and B corresponds to 3a and 3b).*

The next base to be investigated for modification under the dihydroxylation (1000 eq.) conditions was thymine itself. This nucleobase was previously found to undergo reaction to the diol of thymine (**44**) in 87% conversion in the *bis*-TBS protected deoxynucloside system. Upon treatment with 1000 eq. of dihydroxylation reagents the modified poly(dA)$T_{(14)}$ oligonucleotide (T-D) was found to produced a single residual current level (Figure 4 – 10A, left). This was indicative of either complete modification to a new product or no reaction having occurred at all. Pleasingly, upon addition of unmodified poly(dA)$T_{(14)}$ two distinct block levels were observed (Figure 4 – 10A, right), with the modified base corresponding to the lower residual current. The control experiment (Figure 4 – 10B) confirmed that the unmodified strand produced a single block level when added to the *cis* compartment (T) and that two distinct block levels were observed upon subsequent addition of modified poly(dA)$T_{(14)}$ (T-D). The electrical recording data indicated that the $\Delta I_{RES}^{(T\text{-}D\ -\ T)} = -0.41 \pm 0.00\%$ (n = 3, Table 4 – 10). These findings

supported those previously observed for the *bis*-TBS protected model system, as complete conversion would be expected with an excess of dihydroxylation reagents.

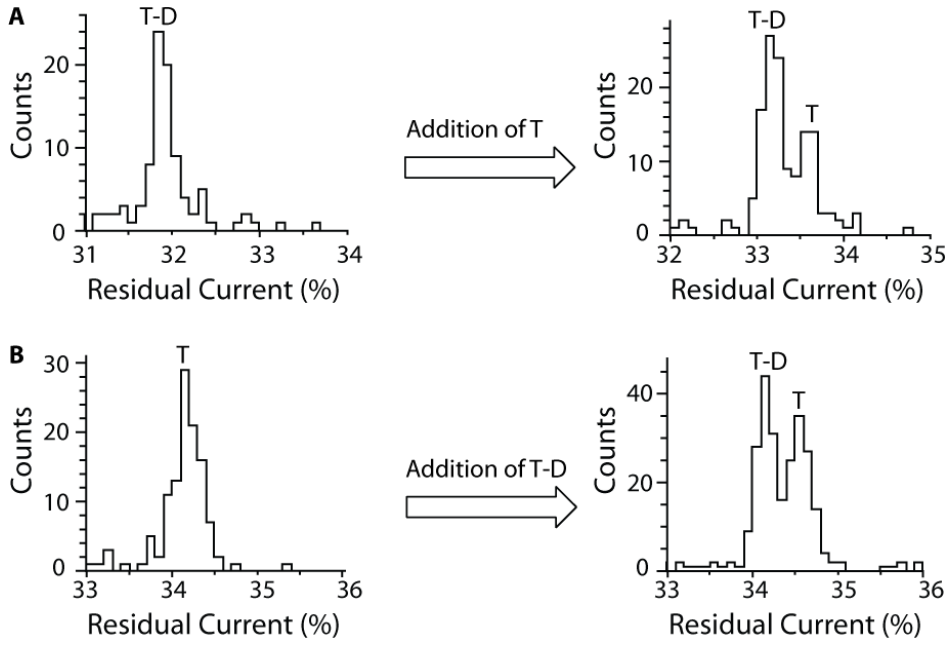


Figure 4 – 10 – *The NN pore probed with reacted and unreacted poly(dA)$T_{(14)}$. A) The histogram on the left shows only the presence of T-D – this suggested that T had either undergone complete conversion to the product or that no reaction had occurred. The histogram on the right corresponds to when T was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – a new residual current level was detected upon addition of T indicating this peak corresponded to the un-modified poly(dA)$T_{(14)}$. Therefore, complete modification of thymine had occurred under the reaction conditions tested. B) The control experiment on the left shows only the presence of T in the cis compartment – a single block level was observed The histogram on the right corresponds to when T-D was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – the appearance of a new lower residual current block level corresponded to product species (T-D).*

Entry	I_O (pA)	I_{RES} T (pA)	I_{RES} T-D (pA)	ΔI_{RES} (T-D – T) %
1a[#]	182	-	32.90	-
1b[#]	183	33.59	33.17	−0.42
2a[#]	169	34.19	-	-
2b[#]	168	34.56	34.15	−0.41
3b	171	34.35	33.94	−0.41
Mean	175	34.17	33.54	−0.41
SD	7	0.42	0.60	0.00

Table 4 – 10 – *Residual current data for the NN pore probed with reacted poly(dA)$T_{(14)}$ (T-D) and un-reacted poly(dA)$T_{(14)}$ (T) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment,*

where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case T-D) and 'b' denotes when both oligonucleotides are present in the cis compartment (T and T-D). The # indicates the experiments which correspond to the histograms in Figure 4 – 10 (A corresponds to 1a and 1b and B corresponds to 2a and 2b).

Finally, the other pyrimidine bases 5mC and C were investigated for modification by the dihydroxylation conditions (1000 eq.). Modified poly(dA)5mC$_{(14)}$ (5mC-D) was observed to produce a single block level upon addition to the *cis* compartment (Figure 4 – 11A, left). Addition of unmodified poly(dA)5mC$_{(14)}$ (5mC) resulted in two distinct current block levels, with the modified base producing a greater block when immobilised in the αHL pore (Figure 4 − 11A, right). Subsequent control experiments showed that the unmodified poly(dA)5mC$_{(14)}$ (5mC) produced a single block level, and that a new distinguishable block level was detected upon addition of the modified oligonucleotide (5mC-D) (Figure 4 – 11B). Similar reactivity was observed for C when exposed to the optimised dihydroxylation conditions (Figures 4 – 11C and D). Although both bases exhibited complete modification under the dihydroxylation conditions, they produced different ΔI_{RES} values when the modified and unmodified residual block levels were compared for each ssDNA sequence ($\Delta I_{RES}^{(5mC\text{-}D\,-\,5mC)} = -0.44 \pm 0.06\%$ (n = 3) (Table 4 – 11A,B) and $\Delta I_{RES}^{(C\text{-}D\,-\,C)} = -0.29 \pm 0.01\%$ (n = 3) (Table 4 – 11C,D) respectively). From previously obtained results, these two bases were not expected to be modified under the dihydroxylation conditions. The lack of selectivity observed in the electrical recording experiments was considered to be owing to the large excess of reagent utilised in the reaction conditions.

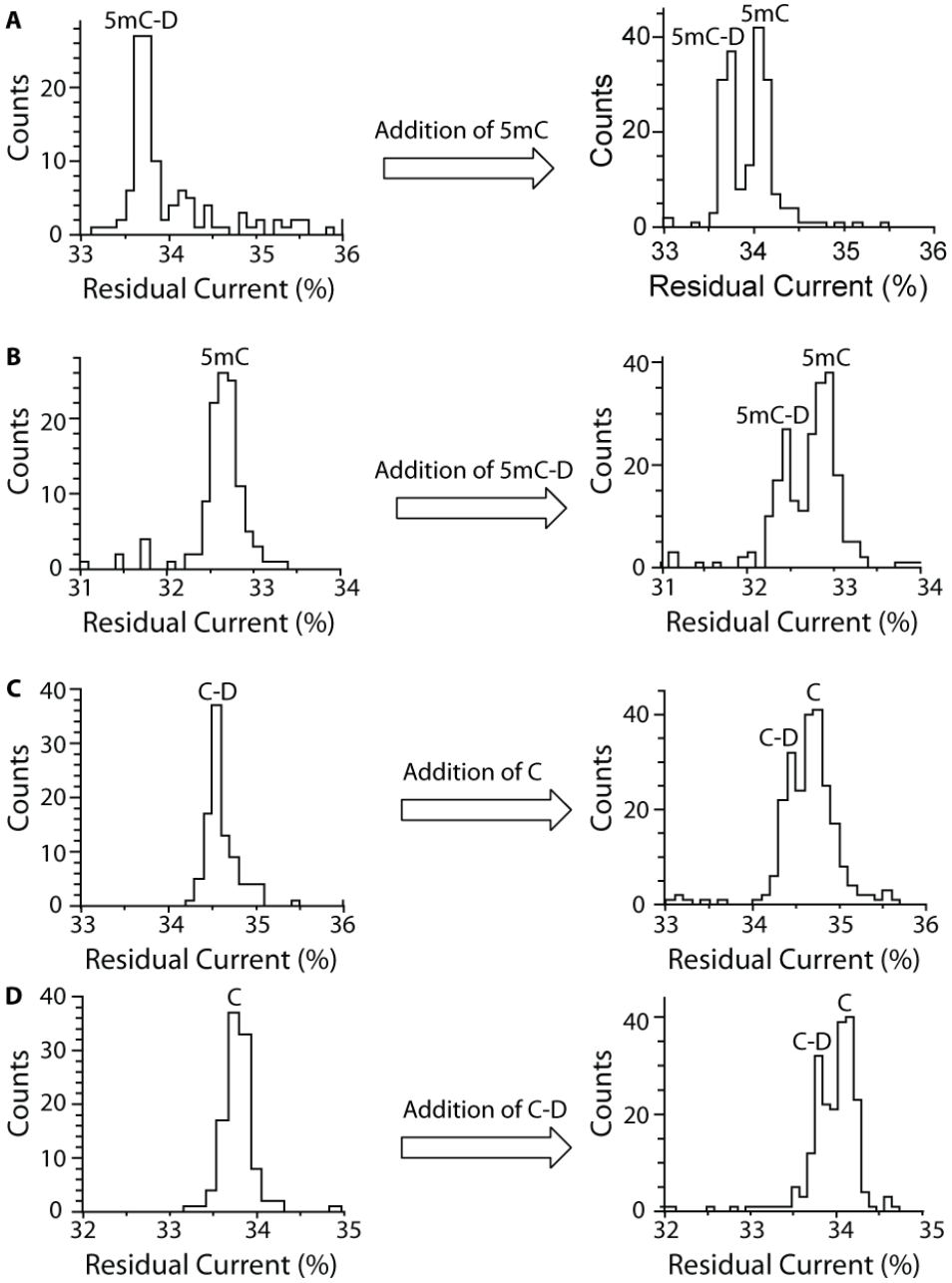


Figure 4 – 11 – *The NN pore probed with reacted and unreacted poly(dA)5mC$_{(14)}$ (A and B) and poly(dA)C$_{(14)}$ (C and D). A) The histogram on the left shows only the presence of 5mC-D – this suggested that 5mC had either undergone complete conversion to the product or that no DNA modification reaction had occurred. The histogram on the right corresponds to when 5mC was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – a new residual current level was observed upon addition of 5mC indicating this peak corresponded to the unmodified poly(dA)5mC$_{(14)}$, with the modified base (5mC-D) producing a lower residual current. Therefore, complete modification of 5-methylcytosine had occurred under the reaction conditions tested. B) The control experiment on the left shows only the presence of 5mC in the cis compartment – a single block level was observed. The histogram on the right corresponds to when 5mC-D was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new lower residual current block level corresponded to product species (5mC-D). C) The histogram on the left shows only the presence of C-D – this suggested*

that C had either undergone complete conversion to the product or that no DNA modification reaction had occurred. The histogram on the right corresponds to when C was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – a new residual current level was observed upon addition of C indicating this peak corresponded to the unmodified poly(dA)$C_{(14)}$, with the modified base (C-D) producing a lower residual current. Therefore, complete modification of cytosine had occurred under the reaction conditions tested. D) The control experiment on the left shows only the presence of C in the cis compartment – a single block level was observed. The histogram on the right corresponds to when C-D was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new lower residual current block level corresponded to product species (C-D).

Entry	I_O (pA)	I_{RES} 5mC (pA)	I_{RES} 5mC-D (pA)	ΔI_{RES} (5mC-D – 5mC) %
1a#	165	-	33.71	-
1b#	165	34.08	33.71	−0.37
2a#	153	32.66	-	-
2b#	153	32.89	32.43	−0.46
3b	175	32.37	31.90	−0.48
Mean	163	33.00	32.93	−0.44
SD	9	0.75	0.92	0.06

Table 4 – 11A,B – *Residual current data for the NN pore probed with reacted poly(dA)5m$C_{(14)}$ (5mC-D) and un-reacted poly(dA)5m$C_{(14)}$ (5mC) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case 5mC-D) and 'b' denotes when both oligonucleotides are present in the cis compartment (5mC and 5mC-D). The # indicates the experiments which correspond to the histograms in Figure 4 – 11 (A corresponds to 1a and 1b and B corresponds to 2a and 2b).*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} C-D (pA)	ΔI_{RES} (C-D – C) %
1b	166	34.74	34.46	−0.28
2b	170	34.49	34.21	−0.28
3a#	175	-	34.54	-
3b#	178	34.66	34.36	−0.29
4a#	168	33.78	-	-
4b#	169	34.12	33.81	−0.31
Mean	171	34.35	34.28	−0.29
SD	5	0.40	0.29	0.01

Table 4 – 11C,D – *Residual current data for the NN pore probed with reacted poly(dA)$C_{(14)}$ (C-D) and un-reacted poly(dA)$C_{(14)}$ (C) (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 3a and 3b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case C-D) and 'b' denotes when both oligonucleotides are*

present in the cis compartment (C and C-D). The # indicates the experiments which correspond to the histograms in Figure 4 – 11 (C corresponds to 3a and 3b and D corresponds to 4a and 4b).

In an effort to regain the selectivity that was previously observed in the *bis*-TBS protected deoxynucleotide system, the number of equivalents of reagent used to modify the bases was reduced to 100 and 10 equivalents, respectively (Table 4 – 12). It was proposed that if the equivalents of reagent were reduced, the previously observed less reactive bases (5mC and C, which did not react in the *bis*-TBS protected model system (subsection 3.2.6)) would not be dihydroxylated and only thymine would be modified under the reaction conditions. Initially, when 1000 eq. of dihydroxylation promoting reagent had been employed, complete product conversion was observed for each of the pyrimidine bases tested. Reduction of the number of equivalents of dihydroxylation reagents resulted in a general trend of reduction in overall product conversion. This was established because when modified oligonucleotides were added to the cis compartment, it was possible to observe two current block levels (one corresponding to the product and the other corresponding to the unmodified nucleobase) suggesting incomplete modification of the nucleobase had occurred (Figure 4 – 12). The ΔI_{RES} value for each oligonucleotide (poly(dA)$T_{(14)}$, poly(dA)5m$C_{(14)}$ and poly(dA)$C_{(14)}$) was, within experimental error, the same for each set of experimental conditions investigated. This indicated that the same dihydroxylated products of T, 5mC and C were detected under each of the modification conditions tested (e.g. 1000, 100 and 10 equivalents of dihydroxylation reagents respectively, see Table 4 − 12). Unfortunately, reduction of the number of equivalents of reagent added to modify the DNA did not result in selective modification of thymine, as 5-methylcytosine and cytosine still exhibited a reaction.

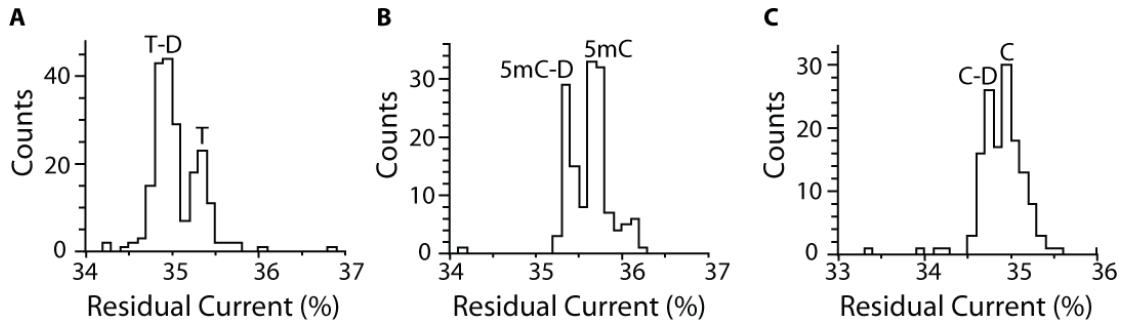


Figure 4 – 12 – *Example histograms corresponding to the modification of poly(dA)$T_{(14)}$, poly(dA)5m$C_{(14)}$ and poly(dA)$C_{(14)}$ when treated with 10 equivalents of dihydroxylation promoting reagents. Histograms A-C show the presence of only the modified oligonucleotide in the cis compartment (A – T-D, B – 5mC-D and C – C-D), suggesting that complete modification of the bases was not observed because block levels corresponding to the unmodified strands were detected.*

Entry	Strand Sequence	Equivalents of Dihydroxylation Reagents	$\Delta I_{RES}^{(N\text{-}D\ -\ N)}$ ± S.D (%)
1	poly(dA)$T_{(14)}$	1000	−0.41 ± 0.00
2	poly(dA)5m$C_{(14)}$	1000	−0.44 ± 0.06
3	poly(dA)$C_{(14)}$	1000	−0.29 ± 0.01
4	poly(dA)$T_{(14)}$	100	−0.37 ± 0.01
5	poly(dA)5m$C_{(14)}$	100	−0.38 ± 0.02
6	poly(dA)$C_{(14)}$	100	−0.26 ± 0.05
7	poly(dA)$T_{(14)}$	10	−0.37 ± 0.05
8	poly(dA)5m$C_{(14)}$	10	−0.33 ± 0.02
9	poly(dA)$C_{(14)}$	10	−0.23 ± 0.02

Table 4 – 12 – *$\Delta I_{RES}^{(N\text{-}D\ -\ N)}$ values (n ≥ 3) for the NN pore probed with reacted and unreacted poly(dA)$T_{(14)}$ (T-D and T), poly(dA)5m$C_{(14)}$(5mC-D and 5mC) and poly(dA)$C_{(14)}$ (C-D and C) that had been subjected to either 1000, 100 or 10 equivalents of dihydroxylation promoting reagents.*

4.2.4 – Mass Spectrometry Analysis of the Modified ssDNA Oligonucleotides which were Produced upon Treatment under Oxidising Conditions (Carell's Conditions and the Optimised Dihydroxylation Conditions)

In an effort to identify the product species produced by the two modification reactions investigated and determine the reason for the observed discrepancies between the model bis-TBS protected system and ssDNA, for the citric acid dihydroxylation reaction, mass spectrometry analysis was performed on the modified oligonucleotides. Analysis of the poly(dA)$T_{(14)}$, poly(dA)$C_{(14)}$ and poly(dA)5m$C_{(14)}$ oligonucleotides, which had been subjected to Carell's oxidation conditions, showed MS peaks corresponding to an increase

in mass of +18 Da (Figure 4 – 13 shows the MS spectra for unmodified and modified poly(dA)$T_{(14)}$). This corresponded to the mass of a single oxygen atom (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da).

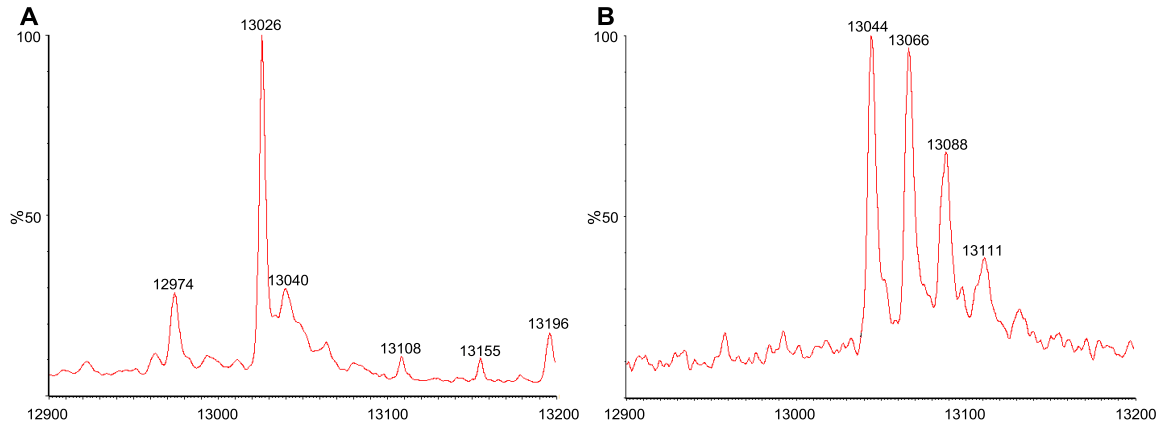


Figure 4 – 13 – *MS spectra corresponding to unmodified and modified poly(dA)$T_{(14)}$ before and after treatment with Carell's oxidation conditions. A) Corresponds to unmodified poly(dA)$T_{(14)}$ (T) (calculated mass = 13027) and B) corresponds to modified poly(dA)$T_{(14)}$ (T-CR), the 100% peak corresponds to an increase in mass of 18 Da which corresponds to the addition of one oxygen atom (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da).*

Analysis of poly(dA)$_{40}$, poly(dA)$G_{(14)}$, poly(dA)$T_{(14)}$, poly(dA)$C_{(14)}$ and poly(dA)5m$C_{(14)}$, which had been subjected to the citric acid dihydroxylation conditions, showed peaks in the MS spectra corresponding to an increase in mass of between +30-33 Da in all cases. This corresponded to the mass of two oxygen atoms (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). As the same increase in mass was observed for each of the oligonucleotides, which were exposed to the dihydroxylation conditions, it was likely that this mass increase corresponded to modification at a region of the oligonucleotide that was common to all strands, despite the fact that this increase in mass could also have correspond to dihydroxylation of the DNA bases.

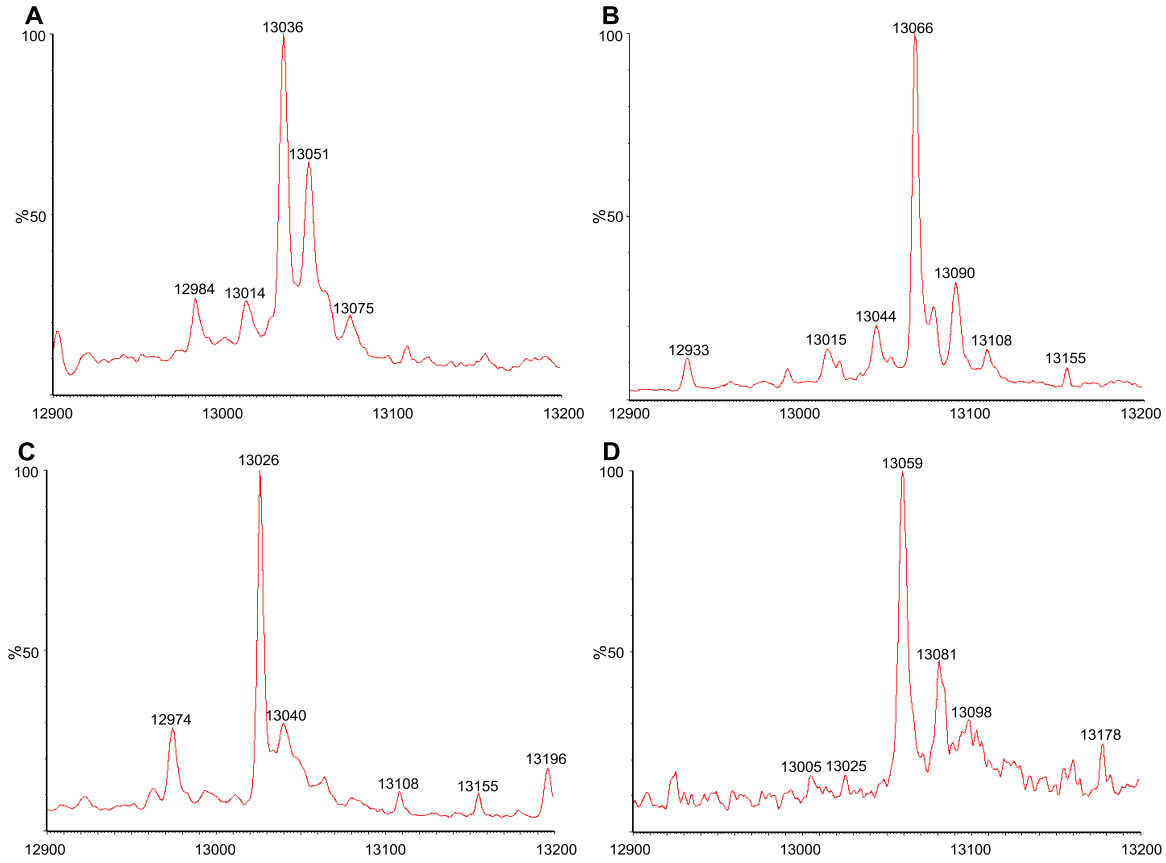


Figure 4 – 14 – *MS spectra corresponding to unmodified and modified poly(dA)$_{40}$ and poly(dA)T$_{(14)}$ before and after treatment with the optimised dihydroxylation conditions. A) Corresponds to unmodified poly(dA)$_{40}$ (A) (calculated mass = 13036) and B) corresponds to modified poly(dA)$_{40}$ (A-CR), the 100% peak corresponds to an increase in mass of 30 Da which corresponds to the addition of two oxygen atoms (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). C) Corresponds to unmodified poly(dA)T$_{(14)}$ (T) (calculated mass = 13027) and D) corresponds to modified poly(dA)T$_{(14)}$ (T-CR), the 100% peak corresponds to an increase in mass of 33 Da which corresponds to the addition of two oxygen atoms (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da).*

A possible explanation for the results observed was that under the oxidation conditions, the sulfur moiety of the biotin had been oxidised to the corresponding sulfoxide and/or sulfone. The oxidation of this moiety would not have prevented binding of the biotin to streptavidin, however, it was possible that the binding conformation might have been altered.[244] If the oxidised biotin was bound to streptavidin in such a way that it resulted in a slight change in the overall length of the linker, then it would have altered the positioning of the base (N) at position 14 relative to the recognition site (R_2). In subsequent electrical

recording experiments of the modified oligonucleotide it would have appeared that a DNA base modification had occurred, when in actual fact it was an alteration in the location of the base at position 14 in relation to the recognition site. In an effort to confirm that oxidation of the biotin had occurred, under the conditions investigated, a model biotin linker was exposed to both Carell's oxidation conditions and the optimised citric acid dihydroxylation conditions (Figure 4 – 13). In both cases, signals corresponding to the mass of the linker plus one and two additional oxygens were detected by MS analysis. This supports the proposed hypothesis that these conditions resulted in oxidation of the sulfur moiety of the biotin linker.

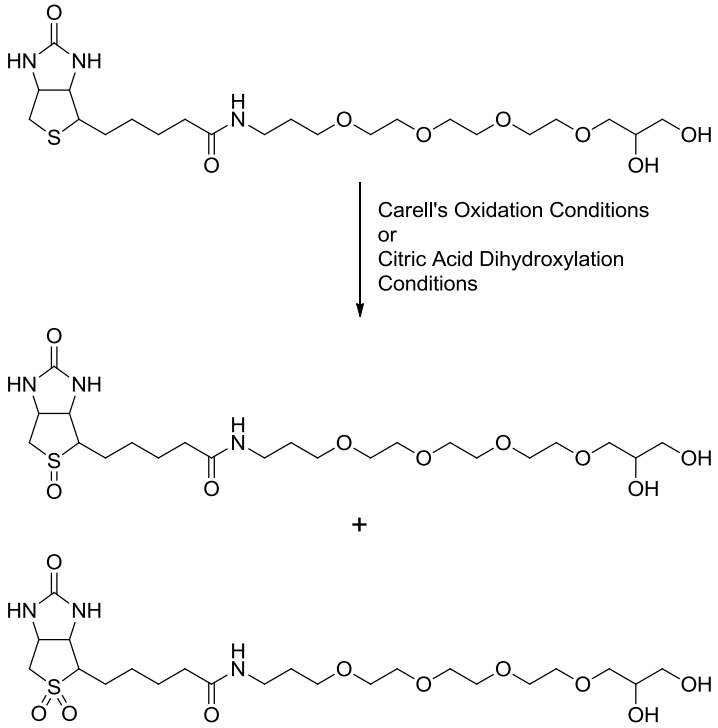


Figure 4 – 13 – *Treatment of the model biotin linker with both Carell's oxidation conditions and the citric acid dihydroxylation conditions. Both experiments produced signals in the MS spectra which correspond to the addition of one or two oxygens (MS analysis detected a signal at 546, corresponding to the mass of the linker containing a sulfoxide moiety plus sodium and a signal at 562, corresponding to the mass of the linker containing a sulfone moiety plus sodium).*

Evidence in the literature to support the proposed binding conformational change of the oxidised biotin comes from work carried out by Qureshi and Garlick.[245, 246] Qureshi *et al.*

showed that the sulfur of biotin formed a hydrogen bonding interaction with the Thr-90 residue in streptavidin and when the threonine was mutated to an alanine, the binding affinity of biotin to streptavidin decreased.[245] Therefore if the sulfur was oxidised, this would disrupt this hydrogen bonding interaction and could result in a slightly different binding conformation. Garlick *et al.* investigated the binding affinity of oxidised biotin to avidin (another protein known to bind biotin). It was observed that a biotin moiety which had been oxidised to the corresponding sulfoxide exhibited a decreased binding affinity for avidin. It was also proposed that the resultant sulfoxide oxygen would exhibit an unfavourable interaction between three tyrosine residues when bound to streptavidin.[246] This would support the theory that the oxidised sulfur moiety could promote an alteration in biotin's binding conformation, in order to minimise these unfavourable interactions.

The proposed hypothesis that the oxidation of the sulfur moiety of the biotin linker, had occurred under the conditions investigated, would account for the reactivity displayed by the poly(dA)$N_{(14)}$ oligonucleotides. The poly(dA)$T_{(14)}$ and poly(dA)$C_{(14)}$ oligonucleotides exhibited oxidation under both sets of conditions tested. If oxidation of the sulfur resulted in a similar conformational binding change for each of the reaction conditions then this would account for why similar ΔI_{RES} values were obtained for thymine and cytosine, for both sets of reaction conditions investigated, despite different products being predicted ($\Delta I_{RES}T = -0.41 \pm 0.08\%$ for Carell's conditions and $\Delta I_{RES}T = -0.41 \pm 0.00\%$ for the dihydroxylation reaction, $\Delta I_{RES}C = -0.29 \pm 0.03\%$ for Carell's conditions and $\Delta I_{RES}C = -0.29 \pm 0.01\%$ for the dihydroxylation reaction, respectively). The observed block levels for the poly(dA)5m$C_{(14)}$ oligonucleotide were not identical for both oxidation conditions investigated, however, the reason for these observed differences is not fully understood. Oxidation of the biotin linker of the poly$(dA)_{40}$ oligonucleotide would not have been expected to result in the production of a different residual current level when

compared to the unmodified oligonucleotide. This was because the sequence at the detection region of R_2 (bases 13-15) would not have been altered, in a strand where the entire sequence was comprised of a single DNA base (adenine in this case).[59] The poly(dA)$G_{(14)}$ oligonucleotide was observed to undergo oxidation upon exposure to the dihydroxylation conditions, however, no new block level was observed for this modified oligonucleotide. The precise reason for the lack of detection of a new residual current level when the modified oligonucleotide was immobilised in αHL is unknown. Therefore, it was proposed that the poly(dA)$N_{(14)}$ oligonucleotides were unreactive under both conditions investigated and only oxidation of the biotin linker was observed for the oligonucleotides tested.

4.3 – Conclusions

Owing to the fact that mutants of αHL were able to distinguish cytosine from its epigenetic modifications, despite there being little difference in their chemical structures (e.g. a single Me group at C5 for 5mC and a hydroxymethyl group for hmC), it was proposed that αHL could be used to monitor chemical reactions on ssDNA. This theory was tested by investigating two sets of oxidation conditions developed for the selective modification of the DNA bases – Carell's oxidation conditions and the optimised citric acid dihydroxylation reaction (Chapter 3).[174] By treatment of poly(dA)$N_{(14)}$ (where N = C, T, A, G or 5mC) oligonucleotides with a chemical reaction mixture, purification of the DNA from the excess reagents and subsequent concentration of the sample it was possible to test the NN αHL mutant's discriminatory ability for the reaction products.

Differences in the expected reactivity of the DNA bases were observed for the oxidation conditions tested (Chapter 3).[174] Therefore, the modified oligonucleotides were further investigated by MS analysis. The oligonucleotides subjected to Carell's oxidation reaction exhibited an increase in mass of +18 Da. Those that were treated under the

dihydroxylation conditions showed an increase in mass of between +30-33 Da. Therefore, it was proposed that the reaction conditions had resulted in oxidation of the sulfur moiety in the biotin linker (to a sulfoxide or sulfone) instead of the desired modification of the DNA bases. Evidence to support this hypothesis was obtained by exposing the model biotin linker to the oxidation conditions investigated, for which MS signals corresponding to the addition of one and two oxygen's were detected. The oxidation of the sulfur moiety of biotin could have caused a change its binding conformation with streptavidin, therefore, leading to an alteration in the overall length of the linker. This would have resulted in a change in the location of the DNA base at position 14 relative to the recognition site, which could have manifested itself as an observable change in block level. Literature evidence which supported the hypothesis that oxidation of the sulfur moiety could have produced an alteration in the binding conformation of biotin to streptavidin was also noted.[245, 246]

In conclusion, it was not possible to investigate oxidiation reactions in the poly(dA)$N_{(14)}$ biotin•streptavidin immobilised system because oxidation of the biotin linker was thought to have occurred, instead of direct modification of the DNA bases themselves.

Chapter 5

Investigations by Electrical Recording

Experiments into the Modification of DNA by the Cancer Drug Temozolomide

Chapter 5 – Investigations by Electrical Recording Experiments into the Modification of DNA by the Cancer Drug Temozolomide

5.1 – Introduction

Owing to difficulties encountered monitoring oxidative chemical reactions of the DNA bases (Chapter 4), it was hypothesised that modified DNA bases, which resulted from non-oxidative chemical damage, could perhaps be identified using electrical recording experiments. This would facilitate the monitoring of the effects of various drug molecules on the structure of DNA.

Many drugs, used in cancer treatment, promote their therapeutic effect by altering the structure of DNA in a variety of ways (for example by alkylating the bases themselves, causing DNA strand breaks or by producing DNA cross-links).[247] It would not be easy to effectively monitor DNA cross-linking or strand cleavage by electrical recording experiments, however, electrical recording experiments could be used to monitor alkylation of the bases. Before selecting a drug for investigation some additional factors needed to be considered, such as the method by which the pro-drug is converted to its active form. Some drugs require an additional *in vivo* modification to convert them into an active agent e.g. hepatic metabolism (e.g. dacarbazine)[248] or activation by complexation with dsDNA (e.g. the duocarmycins which are activated by undergoing a DNA binding-induced conformational change).[247] Both of these scenarios would be difficult to replicate in an *in vitro* model system that would also be suitable for electrical recording experiments. Therefore, the cancer drug temozolomide was chosen, for further exploration by electrical recording experiments, because it is known to act by alkylation of DNA and the conversion of the prodrug to the active methylating agent is controlled solely by the pH

of the surrounding solution. A brief introduction to temozolomide and its mechanism of action *in vivo* is detailed below.

Temozolomide is a monofunctional alkylating prodrug that was first discovered in the mid 1990s.[248] It was initially investigated because the original lead compound, mitozolomide, produced only limited clinical activity and resulted in unpredictable and severe myelosuppression (a decrease in the ability of bone marrow to produce blood cells), despite its potent anti-tumour activity in a number of rodent tumour models.[248, 249] Temozolomide was chosen for further study because it demonstrated excellent bioavailability, exhibited a different profile of antitumor activity against mouse tumours[250] and was less toxic than mitozolomide.[249]

Temozolomide is stable under acidic pH but decomposes to the active methylating agent, above pH 7, by the mechanism shown in Figure 5 – 1.[251-253] The C4 position of temozolomide (marked with a * in Figure 5 – 1) carries the greatest positive charge and is, therefore, most susceptible to nucleophilic attack by water. The base-catalysed attack of water has been observed to be the rate-determining step in the overall decomposition of temozolomide.[252] The methyltriazen-1-yl imidazole-4-carboxamide (MTIC) which is produced by this reaction is rapidly degraded to the inactive carboxcyclic acid derivative, 5-aminoimidazole-4-carboxamide (AIC) and the active methylating agent methyldiazonium cation.

