



The
University
Of
Sheffield.

Access to Electronic Thesis

Author: Oliver Wilkinson
Thesis title: Studies of Alkyltransferase-like (ATL) Proteins
Qualification: PhD

This electronic thesis is protected by the Copyright, Designs and Patents Act 1988. No reproduction is permitted without consent of the author. It is also protected by the Creative Commons Licence allowing Attributions-Non-commercial-No derivatives.

This thesis was embargoed until November 2016.

If this electronic thesis has been edited by the author it will be indicated as such on the title page and in the text.

Oliver Wilkinson



**Studies of Alkyltransferase-like
(ATL) Proteins**

**A thesis submitted to the University of Sheffield for the
degree of Doctor of Philosophy in the Faculty of Science,**

November 2011

Supervisor: Dr D. M. Williams

Contents

Abstract	i
Acknowledgements	iv
Abbreviations	vi
Collaborators	x
1.0 Introduction	1
1.1 DNA Structure.....	1
1.2 DNA Replication.....	4
1.3 Transcription and Translation.....	5
1.4 Protein Structure.....	7
1.5 DNA Damage and Mutation.....	10
1.5.1 DNA Alkylation Damage.....	11
1.6 DNA Repair.....	13
1.6.1 Homologous Recombination (HR) and Non-homologous End Joining (NHEJ).....	14
1.6.2 Direct Repair (Damage Reversal).....	15
1.6.3 Mismatch Repair (MMR).....	16
1.6.4 Base-excision Repair (BER).....	18
1.6.5 Nucleotide-excision Repair (NER).....	19
1.7 O ⁶ -Methylguanine-DNA methyltransferase (MGMT).....	23
1.8 Alkyltransferase-like (ATL) Proteins.....	27
1.9 Structure of Atl1 and Atl1-DNA Complex.....	31
1.10 Substrate Specificity of ATL Proteins	38
1.11 ATL Proteins and Nucleotide-Excision Repair (NER).....	43

1.12 Evolution of Alkyltransferases.....	49
2.0 Research Aims.....	52
2.1 Substrate Specificity of ATL Proteins.....	52
2.2 ATL Proteins and NER.....	53
3.0 Synthesis and Modification of O⁶-alkylguanine- containing Oligodeoxyribonucleotides (ODNs).....	55
3.1 Introduction to ODN Synthesis and Modification.....	55
3.1.1 Synthesis of Oligodeoxyribonucleotides.....	55
3.1.2 Post-DNA Synthesis Modification of ODNs.....	60
3.2 ODN Substrates for ATL-DNA Binding Assays.....	62
3.2.1 Selection of Fluorescent Dye.....	63
3.2.2 Synthesis of 5'-SIMA-labelled ODNs	64
3.3 ODN Substrates for <i>S.pombe</i> Affinity Purification Assays.....	78
3.3.1 Preparation of 219-mer ODN Substrates by Primer Extension.....	79
3.3.2 Preparation of 102-mer ODN Substrates by Annealing and Ligation.....	81
3.4 ODNs for Fluorescence-based MGMT Activity Assay.....	85
3.5 ODNs for Structural Studies of O ⁶ -carboxymethylguanine- Containing DNA	87
4.0 Fluorescence-based Methods for Studying ATL and AGT Proteins.....	91
4.1 Introduction to Fluorescence-based Methods.....	92
4.1.1 Fluorescence Anisotropy.....	92
4.1.2 Total Fluorescent Emission Intensity (TFEI).....	94
4.1.3 Förster Resonance Energy Transfer (FRET).....	96
4.1.4 Molecular Beacons.....	98
4.2 Measurement of ATL-DNA Dissociation Constants (K_D values).....	100