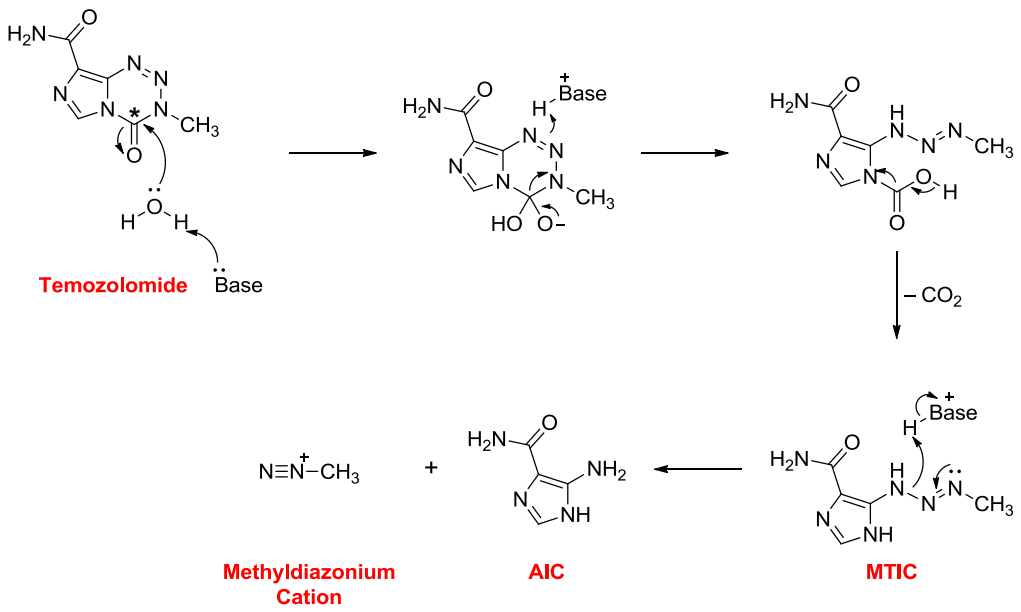


Figure 5 – 1 – *The molecular mechanism of decomposition of temozolomide to its active methylating agent – the methyldiazonium cation. Figure adapted from Newlands et al. 1997.*[248]

Due to the high reactivity of the methyldiazonium cation, it has not been observed to facilitate selective methylation of one particular base (Figure 5 – 2). Temozolomide methylates the N-7 of guanine (70% of adducts), the O-6 of guanine (5% of adducts), the N-1 and N-3 of adenine and the N-3 of cytosine (all three combine to make up the additional 25% of adducts).[253]

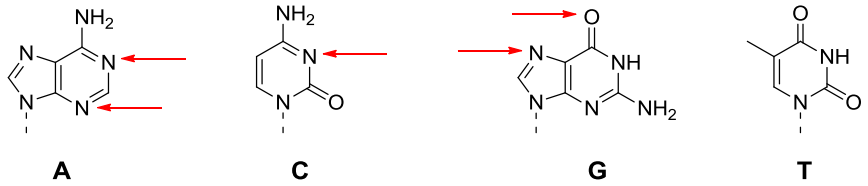


Figure 5 – 2 – *Diagram indicating the temozolomide methylation sites on the canonical bases adenine, cytosine, guanine and thymine.*

There are a number of methods by which the methylated DNA, resulting from temozolomide treatment, can be repaired by cells: 1) O^6-alkylguanine-DNA alkyltransferase (AGT), 2) DNA-mismatch repair (MMR) enzymes and 3) base excision repair, assisted by poly(ADP-ribose) polymerase. AGT is the primary method by which

the methyl group at the O-6 position of guanine is removed. The enzyme transfers the methyl moiety onto an internal cysteine in its catalytic domain, which effectively reverses the cytotoxic effects of this lesion. If the O^6-methylguanine base is not removed by AGT it is known to form mismatched base pairs with thymine.[248] Therefore, in cell lines which contain low-levels of AGT, DNA-mismatch repair (MMR) enzymes attempt to correct the mismatch by excising the thymine base in the daughter strand. Finally, base excision repair, assisted by poly(ADP-ribose) polymerase, removes the N^7-guanine and N^3-adenine residues.[248, 254]

The observed therapeutic effect upon temozolomide treatment is cell apoptosis. The O^6-guanine moiety is the base thought to be responsible for mediating cell death, despite its low overall abundance.[255] This is because cell lines which exhibit high levels of the AGT (the enzyme responsible for removing the O^6-guanine residue) are found to be resistant to temozolomide.[255, 256] If the O^6-methyguanine lesion is not repaired by AGT, it results in the repeated mismatching of a thymine base in the corresponding daughter strand. Therefore, the MMR enzymes attempt to continually repair the mismatch leading to breaks and permanent nicks in the daughter strand which consequently results in apoptosis.[254]

Temozolomide is generally administered as an oral drug, owing to its stability under acid conditions.[253, 257] Due to its relatively short half-life *in vivo* (which was measured to be approximately 1.8 hours),[258] temozolomide is normally administered at a dose of 150 to 200 mg/m^2 per day, on days 1 to 5 of a 28 cycle.[257] This regime has been widely utilised for the treatment of patients suffering with recurrent high-grade glioma and advanced metastatic melanoma since the late 1990's.[253, 254, 257] Temozolomide has been particularly successful in the treatment of gliomas due to the fact that it can cross the blood brain barrier, without decomposing to its active agent. Several more recent clinical trials

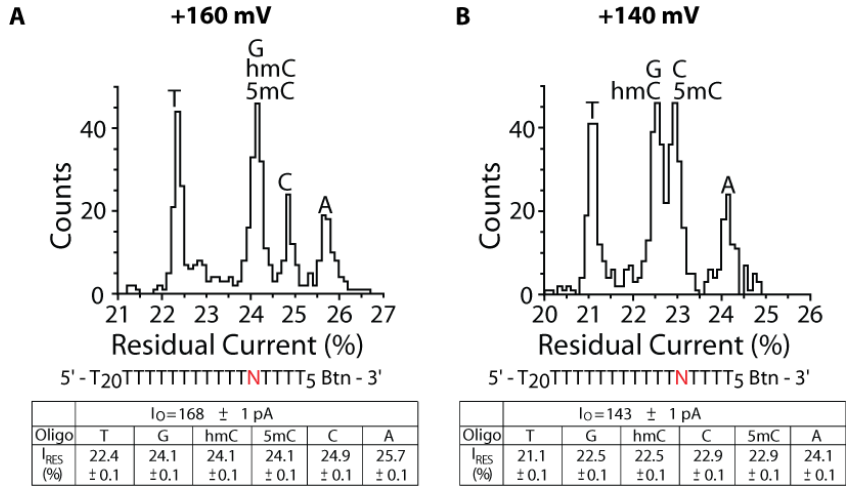
have been investigating methods of overcoming temozolomide resistance, promoted by high levels of AGT. These trials have either employed combination therapies with other drugs which inhibit AGT (such as O^6-benzylguanine)[256] or utilised a dose-dense temozolomide regime (where the drug is administered over a longer period of time with less time between treatments e.g. treatment for 21 days out of 28 instead of 5 days out 28).[257] However, these combination treatments require further investigation as some have been linked with an increased incidence of secondary leukemias.[254]

5.2 – Results and Discussion

5.2.1 – Discrimination of Unmodified Poly(dT)$N_{(9)}$ Oligonucleotides by the NNY αHL Mutant

It was necessary to identify a DNA background that had not previously been observed to undergo any reaction with temozolomide, in order to successfully investigate the methylation of the DNA bases by this cancer drug. A poly$(dT)_{40}$ oligonucleotide background was chosen because temozolomide was not observed to methylate this base.[253] Previous work by Purnell *et al.* detected improved base discrimination, at the constriction site of WT αHL when probed with poly$(dT)_{40}$ oligonucleotides.[58] Therefore, the single base substitutions in the poly$(dT)_{40}$ oligonucleotides were made at position 9, which is known to reside within recognition point R_1, located at the central constriction of αHL.[60] As the NNY mutant of αHL had previously exhibited the best base discriminating abilities at the constriction site, when probed with poly$(dC)_{40}$ oligonucleotides, this mutant was investigated for base discrimination of single base substitutions in the poly$(dT)_{40}$ background. The NNY mutant (protein prepared by E. Mikhailova) was probed with poly(dT)$N_{(9)}$ (where N = C, T, A, G, 5mC or hmC) oligonucleotides at applied potentials of +160 and +140 mV (Figure and Tables 5 – 3) in 1 M KCl, 25 mM Tris, 100 μM EDTA, pH 8.0 buffer (see Chapter 5, section 5.2.2 for the reasons behind the application of a

potential of +140 mV). Despite probing a different αHL mutant, at +160 mV the bases resulted in the same order of increasing block level as was observed by Purnell (T > G > C > A).[58] A similar order of increasing block level was observed at +140 mV. Furthermore, at +160 mV G, 5mC and hmC all produced identical block levels, whereas, at +140 mV the epigenetically modified bases could be distinguished from each other (as G overlaid with hmC and C overlaid with 5mC). The ΔI_{RES} ^{OVERALL} was also noted to have been slightly larger at +160 mV than at +140 mV. However, neither applied potential facilitated six base discrimination when the poly(dT)$N_{(9)}$ oligonucleotides probed the NNY mutant of αHL.



	I_O=168 ± 1 pA					
Oligo	T	G	hmC	5mC	C	A
I_{RES} (%)	22.4 ± 0.1	24.1 ± 0.1	24.1 ± 0.1	24.1 ± 0.1	24.9 ± 0.1	25.7 ± 0.1

	I_O=143 ± 1 pA					
Oligo	T	G	hmC	C	5mC	A
I_{RES} (%)	21.1 ± 0.1	22.5 ± 0.1	22.5 ± 0.1	22.9 ± 0.1	22.9 ± 0.1	24.1 ± 0.1

Figure 5 – 3 – *Base discrimination abilities of the NNY pore when probed with poly(dT)$N_{(9)}$ oligonucleotides with a single substituted base (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3′ biotin-TEG. A) Corresponds to an applied voltage of +160 mV and B) corresponds to an applied voltage of +140 mV. Figures A and B relate to separate experiments.*

Entry	I_O (pA)	I_{RES} poly $(dT)_{40}$ (pA)	I_{RES} C (pA)	ΔI_{RES} C %	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} ^{OVERALL} %
1	164	22.9	25.5	2.6	24.7	1.8	24.7	1.8	26.5	3.6	24.7	1.8	3.6
2[#]	168	22.4	24.9	2.5	24.1	1.8	24.1	1.8	25.7	3.4	24.1	1.8	3.4
3	171	22.7	25.3	2.6	24.5	1.7	24.5	1.7	26.1	3.4	24.5	1.7	3.4
4	164	23.1	25.7	2.6	24.9	1.8	24.9	1.8	26.5	3.5	24.9	1.8	3.5
5	167	23.2	25.6	2.4	25.0	1.8	25.0	1.8	26.7	3.5	25.0	1.8	3.5
Mean	167	22.9	25.4	2.6	24.6	1.8	24.6	1.8	26.3	3.5	24.6	1.8	3.5
SD	3	0.3	0.3	0.1	0.3	0.0	0.3	0.0	0.4	0.1	0.3	0.0	0.1

Table 5 – 3A – *Residual current data for the NNY mutant of αHL probed with poly(dT)$N_{(9)}$ oligonucleotides with a single substituted base (N = C, T, A, G, 5mC or hmC) at position 9*

relative to the 3' biotin-TEG, at an applied voltage of +160 mV (n = 5). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} A, for example, is the difference in residual current between the poly(dT)$_{40}$ and the poly(dT)-A$_{(9)}$ levels (ΔI_{RES} A = I_{RES} A − I_{RES} T). The # indicates the experiment which corresponds to the histogram in Figure 5 – 3A for the NNY mutant.

Entry	I_O (pA)	I_{RES} poly (dT)$_{40}$ (pA)	I_{RES} C (pA)	ΔI_{RES} C %	I_{RES} hmC (pA)	ΔI_{RES} hmC %	I_{RES} 5mC (pA)	ΔI_{RES} 5mC %	I_{RES} A (pA)	ΔI_{RES} A %	I_{RES} G (pA)	ΔI_{RES} G %	ΔI_{RES} OVERALL %
1	144	20.9	22.8	1.9	22.4	1.5	22.8	1.9	24.0	3.2	22.4	1.5	3.2
2	148	20.3	22.1	1.9	21.7	1.4	22.1	1.9	23.3	3.0	21.7	1.4	3.0
3	151	20.2	22.1	1.9	21.7	1.4	22.1	1.9	23.2	3.0	21.7	1.4	3.0
4	142	21.0	23.0	2.0	22.5	1.5	23.0	2.0	24.1	3.1	22.5	1.5	3.1
5#	143	21.1	22.9	1.9	22.5	1.4	22.9	1.9	24.1	3.0	22.5	1.4	3.0
Mean	146	20.7	22.6	1.9	22.1	1.4	22.6	1.9	23.8	3.1	22.1	1.4	3.1
SD	4	0.4	0.4	0.0	0.4	0.0	0.4	0.0	0.5	0.1	0.4	0.0	0.1

Table 5 – 3B – *Residual current data for the NNY mutant of αHL probed with poly(dT)N$_{(9)}$ oligonucleotides with a single substituted base (N = C, T, A, G, 5mC or hmC) at position 9 relative to the 3' biotin-TEG, at an applied voltage of +140 mV (n = 5). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} A, for example, is the difference in residual current between the poly(dT)$_{40}$ and the poly(dT)-A$_{(14)}$ levels (ΔI_{RES} A = I_{RES} A − I_{RES} T). The # indicates the experiment which corresponds to the histogram in Figure 5 – 3B for the NNY mutant.*

5.2.2 – Electrical Recording Studies of the Methylation of the DNA Bases by the Action of the Cancer Drug Temozolomide

In order to investigate the methylation of DNA by temozolomide, suitable reaction conditions needed to be employed such that temzolomide would decompose to its active agent (methyldiazonium cation) and promote a reaction. Stability studies of the decomposition of temozolomide revealed its half-life to be 1.83 h in 0.1 M potassium phosphate buffer at pH 7.4 at 37 °C.[252] Therefore, these conditions were used for the modification of poly(dT)N$_{(9)}$ (where N = C, T, A or G) oligonucleotides, where a one hundred fold excess of temozolomide (when compared to the oligonucleotide concentration) was employed to maximise the opportunity of methylation occurring. The ssDNA was incubated with temozolomide for 18 h to allow for complete decomposition of the drug over the reaction timeframe. Therefore, the final conditions were poly(dT)N$_{(9)}$

(50 μM) (where N = C, T, A or G) incubated with 5 mM temozolomide in 0.1 M potassium phosphate buffer, pH 7.4, at 37 °C for 18 h.

Modified and unmodified poly(dT)$_{40}$ were initially used to probe the NNY mutant of αHL (at +140 and +160 mV) to establish whether methylation of thymine was observed under the conditions investigated. Electrical recording experiments were performed in a similar manner to those previously described, for the investigation of the modification of DNA by a variety of chemical reagents (Chapters 4). The modified poly(dT)$_{40}$ (T-TM1) was pre-incubated with excess streptavidin (10 μL of 1 mg/mL stock solution in 10 mM sodium phosphate, 150 mM sodium chloride, pH 7.2) and then added to the *cis* compartment (Figure and Table 5 – 4A, left). At an applied potential of +140 mV, a single residual current level was observed, indicating that either no modification had occurred or complete conversion to a new product, had resulted. Pleasingly, upon addition of unmodified poly(dT)$_{40}$ (T), to the *cis* compartment, the same current level with increased count number was produced, indicating that no methylation of thymine had occurred (Figure 5 – 5A, right). To confirm that the unmodified poly(dT)$_{40}$ (T) produced a single block level, a control experiment was performed where this oligonucleotide was added to the chamber first and T-TM1 second (Figure 5 – 5B). The control experiment also resulted in the observation of a single block level when both modified and unmodified poly(dT)$_{40}$ samples were present in the *cis* compartment. A single residual current level was also observed for the modified and unmodified poly(dT)$_{40}$ oligonucleotide, at an applied potential of +160 mV. Under the reaction conditions investigated, no modification of thymine (the base which comprises the majority of the poly(dT)N$_{(9)}$ (where N = C, T, A or G) oligonucleotide sequences) was observed, which was supported by previous data in the literature on the base selectivity of methylation by temozolomide.[253] Therefore, as thymine is not modified under the reaction conditions, it will provide an unreactive background

oligonucleotide sequence upon which methylation of the bases C, A and G, by temozolomide, can be investigated.

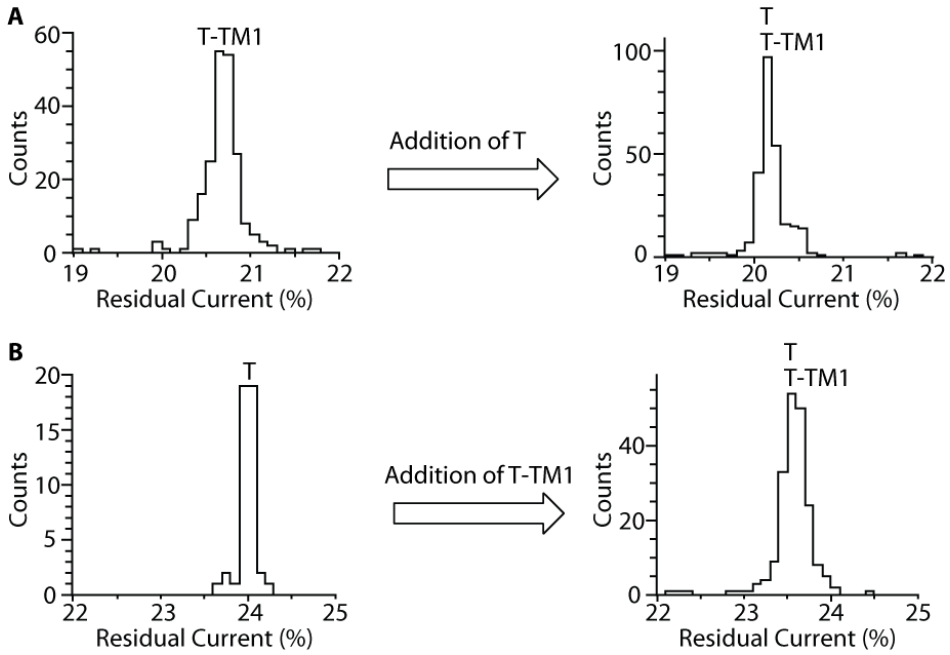


Figure 5 – 4 – *The NNY pore probed with reacted poly(dT)$_{40}$ (T-TM1) and un-reacted poly(dT)$_{40}$ (T) at an applied potential of +140 mV. A) The histogram on the left shows only the presence of T-TM1 in the cis compartment. The histogram on the right corresponds to when T was added to the cis compartment and a single residual current level with increased count number was observed when both modified (T-TM1) and unmodified oligonucleotides (T) were present in the cis compartment. This suggests that no reaction had occurred (these histograms correspond to different experiments). B) The control experiment on the left shows only the presence of T in the cis compartment – a single residual current level was observed. The histogram on the right corresponds to when T-TM1 was added to the cis compartment and again a single residual current level with increased count number was detected when both the modified and unmodified oligonucleotides were present in the cis compartment (these histograms correspond to different experiments). Therefore, this suggested that no DNA modification reaction had occurred.*

Entry	I_O (pA)	I_{RES} T (pA)	I_{RES} T-TM1 (pA)	ΔI_{RES} (T-TM1 – T) %
1a#	143	-	20.70	-
2b#	145	20.16	20.16	0.00
3a#	155	24.00	-	-
4b#	148	23.59	23.59	0.00
5b	152	22.72	22.72	0.00
6a	153	23.66	-	-
6b	152	23.68	23.68	0.00
7b	156	23.45	23.45	0.00
Mean	150	23.04	22.38	0.00
SD	5	1.33	1.56	0.00

Table 5 – 4 – *Residual current data for the NNY pore probed with reacted poly(dT)$_{40}$ (T-TM1) and un-reacted poly(dT)$_{40}$ (T) at an applied potential of +140 mV (n = 5). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 6a and 6b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case T) and 'b' denotes when both oligonucleotides are present in the cis compartment (T and T-TM1). The # indicates the experiments which correspond to the histograms in Figure 5 – 4 (A corresponds to 1a# and 2b# and B corresponds to 3a# and 4b#).*

Poly(dT)$G_{(9)}$ was the next oligonucleotide to be investigated for methylation by temzolomide because the drug's therapeutic effect is mediated by methylation of guanine at the O-6 position.[253] Upon addition of modified poly(dT)$G_{(9)}$ (G-TM1) to the *cis* compartment, a single block level was observed at an applied potential of +140 mV (Figure 5 – 5A, left). This meant that either 1) complete modification to a new product had occurred or 2) no reaction had been observed. To establish which outcome was correct, additional unmodified poly(dT)$G_{(9)}$ (G) was added to the *cis* compartment and two distinct residual current levels were produced (Figure 5 – 5A, right). The control experiment, where G was added to the *cis* compartment first and G-TM1 second, also resulted in the observation of two distinct residual current levels when both modified and unmodified poly(dT)$G_{(9)}$ samples were present. The electrical recording data indicated that the $\Delta I_{RES}^{(G\text{-}TM1\ -\ G)}$ = 0.26 ± 0.02% (n = 4) (Table 5 – 5). This suggested that complete conversion to a new product had been observed. The applied potential used for these experiments was +140 mV, because at +160 mV the $\Delta I_{RES}^{(G\text{-}TM1\ -\ G)}$ was slightly smaller

($\Delta I_{RES}^{(G\text{-}TM\ -\ G)}$ = 0.22 ± 0.01% at +160 mV) and the signal resolution was not as clear. Therefore, the applied potential of +140 mV was used for all subsequent experiments investigating methylation by temozolomide.

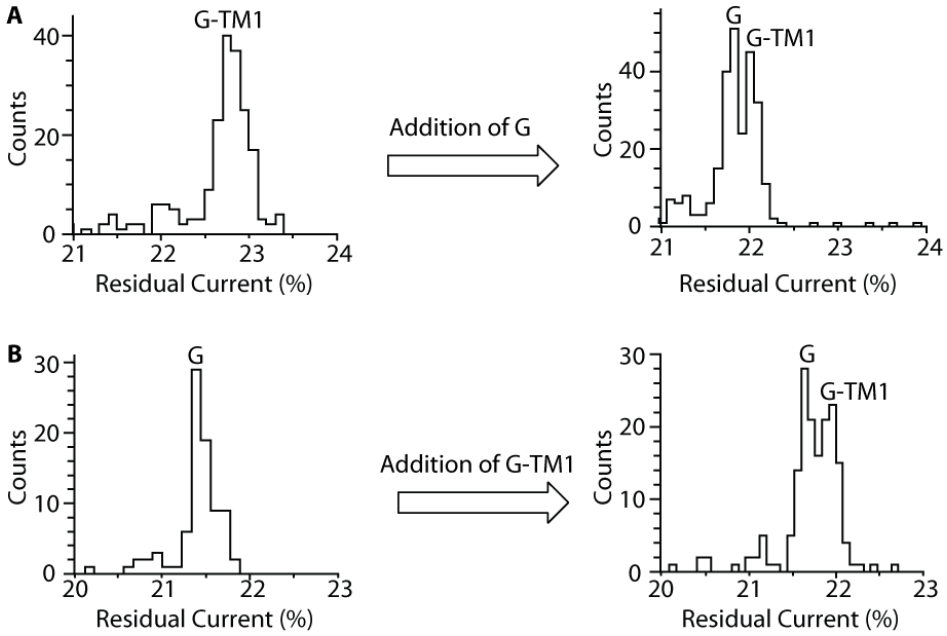


Figure 5 – 5 – *The NNY pore probed with reacted poly(dT)$G_{(9)}$ (G-TM1) and un-reacted poly(dT)$G_{(9)}$ (G) at an applied potential of +140 mV. A) The histogram on the left shows only the presence of G-TM1 – this suggests that either G had undergone complete conversion to the product or no DNA modification reaction had occurred. The histogram on the right corresponds to when G was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, – a new residual current level was observed upon addition of G indicating this peak corresponded to the unmodified poly(dT)$G_{(9)}$, with the modified base (G-TM1) producing a higher residual current (these histograms correspond to two different experiments). Therefore, complete modification of guanine had occurred under the reaction conditions investigated. B) The control experiment on the left shows only the presence of G in the cis compartment – a single block level was observed. The histogram on the right corresponds to when G-TM1 was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new greater residual current block level corresponds to product species (G-TM1).*

Entry	I_O (pA)	I_{RES} G (pA)	I_{RES} G-TM1 (pA)	ΔI_{RES} (G-TM1 – G) %
1a#	150	21.44	-	-
1b#	144	21.65	21.93	0.28
2a#	146	-	22.81	-
3b#	152	21.79	22.04	0.25
4b	146	22.18	22.44	0.26
5b	146	22.25	22.49	0.24
Mean	147	21.86	22.34	0.26
SD	3	0.35	0.36	0.02

Table 5 – 5 – *Residual current data for the NNY pore probed with reacted poly(dT)$G_{(9)}$ (G-TM1) and un-reacted poly(dT)$G_{(9)}$ (G) at an applied potential of +140 mV (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case G-TM1) and 'b' denotes when both oligonucleotides are present in the cis compartment (G and G-TM1). The # indicates the experiments which correspond to the histograms in Figure 5 – 5 (A corresponds to 2a# and 3b# and B corresponds to 1a# and 1b#).*

Temozolomide is also known to methylate adenine in two positions (N-1 and N-3) and cytosine in only one (N-3). The methylated products of these two bases were also investigated by electrical recording experiments at an applied potential of +140 mV. Reacted poly(dT)$A_{(9)}$ (A-TM1) exhibited two distinct block levels, which indicated that either complete conversion to the two products had occurred or incomplete modification had resulted and some residual unmodified poly(dT)$A_{(9)}$ was still present (Figure 5 – 6A, left). Addition of unmodified poly(dT)$A_{(9)}$ resulted in an increase in the count number of the histogram peak corresponding to a lower residual current (Figure 5 – 6A, right). Subsequent control experiments showed that the unmodified poly(dT)$A_{(9)}$ produced a single block level, and that a new distinguishable block level was detected upon addition of the modified oligonucleotide (A-TM1) (Figure 5 – 6B). A much larger ΔI_{RES} ($\Delta I_{RES}^{(A\ TM1\ -\ A)}$) was produced by modification of the poly(dT)$A_{(9)}$ oligonucleotide in comparison to that produced by poly(dT)$G_{(9)}$ ($\Delta I_{RES}^{(A\text{-}TM1\ -\ A)} = 0.66 \pm 0.04\%$ (n = 5) at +140 mV). Similar reactivity was observed for C when exposed to temozolomide, with incomplete conversion to a product again being observed (Figure 5 – 6C, D). The

$\Delta I_{RES}^{(C\text{-}TM1\ -\ C)}$ corresponding to the modification of the poly(dT)$C_{(9)}$ oligonucleotide ($\Delta I_{RES}^{(C\text{-}TM1\ -\ C)}$ = 0.97 ± 0.03% (n = 5) at +140 mV) was larger than that observed for both poly(dT)$G_{(9)}$ and poly(dT)$A_{(9)}$. Therefore, from the results obtained it was possible to detect residual current levels which corresponded to the modified oligonucleotides G-TM1, A-TM1 and C-TM1, by electrical recording experiments, all of which produced a higher residual current level than the corresponding unmodified oligonucleotides.

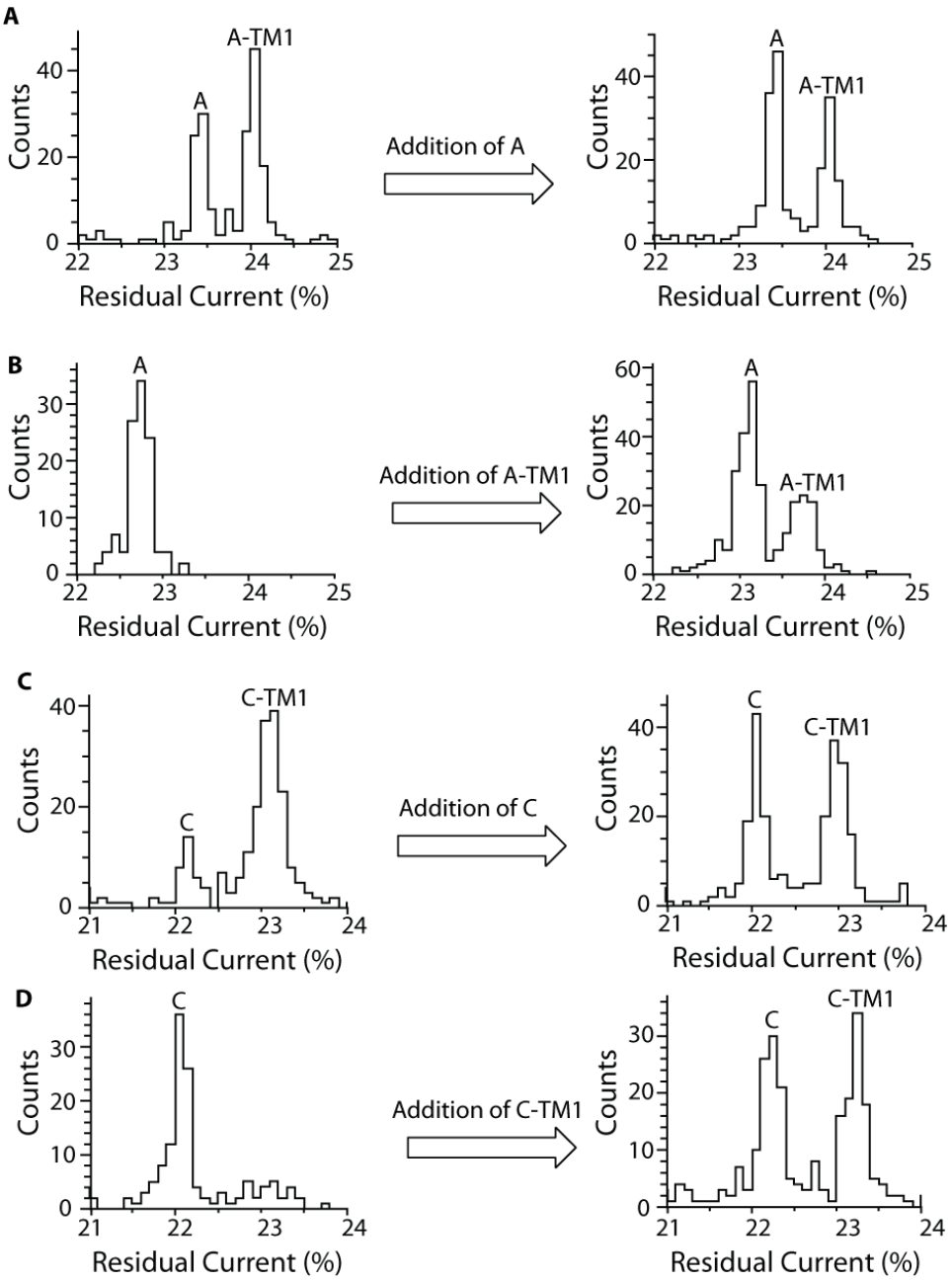


Figure 5 – 6 – *The NNY pore probed with reacted and un-reacted poly(dT)$A_{(9)}$ and poly(dT)$C_{(9)}$ at an applied potential of +140 mV. A) The histogram on the left shows only the presence of A-TM1 – this suggests that A-TM1 had undergone incomplete conversion*

to a product. The histogram on the right corresponds to when A was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, – one of the residual current levels increased its count number upon addition of A indicating this peak corresponds to the un-modified poly(dT)$A_{(9)}$. B) The control experiment on the left shows only the presence of A in the cis compartment – a single block level was observed. The histogram on the right shows when A-TM1 was added to the cis compartment, the appearance of a new higher residual current level corresponded to the product species (A-TM1). C) The histogram on the left shows only the presence of C-TM1 in the cis compartment – this suggests that C-TM1 had undergone incomplete conversion to the product. The histogram on the right shows both C-TM1 and C present in the cis compartment – one of the residual current levels increased its count number upon addition of C indicating this peak corresponds to the un-modified poly(dT)$C_{(9)}$. D) The control experiment on the left shows only the presence of C in the cis compartment – a single block level was observed. The histogram on the right showed when C-TM1 was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new greater residual current level corresponds to product species (C-TM1) (these histograms correspond to two different experiments).

Entry	I_O (pA)	I_{RES} A (pA)	I_{RES} A-TM1 (pA)	ΔI_{RES} (A-TM1 – A) %
1a#	144	23.41	24.04	0.63
1b#	145	23.41	24.04	0.63
2b	149	32.12	23.78	0.66
3b	152	23.14	23.79	0.65
4b	152	22.88	23.60	0.73
5a#	145	22.74	-	-
5b#	145	23.10	23.74	0.63
Mean	147	23.12	23.83	0.65
SD	4	0.25	0.17	0.04

Table 5 – 6A,B – *Residual current data for the NNY pore probed with reacted poly(dT)$A_{(9)}$ (A-TM1) and un-reacted poly(dT)$A_{(9)}$ (A) at an applied potential of +140 mV (n = 5). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case A-TM1) and 'b' denotes when both oligonucleotides are present in the cis compartment (A and A-TM1). The # indicates the experiments which correspond to the histograms in Figure 5 – 6 (A corresponds to 1a# and 1b# and B corresponds to 5a# and 5b#).*

Entry	I_O (pA)	I_{RES} C (pA)	I_{RES} C-TM1 (pA)	ΔI_{RES} (C-TM1 – C) %
1a#	143	22.14	23.10	0.96
1b#	147	22.05	22.99	0.93
2b	151	21.91	22.86	0.95
3a	148	22.08	-	-
3b	146	22.58	23.59	1.01
4a#	150	21.89	-	-
5b#	151	22.23	23.23	1.00
6b	144	22.68	23.66	0.98
Mean	148	22.20	23.24	0.97
SD	3	0.29	0.33	0.03

Table 5 – 6C,D – *Residual current data for the NNY pore probed with reacted poly(dT)$C_{(9)}$ (C-TM1) and un-reacted poly(dT)$C_{(9)}$ (C) at an applied potential of +140 mV (n = 5). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case C-TM1) and 'b' denotes when both oligonucleotides are present in the cis compartment (C and C-TM1). The # indicates the experiments which correspond to the histograms in Figure 5 – 6 (C corresponds to 1a# and 1b# and D corresponds to 4a# and 5b#).*

In order to confirm that methylation of the bases had occurred, MS analysis was performed on the reacted samples of ssDNA – poly(dT)$_{40}$, poly(dT)$A_{(9)}$, poly(dT)$C_{(9)}$ and poly(dT)$G_{(9)}$ (Figure 5 – 7 shows the modified and unmodified oligonucleotide MS for poly(dT)$_{40}$ and poly(dT)$G_{(9)}$). Surprisingly, no methylation was detected for any of the oligonucleotide samples tested, only unmodified starting material was observed (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). On further examination of the initial conditions employed for temozolomide modification, it was noted that Denny *et al.* had observed methylation of both the phosphate buffer and solvent (water) by temozolomide.[252] Therefore, it was likely, that owing to the twenty-fold excess of phosphate buffer present in the reaction solution, in comparison to the temozolomide concentration, that the buffer was being preferentially methylated over the DNA bases. This would explain why no methylation was detected by MS analysis. Future experiments for studying this methylation reaction would need to be carried out in the presence of a buffering agent that would not react with temozolomide e.g. Tris buffer.

However, this does not explain why the electrophysiology results appear to indicate that the DNA had been modified, whereas, the MS analysis clearly showed that this was not the case. Further investigation into the anomalous electrophysiology results was required.

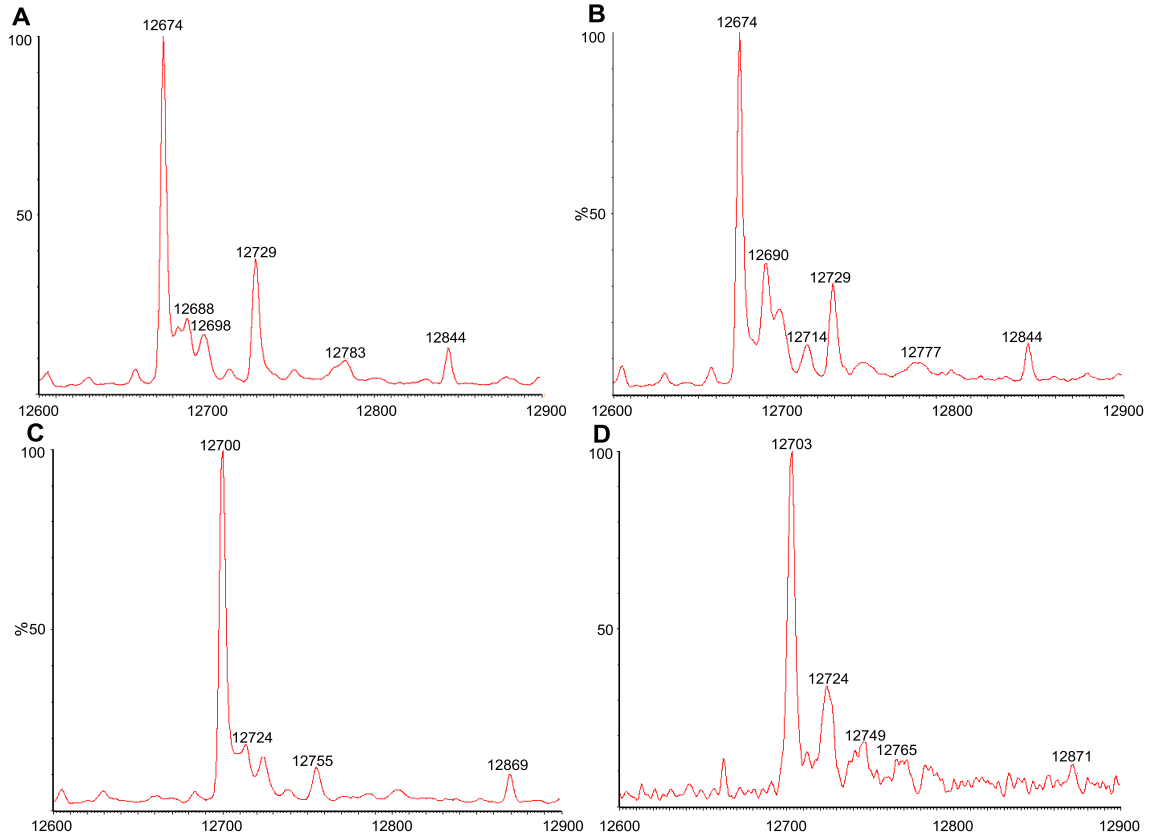


Figure 5 – 7 – *MS spectra corresponding to unmodified and modified poly(dT)$_{40}$ and poly(dT)G$_{(9)}$ before and after treatment with temozolomide. A) Corresponds to unmodified poly(dT)$_{40}$ (T) (calculated mass = 12674) and B) corresponds to modified poly(dT)$_{40}$ (T-TM1), the 100% peaks for A and B correspond to the calculated mass of the unmodified oligonucleotide (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). C) Corresponds to unmodified poly(dT)G$_{(9)}$ (calculated mass = 12699) and D) corresponds to modified poly(dT)G$_{(9)}$, the 100% peaks for C and D correspond to the calculated mass of the unmodified oligonucleotide (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da).*

5.2.3 – Investigations into the Identification of the Species Producing an Unknown Residual Current Signal in the Electrical Recording Experiments

In an effort to establish what species (with identical mass to the starting material) might be causing the emergence of a new residual current block level in the poly(dT)N$_{(9)}$ (where N = G, A and C) reactions with temozolomide, a number of control reactions were carried

out. The control reactions were performed on the poly(dT)$A_{(9)}$ oligonucleotide as there was a large ΔI_{RES} observed between the signal corresponding to unmodified poly(dT)$A_{(9)}$ and the unknown species (X).

Initially, it was investigated if a particular step in the reaction and purification process was producing species X. The reaction procedure itself could be split into three steps 1) heating at 37 °C for 18 h with temozolomide, 2) purification by Econo Pac 10DG columns (Biorad) and 3) concentration of the DNA sample by solvent evaporation under vacuum. Steps 1-3 were repeated on the poly(dT)$A_{(9)}$ oligonucleotide (A), in the absence of temozolomide. Under these reaction conditions, complete conversion to species X was observed (Table 5 – 8). These conditions were also tested on poly(dT)$C_{(9)}$ and incomplete conversion to the unknown species was also noted (ratio of 9:1 X:unmodified poly(dT)$C_{(9)}$ (n = 1)). As species X was detected in the absence of temozolomide, these steps were further divided and individually examined for the production of X. Each of the three steps resulted in incomplete conversion to the unknown species X, with all experiments observing the same $\Delta I_{RES}^{(\text{species X – poly(dT)A(9)})}$ value (within experimental error, Table 5 – 8). It was not possible to test step 2 individually because the DNA concentration, which resulted from purification by the minimum dilution protocol, was too dilute for use in electrical recording experiments (< 5 μM). It was also investigated if species X was produced when the DNA sample was stored at room temperature in pH 8 Tris buffer solution for one week. Under these conditions, incomplete conversion to X was observed (ratio of 1:1 X:unmodified poly(dT)$A_{(9)}$ (n = 1), Figure and Table 5 – 8). Therefore, it has been established that the production of species X occurred gradually over time, under a variety of conditions, and that this modification of the oligonucleotide did not result in a detectable change in mass.

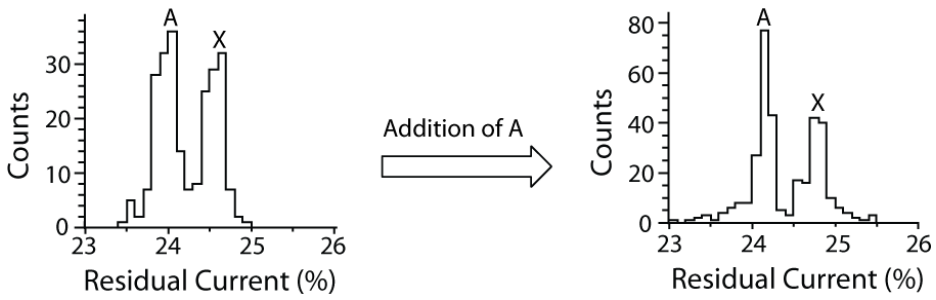


Figure 5 – 8 – *Representative histograms showing the incomplete conversion of poly(dT)$A_{(9)}$ to species X. The histograms shown correspond to incubation of poly(dT)$A_{(9)}$ at room temperature, in pH 8 Tris buffer, for one week (Table 5 – 8, entry 6). The histogram on the left shows the distribution of residual current block levels detected when only the modified poly(dT)$A_{(9)}$ sample was present in the cis compartment – this suggests that A had undergone incomplete conversion to species X, with X producing a higher residual current block level (ratio of 1:1 X:unmodified poly(dT)$A_{(9)}$ (n = 1)). The histogram on the right corresponds to when both modified and unmodified poly(dT)$A_{(9)}$ are present in the cis compartment – one of the residual current levels increased its count number upon addition of A indicating this peak corresponds to the unmodified poly(dT)$A_{(9)}$.*

Entry	Oligonucleotide Sequence (poly(dT)$N_{(9)}$)	Reaction Conditions	ΔI_{RES} (Species X – poly(dT)$N_{(9)}$) ± SD %
1	poly(dT)$A_{(9)}$	Steps 1, 2 and 3 (no temozolomide)	0.64 ± 0.03
2	poly(dT)$C_{(9)}$	Steps 1, 2 and 3 (no temozolomide)	0.99 ± 0.03
3	poly(dT)$A_{(9)}$	Step 1	0.64 ± 0.06
4	poly(dT)$A_{(9)}$	Steps 2 and 3	0.68 ± 0.06
5	poly(dT)$A_{(9)}$	Step 3	0.61 ± 0.05
6[#]	poly(dT)$A_{(9)}$	Room temperature, pH 8 Tris buffer, one week	0.65 ± 0.02

Table 5 – 8 – *The different control experiments that were performed to try and determine to what extent the reaction conditions resulted in the production of species X. The [#] indicates the experiment which corresponds to the histogram in Figure 5 – 8 for the NNY mutant. The ΔI_{RES} values quoted are for n ≥ 3 experiments.*

It was hypothesised that species X was produced as a result of phosphate migration of the 3'-phosphate, attached to the first thymine in the DNA sequence, one carbon along the biotin linker (Figure 5 – 9A). The previously obtained MS data would support phosphate migration (the proposed mechanism is shown in Figure 5 – 9B) because it does not result in an observable change in mass. In an attempt to investigate this process further by ^{1}H and ^{31}P NMR, a very short oligonucleotide was purchased 5'-TT-Biotin TEG 3' (ATDBio). Owing to purification issues, the oligonucleotide sample received was a 50:50 mixture of

the desired product plus an oligonucleotide containing an additional two thymines. The exact location of the additionally attached thymines could not be determined by ^{1}H NMR spectroscopy, therefore, it was not possible to utilise this system to monitor phosphate migration by ^{1}H and ^{31}P NMR experiments. Further investigation of the phosphate migration theory, by examining the DNA fragmentation patterns of modified oligonucleotides produced in MS, was not practical because these experiments were not feasible on the 40nt ssDNA fragments used. However, there was literature precedent for acid/base catalysed phosphate migration of the 3'-phosphate of RNA migrating onto the hydroxyl at C-2 position of the ribose ring.[259] Therefore, it would not be unreasonable to predict that migration one carbon along the biotin linker would also be possible.