4.3 FRET Analysis of AtI1-DNA Binding.....	110
4.4 FRET Analysis of ODN Hybridisation.....	115
4.5 Fluorescence-based MGMT Activity Assay.....	119
5.0 Damaged DNA Recognition by ATL Proteins.....	126
5.1 Protein Recognition of DNA Damage.....	126
5.1.1 Detection of Damaged Bases by Nucleotide-flipping Proteins.....	126
5.1.2 Active-site Interactions of Nucleotide-flipping Proteins.....	131
5.2 Base Recognition by ATL Proteins.....	139
5.2.1 Recognition of Single-Stranded ODN Substrates by AtI1.....	141
5.2.2 Recognition of Single-Stranded ODN Substrates by TTHA1564.....	146
5.2.3 Recognition of Double-Stranded Substrates.....	151
5.3 Mutagenesis Studies of AtI1.....	155
5.4 Molecular Electrostatic Potentials of Nucleobases.....	160
5.5 Structural Studies of AtI1.....	165
5.6 Mechanism of Base Recognition by ATL Proteins.....	169
5.7 Tables of Binding Data.....	176
6.0 Isolation and Identification of AtI1 and Interacting Proteins.....	178
6.1 Introduction to Affinity Purification.....	178
6.2 Introduction to Protein Identification by Mass Spectrometry.....	181
6.2.1 Mass Spectrometry-based Proteomics.....	181
6.2.2 Interpretation of MS Data.....	185
6.3 Affinity-based Isolation of AtI1 and Interacting Proteins.....	187

7.0 Attempted Synthesis of a Crosslinker	205
7.1 Introduction to Crosslinking of DNA to Proteins.....	205
7.2 Disulfide Crosslinkers.....	210
7.3 Maleimide Crosslinkers	216
7.4 Oxidation of ODN Containing S ⁶ -methylpurine	227
8.0 Conclusions and Future Work	230
8.1 DNA Recognition by ATL Proteins.....	230
8.2 Affinity-based Isolation of AtI1 and Interacting Proteins.....	233
8.3 Attempted Synthesis of a Crosslinker	233
8.4 Fluorescence-based Assays	234
8.4.1 MGMT Activity Assay.....	234
8.4.2 FRET Analysis of ODN Hybridisation.....	235
8.5 Synthesis and Modification of ODNs.....	236
9.0 Experimental	238
9.1 Chemical Synthesis.....	238
9.2 Synthesis of ODNs.....	247
9.3 Post-DNA Synthesis Modification Chemistry.....	258
9.4 Purification of ODNs	263
9.5 Characterisation of ODNs.....	266
9.6 Preparation of Double-stranded ODNs.....	268
9.6.1 Preparation of Long (219-mer) ODNs by Primer Extension.....	268
9.6.2 Preparation of 102-mer ODN by Ligation.....	268
9.6.3 Preparation of dsDNA Substrates for EMSA.....	269
9.7 Other ODN Reactions.....	269
9.7.1 ODN Digestion into Nucleosides.....	269

9.7.2 Oxidation of ODN Containing S ⁶ -methylpurine	269
9.7.3 Reaction of ODN containing O ⁶ -aminoethylguanine with <i>N</i> -methoxycarbonylmaleimide.....	269
9.8 Polyacrylamide Gel Electrophoresis.....	270
9.8.1 SDS PAGE for Protein Analysis.....	270
9.8.2 Non-denaturing PAGE for ODN Analysis	271
9.8.3 Electrophoretic Mobility Shift Assays (EMSA).....	271
9.9 Protein Expression and Purification.....	271
9.9.1 Expression of Recombinant MBP-Atl1 Proteins in <i>E.coli</i>	271
9.9.2 Purification of Atl1 (Wild-type and Mutants)	272
9.9.3 Expression of Recombinant MBP-TTHA1564 Protein in <i>E.coli</i>	273
9.9.4 Purification of MBP-TTHA1564	274
9.10 Site-directed Mutagenesis of Atl1.....	275
9.11 Fluorescence-based Assays.....	276
9.11.1 Fluorescence-based ATL-DNA Binding Assays.....	276
9.11.2 FRET Analysis of Atl1-DNA Binding.....	278
9.11.3 FRET Analysis of ODN Hybridisation.....	279
9.11.4 Fluorescence-based MGMT Activity Assay.....	279
9.12 Affinity-based Isolation of Atl1.....	279
9.12.1 <i>S.pombe</i> Growth.....	279
9.12.2 Preparation of Streptavidin Beads (ODN Binding).....	280
9.12.3 Pulldown Assays (1).....	280
9.12.4 Pulldown Assays (2).....	281
9.12.5 Atl1 Quantification by ELISA.....	282
9.13 Tandem Mass Spectrometry (MS/MS) Experiments.....	282
9.13.1 Sample Preparation.....	283
9.13.2 LC MS/MS Experimental Set-up.....	284