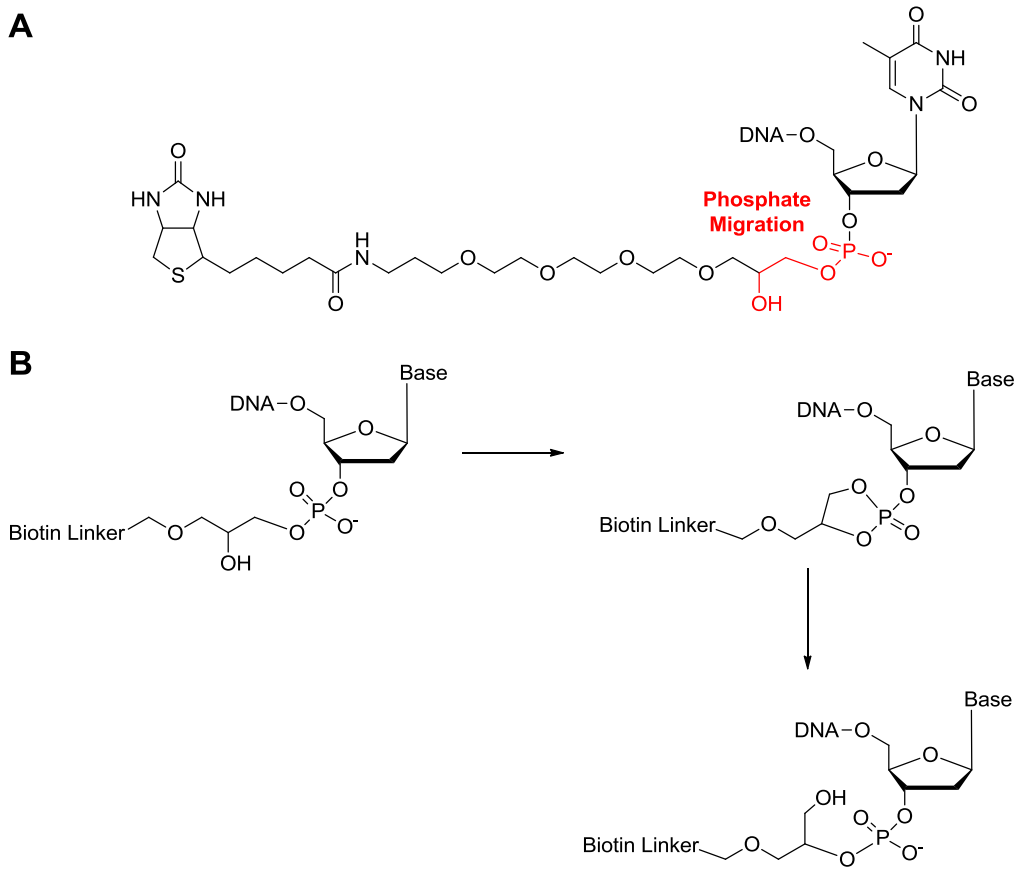


Figure 5 – 9 – *A) the structure of the biotin linker attached to the first base of the ssDNA oligonucleotide, the phosphate moiety which is proposed to undergo migration is highlighted in red. B) Proposed mechanism for the migration of a phosphate group one carbon along the biotin linker chain.*

The phosphate migration hypothesis could have explained the electrical recording experimental results obtained for poly(dT)$A_{(9)}$, poly(dT)$C_{(9)}$ and poly(dT)$G_{(9)}$ because it would have resulted in a change in the overall length of the biotin linker. This in turn would have slightly displaced the base, substituted at position 9, in relation to recognition site R_1, altering the amount of residual current flowing through the αHL mutant. This process would have been unlikely to alter the residual current level for the poly(dT)$_{40}$ oligonucleotide because the sequence at the detection region of R_1 (bases 8-10) would not have altered, in a strand where all the bases are identical. Therefore, from the data obtained it seemed that the proposed phosphate migration was likely to be responsible for the production of the previously unknown residual current level (labelled as species X).

5.2.4 – Investigations into a Method for the Prevention of Phosphate Migration

In order to prevent migration onto the hydroxyl of the adjacent carbon, a protected biotin linker was utilised for the immobilisation of poly(dT)$N_{(9)}$* (where N = C, T, A or G and the * indicates the presence of the protected biotin linker) oligonucleotides inside the αHL protein pore (Figure 5 – 10).

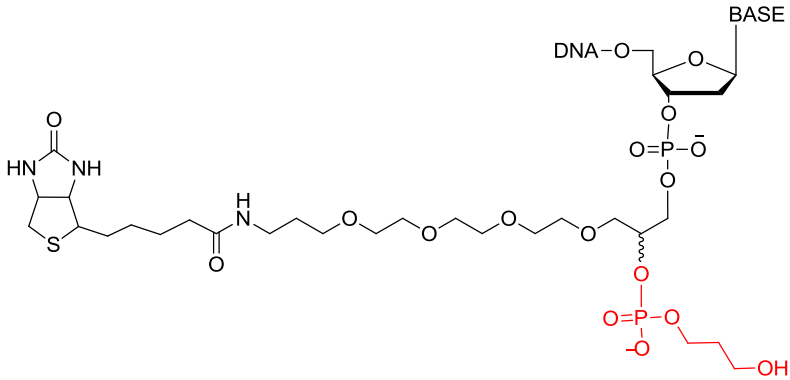
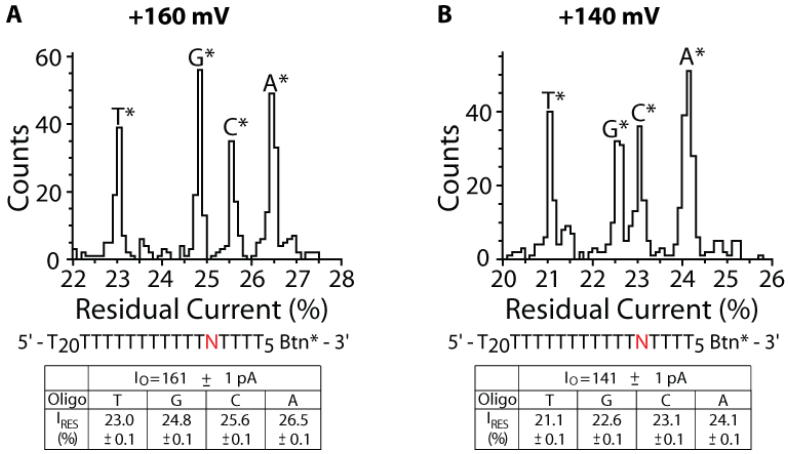


Figure 5 – 10 – *The structure of the new biotin linker which can bind to streptavidin and immobilise ssDNA inside the αHL protein pore. The protecting group employed is highlighted in red.*

In order to investigate the modification of the DNA oligonucleotides by temozolomide, it was first necessary to establish whether the addition of this protecting group affected the

base discrimination properties of the NNY mutant and if the substitution of different bases at position 9 was still within recognition site R_1. The oligonucleotides poly(dT)$N_{(9)}$* (where N = C, T, A or G) were used to probe the base discriminating abilities of the NNY αHL mutant, at applied voltages of +140 mV and +160 mV (Figure 5 – 11). The new linker did not alter base discrimination and the ΔI_{RES} values were the same as observed for immobilisation by the previous biotin linker (within experimental error, see Figures and Tables 5 – 3 and 5 – 11).



	I_O=161 ± 1 pA			
Oligo	T	G	C	A
I_{RES} (%)	23.0 ± 0.1	24.8 ±0.1	25.6 ±0.1	26.5 ± 0.1

	I_O=141 ± 1 pA			
Oligo	T	G	C	A
I_{RES} (%)	21.1 ± 0.1	22.6 ±0.1	23.1 ±0.1	24.1 ± 0.1

Figure 5 – 11 – *Base discrimination abilities of the NNY pore when probed with poly(dT)$N_{(9)}$* oligonucleotides with a single substituted base (N = C, T, A or G) at position 9 relative to the 3' biotin protected TEG. A) Corresponds to an applied voltage of +160 mV and B) corresponds to an applied voltage of +140 mV. Figures A and B correspond to the same experiment.*

Entry	I_O (pA)	I_{RES} poly(dT)$_{40}$* (pA)	I_{RES} C* (pA)	ΔI_{RES} C* %	I_{RES} A* (pA)	ΔI_{RES} A* %	I_{RES} G* (pA)	ΔI_{RES} G* %	ΔI_{RES} OVERALL %
1	162	22.8	25.3	2.5	26.2	3.3	24.6	1.8	3.3
2	159	22.1	24.8	2.8	25.7	3.6	23.9	1.8	3.6
3#	161	23.0	25.6	2.6	26.5	3.4	24.8	1.8	3.4
Mean	161	22.6	25.3	2.6	26.1	3.5	24.4	1.8	3.5
SD	1	0.5	0.4	0.1	0.4	0.2	0.5	0.0	0.2

Table 5 – 11A – *Residual current data for the NNY mutant of αHL probed with poly(dT)$N_{(9)}$* oligonucleotides with a single substituted base (N = C, T, A or G) at position 9 relative to the 3' biotin protected TEG, at an applied voltage of +160 mV (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} A, for example, is the difference in residual current between the poly(dT)$_{40}$* and the poly(dT)-$A_{(9)}$* levels (ΔI_{RES} A* = I_{RES} A* − I_{RES} T*). The # indicates the experiment which corresponds to the histogram in Figure 5 – 11A for the NNY mutant.*

Entry	I_O (pA)	I_{RES} poly(dT)$_{40}$* (pA)	I_{RES} C* (pA)	ΔI_{RES} C* %	I_{RES} A* (pA)	ΔI_{RES} A* %	I_{RES} G* (pA)	ΔI_{RES} G* %	ΔI_{RES} OVERALL %
1	140	20.9	22.9	2.0	23.9	3.0	22.4	1.5	3.0
2	139	20.0	22.2	2.2	23.3	3.3	21.6	1.6	3.3
3$^{\#}$	141	21.1	23.1	2.0	24.1	3.1	22.6	1.5	3.1
Mean	140	20.7	22.7	2.1	23.8	3.1	22.2	1.5	3.1
SD	1	0.6	0.5	0.1	0.4	0.2	0.5	0.1	0.2

Table 5 – 11B – *Residual current data for the NNY mutant of αHL probed with poly(dT)$N_{(9)}$* oligonucleotides with a single substituted base (N = C, T, A or G) at position 9 relative to the 3' biotin protected TEG at an applied voltage of +140 mV (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. The ΔI_{RES} A*, for example, is the difference in residual current between the poly(dT)$_{40}$* and the poly(dT)-$A_{(9)}$* levels (ΔI_{RES} A* = I_{RES} A* − I_{RES} T*). The $^{\#}$ indicates the experiment which corresponds to the histogram in Figure 5 – 11B for the NNY mutant.*

The poly(dT)$N_{(9)}$* (where N = C, T, A or G) oligonucleotides were then stored at room temperature for a week to see if the protected biotin TEG was able to prevent phosphate migration under these conditions. Pleasingly, none of the oligonucleotides underwent modification to produce a product species under these conditions. This provided a system that could be used to study the ability of temozolomide to methylate the DNA bases by electrical recording experiments.

5.2.5 – Electrical Recording Studies of the Methylation of the DNA Bases by the Action of the Cancer Drug Temozolomide Utilising the Protected Biotin•Streptavidin Immobilisation System

The conditions used previously for the modification of the DNA bases by temozolomide did not result in any DNA methylation being observed because the buffer used to stabilise the pH (0.1 M phosphate buffer pH 7.4) was preferentially methylated by temozolomide.[252] Therefore, new conditions were developed for the modification of the poly(dT)$N_{(9)}$* (where N = C, T, A or G) oligonucleotides. The reaction was carried out at room temperature utilising a saturated solution of temozolomide (to maximise the opportunity of observing methylation of the DNA bases) in pH 8.0 Tris buffer (12.5 mM).

Thymine was initially investigated for modification by temozolomide, to establish whether methylation of the background DNA sequence occurred. Under the improved modification conditions, a single block level was observed for modified poly(dT)$_{40}$* (T-TM2*) at an applied potential of +140 mV in electrical recording experiments (Figure 5 – 12A, left). Subsequent addition of unmodified poly(dT)$_{40}$* resulted in a single residual current level with an increased count number (Figure 5 – 12A, right). The corresponding control experiment where unmodified poly(dT)$_{40}$* was added to the *cis* compartment first and the modified oligonucleotide second, produced a single residual current level (Figure 5 − 12B). This meant that the oligonucleotide background sequence (which consists of thymine bases in this case) did not appear to be modified under the new reaction conditions and that this system was suitable for investigating the methylation of the other three bases C, A and G.

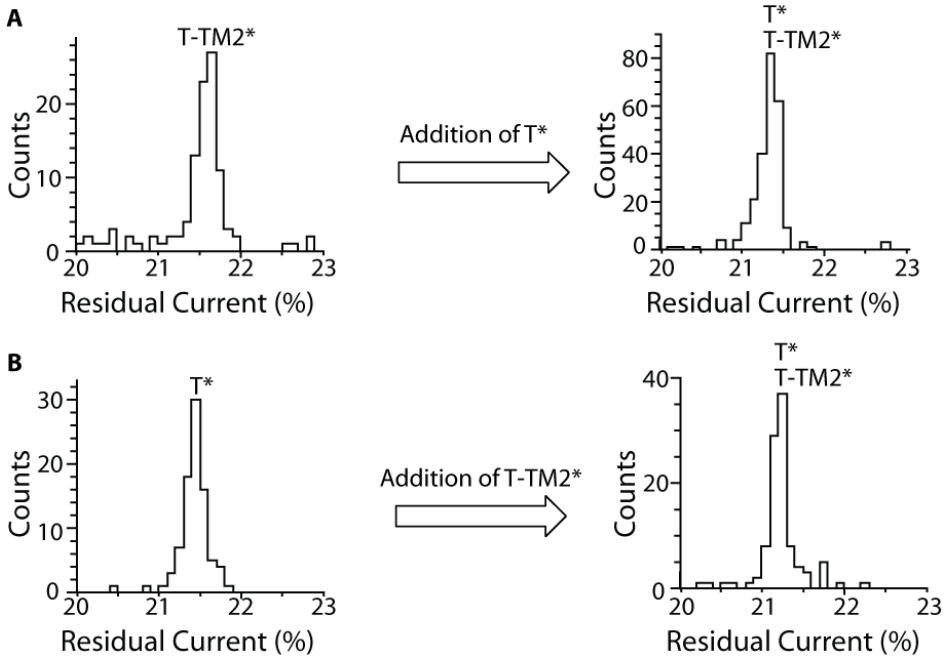


Figure 5 – 12 – *The NNY pore was probed with reacted poly(dT)$_{40}$* (T-TM2*) and un-reacted poly(dT)$_{40}$* (T*) at an applied potential of +140 mV. A) The histogram on the left shows only the presence of T-TM2* in the cis compartment. The histogram on the right corresponds to when T* was added to the cis compartment and a single residual current level with increased count number was observed when both modified (T-TM2*) and unmodified oligonucleotides (T*) were present in the cis compartment. This suggests that no reaction had occurred. B) The control experiment on the left shows only the presence of T* in the cis compartment – a single block level is observed in the histogram. The*

histogram on the right corresponds to when T-TM2 was added to the cis compartment and a single residual current level with increased count number was observed when both modified (T-TM2*) and unmodified oligonucleotides (T*) were present in the cis compartment. This suggests that no reaction had occurred.*

Entry	I_O (pA)	I_{RES} T* (pA)	I_{RES} T-TM2* (pA)	ΔI_{RES} (T-TM2* – T*) %
1a	133	-	20.98	-
1b	135	20.90	20.90	0.00
2a#	133	21.44	-	-
2b#	132	21.21	21.21	0.00
3a#	134	-	21.60	-
3b#	132	21.36	21.36	0.00
Mean	133	21.23	21.21	0.00
SD	1	0.24	0.28	0.00

Table 5 –12 – *Residual current data for the NNY pore probed with reacted poly(dT)$_{40}$* (T-TM2*) and un-reacted poly(dT)$_{40}$* (T*) at an applied potential of +140 mV (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case T-TM2*) and 'b' denotes when both oligonucleotides are present in the cis compartment (T* and T-TM2*). The # indicates the experiments which correspond to the histograms in Figure 5 – 12 (A corresponds to 3a# and 3b# and B corresponds to 2a# and 2b#).*

By employing the new temozolomide modification conditions, the methylation of adenine, guanine and cytosine was also investigated. At an applied potential of +140 mV, modified poly(dT)G$_{(9)}$* (G-TM2*) produced two distinct block levels (Figure 5 – 13A, left), suggesting either incomplete conversion to the product species (ratio of modified:unmodified 4:6 or 6:4 (n = 1)) or complete conversion to two different products. Addition of unmodified poly(dT)G$_{(9)}$* (G*) to the *cis* compartment resulted in the lower residual current signal increasing in count number (Figure 5 – 13A, right). The control experiment, where unmodified G* was added to the *cis* compartment first, confirmed that G* resulted in a single block level and that a new residual current signal was subsequently detected upon addition of G-TM2* (Figure 5 – 13B). Electrical recording experiments observed the $\Delta I_{RES}^{(G\text{-}TM2^*-G^*)} = 0.19 \pm 0.05\%$ (n = 5, Table 5 – 13). This indicated that it

was possible to use electrical recording experiments to distinguish the modified oligonucleotide (G-TM2*) from unmodified poly(dT)$G_{(9)}$*.

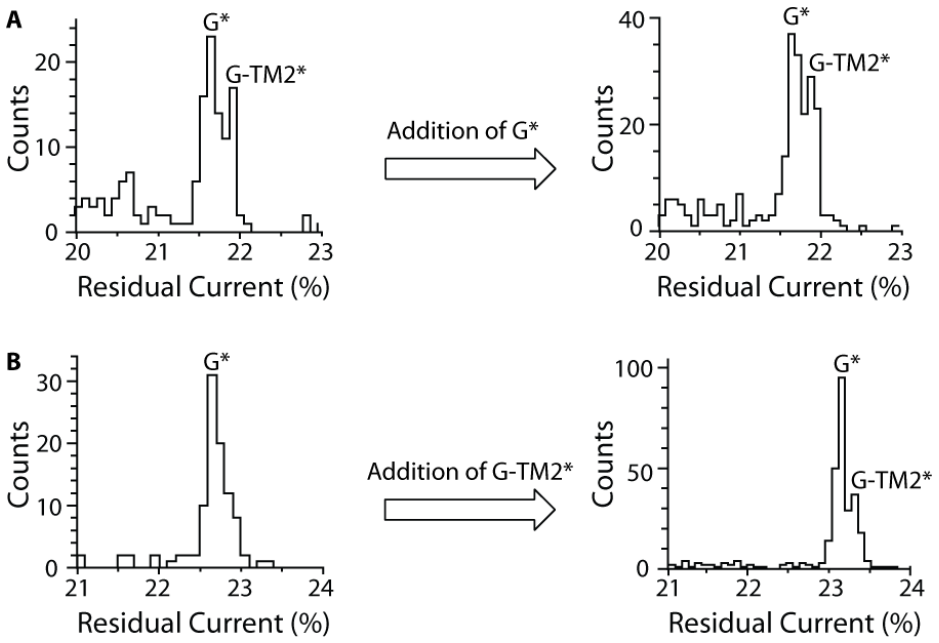


Figure 5 – 13 – *The NNY pore probed with reacted poly(dT)$G_{(9)}$* (G-TM2*) and un-reacted poly(dT)$G_{(9)}$* (G*) at an applied potential of +140 mV. A) The histogram on the left shows only the presence of G-TM2* in the cis compartment – this suggests that guanine has undergone incomplete conversion to the product species. The histogram on the right corresponds to when G* was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, – one of the residual current levels increased its count number upon addition of G* indicating that this peak corresponds to the un-modified poly(dT)$G_{(9)}$*. B) The control experiment on the left showed only the presence of G* in the cis compartment – a single block level was observed. The histogram on the right corresponds to when G-TM2* was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new greater residual current block level signal corresponds to product species (G-TM2*).*

Entry	I_O (pA)	I_{RES} G* (pA)	I_{RES} G-TM2* (pA)	ΔI_{RES} (G-TM2* – G*) %
1a#	134	21.64	21.88	0.25
1b#	134	21.68	21.91	0.23
2a#	137	22.69	-	-
2b#	137	23.15	23.30	0.15
3a	138	22.72	22.92	0.20
3b	138	22.69	22.92	0.23
4a	140	22.58	22.74	0.16
4b	141	22.60	22.71	0.11
5b	144	22.95	23.15	0.21
Mean	138	22.52	22.69	0.19
SD	3	0.52	0.53	0.05

Table 5 – 13 – *Residual current data for the NNY pore probed with reacted poly(dT)$G_{(9)}$* (G-TM2*) and un-reacted poly(dT)$G_{(9)}$* (G*) at an applied potential of +140 mV (n = 5). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case G-TM2*) and 'b' denotes when both oligonucleotides are present in the cis compartment (G* and G-TM2*). The # indicates the experiments which correspond to the histograms in Figure 5 – 13 (A corresponds to 1a# and 1b# and B corresponds to 2a# and 2b#).*

When adenine was investigated by electrical recording experiments, for methylation by temozolomide, incomplete conversion to the methylated products was also observed. When added to the *cis* compartment, A-TM2* produced two distinguishable block levels, one of which increased in count number upon addition of A* (Figure 5 – 14A). The product was observed to produce a higher residual current signal with a $\Delta I_{RES}^{(A\text{-}TM2^*-A^*)} = 0.64 \pm 0.06\%$ (n = 3). Control experiments confirmed that poly(dT)$A_{(9)}$* (A*) produced a single block level and that addition of A-TM2* resulted in the detection of a new and distinguishable residual current signal (Figure 5 – 14B). Finally, methylation of cytosine was investigated by electrical recording experiments. Modified poly(dT)$C_{(9)}$* (C-TM2*) exhibited two distinguishable block levels, the lower residual current signal was found to correspond to unmodified poly(dT)$C_{(9)}$* as it increased in count number upon addition of the unmodified strand (C*) (Figure 5 – 14C). The control experiment confirmed that the unmodified strand produced a single block level (Figure 5 – 14D). The electrical recording experiments established $\Delta I_{RES}^{(C\text{-}TM2^*-C^*)} = 1.07 \pm 0.19\%$ (n = 4). In all

cases, where modification was observed, incomplete conversion to the product was detected, with the modified oligonucleotides allowing more current to flow through the αHL mutant than the corresponding unmodified oligonucleotides. The bases which were observed, by electrical recording experiments, to be modified by temozolomide (A, C and G) were those previously noted in the literature to undergo methylation by this cancer drug.[253]

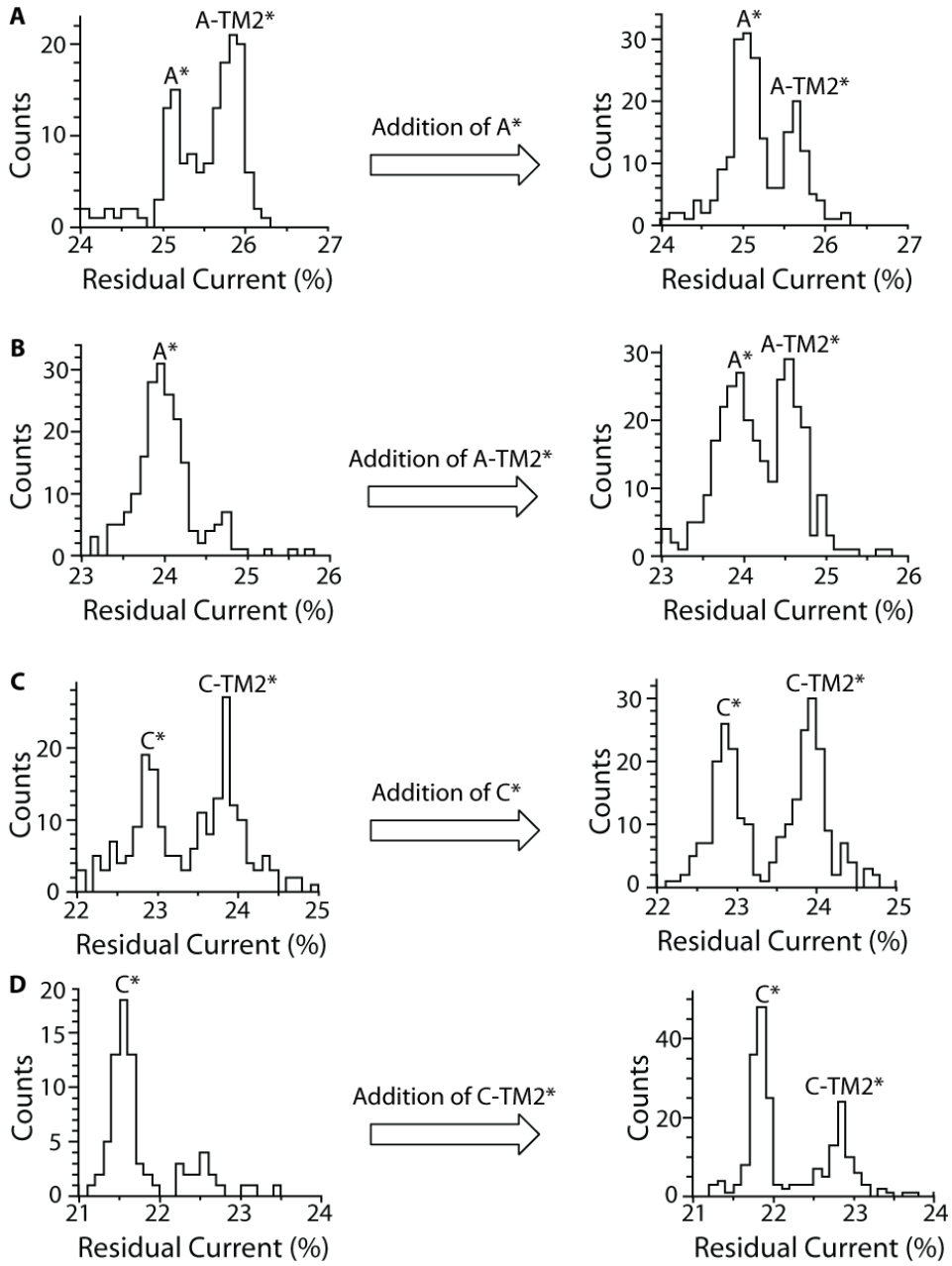


Figure 5 – 14 – *The NNY pore probed with reacted and un-reacted poly(dT)$A_{(9)}$* and poly(dT)$C_{(9)}$* at an applied potential of +140 mV. A) The histogram on the left shows only the presence of A-TM2* in the cis compartment – this suggests that adenine had*

undergone incomplete conversion, with the product producing a higher residual current. The histogram on the right corresponds to when A was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, – one of the residual current levels increased its count number upon addition of A* indicating this peak corresponds to the un-modified poly(dT)$A_{(9)}$*. B) The control experiment on the left shows only the presence of A* in the cis compartment – a single block level was observed. The histogram on the right corresponds to when A-TM2* was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new higher residual current block level corresponded to product species (A-TM2*). C) The histogram on the left shows only the presence of C-TM2* in the cis compartment – this suggests that cytosine had undergone incomplete conversion, with the product producing a higher residual current level. The histogram on the right corresponds to when C* was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment – one of the residual current levels increased its count number upon addition of C* indicating this peak corresponds to the un-modified poly(dT)$C_{(9)}$*. D) The control experiment on the left shows only the presence of C* in the cis compartment – a single block level was observed. The histogram on the right corresponds to when C-TM2* was added to the cis compartment, resulting in the presence of both modified and unmodified oligonucleotides in the cis compartment, the appearance of a new greater residual current block level corresponded to product species (C-TM2*).*

Entry	I_O (pA)	I_{RES} A* (pA)	I_{RES} A-TM2* (pA)	ΔI_{RES} (A-TM2* – A*) %
1a	133	24.49	25.08	0.59
1b	133	24.31	24.96	0.65
2a	134	24.47	25.07	0.59
3a$^{\#}$	133	25.11	25.83	0.72
3b$^{\#}$	132	25.05	25.64	0.59
4a$^{\#}$	130	23.97	-	-
4b$^{\#}$	130	23.89	24.59	0.70
Mean	132	24.47	25.19	0.64
SD	2	0.48	0.46	0.06

Table 5 – 14A,B – *Residual current data for the NNY pore probed with reacted poly(dT)$A_{(9)}$* (A-TM2*) and un-reacted poly(dT)$A_{(9)}$* (A*) at an applied potential of +140 mV (n = 3). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case A-TM2*) and 'b' denotes when both oligonucleotides are present in the cis compartment (A* and A-TM2*). The $^{\#}$ indicates the experiments which correspond to the histograms in Figure 5 – 14 (A corresponds to 3a$^{\#}$ and 3b$^{\#}$ and B corresponds to 4a$^{\#}$ and 4b$^{\#}$).*

Entry	I_O (pA)	I_{RES} C* (pA)	I_{RES} C-TM2* (pA)	ΔI_{RES} (C-TM2* – C*) %
1a	134	22.81	24.26	1.45
1b	137	22.81	23.76	0.95
$2a^{\#}$	134	21.54	-	-
$2b^{\#}$	134	21.83	22.84	1.01
$3a^{\#}$	135	22.89	23.85	0.95
$3b^{\#}$	135	22.87	23.92	1.05
4b	141	22.63	23.61	0.97
Mean	136	22.48	23.71	1.07
SD	3	0.56	0.48	0.19

Table 5 – 14C,D – *Residual current data for the NNY pore probed with reacted poly(dT)$C_{(9)}$* (C-TM2*) and un-reacted poly(dT)$C_{(9)}$* (C*) at an applied potential of +140 mV (n = 4). The I_O and I_{RES} given for each oligonucleotide are mean values taken from Gaussian fits of event histograms, for individual experiments. Experiment number 1a and 1b correspond to the same experiment, where 'a' denotes the residual current data for the first oligonucleotide to be added to the cis compartment (in this case C-TM2*) and 'b' denotes when both oligonucleotides are present in the cis compartment (C* and C-TM2*). The $^{\#}$ indicates the experiments which correspond to the histograms in Figure 5 – 14 (C corresponds to $3a^{\#}$ and $3b^{\#}$ and B corresponds to $2a^{\#}$ and $2b^{\#}$).*

5.2.6 – Mass Spectrometry Analysis of the Modified Poly(dT)$N_{(9)}$* Oligonucleotides Produced Upon Reaction with Temozolomide

In order to confirm methylation of the DNA bases (adenine, guanine and cytosine) by temozolomide, MS analysis was carried out on the modified oligonucleotides. In all cases, a peak for the mass of the unmodified oligonucleotide and a peak corresponding to an increase in mass of +15 Da were observed (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). For example, Figure 5 – 15 shows the MS spectra of the unmodified and unmodified poly$(dT)_{40}$* and poly$(dT)G_{(9)}$* oligonucleotides. The unmodified sample of poly$(dT)_{40}$* (T*) (Figure 5 – 15A) shows a sharp signal corresponding to the mass of the unmodified oligonucleotide (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). The modified sample of poly$(dT)_{40}$* (T-TM2*) (Figure 5 – 15B) exhibited peaks which correspond to the unmodified strand (12815) and a signal corresponding to the addition of a methyl group (12830). This sample exhibited several other signals which did not

correlate exactly to the addition of multiple methyl groups, however, with the observed resolution and quality of the spectra, the possibility of multiple methylations of the oligonucleotide cannot be discounted. Again for the poly(dT)$G_{(9)}$* a sharp signal corresponding to the mass of the unmodified oligonucleotide (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da) was detected. The modified oligonucleotide (G-TM2*) shows a signal for the starting material (12839) and another signal corresponding to the addition of one methyl group (12854). A similar quality spectrum was also observed for this and for the other the oligonucleotide sequences investigated (poly(dT)$A_{(9)}$* and poly(dT)$C_{(9)}$*).

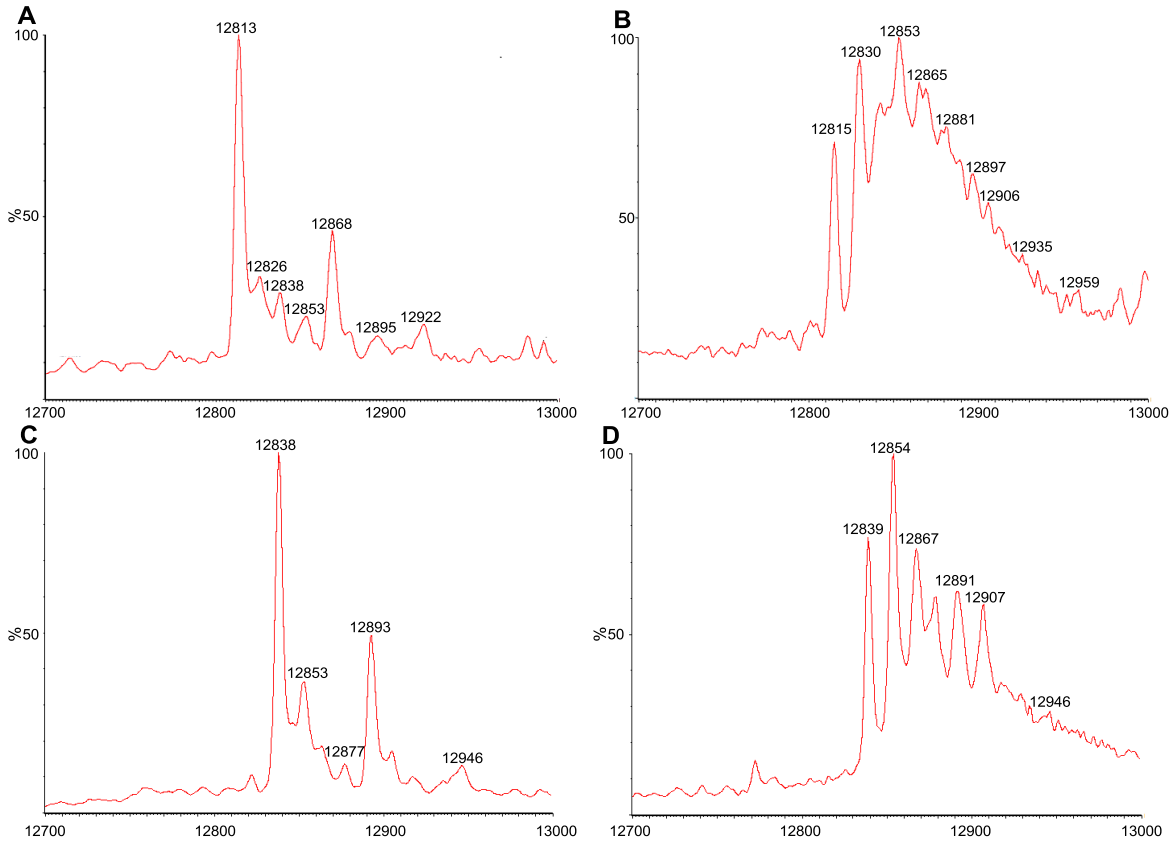


Figure 5 – 15 *– MS spectra corresponding to unmodified and modified poly(dT)$_{40}$* and poly(dT)$G_{(9)}$* before and after treatment with a saturated solution of temozolomide. A) Corresponds to unmodified poly(dT)$_{40}$* (T*) (calculated mass = 12813) and B) corresponds to modified poly(dT)$_{40}$* (T-TM2*), this MS spectrum shows signals for unmethylated and methylated DNA (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da). C) Corresponds to unmodified poly(dT)$G_{(9)}$* (calculated mass = 12836) and D) corresponds to modified poly(dT)$G_{(9)}$* (G-TM2*), this*

MS spectrum shows signals for unmethylated and methylated DNA (within experimental error of the machine, for masses ~10-15 thousand daltons, which is ± 3 Da).

After treatment with temozolomide, methylation was expected for bases A, C and G only, as temozolomide is not known to methylate thymine.[253] It is possible that methylation occurred at bases A, C and G, however, it was proposed that methylation was also occurring at a region of the oligonucleotide which was common to all four of the DNA strands because methylation of thymine was detected. It was possible that temozolomide could have methylated either the backbone phosphates of the ssDNA or the biotin linker.

There is literature precedent for methylation of the backbone phosphates because methylated phosphotriesters were detected by Mizuno *et al.* upon treatment of calf thymus DNA with mitozolomide (which is a similar pro-drug to temozolomide that produces the same reactive intermediate).[260] Although only the addition of a single methyl group can be clearly established by the MS spectra (Figure 5 – 15), owing to the quality and resolution of the spectra obtained, for the modified oligonucleotides after temozolomide treatment (Figure 5 – 15 B and D), it is possible that multiple methylation of the strands could have occurred. If methylation of the phosphate backbone of DNA had occurred on the oligonucleotide strands, it could have produced an overall change in the conformation of the ssDNA, when it was immobilised in the pore, manifesting as a change in the residual current level detected. Another possible explanation could be that if methylation was observed within the region of the recognition site, then the presence of the methyl group itself could produce a new I_{RES} level. It is not known which of the two hypothesise is correct as it was not possible to determine the location(s) of the methyl group(s). Methylation of the phosphate backbone would explain the electrical recording data observed when poly(dT)$A_{(9)}$*, poly(dT)$C_{(9)}$* or poly(dT)$G_{(9)}$* were immobilised inside αHL. The fact that a new block level was not observed for poly$(dT)_{40}$* could be explained if methylation of the backbone had not occurred near recognition site (R_1), however, if

methylation did promote a conformational change then a different residual current signal would also have been expected for the modified poly(dT)$_{40}$* oligonucleotide (T-TM2*).

Another possible theory, which would account for the observed MS spectra, was that temozolomide had methylated the biotin linker used to immobilise the ssDNA oligonucleotides. In an effort to determine whether methylation of the biotin linker had occurred, a model biotin linker compound was exposed to a saturated solution of temozolomide (Figure 5 – 16). MS analysis observed a signal corresponding to the mass of the unmodified linker plus a sodium atom, with no signals corresponding to methylation of the linker being detected, therefore, this hypothesis was discounted.

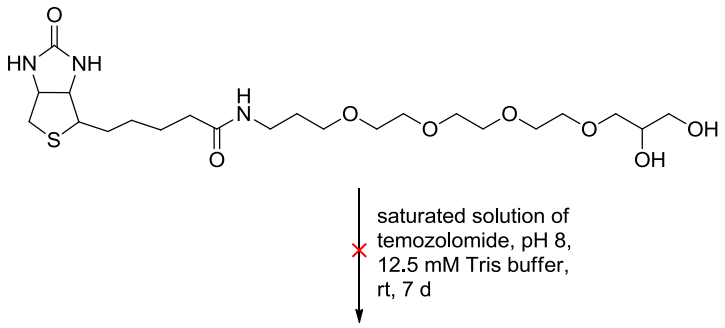


***Figure 5 – 16** – Treatment of the model biotin linker with a saturated solution of temozolomide did not result in methylation of the model compound (MS analysis detected a signal of 530, corresponding to the mass of the linker plus sodium).*

From the data obtained by electrical recording experiments and MS analysis it is proposed that the observed methylation of the poly(dT)N$_{(9)}$* oligonucleotides (where N = C, T, A or G) was owing to both methylation of the A, C and G bases, as expected,[253] and additional methylation of the backbone phosphates.