9.12.3 MS/MS Data Analysis.....	285
11.0 References.....	286

Abstract

ATL proteins are highly conserved homologues of the direct DNA damage reversal proteins O^6 -alkylguanine-DNA-transferases (AGTs). AGTs repair highly toxic and mutagenic O^6 -alkylguanine lesions in DNA by transfer of the alkyl group to an active site cysteine, which in ATL proteins is replaced by tryptophan or alanine. ATL proteins bind tightly to DNA containing O^6 -alkylguanine lesions but cannot engage in direct repair. Complexes formed between ATLs and damaged DNA are thought to be processed via the nucleotide-excision repair (NER) pathway. This thesis investigates the recognition of O^6 -alkylguanine-containing DNA by two ATL proteins, AtI1 from *S.pombe* and TTHA1564 from *T.thermophilus*.

Oligodeoxyribonucleotides (ODNs) bearing a 5'-fluorescent (SIMA-HEX) label and containing a wide range of O^6 -alkylguanine and related modified purine bases have been prepared by post-synthesis displacement chemistry from a synthetic ODN precursor containing 2-amino-6-methylsulfonyl-purine. Using fluorescence anisotropy and fluorescence intensity measurements, dissociation constants (K_D values) of complexes with AtI1 and TTHA1564 were determined. Both ATL proteins recognise ODNs containing a wide range of O^6 -alkylguanine lesions with high affinity (K_D values range between 0.3 and 3.6 nM) including many that are poorly or not repaired by the human AGT protein MGMT. AtI1 displays an approximate 10-fold difference in its affinity for ODNs containing O^6 -alkylguanines with large, bulky lesions compared to O^6 -methylguanine and in general shows a much

higher discrimination between O^6 -alkylguanine and guanine compared to that achieved by TTHA1564. In contrast the ability of TTHA1564 to distinguish between O^6 -alkylguanines and guanine is decreased and furthermore this ATL recognises these ODNs containing O^6 -alkylguanines with similar affinity regardless of the size of the alkyl group. Both ATL proteins recognise ODNs containing 2,6-diaminopurine, whilst an ODN containing 6-methoxypurine (O^6 -methylhypoxanthine) is a very poor substrate for AtI1.

Structural studies (X-ray crystallography) were carried out with AtI1 and ODN duplexes containing 2,6-diaminopurine and 2-aminopurine. These structures display base-flipping, DNA bending and characteristics largely identical to previous structures of AtI1 in complex with O^6 -alkylguanine-containing DNA. A key feature of these structures is a cationic- π interaction between the pyrimidine ring of the flipped purine base and an active site arginine residue (R69). AtI1 R69A and R69F mutants show a decreased ability to distinguish between O^6 -alkylguanines and guanine, consistent with a key recognition element in which R69 acts as a molecular probe of the electrostatic potential of the flipped-base within the active site of the protein.

In an attempt to isolate AtI1 along with interacting proteins, affinity chromatography (pull-down) assays with a number of double-stranded ODNs prepared by different approaches were performed, with subsequent analysis by mass spectrometry-based proteomics. In addition, molecular beacon ODNs were used in the development of a novel, non-radioactive AGT activity assay, and Dickerson dodecamers containing O^6 -carboxymethylguanine residues were prepared for use in structural studies of lesion-containing DNA.

Finally, a number of routes have been explored to incorporate a reactive chemical group at the O6-position of guanine, in an attempt to prepare an ODN capable of forming a covalent crosslink with the At11 W56C mutant protein.

Acknowledgements

Firstly, a massive thank you must go to my supervisor David Williams whose support and guidance have been invaluable during the course of my research work. David and I have known each other for some time and I feel that our relationship has been both productive and enjoyable. Thanks also to Chris Millington who helped me to find my feet when I first arrived in the lab and was a source of helpful and friendly advice until his departure from the group. Rihito Morita also provided me with much-appreciated help when I first arrived in the lab in Osaka. I am grateful to Jim Reid who has been my independent advisor and without the use of whose fluorimeter a significant proportion of this research would not have been completed, and to Jane Grasby for advice on my project. Thanks to the people with whom I've had the pleasure of sharing a lab or office for the last three years: Mike, Chris, Nikesh, Ash, John, David, Kabir, Hend, Jack and Amanda. Thanks also to Elaine Frary for her skills using the DNA synthesiser, and to Jakob and David at DNA Technology for providing an excellent service.