5.3 – Conclusions

Due to the oxidation of the biotin linker, that was detected by electrical recording experiments, when poly(dA)N$_{(14)}$ oligonucleotides were subjected to two different oxidation conditions (Carell's oxidation reaction and the optimised citric acid dihydroxylation reaction (Chapter 4)),[174] it was not possible to investigate oxidation of the

DNA bases using the biotin•streptavidin immobilisation method. However, it was proposed that chemical modification of the DNA bases could still be investigated, using this system, if non-oxidising conditions were employed. To this end, the cancer drug temozolomide, which had been observed to methylate adenine, cytosine and guanine, was examined. Poly(dT)$N_{(9)}$ (where N = A, C, G or T) oligonucleotides were probed for reaction with temozolomide because thymine, the base which comprises the majority of the poly(dT)$N_{(9)}$ oligonucleotide sequences, was not observed to be methylated by the drug[253] and, therefore, would have provided an un-reactive background control.

The conditions initially investigated for the methylation of the bases was unsuccessful in promoting the methylation reaction. This was likely owing to the occurrence of preferential methylation of the phosphate buffer occurring because the experimental conditions tested had been previously observed to methylate both the solvent and the phosphate use to buffer the reaction.[252] However, electrical recording experiments detected modification of the poly(dT)$A_{(9)}$, poly(dT)$C_{(9)}$ and poly(dT)$G_{(9)}$ oligonucleotides under the conditions tested. It was proposed that the oligonucleotides underwent phosphate migration of the DNA strand one carbon along the biotin linker chain. This would have resulted in an alteration in the linker length, which could have produced a new residual current level for the poly(dT)$A_{(9)}$, poly(dT)$C_{(9)}$ and poly(dT)$G_{(9)}$ oligonucleotides as the location of the base at position 9 would have been altered in relation to recognition site R_1. This hypothesis was also supported by MS analysis of the modified oligonucletides (Figure 5 – 9).

In an effort to prevent phosphate migration, a protected biotin linker was employed (structure shown in Figure 5 – 10). Once it was confirmed that immobilisation using the new linker did not affect the residual current signals produced by the ssDNA, the oligonucleotides (poly(dT)$N_{(9)}$* where N = A, C, G or T) were treated with a saturated

solution of temozolomide. Selective methylation of adenine, cytosine and guanine was thought to have occurred, however, MS analysis showed methylation of all four strands treated with the cancer drug. As thymine was not predicted to have been methylated by temozolomide it was proposed that methylation of the backbone phosphates had also occurred. There is literature precedent for the methyaltion of the backbone phosphates as methylated phosphotriesters were detected by Mizuno *et al.* upon treatment of calf thymus DNA with mitozolomide (a similar pro-drug to temozolomide that produces the same reactive intermediate).[260] Therefore, the observed residual current levels for modified poly(dT)$A_{(9)}$*, poly(dT)$C_{(9)}$* and poly(dT)$G_{(9)}$* oligonucleotides were thought to have occurred as a result of methylation of C, A and G and additional methylation of the backbone phosphates by temozolomide. Methylation of the poly$(dT)_{40}$* oligonucleotide was thought to have not resulted in a different residual current level because methylation had occurred outside of the recognition site, where it is unlikely that it would have been detected by electrical recording experiments.

Therefore, it was proposed that investigations into the modification of the DNA bases, by the cancer drug temozolomide, had exhibited methylation of both the phosphate backbone and the DNA bases themselves.

Chapter 6

Conclusions and Future Work

Chapter 6 – Conclusions and Future Work

6.1 – Conclusions

Work included in this thesis has concentrated on the discrimination of unmodified DNA bases, as well as the investigation, optimisation and detection of selective chemical modifications of the individual nucleobases in ssDNA, using ionic current blocking through a biological nanopore (αHL). It has been established that two of the αHL mutants that were probed for six-base discrimination (NNW and NNY) had the ability to distinguish cytosine from its epigenetic modifications 5-methylcytosine and 5-hydroxymethylcytosine. The NNY mutant was further optimised and shown to distinguish a distinct residual current level for each of the six bases, at an applied potential of +110 mV, when immobilised within recognition site R_1.[184] Consequently, this proved that αHL is capable of detecting all the four standard bases along with the epigenetic modifications 5mC and hmC when they are immobilised using the biotin•streptavidin complex.

A model system (the bis-TBS protected deoxynucleoside) was developed in order to investigate the reactivity and selectivity of a number of chemical reagents for the modification of the DNA bases. The citric acid dihydroxylation and subsequent condensation of an osmate ester was a two-step reaction process that was observed to be selective for *bis*-TBS protected thymidine and was subsequently optimised for use on ssDNA. However, treatment of poly(dA)$N_{(14)}$ (where N = A, C, G, T or 5mC) with two different oxidising conditions (Carell's conditions[174] for the selective oxidation of the C5-C6 bond of 5mC and the citric acid dihydroxylation) resulted in oxidation of the sulfur moiety within the biotin linker, instead of DNA base modification. It was, therefore,

established that it was not possible to investigate oxidising modification conditions, employing the biotin•streptavidin immobilisation strategy.

Finally, the methylation of the DNA bases, by the cancer drug temozolomide, was probed by electrical recording experiments, as these conditions would not have been expected to result in oxidation of the biotin linker. Initially, the conditions which were employed to promote methylation were observed to preferentially methylate the phosphate buffer, used to maintain the pH. Although no chemical modification was detected by MS analysis, modification of the oligonucleotides was detected by electrical recording experiments. Further investigations illustrated that the probable cause, for the detection of new residual current levels, was likely to be the phosphate migration of the ssDNA along the biotin linker. A new protected biotin linker was subsequently utilised for the immobilisation of ssDNA and the methylation of the bases, by a saturated solution of temozolomide, was investigated. Electrical recording data suggested that selective methylation of A, C and G had occurred, however, MS analysis showed methylation of all four poly(dT)$N_{(9)}$* (where N = A, C, G or T) oligonucleotides. It was hypothesised that additional methylation of the phosphate backbone had occurred. Therefore, it was proposed that investigations into the modification of the DNA bases, by the cancer drug temozolomide, had observed both methylation of the phosphate backbone and the DNA bases themselves.

As a result of the work carried out herein, it was established that the αHL nanopore is capable of distinguishing all six bases, when they are immobilised using the stalling biotin•streptavidin complex. A model system was effectively employed for the investigation of selective chemical modification reactions, two of which were further optimised for the modification of ssDNA. The biotin•strepatividin immobilisation strategy was employed for the investigation of two selective oxidation reaction conditions, however, the sulfur moiety of the biotin linker itself was observed to react under both

reaction conditions investigated. This meant that another immobilisation strategy would need to be employed, in order to successfully investigate oxidation reactions in the future. The cancer drug temozolomide was utilised in an attempt to modify the DNA nucleobases. The ssDNA attached to the biotin linker was observed to undergo phosphate migration, therefore, a protected biotin linker was employed for all further experiments. Upon treatment of ssDNA (containing the protected biotin linker) with a saturated solution of temozolomide, methylation of A, C and G as well as the backbone phosphates was observed.

6.2 – Future Work

It has been demonstrated in this thesis that by utilising nanopore technology, it is possible to discriminate the four canonical bases as well as the epigenetically modified bases 5mC and hmC.[184] Consequently, if this technology can be integrated into a practicable nanopore device, then it should be possible for αHL to identify each of the bases individually. Currently, there are a number of challenges which need to be overcome before nanopores could be effectively employed for the sequencing of DNA (Section 1.1.6).[25] The next major challenge is to significantly reduce the rate at which ssDNA currently translocates (~1-20 μs per base) so that individual nucleotide transitions can be detected as they pass through the pore.[48, 49] One potential method for controlling unidirectional processive motion of DNA through the αHL pore, is by the employment of enzymes such as DNA helicases, [85, 86] polymerases[53-55] or DNA exonucleases.[62, 63] Another viable approach could involve the detection of DNA's component nucleoside monophosphates as they are released (by the action of a DNA exonuclease covalently attached to the nanopore) and translocated through the pore.[63] Both of the above methods, for the slowing of DNA translocation, are currently under intense investigation by a number of research groups and

companies and it is hoped that, in the future, nanopores will provide a fast and effective method of sequencing entire human genomes.[53, 54, 63]

The biotin•streptavidin complex has provided an excellent system for the immobilisation of DNA, to enable the study of the base discriminating ability of the αHL pore. However, this immobilisation method has not been effective for the investigation of chemical modifications of the DNA bases, as it has been modified under the chemical reaction conditions. This has resulted in the detection of chemical modifications, by electrical recording experiments, which have not occurred to the DNA bases themselves e.g. oxidation of the sulfur moiety of biotin. This meant that it was not possible to effectively monitor chemical modifications of DNA by electrical recording, when this immobilisation strategy was used, because the reactivity of the linker itself affected the residual current signal observed for the modified oligonucleotides. Therefore, it is necessary to develop a new method for immobilising the DNA. One option would be to attach the biotin moiety after the chemical modification reaction has occurred and the other would be to utilise a different immobilisation strategy.

There are a number of different methods for the attachment of biotin to the 3' or 5' end of DNA or even to a DNA base within the ssDNA sequence.[261-263] As the biotin moiety will need to be attached to the ssDNA after the selective modification reaction has occurred, there are several important factors which require consideration before a method of attachment is selected. Firstly, it would be necessary to establish biotin attachment conditions which promote the incorporation of a single biotin moiety at a predetermined location (e.g. the 3' end of the ssDNA). If incorporation of more than one biotin functionality was possible, at several sites along the DNA oligonucleotide, then this would potentially result in either the detection of multiple block levels in the electrical recording experiments or the complete prevention of DNA capture. The employment of completely

selective biotin attachment conditions would be vital, so that modification of the DNA bases did not occur under the reaction conditions. For example attachment of biotin *via* the reaction of a *N*-hydroxysuccinimide biotin analogue with an NH_2 functionality would need to be carefully controlled so as to avoid biotin incorporation at multiple positions in the DNA strand (as cytosine, adenine and guanine all contain the NH_2 functionality). Finally, it would be important to establish reaction conditions that promoted quantitative conversion to the biotinylated oligonucleotides as incomplete modification would prevent detection of the non-biotinylated ssDNA in electrical recording experiments.

The enzyme terminal deoxynucleotidyl transferase (TdT) could offer a possible method for biotin attachment to the 3' end of DNA.[261, 262] The main advantages of this approach are that it can be used to incorporate a single biotin moiety, a template DNA sequence is not required for the enzyme to function and the enzyme is not significantly affected by the DNA sequence. However, enzyme incorporation is generally performed on a small scale, which would limit the quantity of DNA that could be biotinylated using this method.

A number of methods, other than the biotin•streptavidin approach, have been utilised for the immobilisation of DNA or the formation of DNA αHL rotaxanes.[50-55, 57, 64, 78] The two main methods employ either a DNA hairpin[50, 51, 57, 64] or a short complementary dsDNA fragment at one end of a long ssDNA oligonucleotide (Figure 6 – 1).[53-55, 64, 78] These immobilisation techniques would provide a system that would consist entirely of DNA and contain no additional chemical functionality. As a consequence of this, any modifications observed by MS and electrical recording analysis would be due to DNA modification alone, and not modification of the linker required for immobilisation. However, there are various technical challenges that would need to be overcome for this method to be utilised for the immobilisation of DNA. There is literature precedent, from the work of Ghadiri and co-workers, for the use of the DNA hairpin immobilisation technique for determining

single nucleobase recognition in αHL.[50] However, despite being able to observe single nucleobase resolution, it was discovered that variations in the hairpin structure shifted the alignment of the captured DNA strand relative to the recognition site. Furthermore, Purnell *et al.* observed multiple dynamic current levels when a DNA hairpin was immobilised within the αHL pore.[57] This was also observed by Akeson and co-workers, who illustrated that the block level produced by an immobilised DNA hairpin depended on the length and base pairing of the hairpin.[51] Therefore, the sequence used for either the hairpin or a short complementary DNA fragment would need to be carefully designed to minimise the production of multiple block levels when immobilised in the pore. It has also been observed that the DNA double helix can be un-zipped if sufficiently high potentials are applied across the bilayer in the presence of immobilised strands.[54, 78] Therefore, another important experimental consideration would be the employment of sufficiently low potentials so as to facilitate ssDNA capture but prevent the un-zipping of the DNA double helix. However, this should not limit the use of this particular immobilisation technique for the electrical recording experiments discussed here because experiments carried out by Ghadiri and co-workers shows that hairpins containing 12 base pairs could be used to immobilise DNA in the αHL pore at applied potentials up to +170 mV.[50]

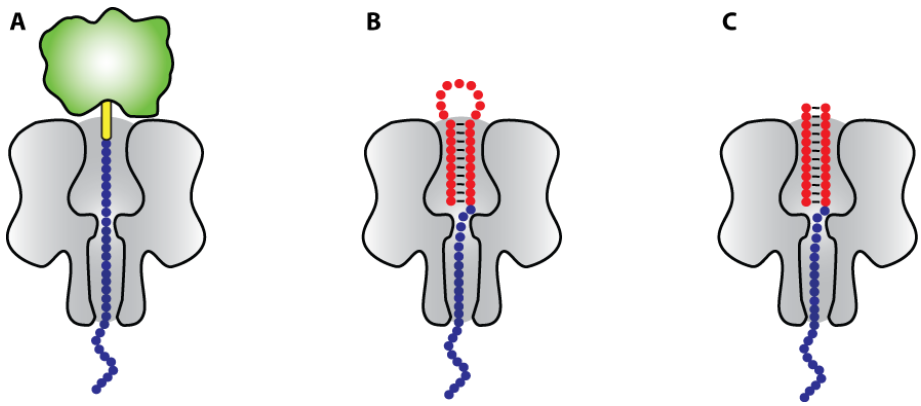


Figure 6 – 1 – *Cartoon representation of three possible methods of immobilisation of DNA within the αHL pore: A – biotin•streptavindin, B – DNA hairpin and C – complementary dsDNA fragment.*

If either of the described immobilisation techniques could be developed into a viable model system, a number of different modification conditions could be investigated. Selective chemical reactions, such as the citric acid dihydroxylation of thymine, could be developed on the *bis*-TBS protected deoxynucleosides and then tested for modification of ssDNA by electrical recording experiments. The effects of many drug compounds, such as temozolomide, could also be explored using this technique. This system could be particularly useful for the investigation of how new drugs promote their therapeutic effect *in vivo*, when their mechanism of action is unknown. Furthermore, if a practicable nanopore sequencing device is developed, then it could be used to monitor cancer treatment, by drugs such as temozolomide, as it would be possible to detect modified bases directly by sequencing of DNA extracted from cancer cells. This is currently not possible using other sequencing techniques which can only identify the standard bases adenine, cytosine, guanine and thymine.[14]

Chapter 7

Materials and Methods

Chapter 7 – Materials and Methods

7.1 – Molecular Biology

All of the αHL genes and constructs were assembled in the pT7 expression vector. Heptameric αHL E111N/K147N (NN), E111N/K147N/M113I (NNI), E111N/K147N/M113L (NNL), E111N/K147N/M113V (NNV), E111N/K147N/M113R (NNR), E111N/K147N/M113W (NNW) and E111N/K147N/M113Y (NNY) were produced as described.[39] All αHL mutants were kindly prepared by E. Mikhailova. The mutant genes were prepared by using a kit for site-directed mutagenesis (QuikChange II XL, no. 200533-5, Stratagene). The DNA sequences were verified by sequencing at the Source BioScience LifeSciences centre, located in the Department of Biochemistry, at the University of Oxford. αHL heptamers were stored at –80 °C in 10.0 mM Tris buffer, 100 μM EDTA, pH 8.00.

7.2 – Planar Bilayer Recordings

7.2.1 – Materials

In nearly all cases, single-channel recording experiments were carried out in 1 M KCl, 25 mM Tris, 100 μM EDTA, pH 8.0 buffer. To reduce contaminants in the buffer solution the potassium chloride was of the highest purity available (less than 0.0005% trace metals, Fluka catalog no. 05257).

Custom synthesised oligonucleotides were obtained as lyophilized pellets from Sigma Genosys and ATDBio Ltd. As pure oliognucleotides are vital for nanopore sequencing experiments the samples were all desalted and HPLC purified (with the exception of the biotinylated oligonucleotide that was two nucleotides in length) by the company which provided them. All single stranded DNA (ssDNA) had a biotinyl group attached at the 3'

end through either a triethyleneglycol (TEG) linker or a modified triethyleneglycol linker with an additional protecting group on the free hydroxyl of the biotin linker (Figure 7 – 1). The * symbol denotes oligonucleotides containing the protected biotin-TEG at the 3' end. ssDNA samples were dissolved in 25 mM Tris, pH 8.0, 100 µM EDTA to a final concentration of 100 µM and stored at –20 °C.

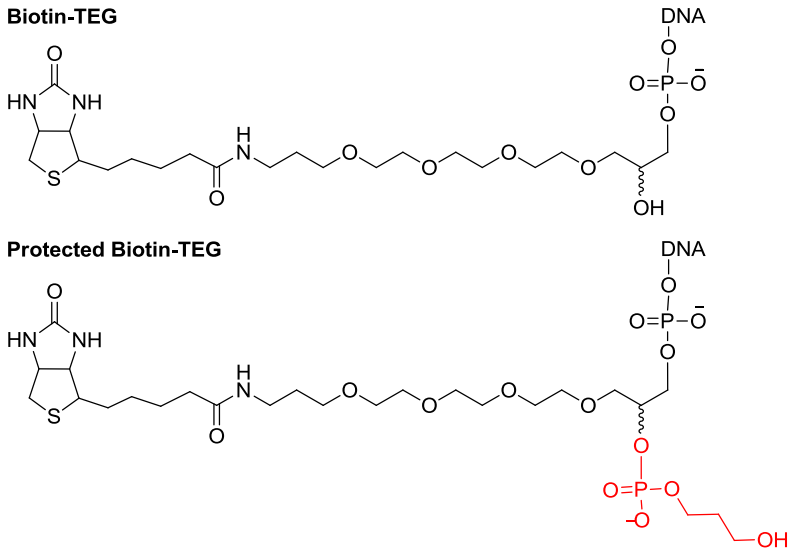


Figure 7 – 1 *– The chemical structure of the two types of biotin-TEG linker used to biotinylate the 3' end of the ssDNA oligonucleotides. The length of the linker, from biotin to the 3' phosphate group of the first DNA base, was calculated to be ~3nm using the program pymol.*

7.2.2 – Electrical Recording Equipment

Current recordings were performed using a patch clamp amplifier (Axopatch 200B, Molecular Devices), with a Digidata 1440A 16-bit digitizer (Molecular Devices). Experiments were carried out in the voltage-clamp mode, where the applied potential is fixed and the transmembrane current generated by the that voltage drop was measured. The measured transmembrane current is proportional to the conductance of the biological nanopore at a given voltage. It was important to be able to control the applied potential as many of the interactions between DNA and αHL are voltage-dependent.

7.2.3 – Faraday Cage

All experiments were conducted inside a Faraday cage in order to shield the active electrodes from external electrical noise. The cage consisted of a steel box with rigid side walls that was grounded *via* the common ground in the patch clamp amplifier.

7.2.4 – Electrodes

In all experiments Ag/AgCl electrodes (Figure 7 – 2) were used – these were known as the 'ground electrode' and the 'active electrode' which were connected to the *cis* and *trans* compartments respectively. For the experiments discussed in this thesis, the *cis* compartment is defined as the compartment which connected the 'ground electrode' to the recording equipment, the side of the bilayer to which protein was added in electrical recording experiments and side of the bilayer into which the cap domain of αHL protruded.[36, 39] The *trans* compartment is defined as the compartment which connected the 'active electrode' to the recording equipment and the side of the bilayer into which the stem domain of αHL protruded.[36] This particular type of electrode was used because it gave long-term chemical stability, allowed reversible electrochemical behaviour, gave predictable junction potentials and a low-noise electrical performance.[183] To prepare the electrodes we used 1.5 mm diameter silver wire (Sigma-Aldrich), which was cleaned with water and scrubbed with an abrasive material (e.g. glass paper), to increase the surface area of the electrode. The electrodes were then treated overnight with a 30% (w/v) NaClO solution (Fisher Scientific) to chlorinate them with a thick-coating of AgCl. To prolong the electrode life-time and prevent UV conversion from AgCl to Ag, the electrodes were rinsed with water after use and stored in the dark.

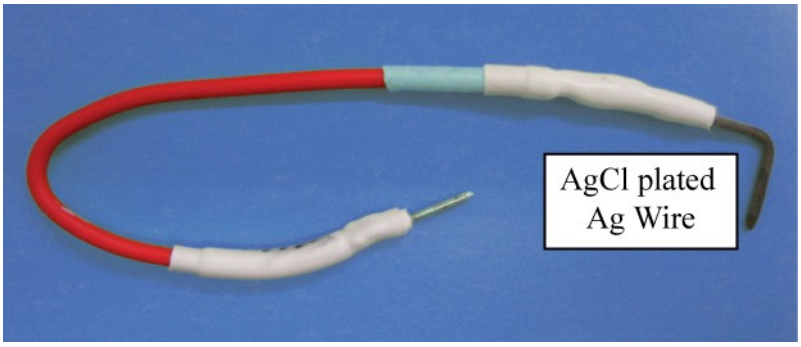


Figure 7 – 2 – *Photograph of a Ag/AgCl electrode that has been pre-treated with hypochlorite solution. Figure adapted from Maglia et al. 2010.*[183]

7.2.5 – Chamber

The open planar-bilayer chambers that were used for all single channel recording experiments were custom made from Delrin (see Figure 7 – 3). The chamber consisted of two main compartments (*cis* and *trans*), each with a volume of 500 µL which were separated by a 25 µm-thick polytetrafluoroethylene film (PTFE, Goodfellow, UK). With this set-up both compartments were easily accessible for the purposes of adding reagents and perfusion of the solution (see Chapter 7, section 7.2.7).

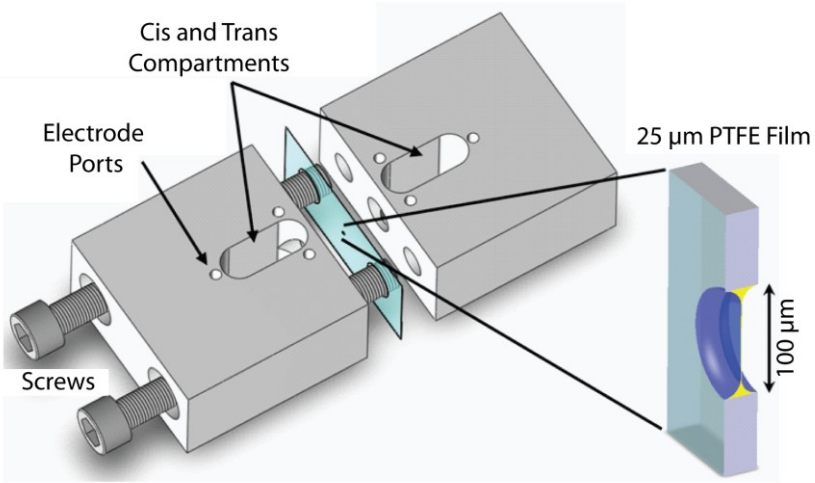


Figure 7 – 3 – *Custom-built planar-bilayer chamber (left), with two compartments (each holding 500 µL) that are separated by a 25 µm PTFE film (in expanded view on the right). Clamped between the two halves of the chambers, by two screws, is the PTFE film which is sealed with silicon glue to provide water-tightness. In the middle of the PTFE is a small aperture, approximately 100 µm in diameter, across which bilayers are formed. This figure was adapted from Maglia et al. 2010.*[183]

The only connecting pathway between the *cis* and *trans* compartments was through a central aperture in the PTFE film, across which the bilayer was formed. To form the aperture the PTFE must first be weakened, in the desired location, by gently indenting the film with a sharp needle. The film was then mounted between the two electrodes of a 30 kV spark source (Ealing Corporation, Rocklin, California) and a series of 50 μs sparks applied at 1 Hz. The sparks continually cause localized melting of the PTFE centred on the weakened area, resulting in a perfectly round aperture of variable sized depending on sparking time.[264] The diameter of the aperture was measured using a microscope with a graticle and was generally found to be ~ 100 μm.[265] The chamber was cleaned between experiments by rinsing with water, then distilled water and finally ethanol. The ethanol was removed by drying the chamber under a stream of nitrogen gas.

7.2.6 – Preparing Bilayers

In order to form planar lipid bilayers, the approach developed by Montal and Mueller[266] was employed. The aperture was pre-treated with ~ 5 μL hexadecane solution (10% (v/v) in pentane) and the pentane allowed to evaporate. This resulted in the oil forming an annulus around the rim of the aperture, which allowed the bilayer to form an interface with the plastic film (Plateau-Gibbs border).[267] To both the *cis* and *trans* compartments were added 500 μL of buffered, aqueous electrolyte solution (typically 1 M KCl, 25 mM Tris, 100 μM EDTA, pH 8.0) and the Ag/AgCl electrodes were connected *via* the electrode ports (Figure 7 – 3), such that the *cis* was connected to the ground. At this point, there was no bilayer formed over the aperture, therefore, it should be possible to measure an electrical current. In the case where oil was blocking the aperture (i.e. no current was measured) this was removed by gentle pipetting of the solution at the aperture. A small quantity (~5 μL) of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC, Avanti Polar Lipids), at a concentration of 10 mg mL^{-1} (w/v) in pentane, was added to the surface of the

aqueous phase of both compartments. The pentane was again allowed to evaporate, forming a phospholipid monolayer at the solution air interface, then the solution was pipetted from the bottom so that the level of the lipid monolayer, at the water-air interface, was lowered past the level of the aperture forming a monolayer with the hydrophobic lipid tails towards the air. The solution level was then slowly raised back above the aperture folding the second monolayer across the aperture forming a bilayer. This process was repeated in both the *cis* and *trans* compartments until a bilayer formed (Figure 7 – 4). The bilayer formation was detected by a current measurement reading of 0 pA, indicating that an electrical seal was present between the two compartments.

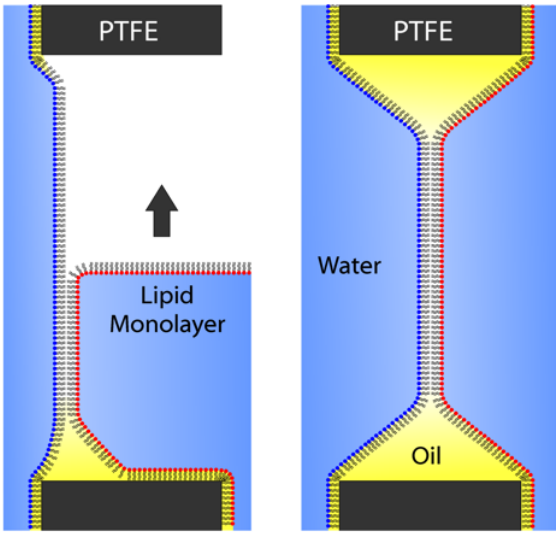


Figure 7 – 4 *– The planar-bilayer was formed by the Montal and Mueller*[266] *approach. The aperture was pre-treated with hexadecane (10% (v/v)) and the bilayer was formed by flowing lipid monolayers over the aperture. This figure was adapted from Maglia et al. 2010.*[183]

7.2.7 – Inserting Pores and Perfusion

In order to initiate αHL pore insertion into the planar bilayer typically 0.2 μL of aliquoted protein (approximate concentration 1 ng/μL^{-1}) was added to the *cis* compartment. The solution was then mixed by pipetting. The insertion of a protein into the bilayer, under an applied potential (typically +200 mV), was observed as a single step-wise change in the

current (Figure 7 – 5). The current step was dependent on the applied voltage, the αHL mutant being used and the concentration of the electrolyte (1 M KCl in most cases). Once a single pore had inserted the *cis* compartment was perfused by manual pipetting (removal of protein-containing buffer and replacement with protein-free buffer solution). This perfusion removes free protein from the bulk solution, reducing the overall protein concentration in the *cis* compartment and decreasing the probability of additional proteins inserting into the bilayer.

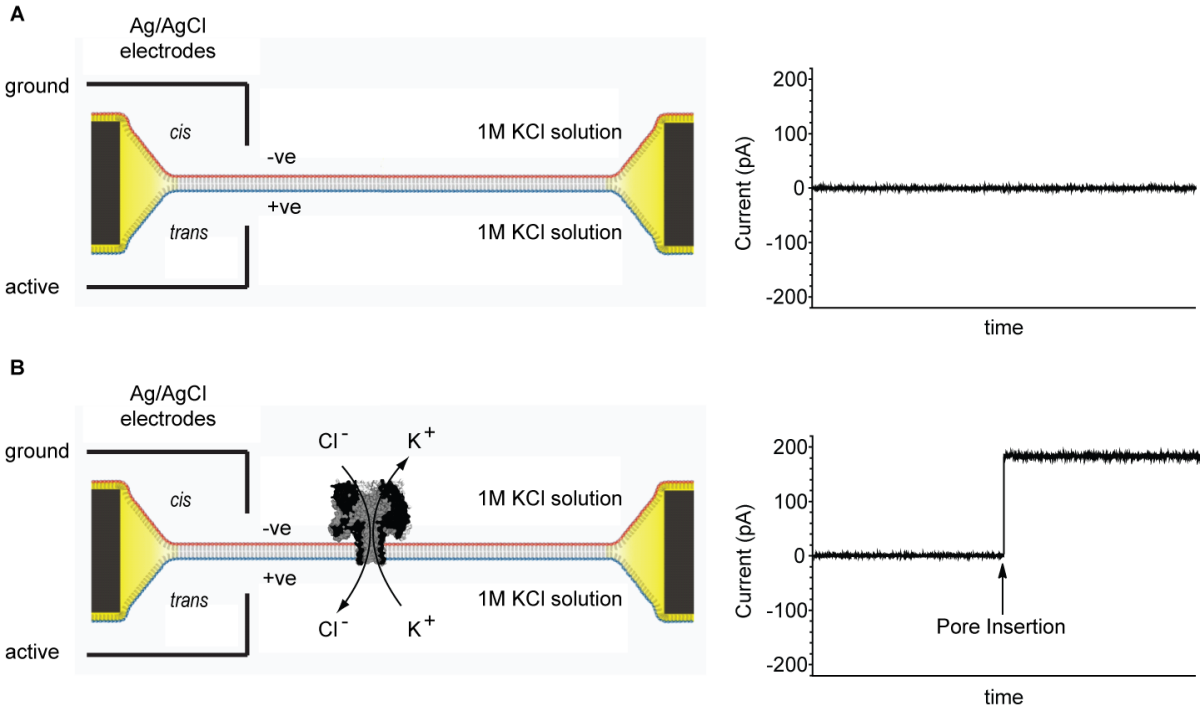


Figure 7 – 5 *– Current measurements observed upon bilayer formation and pore insertion. A) Formation of a planar bilayer is defined by a current reading of zero pA (under positive or negative applied potential) between the cis and trans compartments. This is owing to complete electrical isolation of the two compartments. B) Upon insertion of αHL (grey/black cross-section) the cis and trans compartments become electrically connected. This allows an ionic current to flow between the compartments, which is characterised as a step-wise change in the current reading (applied potential in this instance of +200 mV).*

7.3 – Data Acquisition and Processing

7.3.1 – Data Digitization

The ionic current that flows through a nanopore can be measured using the commercially available Axopatch 200B patch-clamp amplifier (Molecular Devices). This instrument converted the generated continuous analog current signal into an analog voltage signal, which was in turn fed to a 16-bit digitizer (Digidata 1440A, Molecular Devices) that converted it into binary numbers which had a finite resolution. Finally, the digitizer was connected to a computer that converted and displayed the digitized current information. For the single channel recording experiments the output gain, that was applied to the amplifier, was set to 10X (i.e the conversion of 1 pA measured current to 10 mV voltage). This corresponded to a dynamic range of ±1 nA and a resolution of 0.0305 pA per bin.[183]

7.3.2 – Filtering and Sampling

In order to remove the high-frequency electrical noise generated during the experiment, the signal collected by the amplifier must be filtered. The filter employed in this case was a low pass 4-pole Bessel filter which was built into the amplifier. Bessel filters were used in preference to the Butterworth filter because it preserves the shape of the single-channel response (Axon Guide).[268] To keep errors in the current levels estimates to <3% the applied filter cut-off used should be greater than five times the inverse mean event time.[268] If the signal is over filtered then it will lead to inaccurate determination of the current levels.

As well as filtering the signal it is important to consider the sampling of the data. The Nyquist Sampling Theorem states that the rate of sampling should be twice the cut-off frequency of the applied filter for an ideal filter (The Axon Guide).[269] In practice, for a real filter with a non-brick wall transfer function, and a non-sinusoidal signal, a higher

sampling rate is necessary to avoid significant aliasing, therefore, a sampling rate of five times the cut-off frequency was used.[183]

7.3.3 – Acquisition Protocols

The software used to collect and display the data obtained throughout single channel recordings was pClamp 10. (Molecular Devices). This programme can also be used to control the digitizer so that it will apply a potential through the amplifier. The 'gap-free' acquisition mode was used while waiting for a channel to insert into the bilayer. A constant DC voltage is applied, and the current data are passively and continuously digitized, displayed and saved.

The episodic stimulation mode was used to output a voltage signal from the digitizer, and apply it to the bilayer system, while at the same time acquiring the resulting current signal. The voltage protocol shown in Figure 7 – 6 was used to study the immobilization of ssDNA inside the αHL protein.[59-61, 183, 184] Using the Episodic Stimulation mode in the Clampex software (Molecular Devices), a Digidata 1440A digitizer (Molecular Devices) was used to apply repeated voltage steps to drive a DNA-biotin•streptavidin complex into the pore (positive potential) and then eject the complex (negative potential). The changes in the current level, that were observed as a result of the applied voltage protocol, were simultaneously acquired. The protocol can be repeated hundreds of times to record many DNA blocking events.

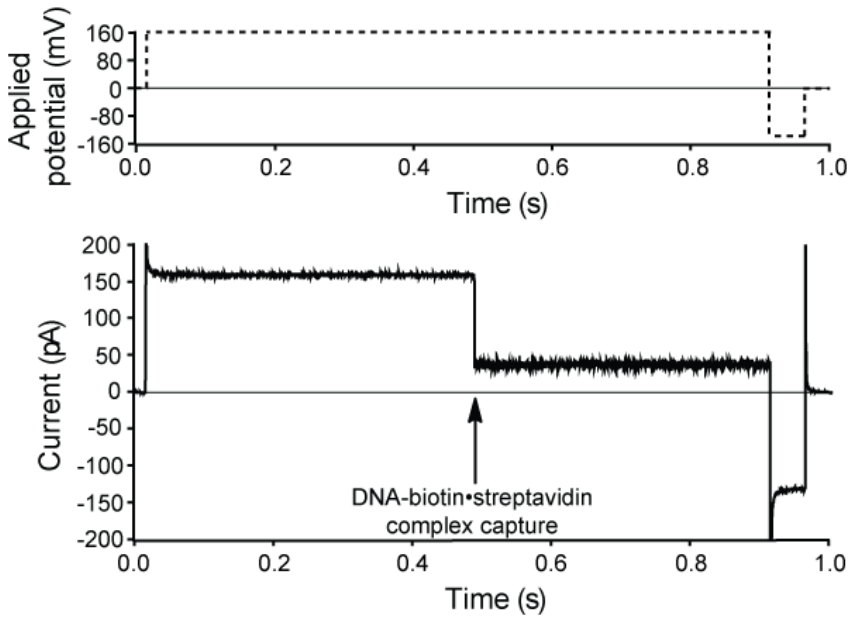


Figure 7 – 6 – *Episodic simulation mode allows both an output voltage to be applied (dashed line), while recording the resulting changes in ionic current (lower – black trace) at the same time. A positive voltage, in this case +160 mV, was applied to the system, resulting in an open pore current level of ~ 160 pA. Under the applied potential a DNA molecule was attracted into the pore and immobilised (through a terminal biotin•streptavidin complex). This is observed as a reduction in the current flowing through the pore. The applied potential is then reversed (−140 mV) and the DNA-biotin•streptavidin complex expelled. The applied potential is then returned to zero. This entire protocol takes one second and is defined as one episodic simulation sweep. The protocol can be repeated hundreds of times. This figure was adapted from Maglia et al. 2010.*[183]

7.3.4 – Analysis of Unmodified DNA-Biotin•Streptavidin Molecules

ssDNA oligonucleotides (2 µL of 100 µM in 25 mM Tris, pH 8.0, 100 µM EDTA), with an appropriate biotinyl group attached at the 3' end were pre-incubated with streptavidin (3 µL of 1 mg/mL stock solution in 10 mM sodium phosphate, 150 mM sodium chloride, pH 7.2, Sigma-Aldrich) at room temperature for at least 5 min (see Chapter 9, appendix 9.4 for the sequences of all oligonucleotides). Each DNA strand was added to the *cis* compartment to a final concentration of typically 200 nM. In a standard experiment, +160 mV was applied to the *trans* side for 900 ms to drive the negatively charged DNA-biotin•streptavidin complex into the pore (Figure 7 – 6). The capture of a DNA strand by the αHL pore was observed as a stepwise decrease in the open pore current level (I_O) to a lower current level known as the block level (I_B) (Figure 7 – 7). A voltage of −140

mV was then applied for 50 ms to eject the immobilized DNA complex from the pore. The applied potential was then stepped to 0 mV for 50 ms. This one-second sequence or sweep was repeated for at least 100 cycles for each DNA species added. The signal was low-pass filtered at 1 kHz and sampled at 5 kHz.

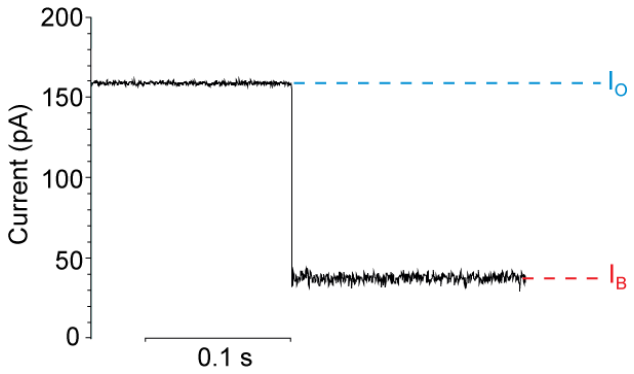


***Figure** 7 – 7 – Typical DNA-biotin•streptavidin block event. The current changed from an open pore event to a blocked pore event and this was analysed by Clampfit software. The program assigns the events and calculates the mean value for I_O (blue) and I_B (red) (indicated by dashed lines).*

Data were analyzed and presented using pClamp (version 10.1, Molecular Devices). Single-channel searches were performed to find the average current level for each ssDNA block event (I_B). It was necessary to estimate pre-defined threshold levels for the open pore (level 1) and the blocked pore level (level 2) respectively. The software detects when the ionic current crosses from a level 1 event (open pore) to a level 2 event (blocked pore) and determines the arithmetic mean current of each event from all acquired samples belonging to a particular event. The I_B values observed from repeated blocks, for a particular DNA sequence, have approximately a Gaussian distribution (Figure 7 – 8). To determine the mean I_B value for each sequence a Gaussian was fitted to the histogram of all of the I_B values from repeated blocks. The current block for each oligonucleotide was expressed as a percentage of the open pore current (I_O): $I_{RES} = (I_B/I_O) \times 100$ (Figure 7 – 8B).