I am grateful to the BBSRC for providing the funding for my three years of study: to be given the opportunity to spend three years pursuing something you are passionate about is a massive privilege and one that I appreciate. In addition, a fellowship from the Japan Society for the Promotion of Science (JSPS) allowed me to travel to and conduct research in Osaka, Japan which was a wonderful experience and one that I am very grateful for.

I would like to mention and extend my thanks to the people that we have collaborated with: to Bernard Connolly whose expertise and advice on fluorescence-based assays was extremely helpful during my time in Newcastle; to Ryoji Masui and Seiki Kuramitsu who allowed me to work in their lab at Osaka University on the *Thermus thermophilus* protein and with whom we formed a new and rewarding collaboration; to John Tainer and Julie Tubbs of the Scripps Institute who solved the structures of AtI1 in complex with ODNs containing 2,6-diaminopurine and 2-aminopurine; and to Akio Takenaka and his group at Iwaki-Meisi University, Japan with whom the structural studies of damaged DNA took place. Thanks to Jags Pandhal who took time out of his busy schedule to teach me about mass spec proteomics and whom I am lucky enough to call a friend. Special thanks go to Geoff Margison with whom we have collaborated very closely and who allowed me to take part in experiments in his lab on many occasions. Also thanks to the members of his group with whom it was a pleasure working: Mandy Watson, Vitaly Latypov, Andy Marriott, Mary Thorncroft and Gail McGown.

Finally, i'd like to thank Mum, Dad, Colin and my brothers Matt, Will and Joe for providing me with help, nourishment and love for as long as I can remember. My lifelong interest in science was cultivated at home from an early age as I was encouraged to question, explore and explain the world around me. Thanks also to Danielle whose support, love and understanding has been integral to my success and happiness for the last two years, and I hope will be for many more to come.

Abbreviations

A	Adenine
AEG	Aminoethyl
AGT	O ⁶ -alkylguanine-DNA-transferase
Ala	Alanine
2-AP	2-aminopurine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATL	Alkylguanine transferase-like protein
AtI1	ATL homologue from <i>S.pombe</i>
BER	Base-excision repair
Bn	Benzyl
Bu	Butyl
C	Cytosine
CMG	Carboxymethyl
Cy	Cyanine
Cys	Cysteine
DABCO	1,4-diazabicyclo[2.2.2]octane
DAP	2,6-diamniopurine
DBU	Diaza(1,3)bicycle[5.4.0]undecane
DCM	Dichloromethane
DIPEA	N,N-diisopropylethylamine
DMT-OFF	Dimethyltrityl off
DMT-ON	Dimethyltrityl on

dNTPs	Deoxyribonucleotide triphosphate
DNA	2'-Deoxyribonucleic acid
dsDNA	Double-stranded DNA
eATL	ATL homologue from <i>E.coli</i>
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
ESI	Electrospray ionisation
Et	Ethyl
EtOH	Ethanol
Et ₂ O	Diethyl ether
FA	Fluorescence Anisotropy
FAM	Fluorescein
FRET	Förster resonance energy transfer
G	Guanine
GGR	Global genome repair
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GST	Glutathione S-transferase
HEX	Hexachlorofluorescein
His	Histidine
HOEt	Hydroxyethyl
HOPr	Hydroxypropyl
HPLC	High performance liquid chromatography
HR	Homologous recombination
Ile	Isoleucine
Leu	Leucine

Lys	Lysine
MAG	Methyladamantyl
MALDI	Matrix-assisted laser desorption ionisation
MBP	Maltose-binding protein
MeCN	Acetonitrile
MeOH	Methanol
Met	Methionine
MGMT	O ⁶ -methylguanine-DNA-methyltransferase
MHx	Methylhypoxathine
µL	Microlitres
mmol	Millimoles
mL	Millilitres
MMPP	Magnesium monoperoxyphthalate
MMR	Mismatch repair
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NER	Nucleotide-excision repair
NHEJ	Non-homologous end joining
NMR	Nuclear magnetic resonance
ODN(s)	Oligodeoxyribonucleotide(s)
PAGE	Polyacrylamide gel electrophoresis
Phe	Phenylalanine
Pob	4-oxo-4-(3-pyridyl)butyl
Pr	Propyl
Pro	Proline
pSRM	Pseudo-selective reaction monitoring
RNA	Ribonucleic acid