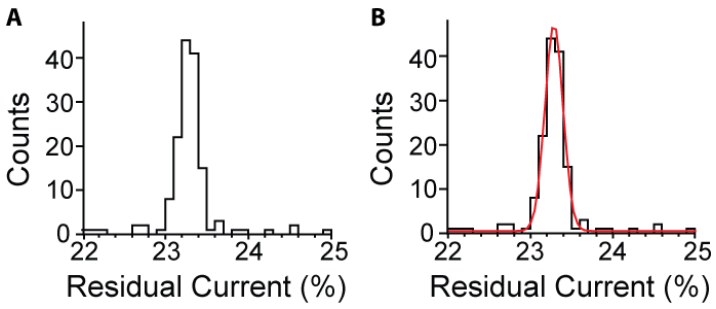


Figure 7 – 8 – *Residual current histograms to determine the average block level of an oligonucleotide sequence. A) Histogram detailing the residual current blocks that were observed for a ssDNA sequence. B) A Gaussian was then fitted to the histogram (illustrated by the red line) to calculate the average block of the oligonucleotide sequence.*

In order to compare several oligonucleotides, a single oligonucleotide species was first added to the *cis* compartment and the current trace required to determine I_B was recorded. A second (and if required, a third, a fourth, a fifth and a sixth) oligonucleotide species was added and additional currents recorded. When these experiments were repeated the oligonucleotides were added to the *cis* compartment in a different order.[59-61, 184] In order to establish if it was possible to discriminate two individual sequences, based solely on single-channel current measurements, the difference between the I_{RES} levels was determined – ΔI_{RES} (Figure 7 – 9). In this case, the signal for poly(dA)$_{40}$, poly(dC)$_{40}$ or poly(dT)$_{40}$ generally acted as the point of reference. For example, if one wished to compare poly(dT)$_{40}$ with poly(dT)-5mC$_{(9)}$ then $\Delta I_{RES}^{(poly(dT)5mC(9) - poly(dT)40)} = I_{RES}^{poly(dT)5mC(9)} - I_{RES}^{poly(dT)40}$. If $\Delta I_{RES}^{(poly(dT)5mC(9) - poly(dT)40)} = 0$ then it was not possible to distinguish the two sequences from each other. However, if $\Delta I_{RES}^{(poly(dT)5mC(9) - poly(dT)40)} \neq 0$ then the oligonucleotides were distinguishable (Figure 7 – 9). The ΔI_{RES} was found to show little experimental variation whereas, the I_B value often showed slight pore-to-pore variation that exceeded ΔI_{RES}.[59]

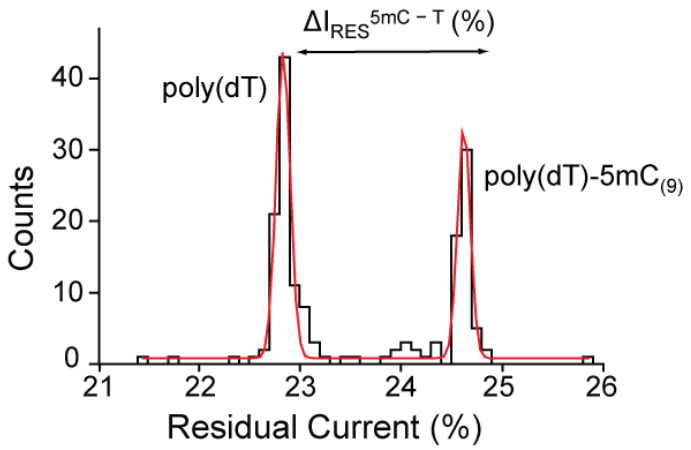


Figure 7 – 9 – *Histogram illustrating how the ΔI_{RES} values were determined. The I_{RES} value for each ssDNA sequence, in this case poly(dT)$_{40}$ and poly(dT)5mC$_{(9)}$, were taken from the Gaussian fits (red line) to the block events, and the ΔI_{RES} value was calculated as the difference between the signals:* $\Delta I_{RES}^{(poly(dT)5mC(9) - poly(dT)40} = I_{RES}^{poly(dT)5mC(9)} - I_{RES}^{poly(dT)40}$ [59, 184].

7.3.5 – Analysis of Chemically Modified DNA-Biotin•Streptavidin Molecules

All experiments were carried out as detailed previously with minor modifications (7.3.1-7.3.4). Samples containing chemically modified DNA had a lower stock-solution concentration (approximately 20-50 μM) than unmodified DNA. In order to observe an equivalent number of blocking events a larger volume of the modified DNA was added to the *cis* compartment. Experiments were performed such that the number of moles of chemically-modified DNA that was added to the *cis* compartment was the same as the number of moles of unmodified DNA.

Chemically modified DNA was also pre-incubated with excess streptavidin (10 μL of 1 mg/mL stock solution in 10 mM sodium phosphate, 150 mM sodium chloride, pH 7.2, Sigma-Aldrich) at room temperature for at least 5 min. This was to maximise the formation of the DNA-biotin•streptavidin complex.

Upon the addition of chemically-modified DNA to the *cis* compartment, the one-second episodic simulation sequence was repeated for at least 200 sweeps (Figure 7 – 6). If one I_B level was observed then this could have corresponded to either 1) a new product having

been produced or 2) no reaction having been detected e.g. the observed block corresponded to unmodified DNA (Figure 7 – 10). To establish if a reaction had occurred unmodified DNA, with an identical sequence, was then added to the *cis* compartment. If subsequently two populations were detected this meant that complete modification of the DNA base had been observed and the block level corresponding to the modified base was identified (Figure 7 – 10A). However, if only one population was observed, when both modified and unmodified DNA samples were present in the *cis* compartment, it was assumed that no reaction had resulted (Figure 7 – 10B, right).

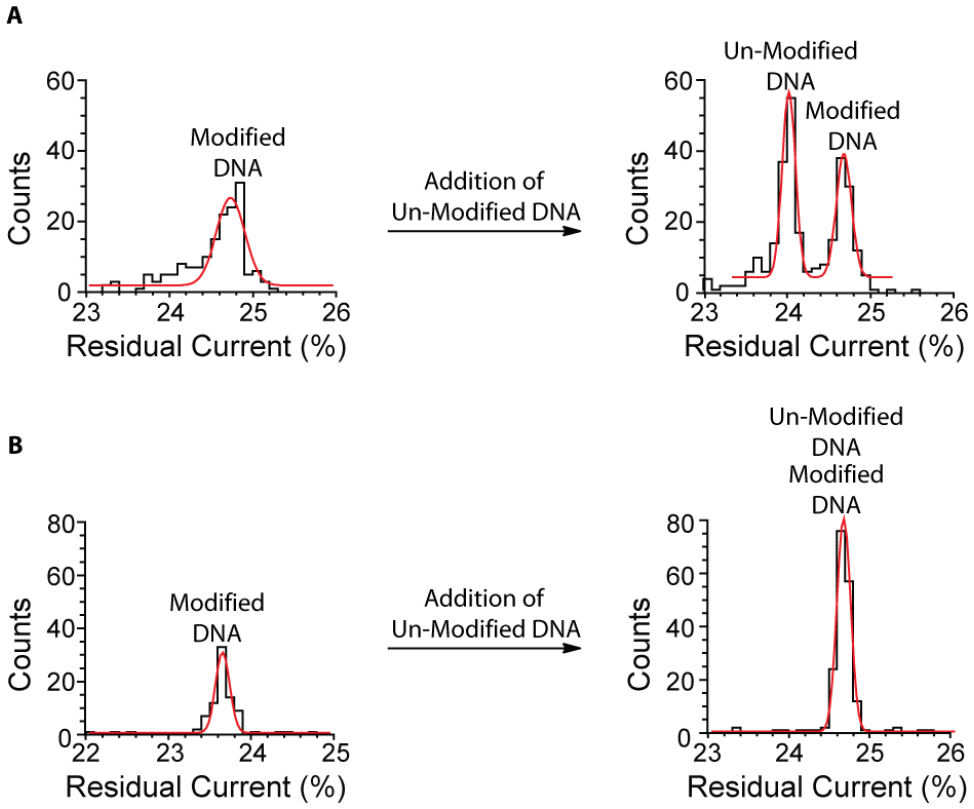


Figure 7 – 10 *– Histograms illustrating how it was possible to determine that modification of a DNA base had been detected by electrical recording experiments. A) These histograms illustrate the scenario where complete modification of a DNA base had occurred. The histogram on the left illustrates when modified DNA only had been added to the cis compartment. Subsequent addition of unmodified DNA lead to the observation of two distinct* I_{RES} *levels, indicating that complete conversion to the chemically-modified base had occurred. B) These histograms illustrate the scenario where no modification of the DNA base had occurred. The histogram on the left illustrates when modified DNA only had been added to the cis compartment. When unmodified DNA was added to the cis compartment only one* I_{RES} *level was detected, indicating no reaction had occurred.*

If two distinguishable I_B levels were noted, when only the chemically modified DNA was present in the *cis* compartment, then two possible scenarios could have explained this observation 1) the DNA oligonucleotide had undergone incomplete modification to form a single product, with unmodified DNA still present in the sample (e.g. a 1:1 mixture modified:unmodified (determined from the event ratio of n = 1 experiments) for the example shown in Figure 7 – 11A, left) or 2) the DNA oligonucleotide had reacted to form two different products (Figure 7 – 11B, left). If scenario 1 was correct then when unmodified DNA, with the same sequence, was added to the *cis* compartment an increase in the count intensity of one of the peaks relative to the other would have been observed (Figure 7 – 11A, right). If scenario 2 was correct then the appearance of a third block level would be expected upon addition of unmodified DNA to the *cis* compartment (Figure 7 − 11B, right).

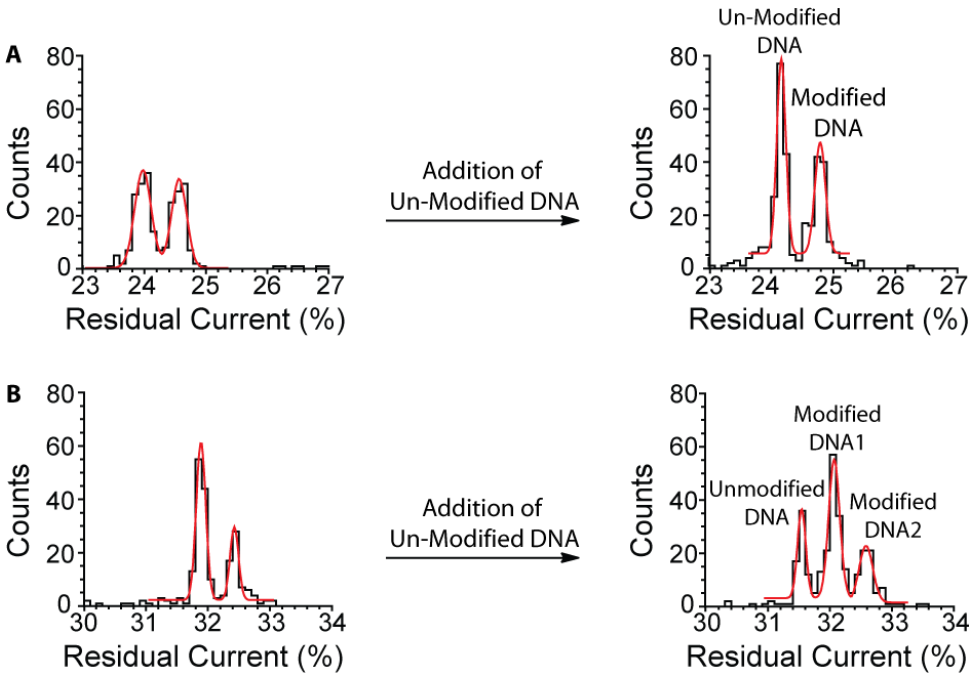


Figure 7 – 11 – *Histograms illustrating how it was possible to determine that modification of a DNA base had been detected by electrical recording experiments. A) These histograms illustrate scenario 1 where incomplete modification of the DNA base occurred producing one single product. The histogram on the left represents when only modified DNA had been added to the cis compartment and partial conversion to the product was observed (1:1 ratio of modified:unmodified, determined from the event ratio of n = 1 experiments). Subsequent addition of unmodified DNA lead to one of the two residual current block levels increasing in count intensity, relative to the other, indicating that this*

level corresponded to unmodified DNA. B) These histograms illustrate scenario 2 where complete modification of the DNA to two different distinguishable products was observed. The histogram on the left represents when only modified DNA had been added to the cis compartment and complete conversion to two different products was observed (2:1 ratio of the modified product species, determined from the event ratio of n = 1 experiments). Subsequent addition of unmodified DNA lead to the appearance of a third residual current block level, indicating that this level corresponded to unmodified DNA and the previously detected levels corresponded to two different product species.

In all cases, when experimental repeats were carried out, the unmodified DNA was added to the *cis* compartment first, followed by the chemically-modified DNA. This was to confirm that the unmodified oligonucleotide did not produce more than one residual block level. The ΔI_{RES} was calculated in the same manner previously outlined (see Chapter 7, section 7.3.4), using the signal for unmodified DNA as the point of reference. Again, the ΔI_{RES} was found to show little experimental variation whereas the I_B value often showed slight pore-to-pore variation that exceeded ΔI_{RES}.

7.4 – Chemical Modification Procedures

7.4.1 – Carell's Oxidation Conditions[174]

Poly(dA)$_{40}$ (20 μL, 100 μM in 25 mM Tris, pH 8.0, 100 μM EDTA) was treated with a 380 μL (100 mM sodium phosphate buffer, pH 5.0) reaction solution containing sodium periodate (1.05 mM) and lithium bromide (4.21 mM). The reaction was sealed and heated at 40 °C for 1 h. The DNA was removed from the chemical reagents by purification using a Econo-Pac® 10 DG disposable chromatography column (minimal dilution protocol) using 25 mM Tris, pH 8.0, 100 μM EDTA as the elution buffer. The concentration of DNA in the column fraction was measured using a NanoDrop® ND-1000 Spectrophotometer. The samples containing DNA were then concentrated under vacuum from approximately 1 mL to 200 μL. The final concentration of the DNA was measured using a NanoDrop® ND-1000 Spectrophotometer. The DNA sample was then stored in the freezer at –20 °C until it was required for experimental use.

The above procedure was repeated for the following DNA sequences poly(dA)$T_{(14)}$, poly(dA)5m$C_{(14)}$, poly(dA)$C_{(14)}$ and poly(dA)$G_{(14)}$.

7.4.2 – Dihydroxylation Conditions

1000 eq. Procedure – Poly$(dA)_{40}$ (20 μL, 100 μM in 25 mM Tris, pH 8.0, 100 μM EDTA) was treated with a 10 μL (1:1 water:*tert*-butanol, pH 6.0) reaction solution containing potassium osmate dihydrate (4.88 mM), 4-methylmorpholine-*N*-oxide (119 mM), potassium citrate (14.9 mM), citric acid (14.9 mM) and EDTA (100 μM). The reaction was sealed and incubated at room temperature for 24 h. The DNA was removed from the chemical reagents by purification using a Econo-Pac® 10 DG disposable chromatography column (minimal dilution protocol) using 25 mM Tris, pH 8.0, 100 μM EDTA as the elution buffer. The concentration of DNA in the column fraction was measured using a NanoDrop® ND-1000 Spectrophotometer. The samples containing DNA were then concentrated under vacuum from approximately 1 mL to 200 μL. The final concentration of the DNA was measured using a NanoDrop® ND-1000 Spectrophotometer. The DNA sample was then stored in the freezer at –20 ˚C until it was required for experimental use.

The above procedure was repeated for the following DNA sequences poly(dA)$T_{(14)}$, poly(dA)5m$C_{(14)}$, poly(dA)$C_{(14)}$ and poly(dA)$G_{(14)}$.

100 eq. Procedure – Poly(dA)$T_{(14)}$ (20 μL, 100 μM in 25 mM Tris, pH 8.0, 100 μM EDTA) was treated with a 10 μL (1:1 water:*tert*-butanol, pH 6.0) reaction solution containing potassium osmate dihydrate (0.48 mM), 4-methylmorpholine-*N*-oxide (11.9 mM), potassium citrate (1.49 mM), citric acid (1.49 mM) and EDTA (100 μM). The reaction was sealed and incubated at room temperature for 24 h. The DNA was removed from the chemical reagents by purification using a Econo-Pac® 10 DG disposable chromatography column (minimal dilution protocol) using 25 mM Tris, pH 8.0, 100 μM

EDTA as the elution buffer. The concentration of DNA in the column fraction was measured using a NanoDrop® ND-1000 Spectrophotometer. The samples containing DNA were then concentrated under vacuum from approximately 1 mL to 200 μL. The final concentration of the DNA was measured using a NanoDrop® ND-1000 Spectrophotometer. The DNA sample was then stored in the freezer at –20 ˚C until it was required for experimental use.

The above procedure was repeated for the following DNA sequences poly(dA)5m$C_{(14)}$ and poly(dA)$C_{(14)}$.

10 eq. Procedure – Poly(dA)$T_{(14)}$ (20 μL, 100 μM in 25 mM Tris, pH 8.0, 100 μM EDTA) was treated with a 10 μL (1:1 water:*tert*-butanol, pH 6.0) reaction solution containing potassium osmate dihydrate (0.048 mM), 4-methylmorpholine-*N*-oxide (1.19 mM), potassium citrate (0.149 mM), citric acid (0.149 mM) and EDTA (100 μM). The reaction was sealed and incubated at room temperature for 24 h. The DNA was removed from the chemical reagents by purification using a Econo-Pac® 10 DG disposable chromatography column (minimal dilution protocol) using 25 mM Tris, pH 8.0, 100 μM EDTA as the elution buffer. The concentration of DNA in the column fraction was measured using a NanoDrop® ND-1000 Spectrophotometer. The samples containing DNA were then concentrated under vacuum from approximately 1 mL to 200 μL. The final concentration of the DNA was measured using a NanoDrop® ND-1000 Spectrophotometer. The DNA sample was then stored in the freezer at –20 ˚C until it was required for experimental use.

The above procedure was repeated for the following DNA sequences poly(dA)5m$C_{(14)}$ and poly(dA)$C_{(14)}$.

7.4.3 – Temozolomide Modification

Procedure 1 – Poly(dT)$_{40}$ (20 μL, 100 μM in 25 mM Tris, pH 8.0, 100 μM EDTA) was treated with a 20 μL (0.1 M potassium phosphate buffer, pH 7.4) reaction solution containing temozolomide (1 mM) and EDTA (100 μM). The reaction was sealed and incubated at 37 ˚C for 18 h. The DNA was removed from the chemical reagents by purification using a Econo-Pac® 10 DG disposable chromatography column (minimal dilution protocol) using 25 mM Tris, pH 8.0, 100 μM EDTA as the elution buffer. The concentration of DNA in the column fraction was measured using a NanoDrop® ND-1000 Spectrophotometer. The samples containing DNA were then concentrated under vacuum from approximately 1 mL to 200 μL. The final concentration of the DNA was measured using a NanoDrop® ND-1000 Spectrophotometer. The DNA sample was then stored in the freezer at –20 ˚C until it was required for experimental use.

The above procedure was repeated for the following DNA sequences poly(dT)A$_{(14)}$, poly(dT)C$_{(14)}$ and poly(dT)-G$_{(14)}$.

Procedure 2 – Poly(dT)$_{40}$* (20 μL, 100 μM in 25 mM Tris, pH 8.0, 100 μM EDTA) was treated with a 20 μL saturated solution of temozolomide (9.7 mg in 800 μL double distilled water). The reaction was incubated at room temperature for 7 days. The DNA sample was then stored in the freezer at –20 ˚C until it was required for experimental use.

The above procedure was repeated for the following DNA sequences poly(dT)A$_{(14)}$*, poly(dT)-C$_{(14)}$* and poly(dT)-G$_{(14)}$*.

Chapter 8

Chemistry Experimental

Chapter 8 – Chemistry Experimental

8.1 – General Techniques

Solvents and Reagents – Dichloromethane was purified prior to use by filtration through two columns of basic activated aluminium oxide (Brockmann I, standard grade, ~150 mesh, 58 Å) as supplied by Sigma Aldrich.[270] All other solvents were used as supplied without further purification. Petrol refers to the fraction of petroleum ether which boils in the range 30-40 °C, unless otherwise stated. Chemical reagents obtained from Acros, Sigma Aldrich, Avocado, Fluka, TCI and Lancaster fine chemical suppliers were used as supplied or were purified according to procedures reported by Perrin and Armarego.[271] Cyclopentadiene was always freshly prepared by pyrolysis of dicyclopentadiene.[272] All reactions were carried out under an atmosphere of argon and all non-aqueous reactions were conducted in flame-dried glassware. Buffer solutions were prepared using water from a Millipore Elix® 10 system, purified further through a 0.22 μm filter by a Millipore Milli-Q system.

Chromatography – Column fractions and reactions were monitored using thin-layer chromatography (TLC), which was run on commercially available Merck Kieselgel 60 F254 0.25 mm pre-coated aluminium backed silica plates. Spots were visualised with UV light and/or by staining with potassium permanganate or vanillin solution. Flash column chromatography was carried out according to the method of Still,[273] using Merck Kieselgel 60 (0.040-0.063 mm) or Merck reverse phase Kieselgel (0.015-0.035 mm) with head pressure supplied by means of bellows.

Nuclear Magnetic Resonance (NMR) Spectroscopy – ^{1}H NMR, ^{13}C NMR and ^{31}P NMR were recorded using either a Bruker AVANCE AV400 (400 MHz) spectrometer, a Bruker AVANCE AVC500 (500 MHz) spectrometer, a Bruker DRX500 (500 MHz) or a Bruker

AVIII 700 (700MHz). Chemical shifts (δ) are quoted in parts per million (ppm), to the nearest 0.01 ppm for ^{1}H and 0.1 ppm for ^{13}C, relative to residual solvent peaks or tetramethylsilane ($SiMe_4$) as an internal standard. Signal splittings are recorded as singlet (s), doublet (d), triplet (t), quartet (q), multiplet (m) and broad (br) respectively. Coupling constants are in the units of Hertz and are uncorrected. Assignments of ^{1}H and ^{13}C were made by COSY (correlated spectroscopy), HMQC (heteronuclear multiple-quantum correlation), HMBC (heteronuclear multiple-bond correlation), HSQC (heteronuclear single-quantum coherence) and DEPT (distortionless enhancement by polarization transfer) where necessary.

Infra-red Spectroscopy – IR spectroscopy was recorded on a Bruker Tensor 27 FT-IR spectrometer. Compounds were analysed as thin films or pressed solids using the Infrared Diamond ATR Module. Absorption maxima are quoted in wavenumbers (cm^{-1}) and relative intensities are quoted as strong (s), medium (m), weak (w) and broad (br).

Mass Spectrometry – Mass spectra (*m/z*) were recorded on a Fisons Platform II, under electrospray ionisation (ESI) conditions. High resolution mass spectrometry data were recorded under ESI conditions on a Bruker Micro Tof (resolution = 5000 FWHM) using a lock-spray source. The lock-mass which is used for calibration was tetraoctyl ammonium bromide in positive ion and sodium dodecyl sulfate in negative ion. The *m/z* values are reported in Daltons, with the percentage abundance in parentheses. The Waters Acquity UPLC coupled to the LCT premier XE mass spectrometer was used to analyse strands of DNA.

Melting Points – Melting points were determined using a Leica VMTG heated-stage microscope and are uncorrected.

8.2 – General Procedures

8.2.1 – Bromination Procedure 1[193]

To a stirred solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in 1,4-dioxane (2.7 mL) and an aqueous solution of sodium acetate buffer (0.5 M, pH 4.7, 4.3 mL) was added a solution of bromine (1.3 mmol) in 1,4-dioxane (5.4 mL) drop-wise. The mixture was stirred for 16 h at room temperature. The reaction was quenched by the addition of an aqueous saturated solution of sodium metabisulfite, which was added until the solution became colourless. The pH was altered to pH 7 by the addition of an aqueous solution of sodium hydroxide (0.50 M). The aqueous layer was then extracted with ethyl acetate (3 × 30 mL) and the combined organic fractions were washed with brine (3 × 30 mL), dried over magnesium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.2 – Bromination Procedure 2[197]

A solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in water:acetonitrile (2:1, 23 mL) was cooled to 0 °C. *N*-bromosuccinimide (**15**, 1.0 mmol) was then added and the resultant solution was stirred for 4 h. The resultant reaction mixture was concentrated *in vacuo* and purified as specified.

8.2.3 – Bromination Procedure 3

To a solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in acetone:water (3:1, 28 mL) was added 2,4,4,6-tetrabromocyclohexa-2,5-dione (**17**, 1.0 mmol) and the reaction stirred at room temperature for 7 h. The reaction was quenched by the addition of solid sodium thiosulfate (100 mg) and the aqueous layer extracted using ethyl acetate (3 × 30 mL). The combined organic fractions were dried over sodium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.4 – Bromination Procedure 4[201]

To a stirred solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in acetone:water (10:1, 15 mL) was added 1,3-bromo-3,5-dimethyl-hydantoin (**18**, 1.0 mmol) portionwise over 5 min. The experiment was placed in the dark and stirred at room temperature for 8 h. The reaction was quenched by the addition of an aqueous saturated solution of ammonium chloride (30 mL) and diluted with water (30 mL). The aqueous layer was then extracted with ethyl acetate (3 × 30 mL) and the combined organic fractions were washed with brine (3 × 30 mL), dried over sodium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.5 – Bromination Procedure 5[202]

To a solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in acetonitrile:water (25:1, 26 mL) was added potassium bromide (2.5 mmol) and Selectfluor® (**22**) (2.0 mmol) and the reaction stirred at room temperature for 24 h. The reaction was then diluted with water (30 mL) and the aqueous layer extracted with ethyl acetate (3 × 30 mL). The combined organic fractions were then washed with an aqueous saturated solution of sodium bicarbonate (3 × 30 mL), an aqueous saturated solution of sodium sulfite (3 × 30 mL) and brine (3 × 30 mL), dried over sodium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.6 – Bromination Procedure 6[203]

To a solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in tetrahydrofuran:water (1:1, 6.0 mL) was added tetrabutylammonium tribromide (1.0 mmol) portionwise over 5 min. The reaction was stirred at room temperature for 16 h before being quenched with water (30 mL) and the aqueous layer extracted with ethyl

acetate (3 × 30 mL). The combined organic fractions were washed with brine (3 × 30 mL), dried over sodium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.7 – Bromination Procedure 7[204]

A solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in *N*,*N*-dimethylformamide:water (7:3, 26 mL) was cooled to 0 °C. Tetrabutylammonium tribromide (1.0 mmol) in *N*,*N*-dimethylformamide (3.7 mL) was added to the reaction mixture and the reaction stirred at room temperature for 2 h. The mixture was then quenched with an aqueous solution of sodium bisulfite (5% wt, 30 mL) and the aqueous layer extracted with ethyl acetate (3 × 30 mL). The combined organic fractions were dried over magnesium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.8 – Dihydroxylation Procedure 1[192]

To a solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in tetrahydrofuran:*tert*-butanol:water (5:5:2, 6.0 mL) was added 4-methylmorpholine-*N*-oxide (2.0 mmol) and osmium tetroxide (0.060 mmol). The reaction was stirred at 45 ˚C for 36 h before an additional portion of osmium tetroxide (0.040 mmol) was added. After a further 12 h the reaction was cooled to 0 ˚C, quenched by the addition of solid sodium sulfite (1.0 mmol) and stirred for 30 min. The mixture was diluted with water (30 mL) and the aqueous layer extracted with dichloromethane (3 × 30 mL). The combined organic fractions were then dried over magnesium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.9 – Dihydroxylation Procedure 2[237]

To a solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in *tert*-butanol:water (1:1, 19 mL) was added potassium ferricyanide(III) (3.0 mmol), potassium carbonate (3.0 mmol), $(DHQD)_2PHAL$ (0.25 mmol), methanesulfonamide

(1.0 mmol) potassium osmate dihydrate (0.050 mmol) and the reaction stirred at room temperature for 5 d. The reaction was then quenched by the addition of solid sodium sulfite (1.0 mmol) and stirred for 30 min. The mixture was then diluted with water (30 mL) and the aqueous layer extracted with ethyl acetate (3 × 30 mL). The combined organic fractions were then washed with an aqueous solution of potassium hydroxide (2.0 M, 3 × 30 mL) and brine (3 × 30 mL), dried over magnesium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.10 – Dihydroxylation Procedure 3[239]

To a solution of *bis*-TBS-protected deoxynucleoside (**2**, **4**, **6**, **8** or **10**, 1.0 mmol) in *tert*-butanol:water (1:1, 9.5 mL) was added citric acid (0.75 mmol), 4-methylmorpholine-*N*-oxide (1.2 mmol), potassium osmate dihydrate (0.050 mmol) and the reaction stirred at room temperature for 36 h. The reaction was then quenched by the addition of solid sodium sulfite (1.0 mmol) and stirred for 30 min. The mixture was then diluted with water (30 mL) and the aqueous layer extracted with ethyl acetate (3 × 30 mL). The combined organic fractions were then washed with brine (3 × 30 mL), dried over magnesium sulfate, filtered, concentrated *in vacuo* and purified as specified.

8.2.11 – Formation of Osmate Esters Procedure 1[241]

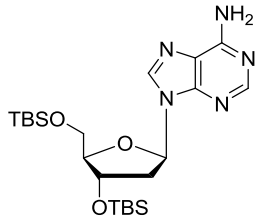
To a stirred solution of diols **44** and **45** (1.0 mmol) in acetone (100 mL) was added a diamine source (1.1 mmol) and a solution of potassium osmate dihydrate (1.1 mmol) in water (25 mL). The pH was altered to pH 7 by the addition of an aqueous solution of hydrochloric acid (1.0 M). The resulting brown solution was stirred at room temperature for 24 h, concentrated *in vacuo* and purified as specified.

8.2.12 – Formation of Osmate Esters Procedure 2

To a solution of osmium tetroxide (280 mg, 1.10 mmol) in acetone (16.4 mL) was added quinuclidine (122 mg, 1.10 mmol) and the reaction stirred for 5 min – the solution was observed to turn bright orange (**53**). Cyclohexene (82.0 mg, 1.00 mmol) was then added and the reaction stirred for a further 5 min – the reaction was observed to turn green. This produced a 1.1 mmol solution of the osmium source **54**. To a solution of the appropriate substrate (**2**, **4**, **8**, **10** or **44** and **45**), 1.0 mmol) in acetone:water (10:3, 110 mL) was added a solution of the osmium source **54** (16 mL in acetone) dropwise. The reaction mixture was stirred at room temperature for 5 min before *N,N,N′,N′*-tetramethylethylenediamine (1.1 mmol) was added – the reaction solution was observed to turn dark brown. The reaction was stirred at room temperature for 24 h, concentrated *in vacuo* and purified as specified.

8.3 – Experimental Details

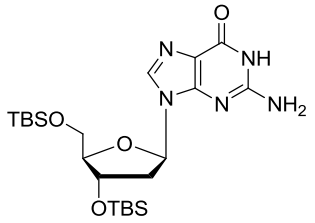
9-((2*R*,4*S*,5*R*)-4-((*tert*-Butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-9*H*-purin-6-amine, 2



To a solution of 2'-deoxyadenosine (**1**) (50 mg, 0.19 mmol) in anhydrous *N*,*N*-dimethylformamide (110 μL) was added *tert*-butyldimethylsilyl chloride (110 mg, 0.74 mmol) and imidazole (89 mg, 1.3 mmol). The mixture was stirred at room temperature for 24 h before being diluted with ethyl acetate (15 mL). The resultant mixture was washed with an aqueous saturated solution of ammonium chloride (3 × 15 mL) and brine (3 × 15 mL), dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude

product was purified by flash column chromatography [SiO_2, 1:99→3:97 methanol:dichloromethane] which afforded bis-TBS-protected **2** (79.0 mg, 89%) as a white solid. **m.p.** (132-133 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3321 br, 3165 br, 2954 s, 2930 s, 2886 m, 2858 m, 1648 s, 1599 s, 1472 m, 1254 s, 1211 s, 837 s; **^{1}H NMR δ_H** (500 MHz, $CDCl_3$) 8.35 (1 H, s, OCHNC*H*), 8.15 (1 H, s, C*H*NCNH$_2$), 6.44 (1 H, t, *J* = 6.5, OC*H*N), 5.80 (2 H, br s, N*H*$_2$), 4.62 (1 H, dt, *J* = 6.7, 2.9, C*H*OTBS), 4.02 (1 H, q, *J* = 3.5, C*H*CH$_2$OTBS) 3.88 (1 H, dd, *J* = 11.0, 4.0, CH_AH$_B$OTBS), 3.78 (1 H, dd, *J* = 11.0, 3.0, CH$_A$*H*$_B$OTBS), 2.64 (1 H, dt, *J* = 13.0, 6.5, NCHCH_AH$_B$), 2.44 (1 H, ddd, *J* = 13.0, 6.0, 4.0, NCHCH$_A$*H*$_B$), 0.92 (9 H, s, CH$_2$OSi(CH$_3$)$_2$C(C*H*$_3$)$_3$), 0.91 (9 H, s, CHOSi(CH$_3$)$_2$C(C*H*$_3$)$_3$), 0.11 (6 H, s, CH$_2$OSi(C*H*$_3$)$_2$), 0.04 (6 H, CHOSi(C*H*$_3$)$_2$); **^{13}C NMR δ_C** (125 MHz, $CDCl_3$) 155.3 (*C*NH$_2$), 152.9 (*C*HNCNH$_2$), 149.6 (NH$_2$C=*C*), 139.0 (OCHN*C*H), 119.9 (N*C*=C), 87.9 (*C*HCH$_2$OTBS), 84.3 (O*C*HN), 71.8 (*C*HOTBS), 62.7 (*C*H$_2$OTBS), 41.3 (NCH*C*H$_2$), 25.9, 25.7 (2 × SiC(*C*H$_3$)$_3$), 18.4, 18.0 (2 × Si*C*(CH$_3$)$_3$), −4.7, −4.8, −5.4, −5.5 (2 × Si(*C*H$_3$)$_2$); **MS** ***m/z*** (ESI$^+$) 480 (100%, [M-H]$^+$); **HRMS** (ESI$^+$) $C_{22}H_{41}N_5O_3Si_2$ requires *M*+*H* 480.2821, found 480.2822 (0.3 ppm); **$[\alpha]_D^{22}$** –6.1 (*c* 1.0, dichloromethane).

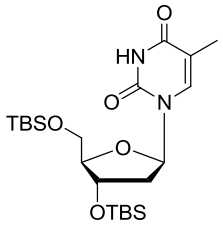
2-Amino-9-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-1*H*-purin-6(9*H*)-one, 4



To a solution of 2'-deoxyguanosine (**3**) (300 mg, 1.12 mmol) in anhydrous *N*,*N*-dimethylformamide (636 μL) was added *tert*-butyldimethylsilyl chloride (677 mg, 4.49 mmol) and imidazole (535 mg, 7.86 mmol). The reaction was stirred at room

temperature for 24 h before being diluted with ethyl acetate (30 mL). The resultant mixture was washed with an aqueous saturated solution of ammonium chloride (3 × 30 mL) and brine (3 × 30 mL) and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 2:98→9:91 methanol:dichloromethane] which afforded bis-TBS protected **4** (491 mg, 88%) as a white solid. **m.p.** (> 325 ºC); **IR** $\boldsymbol{\nu_{max}}$ (thin film)/cm^{-1} 3317 br, 3168 br, 2954 m, 2927 s, 2895 m, 2856 s, 1691 s, 1603 s, 1472 m, 1382 m, 1169 s, 835 s; **^{1}H NMR** $\boldsymbol{\delta_H}$ (500 MHz, $(CD_3)_2SO$) 10.60 (1 H, br s, N*H*C=O), 7.88 (1 H, s, NC*H*=N), 6.46 (2 H, br s, NH_2), 6.10 (1 H, dd, *J* = 7.5, 6.5, OC*H*N), 4.48 (1 H, dt, *J* = 5.5, 2.8, C*H*OTBS), 3.82-3.79 (1 H, m, C*H*CH_2OTBS) 3.69 (1 H, dd, *J* = 11.0, 5.8, C$H_A$$H_B$OTBS), 3.64 (1 H, dd, *J* = 11.0, 4.3, $CH_A$$H_B$OTBS), 2.63 (1 H, ddd, *J* = 13.3, 7.8, 5.5, NCHC$H_A$$H_B$), 2.23 (1 H, ddd, *J* = 13.3, 5.9, 3.4, NCHC$H_A$$H_B$), 0.89 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.87 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.10 (6 H, s, $CH_2OSi(CH_3)_2$), 0.04 (6 H, s, $CHOSi(CH_3)_2$); **^{13}C NMR** $\boldsymbol{\delta_C}$ (125 MHz, $(CD_3)_2SO$) 156.7 (*C*=O), 153.7 (N=*C*NH_2), 151.0 (N*C*=CC=O), 134.9 (N*C*H=N), 116.6 (C=*C*C=O), 87.0 (*C*HCH_2OTBS), 82.1 (O*C*HN), 72.2 (*C*HOTBS), 62.7 (CH*C*H_2OTBS), obscured by (CD_3)$_2$SO (NCH*C*H_2), 25.8, 25.7 (2 × $SiC(CH_3)_3$), 18.0, 17.7 (2 × $SiC(CH_3)_3$), −4.8, −5.0, −5.5, −5.5 (2 × $Si(CH_3)_2$); **MS** ***m/z*** (ESI$^-$) 494 (100%, $[M-H]^-$); **HRMS** (ESI$^+$) $C_{22}H_{41}N_5O_4Si_2$ requires *M*+*Na* 518.2589, found 518.2587 (0.3 ppm).

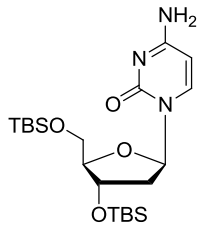
1-((2*R*,4*S*,5*R*)-4-((*tert*-Butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1*H*,3*H*)-dione, 6



To a solution of 2'-deoxythymidine (**5**) (500 mg, 2.06 mmol) in anhydrous *N*,*N*-dimethylformamide (1.89 mL) was added *tert*-butyldimethylsilyl chloride (685 mg, 4.54 mmol) and imidazole (562 mg, 8.25 mmol). The reaction stirred at 50 °C for 6 h, before being diluted with ethyl acetate (30 mL). The resultant mixture washed with an aqueous saturated solution of ammonium chloride (3 × 30 mL) and brine (3 × 30 mL), dried over sodium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 10:90→35:65 ethyl acetate:petrol] which afforded bis-TBS protected **6** (909 mg, 94%) as a white solid. **m.p.** (142-145 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3178 br, 3049 br, 2954 s, 2929 s, 2857 m, 1698 s, 1471 m, 1255 m, 836 s, 778 m; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 9.45 (1 H, s, N*H*), 7.40 (1 H, d, *J* = 1.2, NC*H*), 6.28 (1 H, dd, *J* = 7.6, 6.0, OC*H*N), 4.33 (1 H, dt, *J* =5.7, 2.8, C*H*OTBS), 3.86 (1 H, dd, *J* = 2.4, C*H*CH_2OTBS), 3.80 (1 H, dd, *J* = 11.5, 2.6, C$H_A$$H_B$OTBS), 3.69 (1 H, dd, *J* = 11.5, 2.4, CH_AH_BOTBS), 2.18 (1 H, ddd, *J* = 13.3, 5.8, 2.6, NCHC$H_A$$H_B$), 1.93 (1 H, ddd, *J* = 13.3, 7.6, 6.0, NCHCH_AH_B), 1.83 (3 H, d, *J* = 0.8, CH=CCH_3), 0.85 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.81 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.04 (6 H, d, *J* = 1.2, $CH_2OSi(CH_3)_2$), 0.03 (6 H, d, *J* = 2.8, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 164.1 ($C(CH_3)C$=O), 150.5 (N*C*=O), 135.4 (N*C*H), 110.8 (CH=*C*CH_3), 87.8 (*C*HCH_2OTBS), 84.8 (O*C*HN), 72.2 (*C*HOTBS), 62.9 (*C*H_2OTBS), 41.4 (NCH*C*H_2), 25.9, 25.7 (2 × $SiC(CH_3)_3$), 18.4, 18.0 (2 × $SiC(CH_3)_3$), 12.6 (CH=C*C*H_3), −4.7, −4.9, −5.4, −5.5

(2 × Si(CH_3)$_2$); **MS** ***m/z*** (ESI$^-$) 469 (95%, [M-H]$^-$); **HRMS** (ESI$^+$) $C_{22}H_{42}N_2O_5Si_2$ requires *M+Na* 493.2524, found 493.2526 (–0.4 ppm); $[\alpha]_D^{22}$ +18.0 (*c* 1.0, dichloromethane).

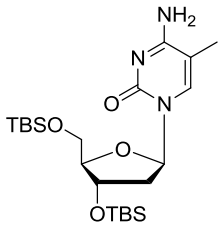
4-Amino-1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)pyrimidin-2(1*H*)-one, 8



To a solution of 2'-deoxycytidine (**7**) (300 mg, 1.31 mmol) in anhydrous *N*,*N*-dimethylformamide (1.20 mL) was added *tert*-butyldimethylsilyl chloride (434 mg, 2.88 mmol) and imidazole (357 mg, 5.24 mmol). The reaction stirred at 50 ºC for 8 h, before being diluted with ethyl acetate (30 mL). The resultant mixture washed with an aqueous saturated solution of ammonium chloride (3 × 30 mL) and brine (3 × 30 mL), dried over sodium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 1:99→5:95 methanol:dichloromethane] which afforded bis-TBS-protected **8** (599 mg, 95%) as a white solid. **m.p.** (94-96 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3330 br, 3135 br, 2930 s, 2858 m, 1627 s, 1488 s, 1407 m, 1255 s, 1115 s, 837 s; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 7.82 (1 H, d, *J* = 7.2, NC*H*=CH), 6.21 (1 H, t, *J* = 5.8, OC*H*N), 5.74 (1 H, d, *J* = 7.2, NCH=C*H*), 4.31 (1 H, q, *J* = 5.2, C*H*OTBS), 3.84-3.81 (2 H, m, C*H*CH_AH$_B$OTBS), 3.70 (1 H, dd, *J* = 12.2, 3.4, CH$_A$$H_B$OTBS), 2.35-2.28 (1 H, m, NCHCH_AH$_B$), 2.03-1.95 (1 H, m, NCHCH$_A$$H_B$), 0.86 (9 H, t, *J* = 3.0, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.81 (9 H, t, *J* = 3.0, $CHOSi(CH_3)_2C(CH_3)_3$), 0.04 (6 H, s, $CH_2OSi(CH_3)_2$), 0.00 (6 H, t, *J* = 3.2, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 166.0 (N=*C*NH$_2$), 156.0 (*C*=O), 140.6 (N*C*H=CH), 94.6 (NCH=*C*H), 87.1 (*C*HCH$_2$OTBS), 85.6 (O*C*HN), 70.4 (*C*HOTBS), 62.0 (*C*H$_2$OTBS), 42.0 (NCH*C*H$_2$), 25.9,

25.7 (2 x SiC(CH_3)$_3$), 18.3, 17.9 (2 × SiC(CH$_3$)$_3$), −4.6, −4.9, −5.5, −5.6 (2 × Si(CH_3)$_2$); **MS** ***m/z*** (ESI$^-$) 454 (100%, [M-H]$^-$); **HRMS** (ESI$^+$) $C_{21}H_{41}N_3O_4Si_2$ requires *M*+*Na* 478.2528, found 478.2532 (–0.9 ppm); **[α]$_D^{22}$** +80.5 (*c* 1.0, dichloromethane).

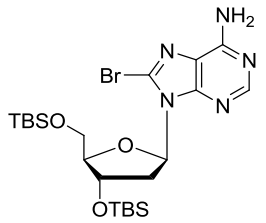
4-Amino-1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5-methylpyrimidin-2(1*H*)-one, 10



To a solution of 2'-deoxy-5-methylcytidine (**9**) (300 mg, 1.24 mmol) in anhydrous *N*,*N*-dimethylformamide (1.14 mL) was added *tert*-butyldimethylsilyl chloride (412 mg, 2.74 mmol) and imidazole (339 mg, 4.97 mmol). The reaction was stirred at room temperature for 4 d, before being diluted with ethyl acetate. The resultant mixture was washed with an aqueous saturated solution of ammonium chloride (3 × 30 mL) and brine (3 × 30 mL), dried over sodium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 2:98→7:93 methanol:dichloromethane] which afforded bis-TBS-protected **10** (524 mg, 90%) as a white solid. **m.p.** (109-111 ºC); **IR** ***ν$_{max}$*** (thin film)/cm^{-1} 3339 br, 3193 br, 2953 s, 2929 s, 2858 m, 1669 s, 1611 m, 1513 s, 1294 m, 1111 m, 836 s; **^{1}H NMR δ$_H$** (400 MHz, $CDCl_3$) 7.52 (1 H, s, NCH), 6.30 (1 H, t, *J* = 6.6, OCHN), 4.34 (1 H, dt, *J* = 6.3, 3.2, CHOTBS), 3.91 (1 H, q, *J* = 2.8, CHCH$_2$OTBS), 3.85 (1 H, dd, *J* = 11.2, 3.0, CH_AH$_B$OTBS), 3.74 (1 H, dd, *J* = 11.2, 2.4, CH$_A$$H_B$OTBS), 2.35 (1 H, ddd, *J* = 13.2, 6.0, 3.6, NCHCH_AH$_B$), 1.97-1.92 (1 H, m, NCHCH$_A$$H_B$), 1.89 (3 H, s, CH=CC$H_3$), 0.90 (9 H, s, CH$_2$OSi(CH$_3$)$_2$C(C$H_3$)$_3$), 0.86 (9 H, s, CHOSi(CH$_3$)$_2$C(CH_3)$_3$), 0.08 (6 H, d, *J* = 2.8,

$CH_2OSi(CH_3)_2$), 0.04 (6 H, d, J = 1.6, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 165.6 ($N=CNH_2$), 155.8 (C=O), 137.8 (NCH), 101.5 ($CH=CCH_3$), 87.6 ($CHCH_2OTBS$), 85.7 (OCHN), 71.8 ($CHOTBS$), 62.7 (CH_2OTBS), 42.1 ($NCHCH_2$), 25.9, 25.8, 25.7 ($2 \times SiC(CH_3)_3$), 18.4, 18.0 ($2 \times SiC(CH_3)_3$), 13.3 ($CH=CCH_3$), −4.6, −4.9, −5.4, −5.5 ($2 \times Si(CH_3)_2$); **MS** ***m/z*** (ESI$^-$) 468 (100%, $[M-H]^-$); **HRMS** (ESI$^+$) $C_{22}H_{43}N_3O_4Si_2$ requires *M+Na* 492.2684, found 492.2687 (–0.6 ppm); **$[\alpha]_D^{22}$** + 49.7 (*c* 1.0, dichloromethane).

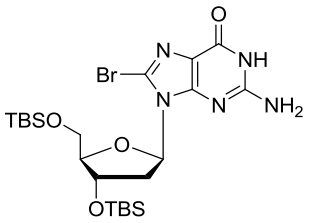
8-Bromo-9-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-9*H*-purin-6-amine, 11



Bis-TBS protected **2** (52 mg, 0.11 mmol) was subjected to standard procedure 8.2.4. The crude product was purified by flash column chromatography [SiO_2, 1:99→3:97 methanol:dichloromethane] which afforded *mono-brominated* **11** (28 mg, 46%) as a brown solid. **m.p.** (134-137 ºC); **IR ν_{max}** (thin film)/cm^{-1} 3322 br, 3178 br, 2954 s, 2929 s, 2885 m, 2857 m, 1647 s, 1600 s, 1461 m, 1321 m, 1293 m, 1255 s, 1109 s, 1080 s, 836 s, 777s; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 8.25 (1 H, s, $CHNCNH_2$), 6.35 (1 H, t, J = 6.8, OC*H*N), 5.76 (2 H, br s, NH_2), 4.87 (1H, dt, J = 6.4, 3.8, C*H*OTBS), 3.98-3.89 (2 H, m, $CHCH_AH_BOTBS$), 3.70-3.64 (2 H, m, CH_AH_BOTBS + $NCHCH_AH_B$), 2.24 (1 H, ddd, J = 13.2, 7.0, 4.2, $NCHCH_AH_B$), 0.94 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.83 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.15 (6 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.00 (3 H, $1 \times CHOSi(CH_3)_2$), −0.04 (3 H, $1 \times CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 154.2 (CNH_2), 152.5 ($CHNCNH_2$), 150.9 ($NH_2C=C$), 128.2 (NCBr), 120.5 ($NC=CCNH_2$),

87.7 (*C*HCH$_2$OTBS), 86.3 (O*C*HN), 72.3 (*C*HOTBS), 62.5 (*C*H$_2$OTBS), 36.7 (NCH*C*H$_2$), 25.8 (2 × SiC(*C*H$_3$)$_3$), 18.3, 18.0 (2 × Si*C*(CH$_3$)$_3$), −4.7, −4.8, −5.4, −5.5 (2 × Si(*C*H$_3$)$_2$); **MS *m/z*** (ESI$^+$) 580, 582 (94%, 100% [M-Na]$^+$); **HRMS** (ESI$^+$) $C_{22}H_{40}BrO_3N_5Si_2$ requires *M+Na* 580.1745, found 580.1730 (2.6 ppm); **[α]$_D^{22}$** –14.5 (*c* 1.0, dichloromethane).

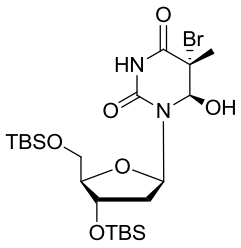
2-Amino-8-bromo-9-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-1*H*-purin-6(9*H*)-one, 12



Bis-TBS-protected **4** (54 mg, 0.11 mmol) was subjected to standard procedure 8.2.1. The crude product was purified by flash column chromatography [SiO$_2$, 1:99→10:90 methanol:dichloromethane], which afforded *mono-brominated* **12** (13 mg, 21%) as a white solid. **m.p.** (140-142 ºC); **IR *ν_{max}*** (thin film)/cm^{-1} 3305 br, 3138 br, 2954 m, 2930 s, 2858 m, 1694 s, 1603 s, 1471 m, 1286 m, 1111 s, 1084 m, 837 s, 778 m; **^{1}H NMR δ_H** (500 MHz, CD$_3$OD) 6.28 (1 H, t, *J* = 7.0, OC*H*N), 4.75 (1 H, dt, *J* = 6.5, 3.4, C*H*OTBS), 3.68-3.84 (2 H, m, C*H$_A$*H$_B$OTBS), 3.77-3.79 (1 H, m, CH$_A$*H$_B$*OTBS), 3.55 (1 H, dt, *J* = 13.3, 6.6, NCHC*H$_A$*H$_B$), 2.20 (1 H, ddd, *J* = 13.4, 7.1, 4.1, NCHCH$_A$*H$_B$*), 0.95 (9 H, s, CH$_2$OSi(CH$_3$)$_2$C(C*H$_3$*)$_3$), 0.86 (9 H, s, CHOSi(CH$_3$)$_2$C(C*H$_3$*)$_3$), 0.17 (6 H, s, CH$_2$OSi(C*H$_3$*)$_2$), 0.02 (6 H, d, *J* = 11.5, CHOSi(C*H$_3$*)$_2$); **^{13}C NMR δ_C** (125 MHz, CD$_3$OD) 158.3 (*C*=O), 154.8 (N=*C*NH$_2$), 154.1 (N*C*=CC=O), 124.0 (N*C*Br), 118.9 (NC=*CC*=*O*), 89.1 (*C*HCH$_2$OTBS), 87.1 (O*C*HN), 73.8 (*C*HOTBS), 64.1 (*C*H$_2$OTBS), 37.7 (NCH*C*H$_2$), 26.3, 26.4 (2 × SiC(*C*H$_3$)$_3$), 19.2, 18.9 (2 × Si*C*(CH$_3$)$_3$), −4.3, −4.4, −5.2, −5.3 (2 × Si(*C*H$_3$)$_2$); **MS *m/z*** (ESI$^-$) 596, 598 (94%, 100% [M-Na]$^+$); **HRMS** (ESI$^+$)

$C_{22}H_{40}BrN_5O_4Si_2$ requires *M+Na* 596.1694, found 596.1715 (−2.9 ppm); $[\alpha]_D^{22}$ +18.7 (*c* 1.0, methanol).

(5*R*,6*R*)-5-Bromo-1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-6-hydroxy-5-methyldihydropyrimidine-2,4(1*H*,3*H*)-dione, 13

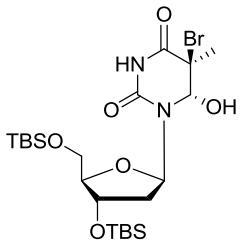


13

Bis-TBS-protected **6** (51 mg, 0.11 mmol) was subjected to standard procedure 8.2.4. The crude product was purified by flash column chromatography [SiO_2, 5:95→20:80 acetone:petrol], which afforded the *bromohydrin* **13** (32 mg, 53%*)* as a white solid. Diastereomer **16** (20 mg, 32%*)* was also isolated as a white solid. **m.p.** (195-196 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3214 br, 3095 br, 2954 s, 2929 s, 2858 m, 1699 s, 1471 m, 1281 m, 1114 m, 836 s, 779 m; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 7.84 (1 H, br s, N*H*), 6.51 (1 H, t, *J* = 6.8, OC*H*N), 5.13 (1 H, d, *J* = 2.6, NC*H*OH), 4.43 (1 H, dt, *J* = 6.8, 4.8, C*H*OTBS), 4.14 (1 H, d, *J* = 2.6, O*H*), 3.86 (1 H, dd, *J* = 11.2, 2.4, CH_AH_BOTBS), 3.80 (1 H, dt, *J* = 4.4, 2.1, C*H*CH_2OTBS), 3.75 (1 H, dd, *J* = 11.2, 1.6, CH_AH_BOTBS), 2.31-2.24 (1 H, m, NCHCH_AH_B), 2.22-2.15 (1 H, m, NCHCH_AH_B), 1.99 (3 H, s, CBrCH_3), 0.93 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.91 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.12 (6 H, d, *J* = 1.2, $CH_2OSi(CH_3)_2$), 0.10 (6 H, d, *J* = 2.8, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 166.6 ($CBrCH_3C$=O), 150.6 (N*C*=O), 85.9 (*C*$CHCH_2$OTBS), 84.8 (O*C*HN), 81.0 (N*C*HOH), 70.8 (*C*HOTBS), 61.8 (*C*H_2OTBS), 53.1 (*C*$BrCH_3$), 38.4 (NCH*C*H_2), 26.0, 25.7 (2 × $SiC(CH_3)_3$), 22.8 ($CBrCH_3$), 18.5, 17.9 (2 × $SiC(CH_3)_3$), −4.5, −4.8, −5.4, −5.5

(2 × Si(*C*H$_3$)$_2$); **MS** ***m/z*** (ESI$^+$) 589, 591 (100%, 100%, [M-Na]$^+$); **HRMS** (ESI$^+$) $C_{22}H_{43}BrN_2O_6Si_2$ requires *M*+*Na* 589.1735, found 589.1735 (0.1 ppm); **[α]$_D^{22}$** +46.4 (*c* 1.0, dichloromethane).

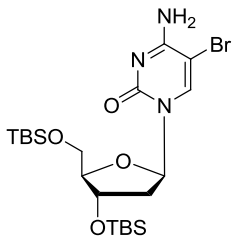
(5*S*,6*S*)-5-Bromo-1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-6-hydroxy-5-methyldihydropyrimidine-2,4(1*H*,3*H*)-dione, 16



16

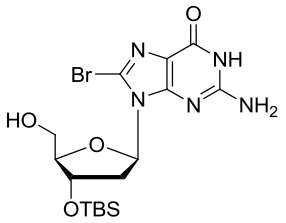
Bromohydrin **16** was isolated as the minor diastereomer in the formation of bromohydrin **13**. **m.p.** (151-153 ºC); **IR** ***v**$_{max}$* (thin film)/cm^{-1} 3222 br, 2955 s, 2930 s, 2858 m, 1703 s, 1464 m, 1253 m, 1116 m, 1031 m, 836 s, 778 m; **^{1}H NMR δ$_H$** (400 MHz, $CDCl_3$) 7.69 (1 H, s, N*H*), 5.84 (1 H, t, *J* = 6.0, OC*H*N), 5.47 (1 H, br s, NC*H*OH), 4.54 (1 H, d, *J* = 2.6, O*H*), 4.40 (1 H, dt, *J* = 6.1, 4.1, C*H*OTBS), 3.98-3.94 (2 H, m, C*H*C*H$_A$*H$_B$OTBS), 3.78 (1 H, dd, *J* = 12.0, 3.2, CH$_A$*H$_B$*OTBS), 2.49-2.43 (1 H, m, NCHC*H$_A$*H$_B$), 2.36 (1 H, ddd, *J* = 13.2, 6.2, 4.6, NCHCH$_A$*H$_B$*), 1.97 (3 H, s, CBrC*H$_3$*), 0.93 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.89 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.13 (6 H, d, *J* = 1.2, $CH_2OSi(CH_3)_2$), 0.08 (6 H, d, *J* = 2.8, $CHOSi(CH_3)_2$); **^{13}C NMR δ$_C$** (100 MHz, $CDCl_3$) 167.1 ($CBrCH_3C$=O), 149.4 (N*C*=O), 87.4 (*C*HCH$_2$OTBS), 87.1 (O*C*HN), 80.9 (N*C*HOH), 71.3 (*C*HOTBS), 63.3 (*C*H$_2$OTBS), 52.9 (*C*BrCH$_3$), 41.2 (NCH*C*H$_2$), 26.0, 25.7 (2 × SiC(*C*H$_3$)$_3$), 22.8 (CBr*C*H$_3$), 18.7, 18.0 (2 × Si*C*(CH$_3$)$_3$), −4.5, −4.8, −5.2, −5.6 (2 × Si(*C*H$_3$)$_2$); **MS** ***m/z*** (ESI$^+$) 589, 591 (100%, 100%, [M-Na]$^+$); **HRMS** (ESI$^+$) $C_{22}H_{43}BrN_2O_6Si_2$ requires *M*+*Na* 589.1735, found 589.1749 (–2.3 ppm); **[α]$_D^{22}$** −29.5 (*c* 1.0, dichloromethane).

4-Amino-5-bromo-1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)pyrimidin-2-2(*1*H)-one, 14



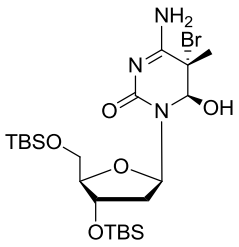
Bis-TBS-protected **8** (50 mg, 0.11 mmol) was subjected to standard procedure 8.2.5. The crude product was purified by flash column chromatography [SiO_2, 1:99→7:93 methanol:dichloromethane], which afforded *mono-brominated* **14** (33mg, 57%) as a white solid. **m.p.** (decomp. 216 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3058 br, 2954 s, 2292 s, 2857 m, 1655 s, 1495 s, 1472 m, 1255 s, 1108 s, 834 s, 777 s; **^{1}H NMR δ_H** (500 MHz, $CDCl_3$) 8.08 (1 H, s, NC*H*=CBr), 7.43 (1 H, br s, NH_AH_B), 6.25 (1 H, t, *J* = 6.3, OC*H*N), 5.63 (1 H, br s, NH_AH_B), 4.36 (1 H, dt, *J* = 6.3, 3.1, C*H*OTBS), 3.98 (1 H, q, *J* = 2.5, C*H*CH_2OTBS), 3.92 (1 H, dd, *J* = 11.5, 2.5, C$H_A$$H_B$OTBS), 3.76 (1 H, dd, *J* = 8.5, 3.0, CH_AH_BOTBS), 2.47 (1 H, ddd, *J* = 13.1, 6.0, 3.8, NCHC$H_A$$H_B$), 2.03-1.98 (1 H, m, NCHCH_AH_B), 0.93 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.89 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.13 (6 H, d, *J* = 5.5, $CH_2OSi(CH_3)_2$), 0.07 (6 H, d, *J* = 3.5, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (125 MHz, $CDCl_3$) 162.1 (N=*C*NH_2), 154.4 (*C*=O), 141.4 (N*C*H=CBr), 88.1 (*C*HCH_2OTBS), 86.9 (NCH=*C*Br), 86.7 (O*C*HN), 71.7 (*C*HOTBS), 62.6 (*C*H_2OTBS), 42.6 (NCH*C*H_2), 26.0, 25.7, 25.6 (2 × SiC(CH_3)$_3$), 18.4, 18.0 (2 × Si*C*$(CH_3)_3$), −4.6, −4.9, −5.4 (2 × Si(CH_3)$_2$); **MS *m/z*** (ESI^+) 556, 558 (95%, 100%, $[M\text{-}Na]^+$); **HRMS** (ESI^+) $C_{21}H_{40}BrN_3O_4Si_2$ requires *M+Na* 556.1633, found 556.1636 (–0.5 ppm); **$[\alpha]_D^{22}$** +28.8 (*c* 1.0, dichloromethane).

2-Amino-8-bromo-9-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(hydroxymethyl)tetrahydrofuran-2-yl)-1*H*-purin-6(9*H*)-one, 19



Bis-TBS-protected **4** (54 mg, 0.11 mmol) was subjected to standard procedure 8.2.4. The crude product was purified by flash column chromatography [SiO_2, 1:99→10:90 methanol:dichloromethane], which afforded *mono-brominated* **19** (19 mg, 37%) as a white solid. **m.p.** (decomp. 300 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3306 br, 3138 br, 2954 m, 2930 s, 2858 m, 1705 s, 1598 s, 1523 m, 1472 m, 1287 m, 1103 s, 1026 m, 834 s, 778 m; **^{1}H NMR** δ_H (500 MHz, CD_3OD) 6.32 (1 H, t, *J* = 7.5, OC*H*N), 4.68 (1 H, t, *J* = 2.5, C*H*OTBS), 4.02-4.00 (1 H, m, C*H*CH_2OH) 3.84 (1 H, dd, *J* = 12.0, 4.0, CH_AH_BOH), 3.73 (1 H, dd, *J* = 12.0, 4.0, CH_AH_BOH), 3.21-3.16 (1 H, m, NCHCH_AH_B), 2.21-2.17 (1 H, m, NCHCH_AH_B), 0.96 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.16 (6 H, s, $CH_2OSi(CH_3)_2$); **^{13}C NMR** δ_C (125 MHz, CD_3OD) 157.9 (*C*=O), 154.8 (N=*C*NH_2), 153.2 (N*C*=CC=O), 123.2 (N*C*Br), 119.0 (NC=*CC*=*O*), 90.0 (*C*HCH_2OH), 88.1 (OCHN), 74.5 (*C*HOTBS), 63.4 (*C*H_2OH), 39.3 (NCH*C*H_2), 26.0 ($SiC(CH_3)_3$), 18.7 ($SiC(CH_3)_3$), −4.8 ($Si(CH_3)_2$); **MS** ***m/z*** (ESI^+) 482, 484 (95%, 100%, $[M\text{-}Na]^+$); **HRMS** (ESI^+) $C_{16}H_{26}BrN_5O_4Si$ requires *M*+*Na* 482.0830, found 482.0833 (−0.8 ppm); **$[\alpha]_D^{22}$** −8.0 (*c* 1.0, methanol).

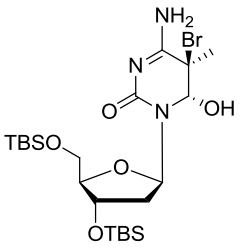
(5*S*,6*R*)-4-Amino-5-bromo-1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-6-hydroxy-5-methyl-5,6-dihydropyrimidin-2(1*H*)-one, 20



20

Bis-TBS-protected **10** (51 mg, 0.11 mmol) was subjected to standard procedure 8.2.4. The crude product was purified by flash column chromatography [SiO_2, 2:98→20:80 acetone:petrol], which afforded *bromohydrin* **20** (35 mg, 56%) as a white solid. Diastereomer **21** (28 mg, 44%*)* was also isolated as a white solid. **m.p.** (106-109 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3207 br, 2954 s, 2929 s, 2857 m, 1718 s, 1461 m, 1254 s, 1091 m, 836 s, 779 m; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 7.80 (2 H, br s, NH_2), 6.50 (1 H, t, J = 6.5, OCHN), 5.03 (1 H, s, NCHOH), 4.42 (1 H, dt, J = 6.8, 4.8, CHOTBS), 3.85 (1 H, dt, J = 11.2, 2.4, CH_AH$_B$OTBS), 3.80 (1 H, dt, J = 4.3, 2.1, CHCH$_2$OTBS), 3.29 (1 H, dd, J = 11.2, 1.6, CH$_A$$H_B$OTBS), 2.27 (1 H, ap dt, J = 13.6, 6.8, NCHCH_AH$_B$), 2.18 (1 H, ddd, J = 13.3, 7.5, 5.6, NCHCH$_A$$H_B$), 2.11 (3 H, s, CBrC$H_3$), 0.93 (9 H, s, CH$_2$OSi(CH$_3$)$_2$C(C$H_3$)$_3$), 0.91 (9 H, s, CHOSi(CH$_3$)$_2$C(CH_3)$_3$), 0.12 (6 H, d, J = 4.4, CH$_2$OSi(CH_3)$_2$), 0.10 (6 H, d, J = 2.4, CHOSi(CH_3)$_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 159.2 (N=CNH$_2$), 149.2 (C=O), 85.8 (CHCH$_2$OTBS), 84.2 (OCHN), 81.6 (NCHOH), 70.9 (CHOTBS), 61.8 (CH$_2$OTBS), 53.5 (CBrCH$_3$), 38.4 (NCHCH$_2$), 26.0, 25.7 (2 × SiC(CH$_3$)$_3$), 24.2 (CBrCH_3), 18.5, 17.9 (2 × SiC(CH$_3$)$_3$), −4.5, −4.8, −5.3, −5.5 (2 × Si(CH$_3$)$_2$); **$[\alpha]_D^{22}$** +22.1 (*c* 1.0, methanol).

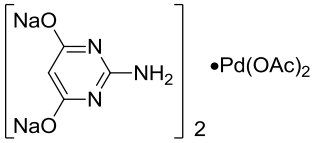
(5R,6S)-4-Amino-5-bromo-1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-6-hydroxy-5-methyl-5,6-dihydropyrimidin-2(1H)-one, 21



21

Bromohydrin **21** was isolated as the minor diastereomer in the formation of bromohydrin **20**. **m.p.** (56-59 ºC); **IR** ν_{max} (thin film)/cm^{-1} 3210 br, 2955 s, 2929 s, 2857 m, 1704 s, 1462 s, 1362 m, 1253 s, 1094 m, 1030 m, 836 s, 778 m; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 7.66 (2 H, br s, NH_2), 5.82 (1 H, t, J = 5.8, OCHN), 5.40 (1 H, s, NCHOH), 4.41-4.38 (1 H, m, CHOTBS), 3.96-3.94 (2 H, m, CHCH_AH$_B$OTBS), 3.78 (1 H, dd, J = 11.0, 1.8, CH$_A$$H_B$OTBS), 2.46 (1 H, ap dt, J = 13.5, 5.9, NCHCH_AH$_B$), 2.36 (1 H, ddd, J = 13.5, 6.3, 4.9, NCHCH$_A$$H_B$), 2.09 (3 H, s, CBrC$H_3$), 0.93 (9 H, s, CH$_2$OSi(CH$_3$)$_2$C(C$H_3$)$_3$), 0.89 (9 H, s, CHOSi(CH$_3$)$_2$C(CH_3)$_3$), 0.13 (6 H, d, J = 6.8, CH$_2$OSi(CH_3)$_2$), 0.08 (6 H, d, J = 5.6, CHOSi(CH_3)$_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 159.5 (N=*C*NH$_2$), 148.2 (*C*=O), 87.3 (*C*HCH$_2$OTBS), 86.9 (O*C*HN), 81.3 (N*C*HOH), 70.8 (*C*HOTBS), 63.1 (*C*H$_2$OTBS), 53.6 (*C*BrCH$_3$), 41.5 (NCH*C*H$_2$), 26.1, 25.7 (2 × SiC(*C*H$_3$)$_3$), 24.1 (CBr*C*H$_3$), 18.7, 18.0 (2 × Si*C*(CH$_3$)$_3$), −4.5, −4.9, −5.2, −5.5 (2 × Si(*C*H$_3$)$_2$); **[α]$_D^{22}$** –9.5 (*c* 1.0, methanol).

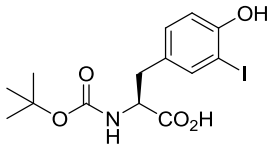
Palladium(II) catalyst 23[212]



A stirred suspension of 2-amino-4,6-dihydroxypyrimidine (13 mg, 0.10 mmol) in an aqueous sodium hydroxide solution (0.10 M, 2.0 mL) was dissolved by heating at 65 ˚C

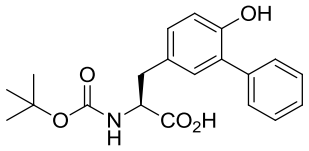
for 5 min. Palladium(II) acetate (11 mg, 0.050 mmol) was then added and the reaction mixture stirred at 65 ˚C for 30 minutes. The reaction mixture was cooled to room temperature and then diluted with distilled water (3.0 mL) which afforded a solution of palladium(II) catalyst **23** (0.01 M).

(*S*)-2-((*tert*-Butoxycarbonyl)amino)-3-(4-hydroxy-3-iodophenyl)propanoic acid, 24[212]



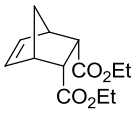
To a stirred solution of 3-iodo-L-tyrosine (400 mg, 1.30 mmol) in water (5.00 mL) at 0 °C was added an aqueous sodium hydroxide solution (5.0 M,). Di-*tert*-butyl dicarbonate (299 mg, 1.37 mmol) in dioxane (5.00 mL) was added dropwise to the reaction mixture at 0 °C. The reaction was then warmed to room temperature over a period of 3 h. The reaction was then quenched with the addition of an aqueous hydrochloric acid solution (1.0 M, 100 mL) and extracted with ethyl acetate (3 × 50 mL). The combined organics fractions were washed with brine (3 × 30 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 1:99→15:85 methanol:dichloromethane] to afford *N-Boc protected* **24** as a white solid (452 mg, 85%). **m.p.** (80-82 ºC); **^{1}H NMR δ_H** (400 MHz, CD_3OD) 7.55 (1 H, br s, CI=C*H*), 7.06 (1 H, d, *J* = 8.0, COH=CHC*H*), 6.76 (1 H, d, *J* = 8.0, COH=C*H*), 4.26-4.23 (1 H, m, NHC*H*), 3.06 (1 H, dd, *J* = 13.0, 4.4, CHCH_AH$_B$), 2.82-2.76 (1 H, m, CHCH$_A$$H_B$), 1.41 (9 H, s, C(C$H_3$)$_3$); **^{13}C NMR δ_C** (100 MHz, CD_3OD) 174.7 (*C*O_2H), 156.7 (*C*OH), 155.8 (NH*C*=O), 140.1 (CI=*C*H), 130.7 (CH_2*C*), 130.5 (COH=CH*C*H), 114.6 (COH=*C*H), 83.5 (*C*I=CH), 79.5 (*C*(CH_3)$_3$), 55.8 (NH*C*H), 36.5 (CH*C*H_2), 27.8 (C(*C*H_3)$_3$); **$[\alpha]_D^{20}$** = +18.8 (c = 1.0, methanol). Spectroscopic data was in good agreement with that of the literature.[212]

(*S*)-2-((*tert*-Butoxycarbonyl)amino)-3-(6-hydroxy-[1,1'-biphenyl]-3-yl)propanoic acid, 26[212]



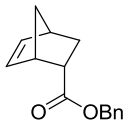
To a solution of boc-3-iodo-L-tyrosine (**24**) (56 mg, 0.14 mmol) in water (3.2 mL) was added phenyl boronic acid (**25**) (25 mg, 0.21 mmol) and dipotassium hydrogen phosphate (0.12 g, 0.69 mmol) and the reaction was stirred for 2 min at 37 ˚C. Formic acid (0.26 μL in 10 μL of water) and palladium(II) catalyst **23** solution (0.01 M, 140 μL) were then added and the resulting mixture was stirred at 37 ˚C for 4 h. The reaction was quenched by the addition of an aqueous solution of hydrochloric acid (1.0 M, 50 mL) before the aqueous layer was extracted with ethyl acetate (3 × 50 mL). The combined organic layers were dried over magnesium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 1:99→8:92 methanol:dichloromethane] which afforded (*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(6-hydroxy-[1,1'-biphenyl]-3-yl)propanoic acid **26** as (34 mg, 70%) as a glassy foam. **^{1}H NMR δ_H** (400 MHz, $(CD_3)_2SO$) 9.41 (1 H, br s, N*H*), 7.53 (2 H, d, *J* = 7.6, 2 × Ar*H*), 7.38 (2 H, t, *J* = 7.4, 2 × Ar*H*), 7.28 (1 H, t, *J* = 7.4, Ar*H*), 7.13 (1 H, s, COHC(Ph)=C*H*), 7.00 (1 H, t, *J* = 10.6, C*H*CH=COH), 6.83 (1 H, d, *J* = 8.4, CHC*H*=COH), 4.05 (1 H, td, *J* = 9.8, 4.4, C*H*CO_2H), 2.95 (1 H, dd, *J* = 13.6, 4.4, NHCHCH_AH_B), 2.76 (1 H, dd, *J* = 13.6, 9.8, NHCHCH_AH_B), 1.31 (9 H, s, 3 x Me); **^{13}C NMR δ_C** (100 MHz, $(CD_3)_2SO$) 174.6 (*C*O_2H), 156.3 (NH*C*=O), 153.6 (*C*OH), 139.5 (CH_2*C*), 132.1 (*C*HC=*C*H), 129.9, 129.5, 128.7 128.0 (Ar*C*), 127.3 (*C*COH), 116.7 (*C*H=COH), 78.9 (*C*$(CH_3)_3$), 56.2 (*C*HCO_2H), 36.6 (CH*C*H_2), 29.0 (C(*C*$H_3)_3$) **$[\alpha]^{20}_D$** = +24.3 (*c* = 1.0, methanol). Spectroscopic data was in good agreement with that of the literature.[212]

(1*R*,3*S*,4*S*)-Diethyl bicyclo[2.2.1]hept-5-ene-2,3-dicarboxylate, 35[220]



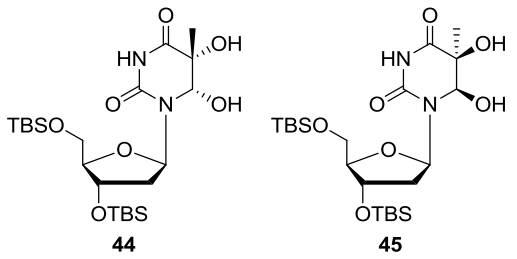
To a solution of diethyl maleate (**33**) (1.86 mL, 14.9 mmol) in water (100 mL) was added freshly cracked cyclopentadiene (**32**) (1.00 mL, 14.9 mmol). The reaction was stirred vigorously for 24 h before the aqueous phase was extracted with ethyl acetate (3 × 50 mL). The combined organic fractions were washed with brine (3 × 50 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo*, which afforded (1*R*,3*S*,4*S*)-diethyl bicycle[2.2.1]hept-5-ene-2,3-dicarboxylate **35** (2.57 g, 73%) as a colourless oil. **IR** ν_{max} (thin film)/cm^{-1} 2982 br, 1724 s, 1178 s, 1032m; **^{1}H NMR** δ_H (400 MHz, $CDCl_3$) 6.12 (1 H, dd, *J* = 5.5, 3.0, HC=C*H*), 5.91 (1 H, dd, *J* = 5.5, 2.8, *H*C=CH), 4.01 (2 H, q, *J* = 7.2, 1 × $CO_2CH_2CH_3$), 3.94 (2 H, qd, *J* = 7.1, 3.1, 1 × $CO_2CH_2CH_3$), 3.22 (1 H, t, *J* = 4.4, 1 × C*H*CO_2Et), 3.10 (1 H, br s, 1 × C*H*CHCO$_2$Et), 2.96 (1 H, br s, 1 × C*H*CHCO$_2$Et), 2.52 (1 H, dd, *J* = 4.4, 1.0, 1 × C*H*CO_2Et), 1.47 (1 H, d, *J* = 8.4, CHCH_AH_BCH), 1.29 (1 H, d, *J* = 8.4, CHCH_AH_BCH), 1.12 (3 H, t, *J* = 7.1, 1 × $CO_2CH_2CH_3$), 1.08 (3 H, *J* = 7.2, 1 × $CO_2CH_2CH_3$); **^{13}C NMR** δ_C (100 MHz, $CDCl_3$) 173.9 (*C*O_2Et), 172.7 (*C*O_2Et), 137.1 (*C*H=CH), 134.7 (CH=*C*H), 60.3 ($CO_2CH_2CH_3$), 60.0 ($CO_2CH_2CH_3$), 47.5 (*C*HCO$_2$Et), 47.4 (*C*HCO$_2$Et), 46.8 (2 × *C*HCHCO$_2$Et), 45.3 (CH*C*H$_2$CH), 13.9 ($CO_2CH_2CH_3$), 13.8 ($CO_2CH_2CH_3$); **MS** ***m/z*** (ESI$^+$) 261 (100%, [M-Na]$^+$); **HRMS** (ESI$^+$) $C_{13}H_{18}O_4$ requires *M*+*Na* 621.1097, found 621.1097 (0.0 ppm).

((1*S*,2*S*,4*S*)-Benzyl bicyclo[2.2.1]hept-5-ene-2-carboxylate, 36[220]



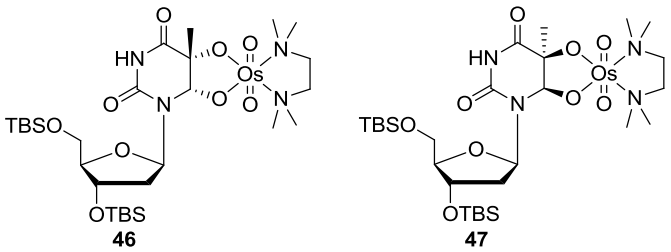
To a solution of benzyl acrylate (**34**) (2.28 mL, 14.9 mmol) in water (100 mL) was added freshly cracked cyclopentadiene (**32**) (1.00 mL, 14.9 mmol). The reaction was stirred vigorously for 24 h before the aqueous phase was extracted with ethyl acetate (3 × 50 mL). The combined organic fractions were washed with brine (3 × 50 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 1:99→3:97 diethyl ether:petrol], which afforded ((1*S*,2*S*,4*S*)-benzyl bicyclo[2.2.1]hept-5-ene-2-carboxylate **36** (2.09 g, 66%) as a colourless oil. **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 7.41-7.33 (5 H, m, 5 x ArH), 6.22 (1 H, dd, *J* = 5.6, 3.0, 1 × C*H*=CH), 5.92 (1 H, dd, *J* = 5.6, 2.8, 1 × CH=C*H*), 5.14 (1 H, d, *J* = 12.4, C$H_A$$H_B$Ph), 5.08 (1 H, d, *J* = 12.4, C$H_A$$H_B$Ph), 3.27 (1 H, br s, C*H*CHC=O), 3.04 (1 H, dt, *J* = 9.3, 4.0, CHC*H*C=O), 2.94 (1 H, br s, C*H*CH_2CHC=O), 1.95 (1 H, ddd, *J* = 11.6, 9.4, 3.8, C$H_A$$H_B$CHC=O), 1.54-1.44 (2 H, m, C$H_A$$H_B$CHC=O + CHC$H_A$$H_B$CH), 1.30 (1 H, d, *J* = 8.4, CHC$H_A$$H_B$CH); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 174.3 (*C*=O), 137.6 (*C*H=CH), 136.2 (CH_2*C*), 132.2 (CH=*C*H), 128.3, 127.9 (5 x Ar C's), 65.8 (*C*H_2Ph), 49.5 (CH*C*H_2CH), 45.6 (*C*HCHC=O), 43.2 (CH_2*C*HC=O), 42.4 (*C*HCH_2CHC=O), 29.1 (*C*H_2CHC=O). Spectroscopic data was in good agreement with that of the literature.[274]

1-((2*R*,4*S*,5*R*)-4-((*tert*-Butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5,6-dihydroxy-5-methyldihydropyrimidine-2,4(1*H*,3*H*)-dione 44 and 45



Bis-TBS-protected **6** (50 mg, 0.11 mmol) was subjected to standard procedure 8.2.10. The crude product was purified by flash column chromatography [SiO_2, 5:95→20:80 ethyl acetate:dichloromethane] which afforded an diols **44** and **45** (49 mg, 89%, d.r. 90:10) as a white solid and as an inseparable diastereomeric mixture. **m.p.** (182-185 ºC); **IR** $\boldsymbol{\nu_{max}}$ (thin film)/cm^{-1} 3389 br, 3102 br, 2954 s, 2929 s, 2858 m, 1717 s, 1462 m, 1227 m, 1120 m, 837 s, 778 m; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 8.08 (1 H, br s, N*H*), 6.27 (1 H, dd, *J* = 7.6, 6.0, OC*H*N), 5.28 (1 H, s, NC*H*OH), 4.45 (1 H, dt, *J* = 6.2, 3.1, C*H*OTBS), 3.91-3.75 (3 H, m, C*H*CH_2OTBS), 2.33 (1 H, ddd, *J* = 13.5, 7.6, 6.2, NCHC$H_A$$H_B$), 2.16 (1 H, dq, *J* = 13.4, 3.1, NCHC$H_A$$H_B$), 1.49 (3 H, s, COHC$H_3$), 0.98 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.94 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.15 (6 H, s, $CH_2OSi(CH_3)_2$), 0.12 (6 H, s, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 173.5 ($C(CH_3)C$=O), 150.2 (N*C*=O), 86.7 (*C*HCH_2OTBS), 84.2 (O*C*HN), 77.6 (N*C*HOH), 71.9 (*C*HOTBS), 64.8 (*C*OCH_3), 63.2 (CH*C*H_2OTBS), 39.2 (NCH*C*H_2), 26.0, 25.9, 25.8, 25.7 (2 × $SiC(CH_3)_3$), 22.7 (COH*C*H_3), 18.4, 18.0 (2 × $SiC(CH_3)_3$), −4.6, −4.9, −5.5, −5.6 (2 × $Si(CH_3)_2$); **MS *m/z*** (ESI^-) 503 (100%, $[M-H]^-$); **HRMS** (ESI^+) $C_{22}H_{44}N_2O_4Si_2$ requires *M*+*Na* 527.2579, found 527.2556 (–4.5 ppm). The ^{1}H and ^{13}C data correspond to the major diastereomer **44**, whose spectroscopic data was found to be in good agreement with the literature.[192]

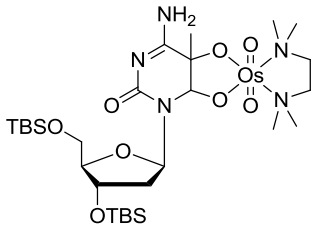
Bis-TBS protected 2'-deoxythymidine TMEDA Osmate Esters, 46 and 47



A mixture of diols **44** and **45** (23 mg, 0.046 mmol) was subjected to standard procedure 8.2.11 (24 h) using the diamine *N,N,N′,N′*-tetramethylethylenediamine. The crude product was purified by flash column chromatography [SiO_2, 2:98→4:96 methanol:dichloromethane], which afforded *bis-TBS-protected 2-deoxythymidine TMEDA osmate esters **46** and **47*** (36 mg, 96%, d.r. 94:6) as a brown solid and as an inseparable diastereomeric mixture. **IR** $\boldsymbol{\nu_{max}}$ (thin film)/cm^{-1} 3507 br, 3208 br, 2954 s, 2929 s, 2897 m, 2858 m, 1699 s, 1462 s, 1284 m, 1026 m, 878 s, 778 m; **^{1}H NMR δ_H** (400 MHz, $CDCl_3$) 7.21 (1 H, br s, N*H*), 6.35 (1 H, dd, *J* = 8.0, 6.0, OC*H*N), 5.22 (1 H, s, NC*H*OOs), 4.30 (1 H, dt, *J* = 6.0, 3.0, C*H*OTBS), 3.80 (1 H, ddd, *J* = 6.8, 4.0, 2.4, C*H*CH_2OTBS), 3.69 (1 H, dd, *J* = 10.8, 4.8, C$H_A$$H_B$OTBS), 3.56 (1 H, dd, *J* = 10.8, 6.4, $CH_A$$H_B$OTBS) 3.13-3.08 (4 H, m, 2 × NCH_2), 2.97 (3 H, s, 1 × NCH_3), 2.92 (3 H, s, 1 × NCH_3), 2.87 (3 H, s, 1 × NCH_3), 2.83 (3 H, s, 1 × NCH_3), 2.44-2.35 (1 H, m, NCHC$H_A$$H_B$), 2.11 (1 H, ddd, *J* = 13.4, 6.2, 3.2, NCHC$H_A$$H_B$), 1.68 (3 H, s, C=OCC$H_3$), 0.89 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.88 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.08 (6 H, s, $CH_2OSi(CH_3)_2$), 0.06 (6 H, s, $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (100 MHz, $CDCl_3$) 173.1 (*C*=OCCH$_3$), 151.1 (N*C*=O), 96.2 (N*C*OOs), 86.0 (*C*HC$H_A$$H_B$OTBS), 84.4 (C=O*C*CH$_3$), 83.9 (OCHN), 72.4 (*C*HOTBS), 64.8, 64.5 (2 × N*C*H_2), 63.7 (*C*H_2OTBS), 52.3, 52.1, 51.9, 51.8 (4 × N*C*H_3), 38.1 (NCH*C*H_2), 26.0, 25.8, (2 × SiC(*C*H_3)$_3$), 21.7 (C=OC*C*H$_3$), 18.4, 18.0 (2 × Si*C*$(CH_3)_3$), −4.6, −4.7, −5.3, −5.4 (2 × Si(*C*H_3)$_2$) **MS *m/z*** (ESI^-) 901 (100%, $[M\text{-}MeCN\text{-}NH_4]^+$); **HRMS** ($ESI^+$) $C_{28}H_{58}O_9N_4OsSi_2$ requires *M+Na* 865.3248, found

865.3267 (–2.0 ppm). The ^{1}H and ^{13}C data correspond to the major diastereomer **46**, whose stereochemistry was inferred from that of diol **44**.

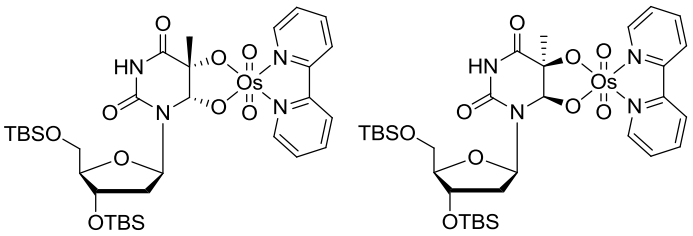
***Bis*-TBS protected 2'deoxymethylcytidine TMEDA Osmate Ester, 48**



To a solution of *bis*-TBS protected deoxynucleoside **10** (25 mg, 0.053 mmol) in dichloromethane (5.3 mL) was added *N,N,N′,N′*-tetramethylethylenediamine (9.00 μL, 0.060 mmol) and the reaction was cooled to −78 ºC. A solution of osmium tetroxide (15 mg, 0.06 mmol) in dichloromethane (1.0 mL) was added dropwise and the reaction was stirred at −78 ºC for 4 h. The reaction mixture was then allowed to warm to room temperature over 16 h before the solvent was removed *in vacuo*. The crude product was purified by flash column chromatography [SiO_2, 2:98→10:90 methanol:dichloromethane], which afforded the *bis-TBS-protected 2'-deoxymethylcytidine TMEDA osmate ester* **48** (42 mg, 95%) as a brown solid and as a single diastereomer. **IR ν_{max}** (thin film)/cm^{-1} 3440 br, 2954 s, 2930 s, 2898 m, 2858 m, 1706 s, 1614 s, 1462 s, 1257 s, 1124 m, 1027 m, 839 s, 778 m; **^{1}H NMR δ_H** (500 MHz, $CDCl_3$) 6.35 (1 H, t, *J* = 7.0, OC*H*N), 5.09 (1 H, s, NC*H*OOs), 4.27 (1 H, t, *J* = 2.8, C*H*OTBS), 3.77-3.74 (1 H, m, C*H*CH_2OTBS), 3.65 (1 H, dd, *J* = 10.5, 4.5, C$H_A$$H_B$OTBS), 3.52 (1 H, dd, *J* = 10.8, 6.3, CHC$H_A$$H_B$OTBS), 3.25-3.07 (4 H, m, 2 x NCH_2), 2.97 (3 H, s, 1 × NCH_3), 2.91 (3 H, s, 1 × NCH_3), 2.88 (3 H, s, 1 × NCH_3), 2.84 (3 H, s, 1 × NCH_3), 2.34-2.30 (1 H, m, NCHC$H_A$$H_B$), 2.05-2.02 (1 H, m, NCHC$H_A$$H_B$), 1.59 (3 H, s, C($NH_2$)CC$H_3$), 0.88 (9 H, s, CH_2OSi(CH_3)$_2$C(CH_3)$_3$), 0.87 (9 H, s, CHOSi(CH_3)$_2$C(CH_3)$_3$), 0.05 (6 H, s, CH_2OSi(CH_3)$_2$), 0.03 (6 H, s, CHOSi(CH_3)$_2$); **^{13}C NMR δ_C** (125 MHz, $CDCl_3$) 174.3 (N=*C*NH_2), 155.9 (N*C*=O), 96.1

(N*C*OOs), 85.9 (C(NH$_2$)*C*CH$_3$) 83.8 (*C*HCH$_2$OTBS), 80.2 (O*C*HN), 72.3 (*C*HOTBS), 64.9, 64.8 (2 × N*C*H$_2$), 63.5 (CH*C*H$_2$OTBS), 53.0, 52.8, 52.1, 51.6 (4 × N*C*H$_3$), 37.9 (NCH*C*H$_2$), 26.0, 25.8 (2 × SiC(*C*H$_3$)$_3$), 23.6 (C(NH$_2$)C*C*H$_3$), 18.4, 17.9 (2 × Si*C*(CH$_3$)$_3$), −4.6, −4.8, −5.3, −5.4 (2 × Si(*C*H$_3$)$_2$) **MS *m/z*** (ESI$^+$) 864 (80%, [M-Na]$^+$); **HRMS** (ESI$^+$) $C_{28}H_{59}N_5O_8OsSi_2$ requires *M*+*H* 842.3588, found 842.3575 (1.8 ppm).

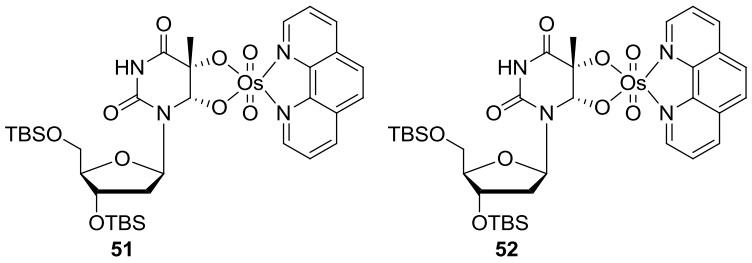
***Bis*-TBS-protected 2'-Deoxythymidine BIPY Osmate Esters, 49 and 50**



A mixture of diols **44** and **45** (23 mg, 0.046 mmol) were subjected to standard procedure 8.2.11 (24 h) using the diamine 2,2'-bipyridine. The crude product was purified by flash column chromatography [SiO$_2$, 1:99→9:91 methanol:dichloromethane], which afforded the *bis-TBS-protected 2-deoxythymidine BIPY osmate esters* **49** *and* **50** as (36 mg, 96%, d.r. 95:5) as a brown solid and as an inseparable diastereomeric mixture. **IR ν_{max}** (thin film)/cm^{-1} 3491 br, 3210 br, 2954 s, 2929 s, 2897 m, 2858 m, 1696 s, 1500 s, 1254 m, 1023 m, 874 s, 775 m; **^{1}H NMR δ_H** (500 MHz, CDCl$_3$) 9.46 (1 H, dd, *J* = 5.5, 1.0, 1 × OsN=C*H*), 9.21 (1 H, dd, *J* = 5.3, 0.8, 1 × OsN=C*H*), 8.36 (2 H, t, *J* = 8.3, 2 × NCC*H*), 8.26-8.21 (2 H, m, 2 × NCCHC*H*), 7.76-7.72 (2 H, m, OsN=CHC*H*), 7.34 (1 H, br s, N*H*), 6.44 (1 H, dd, *J* = 8.0, 6.0, OC*H*N), 5.46 (1 H, s, NC*H*OOs), 4.35 (1 H, dt, *J* = 6.7, 2.9, C*H*OTBS), 3.83 (1 H, ddd, *J* = 6.0, 5.0, 3.0, C*H*CH$_2$OTBS), 3.68 (1 H, dd, *J* = 10.5, 5.0, C*H$_A$*H$_B$OTBS), 3.54 (1 H, dd, *J* = 10.5, 6.5, CH$_A$*H$_B$*OTBS), 2.58 (1 H, ddd, *J* = 13.5, 8.0, 6.0, NCHC*H$_A$*H$_B$), 2.21 (1 H, ddd, *J* = 13.5, 6.0, 3.0, NCHCH$_A$*H$_B$*), 1.81 (3 H, s, C=OCC*H$_3$*), 0.89 (9 H, s, CH$_2$OSi(CH$_3$)$_2$C(C*H$_3$*)$_3$), 0.83 (9 H, s, CHOSi(CH$_3$)$_2$C(C*H$_3$*)$_3$), 0.08 (3 H, s, 1 × CH$_2$OSi(C*H$_3$*)$_2$), 0.06 (3 H, s, 1 × CHOSi(C*H$_3$*)$_2$), 0.00 (3 H, s,

1 × CHOSi(CH_3)$_2$), −0.03 (3 H, s, 1 × CHOSi(CH_3)$_2$); **^{13}C NMR δ_C** (125 MHz, $CDCl_3$) 173.1 (*C*=OCCH$_3$), 156.0, 155.8 (2 × OsN=*C*H), 152.5, 152.2 (2 × OsN*C*) 150.9 (N*C*=O), 141.1, 141.0 (NCCH*C*H), 127.6, 127.4 (2 × OsN=CH*C*H), 123.6, 123.5 (2 × NC*C*H), 95.9 (N*C*HOOs), 86.1 (*C*HCH$_2$OTBS), 84.0 (C=O*C*CH$_3$), 83.6 (O*C*HN), 72.6 (*C*HOTBS), 63.7 (CH*C*H$_2$OTBS), 38.2 (NCH*C*H$_2$), 25.9, 25.8, 25.7 (2 × SiC(*C*H$_3$)$_3$), 22.1 (C=OC*C*H$_3$), 18.3, 18.0 (2 × Si*C*(CH$_3$)$_3$), −4.6, −4.7, (2 × Si(*C*H$_3$)$_2$); **MS *m/z*** (ESI$^+$) 941 (100%, [M-MeCN-NH$_4$]$^+$); **HRMS** (ESI$^+$) $C_{32}H_{50}N_4O_9OsSi_2$ requires *M+Na* 905.2622, found 905.2610 (1.5 ppm). The ^{1}H and ^{13}C data correspond to the major diastereomer **49**, whose stereochemistry was inferred from that of diol **44**.

***Bis*-TBS-protected 2'-Deoxythymidine PHEN Osmate Esters, 51 and 52**



A mixture of diols **44** and **45** (23 mg, 0.046 mmol) were subjected to standard procedure 8.2.11 (24 h) using the diamine 1,10'-phenanthroline. The crude product was purified by flash column chromatography [SiO_2, 1:99→9:91 methanol:dichloromethane], which afforded the *bis-TBS-protected 2-deoxythymidine BIPY osmate esters* **51** and **52** (23 mg, 51%, d.r. 93:7) as a brown solid and as an inseparable diastereomeric mixture. **IR ν_{max}** (thin film)/cm^{-1} 3411 m, 3208 br, 3073 br, 2954 s, 2929 s, 2897 m, 2858 m, 1703 s, 1460 m, 1255 m, 1119 m, 1022 m, 838 s, 778 m; **^{1}H NMR δ_H** (500 MHz, $CDCl_3$) 9.71 (1 H, dd, *J* = 4.8, 1.3, 1 × OsN=C*H*) 9.44 (1 H, dd, *J* = 5.0, 1.0, 1 × OsN=C*H*), 8.74 (1 H, dd, *J* = 8.3, 1.3, 1 × NCHCHC*H*), 8.71 (1 H, dd, *J* = 8.5, 1.0, 1 × NCHCHC*H*), 8.38 (2 H, s, 2 × C=CHCH=C), 8.06-8.01 (2 H, m, 2 × NCHC*H*), 7.42 (1 H, br s, NH), 6.50 (1 H, dd, *J* = 8.0, 6.0, OC*H*N), 5.55 (1 H, s, NC*H*OOs), 4.40 (1 H, dt, *J* = 5.8, 2.9, C*H*OTBS), 3.87

(1 H, ddd, *J* = 6.5, 5.0, 3.0, C*H*CH_2OTBS), 3.72 (1 H, dd, *J* = 10.5, 5.0, C$H_A$$H_B$OTBS), 3.59 (1 H, dd, *J* = 11.0, 7.0, CH_AH_BOTBS), 2.65 (1 H, ddd, *J* = 13.5, 8.0, 6.0, NCHCH_AH_B), 2.27 (1 H, ddd, *J* = 13.5, 6.0, 3.0, NCHCH_AH_B), 1.89 (3 H, s, C=OCCH_3), 0.92 (9 H, s, $CH_2OSi(CH_3)_2C(CH_3)_3$), 0.84 (9 H, s, $CHOSi(CH_3)_2C(CH_3)_3$), 0.12 (3 H, s, 1 × $CH_2OSi(CH_3)_2$), 0.08 (3 H, s, 1 × $CHOSi(CH_3)_2$), 0.02 (3 H, s, 1 × $CHOSi(CH_3)_2$), 0.01 (3 H, s, 1 × $CHOSi(CH_3)_2$); **^{13}C NMR δ_C** (125 MHz, $CDCl_3$) 173.0 (*C*=OCCH$_3$), 151.9 (N*C*=O), 151.3, 150.9 (2 × OsN=*C*H), 147.9, 147.8 (2 × OsN*C*), 139.8, 139.7 (2 × NCHCH*C*H), 131.3 (2 × C=CHCH=C), 128.2, 127.9 (2 × OsN=CH*C*H), 125.8, 125.5 (2 × NC*C*CH), 96.1 (N*C*HOOs), 86.2 (CH_2OTBS), 84.1 (O*C*HN), 83.7 (C=O*C*CH_3), 72.6 (*C*HOTBS), 63.8 (CH*C*H_2OTBS), 38.2 (NCH*C*H_2), 26.0, 25.9, 25.8 (2 × SiC(CH_3)$_3$), 22.2 (C=OC*C*H_3), 18.4, 18.0 (2 × Si*C*$(CH_3)_3$), − 4.5, − 4.6, − 5.3, − 5.3 (2 × Si(CH_3)$_2$); **MS *m/z*** (ESI$^+$) 965 (100%, [M-MeCN-NH_4]$^+$); **HRMS** (ESI$^+$) $C_{34}H_{50}N_5O_8OsSi_2$ requires *M+Na* 929.2860, found 929.2617 (1.5 ppm). The ^{1}H and ^{13}C data correspond to the major diastereomer **51**, whose stereochemistry was inferred from that of diol **44**.

Chapter 9

Appendix

Chapter 9 – Appendix

Appendix 9.1 – Crystallography Data for the Bromohydrin of Thymine (13)

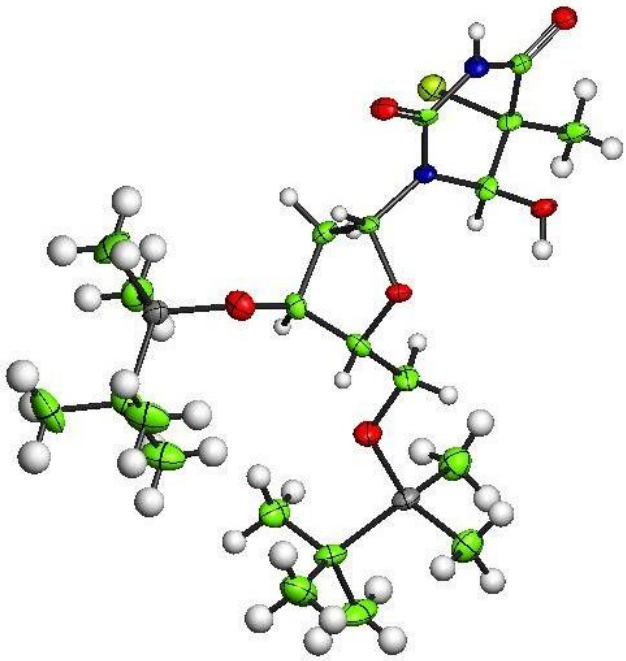


Figure 9 – 1 – *Crystal structure of bromohydrin of thymine (13).*

Crystals of **13** were grown from chloroform. A single crystal having dimensions approximately ***0.02 × 0.2 × 0.2 mm*** was mounted on a glass fibre using perfluoropolyether oil and cooled rapidly to 150 K in a stream of cold N_2 using an Oxford Cryosystems CRYOSTREAM unit. Diffraction data were measured using a Nonius Kappa CCD diffractometer (graphite-monochromated Mo K_α radiation, λ = 0.71073 Å). Intensity data were processed using the DENZO/SCALEPACK package.

Examination of the systematic absences of the intensity data showed the space group to be ***P 1 21 1***. The structure was solved in this space group using the direct-methods program SIR92, which located all non-hydrogen atoms. The H atoms were all located in a difference map, but those attached to carbon atoms were repositioned geometrically. Subsequent full-matrix least-squares refinement was carried out against F^2 using the CRYSTALS program suite. Coordinates and anisotropic thermal parameters were refined

for all non-hydrogen atoms, and the hydrogens were refined with riding constraints. A modified statistical weighting scheme was applied. Refinement converged satisfactorily to give ***R = 0.0591, wR =0.1090***.

Figure 9 – 1 shows a thermal ellipsoid plot (CAMERON) at 50 % probability. A summary of crystallographic data is given below, as are full lists of atomic coordinates, anisotropic thermal parameters and those bond lengths and angles not concerning H atoms.

Crystal identification	004tjd11
Chemical formula	$C_{22}H_{43}Br_1N_2O_6Si_2$
Formula weight	567.67
Temperature (K)	150
Wavelength (Å)	0.7107
Crystal system	Orthorhombic
Space group	P 2 21 21
a (Å)	7.495(10)
b (Å)	35.094(7)
C (Å)	10.979(2)
α (°)	90
β (°)	90
γ (°)	90
Cell volume ($Å^3$)	2887.8(9)
Z	4
Calculated density (Mg/m^3)	1.306
Absorption coefficient (mm^{-1})	1.542
F_{000}	1200.00
Crystal size (mm)	0.02 × 0.2 × 0.2
Description of crystal	Clear Pale Colourless Plate
Absorption correction	Multi-scan
Transmission coefficients (min, max)	–0.76, 1.17
θ range for data collection (°)	5.157≤ θ ≤ 26.405
Index ranges	−9 ≤ h ≤ 9, −40≤ k ≤ 43, −13≤ l ≤ 13
Reflections measured	10150
Unique reflections	5787
Refinement method	Full-matrix least-squares on Fsqd
Parameters refined	299
Weighting scheme	Calc
Goodness of fit	1.0119
R	0.0834
wR	0.1203

Table 9 – 2 – *Crystal structure and refinement details.*

Atom	x	y	z	U_{equiv}
Br(1)	0.00204(10)	0.311953(16)	-0.02860(5)	0.0305
C(2)	0.0649(7)	0.26376(17)	0.0545(5)	0.0226
C(3)	0.2189(7)	0.27156(17)	0.1408(5)	0.0227
O(4)	0.2570(5)	0.23655(11)	0.1948(3)	0.0267
N(5)	0.3753(6)	0.28545(13)	0.0710(4)	0.0197
C(6)	0.4269(8)	0.26667(17)	-0.0316(5)	0.0190
O(7)	0.5718(5)	0.27059(13)	-0.0816(4)	0.0274
N(8)	0.3010(6)	0.24168(14)	-0.0812(4)	0.0237
C(9)	0.1262(8)	0.23732(17)	-0.0468(5)	0.0232
O(10)	0.0307(5)	0.21395(12)	-0.0951(3)	0.0300
C(11)	0.4872(9)	0.31552(15)	0.1205(4)	0.0214
O(12)	0.5274(5)	0.30640(10)	0.2458(3)	0.0228
C(13)	0.5327(8)	0.34079(16)	0.3177(5)	0.0240
C(14)	0.5089(9)	0.37347(15)	0.2268(4)	0.0237
C(15)	0.4030(8)	0.35446(17)	0.1256(5)	0.0271
O(16)	0.6766(5)	0.38531(12)	0.1813(4)	0.0338
Si(17)	0.7443(2)	0.42851(5)	0.14904(15)	0.0307
C(18)	0.8978(9)	0.4449(2)	0.2728(6)	0.0378
C(19)	0.7931(10)	0.4476(2)	0.3924(6)	0.0561
C(20)	1.0472(9)	0.4161(2)	0.2894(7)	0.0570
C(21)	0.9776(13)	0.4838(2)	0.2437(8)	0.0668
C(22)	0.5506(10)	0.4611(2)	0.1330(7)	0.0524
C(23)	0.8702(11)	0.4247(2)	0.0043(6)	0.0586
C(24)	0.3842(8)	0.33848(18)	0.4122(5)	0.0291
O(25)	0.4003(6)	0.37128(12)	0.4889(4)	0.0325
Si(26)	0.3122(2)	0.37313(5)	0.62734(14)	0.0277
C(27)	0.4035(10)	0.3342(2)	0.7234(6)	0.0425
C(28)	0.3822(10)	0.4214(2)	0.6840(6)	0.0334
C(29)	0.3351(10)	0.45242(19)	0.5927(6)	0.0379
C(30)	0.2830(12)	0.4300(2)	0.8049(6)	0.0536
C(31)	0.5837(11)	0.4216(2)	0.7066(8)	0.0520
C(32)	0.0662(8)	0.3684(2)	0.6177(6)	0.0455
C(33)	-0.0979(8)	0.2490(2)	0.1206(6)	0.0318

Table 9 – 3 – *Atomic coordinates and equivalent isotropic thermal parameters ($Å^2$) of non-hydrogen atoms.*

Atom	x	y	z	U_{equiv}
H(31)	0.1831	0.2904	0.1997	0.0270
H(111)	0.5986	0.3164	0.0756	0.0247
H(131)	0.6478	0.3427	0.3569	0.0289
H(141)	0.4412	0.3948	0.2638	0.0284
H(151)	0.4127	0.3675	0.0477	0.0327
H(152)	0.2792	0.3527	0.1475	0.0333
H(193)	0.8692	0.4558	0.4565	0.0840
H(192)	0.6979	0.4653	0.3836	0.0840
H(191)	0.7465	0.4231	0.4117	0.0839
H(201)	1.1241	0.4249	0.3521	0.0839
H(202)	1.1148	0.4138	0.2148	0.0838
H(203)	0.9990	0.3918	0.3121	0.0838
H(213)	1.0608	0.4908	0.3073	0.1020
H(212)	0.8842	0.5026	0.2407	0.1017
H(211)	1.0389	0.4836	0.1661	0.1018
H(223)	0.5903	0.4863	0.1136	0.0737
H(222)	0.4865	0.4621	0.2076	0.0740
H(221)	0.4738	0.4520	0.0701	0.0741
H(231)	0.9202	0.4488	-0.0176	0.0878
H(232)	0.9638	0.4062	0.0113	0.0882
H(233)	0.7900	0.4167	-0.0591	0.0880
H(241)	0.2710	0.3390	0.3698	0.0347
H(242)	0.3956	0.3150	0.4577	0.0351
H(273)	0.3447	0.3342	0.8015	0.0628
H(272)	0.5288	0.3374	0.7362	0.0627
H(271)	0.3841	0.3097	0.6852	0.0627
H(293)	0.3760	0.4773	0.6199	0.0556
H(292)	0.2082	0.4534	0.5792	0.0560
H(291)	0.3912	0.4469	0.5170	0.0561
H(301)	0.3195	0.4550	0.8326	0.0778
H(302)	0.1570	0.4293	0.7904	0.0782
H(303)	0.3148	0.4108	0.8636	0.0780
H(313)	0.6206	0.4465	0.7332	0.0779
H(312)	0.6162	0.4029	0.7670	0.0782
H(311)	0.6435	0.4159	0.6319	0.0779
H(321)	0.0191	0.3627	0.6961	0.0691
H(322)	0.0186	0.3919	0.5883	0.0692
H(323)	0.0381	0.3482	0.5615	0.0690
H(333)	-0.1916	0.2451	0.0631	0.0489
H(332)	-0.1323	0.2677	0.1780	0.0488
H(331)	-0.0698	0.2255	0.1606	0.0489
H(81)	0.3346	0.2297	-0.1463	0.0289
H(41)	0.3112	0.2391	0.2591	0.0403

Table 9 – 4 – *Atomic coordinates and isotropic thermal parameters ($Å^2$) of hydrogen atoms.*

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
Br(1)	0.0241(3)	0.0362(3)	0.0312(3)	0.0049(3)	-0.0034(3)	0.0030(4)
C(2)	0.022(3)	0.031(3)	0.015(3)	-0.001(3)	0.005(2)	0.001(3)
C(3)	0.018(3)	0.029(3)	0.022(3)	-0.001(3)	0.008(3)	-0.001(3)
O(4)	0.029(2)	0.036(3)	0.015(2)	0.0026(18)	-0.0027(18)	-0.002(2)
N(5)	0.015(2)	0.030(3)	0.014(2)	-0.002(2)	0.0034(19)	-0.003(2)
C(6)	0.012(3)	0.030(3)	0.015(3)	0.003(3)	-0.008(3)	0.004(2)
O(7)	0.019(2)	0.042(3)	0.021(2)	-0.003(2)	0.0038(19)	-0.0018(19)
N(8)	0.021(3)	0.030(3)	0.020(2)	-0.003(2)	0.003(2)	-0.001(2)
C(9)	0.027(3)	0.025(3)	0.017(3)	0.002(3)	-0.002(3)	-0.005(3)
O(10)	0.025(3)	0.041(2)	0.024(2)	-0.0011(18)	-0.0002(19)	-0.010(2)
C(11)	0.021(3)	0.032(3)	0.011(2)	-0.002(2)	0.003(3)	-0.007(4)
O(12)	0.024(2)	0.028(2)	0.0159(17)	-0.0028(16)	-0.0031(17)	0.002(2)
C(13)	0.019(4)	0.025(3)	0.028(3)	-0.006(2)	-0.003(3)	-0.002(3)
C(14)	0.017(3)	0.024(3)	0.030(3)	-0.002(2)	0.003(3)	-0.003(3)
C(15)	0.030(3)	0.032(4)	0.019(3)	0.005(3)	0.001(3)	0.001(3)
O(16)	0.022(2)	0.040(3)	0.040(3)	-0.004(2)	0.002(2)	0.000(2)
Si(17)	0.0300(10)	0.0358(10)	0.0262(9)	0.0036(8)	-0.0034(8)	-0.0046(8)
C(18)	0.030(4)	0.042(5)	0.041(4)	-0.008(4)	-0.004(3)	-0.001(3)
C(19)	0.047(5)	0.076(6)	0.045(4)	-0.023(4)	-0.004(4)	0.008(4)
C(20)	0.034(5)	0.079(6)	0.058(5)	-0.014(4)	-0.018(4)	0.009(4)
C(21)	0.052(5)	0.045(4)	0.104(7)	-0.008(4)	-0.016(6)	-0.017(5)
C(22)	0.055(6)	0.046(4)	0.056(5)	0.012(4)	-0.016(4)	0.000(4)
C(23)	0.074(6)	0.072(6)	0.029(4)	0.009(4)	0.014(4)	-0.004(5)
C(24)	0.033(4)	0.028(4)	0.026(3)	-0.007(3)	0.008(3)	-0.001(3)
O(25)	0.040(3)	0.033(3)	0.025(2)	-0.005(2)	0.0052(19)	-0.002(2)
Si(26)	0.0300(9)	0.0346(10)	0.0183(8)	-0.0007(8)	0.0036(7)	0.0034(8)
C(27)	0.042(4)	0.044(5)	0.041(4)	0.009(4)	0.001(3)	0.007(3)
C(28)	0.038(4)	0.042(5)	0.020(3)	-0.002(3)	-0.001(3)	0.006(3)
C(29)	0.048(4)	0.031(4)	0.035(4)	-0.003(3)	0.000(3)	0.008(3)
C(30)	0.091(6)	0.046(5)	0.025(4)	-0.009(3)	0.005(4)	0.014(5)
C(31)	0.053(5)	0.050(5)	0.052(5)	-0.007(4)	-0.016(4)	-0.001(4)
C(32)	0.037(4)	0.056(5)	0.044(4)	0.007(4)	0.004(3)	-0.001(4)
C(33)	0.016(3)	0.059(5)	0.021(3)	-0.001(3)	-0.005(3)	-0.004(3)

Table – 9 – 5 – *Anisotropic thermal parameters ($Å^2$).*

Br(1) – C(2)	1.978(6)	C(14) – C(15)	1.519(8)
C(2) – C(3)	1.519(8)	C(14) – O(16)	1.415(7)
C(2) – C(9)	1.519(8)	O(16) – Si(17)	1.638(5)
C(2) – C(33)	1.511(8)	Si(17) – C(18)	1.872(7)
C(3) – O(4)	1.394(7)	Si(17) – C(22)	1.857(7)
C(3) – N(5)	1.483(7)	Si(17) – C(23)	1.853(7)
N(5) – C(6)	1.361(7)	C(18) – C(19)	1.532(10)
N(5) – C(11)	1.453(7)	C(18) – C(20)	1.519(9)
C(6) – O(7)	1.224(6)	C(18) – C(21)	1.522(10)
C(6) – N(8)	1.399(7)	C(24) – O(25)	1.431(7)
N(8) – C(9)	1.372(7)	O(25) – Si(26)	1.659(4)
C(9) – O(10)	1.212(6)	Si(26) – C(27)	1.857(7)
C(11) – O(12)	1.445(5)	Si(26) – C(28)	1.880(7)
C(11) – C(15)	1.506(8)	Si(26) – C(32)	1.854(7)
O(12) – C(13)	1.442(6)	C(28) – C(29)	1.521(9)
C(13) – C(14)	1.531(7)	C(28) – C(30)	1.551(10)
C(13) – C(24)	1.524(8)	C(28) – C(31)	1.530(9)

Table 9 – 6 – *Bond lengths (Å) (Excluding C-H bonds).*

Br1 – C2 – C3	108.3(4)	C14 – C15 – C11	101.9(5)
Br1 – C2 – C9	104.9(4)	C14 – O16 – Si17	128.6(4)
C3 – C2 – C9	109.7(5)	O16 – Si17 – C18	108.6(3)
Br1 – C2 – C33	108.8(4)	O16 – Si17 – C22	110.4(3)
C3 – C2 – C33	112.1(4)	C18 – Si17 – C22	111.1(3)
C9 – C2 – C33	112.8(5)	O16 – Si17 – C23	106.0(3)
C2 – C3 – O4	105.2(5)	C18 – Si17 – C23	109.4(3)
C2 – C3 – N5	109.7(4)	C22 – Si17 – C23	111.2(4)
O4 – C3 – N5	110.3(4)	Si17 – C18 – C19	109.0(5)
C3 – N5 – C6	119.6(5)	Si17 – C18 – C20	109.6(5)
C3 – N5 – C11	120.1(4)	C19 – C18 – C20	108.4(6)
C6 – N5 – C11	119.8(5)	Si17 – C18 – C21	111.4(5)
N5 – C6 – O7	124.7(6)	C19 – C18 – C21	109.0(6)
N5 – C6 – N8	115.7(5)	C20 – C18 – C21	109.4(6)
O7 – C6 – N8	119.6(6)	C13 – C24 – O25	107.2(5)
C6 – N8 – C9	127.3(5)	C24 – O25 – Si26	122.5(4)
C2 – C9 – N8	115.0(5)	O25 – Si26 – C27	110.2(3)
C2 – C9 – O10	123.7(5)	O25 – Si26 – C28	103.1(3)
N8 – C9 – O10	121.3(5)	C27 – Si26 – C28	111.9(3)
N5 – C11 – O12	108.4(4)	O25 – Si26 – C32	109.9(3)
N5 – C11 – C15	115.5(5)	C27 – Si26 – C32	109.5(3)
O12 – C11 – C15	104.6(4)	C28 – Si26 – C32	112.1(3)
C11 – O12 – C13	110.0(4)	Si26 – C28 – C29	111.2(5)
O12 – C13 – C14	105.5(4)	Si26 – C28 – C30	109.0(5)
O12 – C13 – C24	107.9(4)	C29 – C28 – C30	108.3(6)
C14 – C13 – C24	113.5(5)	Si26 – C28 – C31	109.4(6)
C13 – C14 – C15	102.0(4)	C29 – C28 – C31	109.5(7)
C13 – C14 – O16	110.3(5)	C30 – C28 – C31	109.5(7)
C15 – C14 – O16	109.6(4)		

Table 9 – 7 – *Bond angles (°) (Excluding H atoms).*

Appendix 9.2 – Data for the Additional Citric Acid Dihydroxylation Optimisation Conditions that were Screened

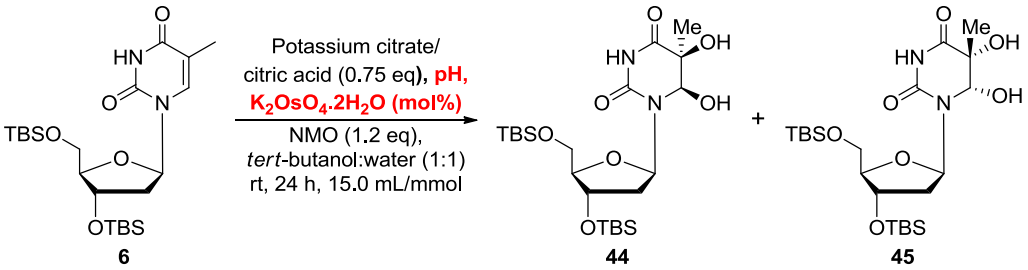


Figure 9 – 8 – *The conditions that were employed in Table 11 – 1 for the citric acid dihydroxylation of* ***6****. The reaction pH and mol% of potassium osmate used were investigated for their effect on the % conversion of* ***6*** *to* ***44*** *and* ***45****.*

Mol% of Potassium Osmate	pH 5		pH 6	
	Yield (%)	d.r. (44:45)	Yield (%)	d.r. (44:45)
5%	72	89:11	87	100:0
10%	89	81:19	58	93:7
15%	84	100:0	71	92:8
25%	76	89:11	66	88:12

Table 9 – 8 – *The calculated % conversions and d.r's for the dihydroxylation reaction of thymine (Figure 9 – 8), upon variation of the buffer, pH and mol% of potassium osmate.*

Appendix 9.3 – ^{1}H NMR Spectra of the short ssDNA Sequence 5' – GGAATGGCAA – 3'

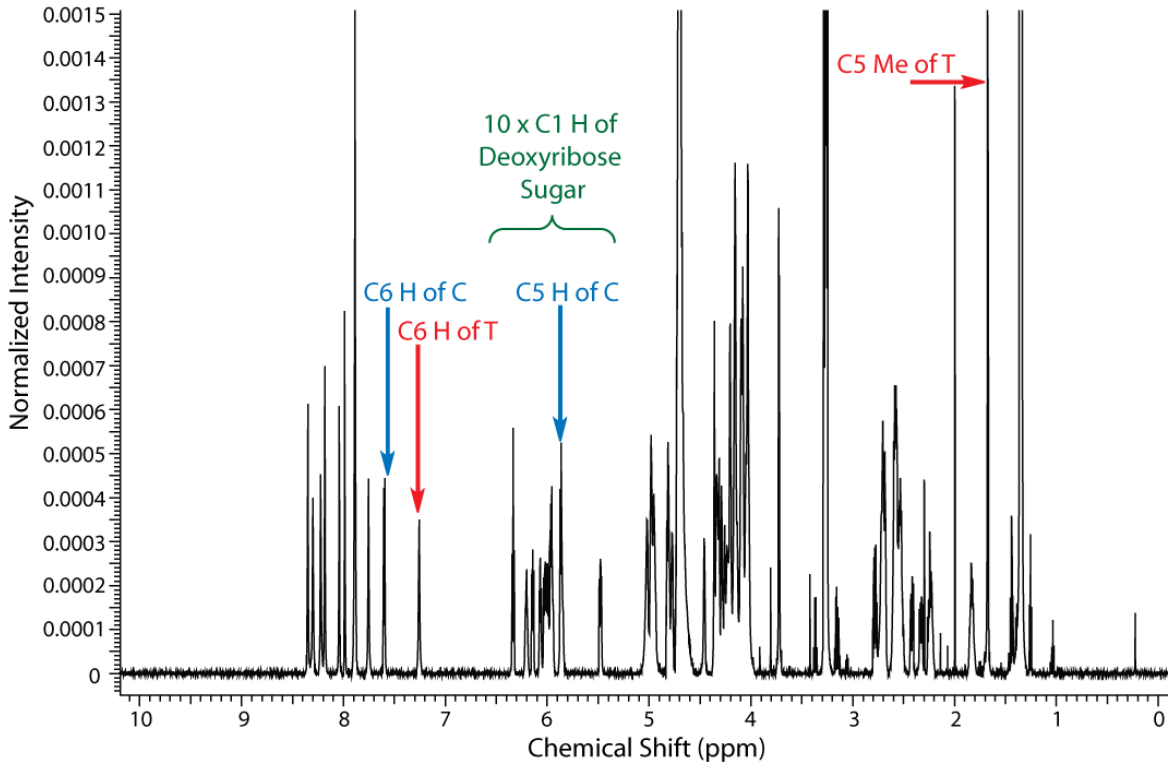


Figure 9 – 9 – *^{1}H NMR spectrum of the short DNA oligonucleotide sequence 5' – GGAATGGCAA – 3', identifying some key proton signals.*

Appendix 9.4 – Oligonucleotide Sequences

Oligonucleotide Name	Oligonucleotide Sequence (5'→3')
poly(dC)$_{40}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-Biotin
poly(dC)A$_{(14)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCC-Biotin
poly(dC)G$_{(14)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCC-Biotin
poly(dC)T$_{(14)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCC-Biotin
poly(dC)5mC$_{(14)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCC[5mC]CCCCCCCCCCCC-Biotin
poly(dC)hmC$_{(14)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCC[hmC]CCCCCCCCCCCC-Biotin
poly(dC)A$_{(9)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCC-Biotin
poly(dC)G$_{(9)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCC-Biotin
poly(dC)T$_{(9)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCC-Biotin
poly(dC)5mC$_{(9)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC[5mC]CCCCCCCC-Biotin
poly(dC)hmC$_{(9)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC[hmC]CCCCCCCC-Biotin
poly(dC)C$_{(14)}$G$_{(13)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCC-Biotin
poly(dC)A$_{(14)}$G$_{(13)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCAGCCCCCCCCCCCC-Biotin
poly(dC)G$_{(14)}$G$_{(13)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCGGCCCCCCCCCCCC-Biotin
poly(dC)T$_{(14)}$G$_{(13)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCTGCCCCCCCCCCCC-Biotin
poly(dC)5mC$_{(14)}$G$_{(13)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCC[5mC]GCCCCCCCCCCCC-Biotin
poly(dC)hmC$_{(14)}$G$_{(13)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCC[hmC]GCCCCCCCCCCCC-Biotin
poly(dC)C$_{(9)}$G$_{(8)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCC-Biotin
poly(dC)A$_{(9)}$G$_{(8)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAGCCCCCCC-Biotin
poly(dC)G$_{(9)}$G$_{(8)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGGCCCCCCC-Biotin
poly(dC)T$_{(9)}$G$_{(8)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTGCCCCCCC-Biotin
poly(dC)5mC$_{(9)}$G$_{(8)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC[5mC]GCCCCCCC-Biotin
poly(dC)hmC$_{(9)}$G$_{(8)}$	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC[hmC]GCCCCCCC-Biotin
poly(dA)$_{40}$	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-Biotin
poly(dA)C$_{(14)}$	AAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAA-Biotin
poly(dA)G$_{(14)}$	AAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAA-Biotin
poly(dA)T$_{(14)}$	AAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAA-Biotin
poly(dA)5mC$_{(14)}$	AAAAAAAAAAAAAAAAAAAAAAAAAAAA[5mC]AAAAAAAAAAAA-Biotin
poly(dT)$_{40}$	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-Biotin
poly(dT)A$_{(9)}$	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTT-Biotin
poly(dT)C$_{(9)}$	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTT-Biotin
poly(dT)G$_{(9)}$	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTT-Biotin
poly(dT)5mC$_{(9)}$	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT[5mC]TTTTTTTT-Biotin
poly(dT)hmC$_{(9)}$	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT[hmC]TTTTTTTT-Biotin
poly(dT)$_{40}$*	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-Protected Biotin
poly(dT)A$_{(9)}$*	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTT-Protected Biotin
poly(dT)C$_{(9)}$*	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTT-Protected Biotin
poly(dT)G$_{(9)}$*	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTT-Protected Biotin

Table 9 – 10 – *The oligonucleotide strand names and sequences referred to throughout Chapters 2-9.*

Chapter 10

References

Chapter 10 – References

1. A. Maxam and W. Gilbert, *Proceedings of the National Academy of Sciences of the United States of America*, 1977, **74**, 560-564.
2. F. Sanger, S. Nicklen and A. R. Coulson, *Proceedings of the National Academy of Sciences of the United States of America*, 1977, **74**, 5463-5467.
3. R. Fleishmann, *et al.*, *Science*, 1995, **269**, 496-512.
4. J. Venter, *Science*, 2001, **292**, 1838-1838.
5. E. Lander, *et al.*, *Nature*, 2001, **409**, 860-921.
6. R. Drmanac, *et al.*, *Science*, 2010, **327**, 78-81.
7. J. C. Venter, *Nature*, 2010, **464**, 676-677.
8. L. Smith, J. Sanders, R. Kaiser, P. Hughes, C. Dodd, C. Connell, C. Heiner, S. Kent and L. Hood, *Nature*, 1986, **321**, 674-679.
9. J. Prober, G. Trainor, R. Dam, F. Hobbs, C. Robertson, R. Zagursky, A. Cocuzza, M. Jensen and K. Baumeister, *Science*, 1987, **238**, 336-341.
10. X. C. Huang, M. A. Quesada and R. A. Mathies, *Analytical Chemistry*, 1992, **64**, 2149-2154.
11. D. Chen, H. Swerdlow, H. Harke, J. Zhang and N. Dovichi, *Journal of Chromatography*, 1991, **559**, 237-246.
12. J. Shendure, G. Porreca, N. Reppas, X. Lin, J. McCutcheon, A. Rosenbaum, M. Wang, K. Zhang, R. Mitra and G. Church, *Science*, 2005, **309**, 1728-1732.
13. M. Margulies, *et al.*, *Nature*, 2005, **437**, 376-380.
14. M. L. Metzker, *Nature Reviews Genetics*, 2010, **11**, 31-46.
15. D. Dressman, H. Yan, G. Traverso, K. Kinzler and B. Vogelstein, *Proceedings of the National Academy of Sciences of the United States of America*, 2003, **100**, 8817-8822.
16. R. Williams, S. Peisajovich, O. Miller, S. Magdassi, D. Tawfik and A. Griffiths, *Nature Methods*, 2006, **3**, 545-550.
17. H. Bayley, *Current Opinion in Chemical Biology*, 2006, **10**, 628-637.
18. N. Rusk, *Nature Methods*, 2011, **8**, 44-44.
19. D. Pushkarev, N. Neff and S. Quake, *Nature Biotechnology*, 2009, **27**, 847-850.
20. M. Levene, J. Korlach, S. Turner, M. Foquet, H. Craighead and W. Webb, *Science*, 2003, **299**, 682-686.
21. J. Eid, *et al.*, *Science*, 2009, **323**, 133-138.
22. B. M. Venkatesan and R. Bashir, *Nature Nanotechnology*, 2011, **6**, 615-624.
23. J. Zhou and B. Xu, *Recent Patents on DNA and Gene Sequences*, 2010, **4**, 192-201.
24. J. Kasianowicz, E. Brandin, D. Branton and D. Deamer, *Biophysical Journal*, 1996, **70**, 13770-13773.
25. D. Branton, *et al.*, *Nature Biotechnology*, 2008, **26**, 1146-1153.
26. A. Meller and D. Branton, *Electrophoresis*, 2002, **23**, 2583-2591.
27. Z. Chen, *et al.*, *Nature Materials*, 2010, **9**, 667-675.
28. T. Mach, C. Chimerel, J. Fritz, N. Fertig, M. Winterhalter and C. Fütterer, *Analytical and Bioanalytical Chemistry*, 2008, **390**, 841-846.
29. M. Kreir, C. Farre, M. Beckler, M. George and N. Fertig, *Lab on a Chip*, 2008, **8**, 587-595.
30. N. Fertig and C. Farre, *Future Medicinal Chemistry*, 2010, **2**, 691-695.
31. V. H. John, T. J. Dale, E. C. Hollands, M. X. Chen, L. Partington, D. L. Downie, H. J. Meadows and D. J. Trezise, *Journal of Biomolecular Screening*, 2007, **12**, 50-60.

32. H. Im, N. J. Wittenberg, A. Lesuffleur, N. C. Lindquist and S.-H. Oh, *Chemical Science*, 2010, **1**, 688-696.
33. B. Venkatesan, J. Polans, J. Comer, S. Sridhar, D. Wendell, A. Aksimentiev and R. Bashir, *Biomedical Microdevices*, 2011, **13**, 671-682.
34. A. Hall, A. Scott, D. Rotem, K. Mehta, H. Bayley and C. Dekker, *Nature Nanotechnology*, 2010, **5**, 874-877.
35. M. Wanunu, W. Morrison, Y. Rabin, A. Grosberg and A. Meller, *Nature Nanotechnology*, 2010, **5**, 160-165.
36. L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley and J. E. Gouaux, *Science*, 1996, **274**, 1859-1865.
37. M. Faller, M. Niederweis and G. Schulz, *Science*, 2004, **303**, 1189-1192.
38. C. Dekker, *Nature Nanotechnology*, 2007, **2**, 209-215.
39. S. Cheley, O. Braha, X. Lu, S. Conlan and H. Bayley, *Protein Science*, 1999, **8**, 1257-1267.
40. G. Maglia, M. Restrepo, E. Mikhailova and H. Bayley, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 19720-19725.
41. X.-f. Kang, L.-Q. Gu, S. Cheley and H. Bayley, *Angewandte Chemie-International Edition*, 2005, **44**, 1495-1499.
42. G. Maglia, M. Henricus, R. Wyss, Q. Li, S. Cheley and H. Bayley, *Nano Letters*, 2009, **9**, 3831-3836.
43. D. Japrung, M. Henricus, Q. Li, G. Maglia and H. Bayley, *Biophysical Journal*, 2010, **98**, 1856-1863.
44. H. Bayley and P. Cremer, *Nature*, 2001, **413**, 226-230.
45. A. Hammerstein, S. Shin and H. Bayley, *Angewandte Chemie-International Edition*, 2010, **49**, 5085-5090.
46. L. Gu, O. Braha, S. Conlan, S. Cheley and H. Bayley, *Nature*, 1999, **398**, 686-690.
47. S. Lu, W. Li, D. Rotem, E. Mikhailova and H. Bayley, *Nature Chemistry*, 2010, **2**, 921-928.
48. A. Meller, L. Nivon, E. Brandin, J. Golovchenko and D. Branton, *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 1079-1084.
49. A. Meller, *Journal of Physics-Condensed Matter*, 2003, **15**, 581-607.
50. N. Ashkenasy, J. Sanchez-Quesada, H. Bayley and M. Ghadiri, *Angewandte Chemie-International Edition*, 2005, **44**, 1401-1404.
51. W. Vercoutere, S. Winters-Hilt, H. Olsen, D. Deamer, D. Haussler and M. Akeson, *Nature Biotechnology*, 2001, **19**, 248-252.
52. J. Sanchez-Quesada, A. Saghatelian, S. Cheley, H. Bayley and M. Ghadiri, *Angewandte Chemie-International Edition*, 2004, **43**, 3063-3067.
53. J. Chu, M. Gonzalez-Lopez, S. Cockroft, M. Amorin and M. Ghadiri, *Angewandte Chemie-International Edition*, 2010, **49**, 10106-10109.
54. F. Olasagasti, K. Lieberman, S. Benner, G. Cherf, J. Dahl, D. Deamer and M. Akeson, *Nature Nanotechnology*, 2010, **5**, 798-806.
55. S. Cockroft, J. Chu, M. Amorin and M. Ghadiri, *Journal of the American Chemical Society*, 2008, **130**, 818-820.
56. J. Kasianowicz, S. Henrickson, H. Weetall and B. Robertson, *Analytical Chemistry*, 2001, **73**, 2268-2272.
57. R. Purnell, K. Mehta and J. Schmidt, *Nano Letters*, 2008, **8**, 3029-3034.
58. R. Purnell and J. Schmidt, *ACS Nano*, 2009, **3**, 2533-2538.
59. D. Stoddart, A. Heron, E. Mikhailova, G. Maglia and H. Bayley, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 7702-7707.

60. D. Stoddart, A. Heron, J. Klingelhoefer, E. Mikhailova, G. Maglia and H. Bayley, *Nano Letters*, 2010, **10**, 3633-3637.
61. D. Stoddart, G. Maglia, E. Mikhailova, A. Heron and H. Bayley, *Angewandte Chemie-International Edition*, 2010, **49**, 556-559.
62. Y. Astier, O. Braha and H. Bayley, *Journal of the American Chemical Society*, 2006, **128**, 1705-1710.
63. J. Clarke, H. Wu, L. Jayasinghe, A. Patel, S. Reid and H. Bayley, *Nature Nanotechnology*, 2009, **4**, 265-270.
64. T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis and J. H. Gundlach, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 20647-20652.
65. Nahid N. Jetha, C. Feehan, M. Wiggin, V. Tabard-Cossa and A. Marziali, *Biophysical Journal*, 2011, **100**, 2974-2980.
66. C. Heinz, H. Engelhardt and M. Niederweis, *Journal of Biological Chemistry*, 2003, **278**, 8678-8685.
67. I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis and J. H. Gundlach, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 16060-16065.
68. A. Aksimentiev and K. Schulten, *Biophysical Journal*, 2005, **88**, 3745-3761.
69. M. Muthukumar and C. Kong, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 5273-5278.
70. H. Liu, S. Qian and H. H. Bau, *Biophysical journal*, 2007, **92**, 1164-1177.
71. G. g. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. g. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen and C. Dekker, *Nano Letters*, 2010, **10**, 3163-3167.
72. J. Li, M. Gershow, D. Stein, E. Brandin and J. Golovchenko, *Nature Materials*, 2003, **2**, 611-615.
73. A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen and C. Dekker, *Nature Materials*, 2003, **2**, 537-540.
74. P. Chen, T. Mitsui, D. Farmer, J. Golovchenko, R. Gordon and D. Branton, *Nano Letters*, 2004, **4**, 1333-1337.
75. A. Aksimentiev, J. Heng, G. Timp and K. Schulten, *Biophysical Journal*, 2004, **87**, 2086-2097.
76. D. Fologea, M. Gershow, B. Ledden, D. McNabb, J. Golovchenko and J. Li, *Nano Letters*, 2005, **5**, 1905-1909.
77. B. McNally, M. Wanunu and A. Meller, *Nano Letters*, 2008, **8**, 3418-3422.
78. J. Nakane, M. Wiggin and A. Marziali, *Biophysical Journal*, 2004, **87**, 615-621.
79. M. Wanunu, J. Sutin, B. McNally, A. Chow and A. Meller, *Biophysical Journal*, 2008, **95**, 4716-4725.
80. C. Cheikh and G. Koper, *Physica A: Statistical Mechanics and its Applications*, 2007, **373**, 21-28.
81. M. Muthukumar, *Journal of Chemical Physics*, 2010, **132**, 195101.
82. K. Luo, I. Huopaniemi, T. Ala-Nissila and S. Ying, *Journal of Chemical Physics*, 2006, **124**, 034714.
83. A. S. Panwar and S. Kumar, *Macromolecules*, 2006, **39**, 1279-1289.
84. D. Fologea, J. Uplinger, B. Thomas, D. McNabb and J. Li, *Nano Letters*, 2005, **5**, 1734-1737.
85. B. Sun, D. S. Johnson, G. Patel, B. Y. Smith, M. Pandey, S. S. Patel and M. D. Wang, *Nature*, 2011, advance online publication.
86. M. R. Singleton, M. R. Sawaya, T. Ellenberger and D. B. Wigley, *Cell*, 2000, **101**, 589-600.

87. J. W. Little, *Journal of Biological Chemistry*, 1967, **242**, 679-686.
88. A. Wolffe and M. Matzke, *Science*, 1999, **286**, 481-486.
89. J. Herman and S. Baylin, *New England Journal of Medicine*, 2003, **349**, 2042-2054.
90. M. Munzel, D. Globisch and T. Carell, *Angewandte Chemie-International Edition*, 2011, **50**, 6460-6468.
91. R. Hotchkiss, *Journal of Biological Chemistry*, 1948, **175**, 315-332.
92. R. Brena, T. Huang and C. Plass, *Journal of Molecular Medicine*, 2006, **84**, 365-377.
93. R. Lister, *et al.*, *Nature*, 2009, **462**, 315-322.
94. M. Gama-Sosa, R. Midgett, V. Slagel, S. Githens, K. Kuo, C. Gehrke and M. Ehrlich, *Biochimica et Biophysica Acta*, 1983, **740**, 212-219.
95. A. Varriale and G. Bernardi, *Genomics*, 2010, **95**, 25-28.
96. T. Vaissiere, *et al.*, *Cancer Research*, 2009, **69**, 243-252.
97. E. Li, T. Bestor and R. Jaenisch, *Cell*, 1992, **69**, 915-926.
98. M. Okano, D. Bell, D. Haber and E. Li, *Cell*, 1999, **99**, 247-257.
99. S. Taylor and P. Jones, *Journal of Cellular Physiology*, 1982, **111**, 187-194.
100. L. Silverman, *et al.*, *Journal of Clinical Oncology*, 2002, **20**, 2429-2440.
101. N. Ahuja, Q. Li, A. Mohan, S. Baylin and J. Issa, *Cancer Research*, 1998, **58**, 5489-5494.
102. T. Vaissiere, C. Cuenin, A. Paliwal, P. Vineis, P. Hainaut, Z. Herceg and G. E. Investigators, *Epigenetics*, 2009, **4**, 221-230.
103. S. Beck and V. Rakyan, *Trends in Genetics*, 2008, **24**, 231-237.
104. E. Oakeley, *Pharmacology & Therapeutics*, 1999, **84**, 389-400.
105. X. Zhang, J. Yazaki, A. Sundaresan, S. Cokus, S. Chan, H. Chen, I. Henderson, P. Shinn, M. Pellegrini, S. Jacobsen and J. Ecker, *Cell*, 2006, **126**, 1189-1201.
106. H. Yamada, K. Tanabe and S. Nishimoto, *Organic & Biomolecular Chemistry*, 2008, **6**, 272-277.
107. K. Tanabe, H. Yamada, T. Ito and S. Nishimoto, *Nucleic Acids Symposium Series*, 2009, **53**, 205-206.
108. M. Frommer, L. McDonald, D. Millar, C. Collis, F. Watt, G. Grigg, P. Molloy and C. Paul, *Proceedings of the National Academy of Sciences of the United States of America*, 1992, **89**, 1827-1831.
109. H. Chen and B. Shaw, *Biochemistry*, 1993, **32**, 3535-3539.
110. S. Cokus, S. Feng, X. Zhang, Z. Chen, B. Merriman, C. Haudenschild, S. Pradhan, S. Nelson, M. Pellegrini and S. Jacobsen, *Nature*, 2008, **452**, 215-219.
111. T. Hsieh, C. Ibarra, P. Silva, A. Zemach, L. Eshed-Williams, R. Fischer and D. Zilberman, *Science*, 2009, **324**, 1451-1454.
112. A. Meissner, *et al.*, *Nature*, 2008, **454**, 766-770.
113. A. Zemach, I. McDaniel, P. Silva and D. Zilberman, *Science*, 2010, **328**, 916-919.
114. K. Tanaka and A. Okamoto, *Bioorganic & Medicinal Chemistry Letters*, 2007, **17**, 1912-1915.
115. G. Wyatt and S. Cohen, *Biochemical Journal*, 1953, **55**, 774-782.
116. S. Kriaucionis and N. Heintz, *Science*, 2009, **324**, 929-930.
117. M. Tahiliani, K. Koh, Y. Shen, W. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. Iyer, D. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930-935.
118. M. Munzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmiller, S. Michalakis, M. Muller, M. Biel and T. Carell, *Angewandte Chemie-International Edition*, 2010, **49**, 5375-5377.
119. D. Globisch, M. Munzel, M. Muller, S. Michalakis, M. Wagner, S. Koch, T. Bruckl, M. Biel and T. Carell, *Plos One*, 2010, **5**, e15367.

120. S. Jin, S. Kadam and G. Pfeifer, *Nucleic Acids Research*, 2010, **38**, e125.
121. L. Shock, P. Thakkar, E. Peterson, R. Moran and S. Taylor, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 3630-3635.
122. M. Ko, *et al.*, *Nature*, 2010, **468**, 839-843.
123. W. Pastor, *et al.*, *Nature*, 2011, **473**, 394-397.
124. C. Nestor, A. Ruzov, R. Meehan and D. Dunican, *Biotechniques*, 2010, **48**, 317-319.
125. S. Tardy-Planechaud, J. Fujimoto, S. Lin and L. Sowers, *Nucleic Acids Research*, 1997, **25**, 553-558.
126. C. Bair and L. Black, *Journal of Molecular Biology*, 2007, **366**, 768-778.
127. C. Song, *et al.*, *Nature Biotechnology*, 2011, **29**, 68-72.
128. Y. Huang, W. Pastor, Y. Shen, M. Tahiliani, D. Liu and A. Rao, *Plos One*, 2010, **5**, e8888.
129. H. Hayatsu and M. Shiragami, *Biochemistry*, 1979, **18**, 632-637.
130. C. Loenarz and C. Schofield, *Chemistry & Biology*, 2009, **16**, 580-583.
131. M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein and M. Drndic, *Journal of the American Chemical Society*, 2010, **133**, 486-492.
132. B. Flusberg, D. Webster, J. Lee, K. Travers, E. Olivares, T. Clark, J. Korlach and S. Turner, *Nature Methods*, 2010, **7**, 461-465.
133. A. Nomura and A. Okamoto, *Biochemistry*, 2011, **50**, 3376-3385.
134. A. Nomura, K. Sugizaki, H. Yanagisawa and A. Okamoto, *Chemical Communications*, 2011, **47**, 8277-8279.
135. S. Ito, A. D'Alessio, O. Taranova, K. Hong, L. Sowers and Y. Zhang, *Nature*, 2010, **466**, 1129-1133.
136. K. Koh, *et al.*, *Cell Stem Cell*, 2011, **8**, 200-213.
137. V. Valinluck, H. Tsai, D. Rogstad, A. Burdzy, A. Bird and L. Sowers, *Nucleic Acids Research*, 2004, **32**, 4100-4108.
138. T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier and T. Carell, *Angewandte Chemie-International Edition*, 2011, **50**, 7008-7012.
139. E. Privat and L. Sowers, *Chemical Research in Toxicology*, 1996, **9**, 745-750.
140. K. Gates, *Chemical Research in Toxicology*, 2009, **22**, 1747-1760.
141. H. Yuki, C. Sempuku, M. Park and K. Takiura, *Analytical Biochemistry*, 1972, **46**, 123-128.
142. P. Lawley and P. Brookes, *Biochemical Journal*, 1963, **89**, 127-138.
143. A. M. Maxam and W. Gilbert, in *Methods in Enzymology*, ed. K. M. Lawrence Grossman, Academic Press, Editon edn., 1980, vol. Volume 65, pp. 499-560.
144. K. Kiakos, A. Sato, T. Asao, P. J. McHugh, M. Lee and J. A. Hartley, *Molecular Cancer Therapeutics*, 2007, **6**, 2708-2718.
145. B. Iverson and P. Dervan, *Nucleic Acids Research*, 1987, **15**, 7823-7830.
146. M. Kai, Y. Ohkura, S. Yonekura and M. Iwasaki, *Analytica Chimica Acta*, 1988, **207**, 243-249.
147. D. A. Dunn, V. H. Lin and I. E. Kochevar, *Biochemistry*, 1992, **31**, 11620-11625.
148. K. Beckman and B. Ames, *Journal of Biological Chemistry*, 1997, **272**, 19633-19636.
149. M. Grisham, D. Jourd'Heuil and D. Wink, *Alimentary Pharmacology & Therapeutics*, 2000, **14**, 3-9.
150. C. Vialas, C. Claparols, G. Pratviel and B. Meunier, *Journal of the American Chemical Society*, 2000, **122**, 2157-2167.

151. A. Joffe, N. E. Geacintov and V. Shafirovich, *Chemical Research in Toxicology*, 2003, **16**, 1528-1538.
152. P. A. Ropp and H. Holden Thorp, *Chemistry & Biology*, 1999, **6**, 599-605.
153. C. Rubin and C. Schmid, *Nucleic Acids Research*, 1980, **8**, 4613-4619.
154. D. Ramaiah, T. Koch, H. Orum and G. Schuster, *Nucleic Acids Research*, 1998, **26**, 3940-3943.
155. L. R. Subbaraman, J. Subbaraman and E. J. Behrman, *Bioinorganic Chemistry*, 1971, **1**, 35-55.
156. K. Nakatani, S. Hagihara, S. Sando, H. Miyazaki, K. Tanabe and I. Saito, *Journal of the American Chemical Society*, 2000, **122**, 6309-6310.
157. A. Nomura and A. Okamoto, *Organic & Biomolecular Chemistry*, 2008, **6**, 3905-3907.
158. M. Beer, S. Stern, D. Carmalt and K. Mohlhenrich, *Biochemistry*, 1966, **5**, 2283-2288.
159. K. Tanaka, K. Tainaka, T. Kamei and A. Okamoto, *Journal of the American Chemical Society*, 2007, **129**, 5612-5620.
160. A. R. Cashmore, D. M. Brown and J. D. Smith, *Journal of Molecular Biology*, 1971, **59**, 359-362.
161. H. Hayatsu, Y. Wataya, K. Kai and S. Iida, *Biochemistry*, 1970, **9**, 2858-2865.
162. K. Negishi, F. Harada, S. Nishimura and H. Hayatsu, *Nucleic Acids Research*, 1977, **4**, 2283-2292.
163. H. Hayatsu, *Biochemistry*, 1976, **15**, 2677-2682.
164. E. Fritzsche, H. Hayatsu, G. Igloi, S. Iida and H. Kossel, *Nucleic Acids Research*, 1987, **15**, 5517-5528.
165. A. Okamoto, *Organic & Biomolecular Chemistry*, 2009, **7**, 21-26.
166. A. Okamoto and K. Tainaka, *Nucleic Acids Symposium Series*, 2005, **49**, 45-46.
167. A. Okamoto, K. Tainaka and T. Kamei, *Organic & Biomolecular Chemistry*, 2006, **4**, 1638-1640.
168. K. Tanaka, K. Tainaka and A. Okamoto, *Bioorganic & Medicinal Chemistry*, 2007, **15**, 1615-1621.
169. K. Tanaka, K. Tainaka, T. Kamei and A. Okamoto, *Nucleic Acids Symposium Series*, 2006, **50**, 141-142.
170. K. Tanaka, K. Tainaka, T. Umemoto, A. Nomura and A. Okamoto, *Journal of the American Chemical Society*, 2007, **129**, 14511-14517.
171. K. Tainaka, K. Tanaka and A. Okamoto, *Nucleic Acids Symposium Series*, 2006, **50**, 129-130.
172. A. Nomura, K. Tainaka and A. Okamoto, *Bioconjugate Chemistry*, 2009, **20**, 603-607.
173. K. Sugizaki, S. Ikeda, H. Yanagisawa and A. Okamoto, *Organic & Biomolecular Chemistry*, 2011, **9**, 4176-4181.
174. S. Bareyt and T. Carell, *Angewandte Chemie-International Edition*, 2008, **47**, 181-184.
175. D. C. Crans and J. J. Smee, in *Comprehensive Coordination Chemistry II*, eds. J. A. McCleverty and T. J. Meyer, Pergamon, Oxford, 2004, pp. 175-239.
176. L. Emmanuvel, T. Shaikh and A. Sudalai, *Organic Letters*, 2005, **7**, 5071-5074.
177. K. Tanabe, H. Yamada and S. Nishimoto, *Journal of the American Chemical Society*, 2007, **129**, 8034-8040.
178. H. Yamada, K. Tanabe and S. Nishimoto, *Bioconjugate Chemistry*, 2008, **19**, 20-23.
179. A. B. Robertson, J. A. Dahl, C. B. Vågbø, P. Tripathi, H. E. Krokan and A. Klungland, *Nucleic Acids Research*, 2011, **39**, e55.

180. Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte and S. Klimasauskas, *Nature Chemical Biology*, 2009, **5**, 400-402.
181. Z. Liutkeviciute, E. Kriukiene, I. Grigaityte, V. Masevicius and S. Klimasauskas, *Angewandte Chemie-International Edition*, 2011, **50**, 2090-2093.
182. M. Akeson, D. Branton, J. Kasianowicz, E. Brandin and D. Deamer, *Biophysical Journal*, 1999, **77**, 3227-3233.
183. G. Maglia, A. Heron, D. Stoddart, D. Japrung and H. Bayley, *Methods in Enzymology, Vol 475: Single Molecule Tools, Pt B*, 2010, **474**, 591-623.
184. E. Wallace, D. Stoddart, A. Heron, E. Mikhailova, G. Maglia, T. Donohoe and H. Bayley, *Chemical Communications*, 2010, **46**, 8195-8197.
185. D. Colquhoun and F. Sigworth, *Single-Channel Recording Chapter 19 -Fitting and Statistical Analysis of Single-Channel records.*, Singer Verlag, Heidelberg, 1995.
186. S. Smith, Y. Cui and C. Bustamante, *Science*, 1996, **271**, 795-799.
187. C. Bustamante, S. Smith, J. Liphardt and D. Smith, *Current Opinion in Structural Biology*, 2000, **10**, 279-285.
188. H. Mok and T. Park, *Bioconjugate Chemistry*, 2006, **17**, 1369-1372.
189. T. Lindahl, *Nature*, 1993, **362**, 709-715.
190. R. K. Grover, S. J. K. Pond, Q. Cui, P. Subramaniam, D. A. Case, D. P. Millar and P. Wentworth, *Angewandte Chemie-International Edition*, 2007, **46**, 2839-2843.
191. P. Potier, A. Abdennaji and J.-P. Behr, *Chemistry – A European Journal*, 2000, **6**, 4188-4194.
192. M. Barvian and M. Greenberg, *Journal of Organic Chemistry*, 1993, **58**, 6151-6154.
193. L. Clima and W. Bannwarth, *Helvatica Chimica Acta*, 2008, **91**, 165-175.
194. L. Valis and H. Wagenknecht, *Synlett*, 2005, 2281-2284.
195. D. Rai, M. Johar, N. Srivastav, T. Manning, B. Agrawal, D. Kunimoto and R. Kumar, *Journal of Medicinal Chemistry*, 2007, **50**, 4766-4774.
196. J. Shi, *et al.*, *Bioorganic & Medicinal Chemistry*, 2005, **13**, 1641-1652.
197. Y. Saito, H. Taguchi, S. Fujii, T. Sawa, E. Kida, C. Kabuto, T. Akaike and H. Arimoto, *Chemical Communications*, 2008, 5984-5986.
198. L. Zhang, *et al.*, *Journal of Medicinal Chemistry*, 2006, **49**, 5352-5362.
199. K. N. Kim, J. Lee, D. H. Kim, J.-S. Yoo and H. J. Kwon, *Bioorganic & Medicinal Chemistry Letters*, 2005, **15**, 77-79.
200. J. Kershaw and C. Winter, un-published work.
201. P. Magnus, N. Sane, B. P. Fauber and V. Lynch, *Journal of the American Chemical Society*, 2009, **131**, 16045-16047.
202. C. Ye and J. Shreeve, *Journal of Organic Chemistry*, 2004, **69**, 8561-8563.
203. E. Gerard and S. Brase, *Chemistry-a European Journal*, 2008, **14**, 8086-8089.
204. S. Lee, B. Jang and Z. Kafafi, *Journal of the American Chemical Society*, 2005, **127**, 9071-9078.
205. C. Price, M. Shipman, N. Rees, M. Elsegood, A. Edwards, W. Clegg and A. Houlton, *Chemistry-a European Journal*, 2001, **7**, 1194-1201.
206. K. Butsch, S. Elmas, N. Sen Gupta, R. Gust, F. Heinrich, A. Klein, Y. von Mering, M. Neugebauer, I. Ott, M. Schafer, H. Scherer and T. Schurr, *Organometallics*, 2009, **28**, 3906-3915.
207. J. Ruiz, M. Villa, V. Rodriguez, N. Cutillas, C. Vicente, G. Lopez and D. Bautista, *Inorganic Chemistry*, 2007, **46**, 5448-5449.
208. R. Butler and A. Coyne, *Chemical Reviews*, 2010, **110**, 6302-6337.
209. L. Botella and C. Najera, *Tetrahedron Letters*, 2004, **45**, 1833-1836.
210. S. Bhattacharya, A. Srivastava and S. Sengupta, *Tetrahedron Letters*, 2005, **46**, 3557-3560.

211. I. Beletskaya and A. Cheprakov, *Chemical Reviews*, 2000, **100**, 3009-3066.
212. J. Chalker, C. Wood and B. Davis, *Journal of the American Chemical Society*, 2009, **131**, 16346-16347.
213. S. Otto and J. Engberts, *Pure and Applied Chemistry*, 2000, **72**, 1365-1372.
214. D. Rideout and R. Breslow, *Journal of the American Chemical Society*, 1980, **102**, 7816-7817.
215. A. Griesbeck, *Tetrahedron Letters*, 1988, **29**, 3477-3480.
216. R. Tona and R. Haner, *Chemical Communications*, 2004, 1908-1909.
217. F. Buller, L. Mannocci, Y. Zhang, C. Dumelin, J. Scheuermann and D. Neri, *Bioorganic & Medicinal Chemistry Letters*, 2008, **18**, 5926-5931.
218. V. Borsenberger and S. Howorka, *Nucleic Acids Research*, 2009, **37**, 1477-1485.
219. A. Boersma, B. Feringa and G. Roelfes, *Organic Letters*, 2007, **9**, 3647-3650.
220. R. Breslow, U. Maitra and D. Rideout, *Tetrahedron Letters*, 1983, **24**, 1901-1904.
221. C. Loncaric, K. Manabe and S. Kobayashi, *Advanced Synthesis & Catalysis*, 2003, **345**, 475-477.
222. T. Akiyama, J. Takaya and H. Kagoshima, *Tetrahedron Letters*, 1999, **40**, 7831-7834.
223. S. Otto, F. Bertoncin and J. B. F. N. Engberts, *Journal of the American Chemical Society*, 1996, **118**, 7702-7707.
224. T. Cohen, A. Mura, D. Shull, E. Fogel, R. Ruffner and J. Falck, *Journal of Organic Chemistry*, 1976, **41**, 3218-3219.
225. B. Trost and D. Keeley, *Journal of Organic Chemistry*, 1975, **40**, 2013-2013.
226. S. Hussain, S. K. Bharadwaj, M. K. Chaudhuri and H. Kalita, 2007, **2007**, 374-378.
227. P. Perlmutter, *"Conjugate Addition Reactions"*, Pergamon Press, Oxford, 1992.
228. G. L. Khatik, R. Kumar and A. K. Chakraborti, *Organic Letters*, 2006, **8**, 2433-2436.
229. Q. A. Rodrigues de Almeida, M. Luiza de Oliveira Pereira, R. B. Coelho, E. Martins de Carvalho, C. R. Kaiser, J. J. Junior and F. Martins de Silva, *Journal of Brazilian Chemical Society*, 2008, **19**, 894-902.
230. S. Garg, R. Kumar and A. Chakraborti, *Tetrahedron Letters*, 2005, **46**, 1721-1724.
231. J. Yadav, B. Reddy and G. Baishya, *Journal of Organic Chemistry*, 2003, **68**, 7098-7100.
232. A. Roy, D. Kundu, S. K. Kundu, A. Majee and A. Hajra, *The Open Catalysis Journal*, 2010, **3**, 34-39.
233. C. Chang, H. Ford and E. Behrman, *Inorganica Chimica Acta-Bioinorganic Chemistry*, 1981, **55**, 77-80.
234. P. Andersson and K. Sharpless, *Journal of the American Chemical Society*, 1993, **115**, 7047-7048.
235. H. Cao, Y. Jiang and Y. Wang, *Nucleic Acids Research*, 2009, **37**, 6635-6643.
236. S. Zuo, R. Boorstein and G. Teebor, *Nucleic Acids Research*, 1995, **23**, 3239-3243.
237. K. Sharpless, *et al.*, *Journal of Organic Chemistry*, 1992, **57**, 2768-2771.
238. H. C. Kolb, M. S. VanNieuwenhze and K. B. Sharpless, *Chemical Reviews*, 1994, **94**, 2483-2547.
239. P. Dupau, R. Epple, A. Thomas, V. Fokin and K. Sharpless, *Advanced Synthesis & Catalysis*, 2002, **344**, 421-433.
240. T. Donohoe, K. Blades, P. Moore, M. Waring, J. Winter, M. Helliwell, N. Newcombe and G. Stemp, *Journal of Organic Chemistry*, 2002, **67**, 7946-7956.
241. T. Donohoe, K. Wheelhouse, P. Lindsay-Scott, P. Glossop, I. Nash and J. Parker, *Angewandte Chemie-International Edition*, 2008, **47**, 2872-2875.
242. J. Wai, I. Marko, J. Svendsen, M. Finn, E. Jacobsen and K. Sharpless, *Journal of the American Chemical Society*, 1989, **111**, 1123-1125.

243. M. Schroder, A. Nielson and W. Griffith, *Journal of the Chemical Society-Dalton Transactions*, 1979, 1607-1611.
244. K. Upadhya, I. Khattak and B. Mullah, *Nucleosides Nucleotides & Nucleic Acids*, 2005, **24**, 919-922.
245. M. H. Qureshi, J. C. Yeung, S.-C. Wu and S.-L. Wong, *Journal of Biological Chemistry*, 2001, **276**, 46422-46428.
246. R. Garlick and R. Giese, *Biochemical Journal*, 1990, **268**, 611-613.
247. S. Wolkenberg and D. Boger, *Chemical Reviews*, 2002, **102**, 2477-2495.
248. E. S. Newlands, M. F. G. Stevens, S. R. Wedge, R. T. Wheelhouse and C. Brock, *Cancer Treatment Reviews*, 1997, **23**, 35-61.
249. M. Stevens and E. Newlands, *European Journal of Cancer*, 1993, **29A**, 1045-1047.
250. M. F. G. Stevens, *et al.*, *Cancer Research*, 1987, **47**, 5846-5852.
251. A. Clark, B. Deans, M. Stevens, M. Tisdale, R. Wheelhouse, B. Denny and J. Hartley, *Journal of Medicinal Chemistry*, 1995, **38**, 1493-1504.
252. B. Denny, R. Wheelhouse, M. Stevens, L. Tsang and J. Slack, *Biochemistry*, 1994, **33**, 9045-9051.
253. M. J. M. Darkes, G. L. Plosker and B. Jarvis, *American Journal of Cancer*, 2002, **1**, 55-80.
254. H. Friedman, T. Kerby and H. Calvert, *Clinical Cancer Research*, 2000, **6**, 2585-2597.
255. J. Baer, A. Freeman, E. Newlands, A. Watson, J. Rafferty and G. Margison, *British Journal of Cancer*, 1993, **67**, 1299-1302.
256. S. Gerson, *Journal of Clinical Oncology*, 2002, **20**, 2388-2399.
257. B. Neyns, A. Tosoni, W. Hwu and D. Reardon, *Cancer*, 2010, **116**, 2868-2877.
258. E. Newlands, *et al.*, *British Journal of Cancer*, 1992, **65**, 287-291.
259. M. Kosonen and H. Lonnberg, *Journal of the Chemical Society-Perkin Transactions 2*, 1995, 1203-1209.
260. N. Mizuno and R. Decker, *Biochemical Pharmacology*, 1976, **25**, 2643-2647.
261. J. Pringle, L. Primrose, C. Kind, I. Talbot and I. Lauder, *Journal of Pathology*, 1989, **158**, 279-286.
262. P. G. Bardin, M. A. Pickett, S. B. Robinson, G. Sanderson, S. T. Holgate and S. L. Johnston, *Histochemistry and Cell Biology*, 1993, **100**, 387-392.
263. R. Pon, *Tetrahedron Letters*, 1991, **32**, 1715-1718.
264. R. Hartshorne, M. Tamkun and M. Montal, *Ion Channel Reconstitution, Chapter 13 The reconstituted Sodium Channel from Brain*, Plenum Press, New York, 1986.
265. M. Mayer, J. Kriebel, M. Tosteson and G. Whitesides, *Biophysical Journal*, 2003, **85**, 2684-2695.
266. M. Montal and P. Mueller, *Proceedings of the National Academy of Sciences of the United States of America*, 1972, **69**, 3561-3566.
267. S. H. White, D. C. Petersen, S. Simon and M. Yafuso, *Biophysical journal*, 1976, **16**, 481-489.
268. S. Silberberg and K. Magleby, *Biophysical Journal*, 1993, **65**, 1570-1584.
269. C. E. Shannon, *Proceedings of the IRE*, 1949, **37**, 10-21.
270. A. Pangborn, M. Giardello, R. Grubbs, R. Rosen and F. Timmers, *Organometallics*, 1996, **15**, 1518-1520.
271. D. D. Perrin and W. L. F. Armarego, *Purification of Laboratory Chemicals*, 3rd edn., Pergamon Press, Oxford, 1988.
272. V. Singh, M. Porinchu, P. Vedantham and P. K. Sahu, *Organic Syntheses*, 2005, **81**, 171-177.
273. W. Still, M. Kahn and A. Mitra, *Journal of Organic Chemistry*, 1978, **43**, 2923-2925.

274. F. M. Pfeffer and R. A. Russell, *Organic & Biomolecular Chemistry*, 2003, **1**, 1845-1851.