SIMA	2-diphenyl-2-dichlorofluorescein
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulphate
Ser	Serine
SPR	Surface plasmon resonance
ssDNA	Single-stranded DNA
T	Thymine
TCA	Tricyclic purine analogue
TCR	Transcription-coupled repair
TFA	Trifluoroacetic acid
TFEI	Total fluorescent emission intensity
Thr	Threonine
Trp	Tryptophan
TTHA1564	ATL homologue from <i>T.thermophilus</i>
Tyr	Tyrosine
UV	Ultra-violet
Val	Valine
vpATL	ATL homologue from <i>V.parahaemolyticus</i>
WCE	Whole cell extract
Ybaz	ATL homologue from <i>E.coli</i> (alternative name)
YES	Yeast extract with supplements

Collaborators

Professor Geoffrey P. Margison

Cancer Research-UK Carcinogenesis Group, Paterson Institute for Cancer Research (Manchester, UK)

Professor Bernard A. Connolly

School of Cell and Molecular Biosciences, University of Newcastle (Newcastle, UK)

Professors Ryoji Masui and Seiki Kuramitsu

Department of Biological Sciences, Graduate School of Science, Osaka University (Osaka, Japan)

Professor John Tainer and Dr. Julie Tubbs

Scripps Research Institute (La Jolla, USA)

Dr. Jagroop Pandhal

ChELSI Institute, Department of Chemical and Biological Engineering, University of Sheffield (Sheffield, UK)

Professor Christopher Hunter

Centre for Chemical Biology, Department of Chemistry, University of Sheffield (Sheffield, UK)

1.0 Introduction

1.1 DNA Structure

Deoxyribonucleic acid (DNA) is a polymeric molecule present in all living systems, existing as the chemical store of genetic information. The blueprint for every constituent part of any organism is encoded for by the DNA in its genome and, as such, DNA is essential for the existence and continuation of life.

DNA is made up of monomer units known as nucleotides. Each nucleotide is composed of a heterocyclic base (guanine, cytosine, thymine or adenine), a sugar ring (2-deoxy- β -D-ribose) and a phosphate diester group which joins one nucleotide to the next one in the chain. The sugar and the base are attached to each other by a β -glycosidic linkage (figure 1.1). The nature of the sugar-phosphate backbone leads to a directionality of the DNA strands (from C-5' to C-3' of the ribose).

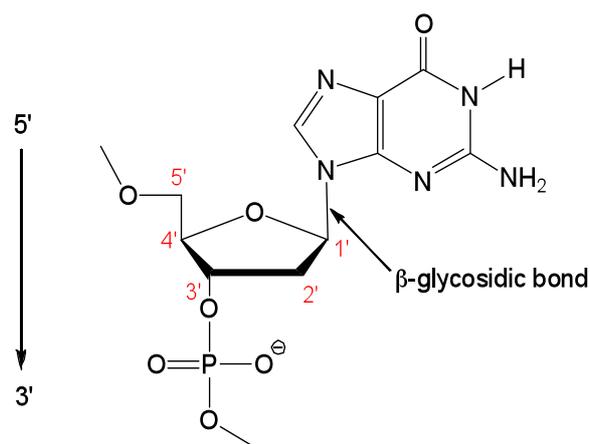


Figure 1.1: A nucleotide monomer unit (containing guanine)

Each of the four bases in DNA selectively hydrogen bond with only one of the other bases; cytosine interacts with guanine and thymine pairs with adenine. This is known as Watson-Crick base pairing (figure 1.2) and it is these interactions that form and stabilise the secondary structure of DNA, the double helix, where two polymeric chains are lined up against each other.(1) The way in which the phosphate backbones twist around each other in the double helix forms the major (wider) and minor (narrower) grooves (figure 1.3).

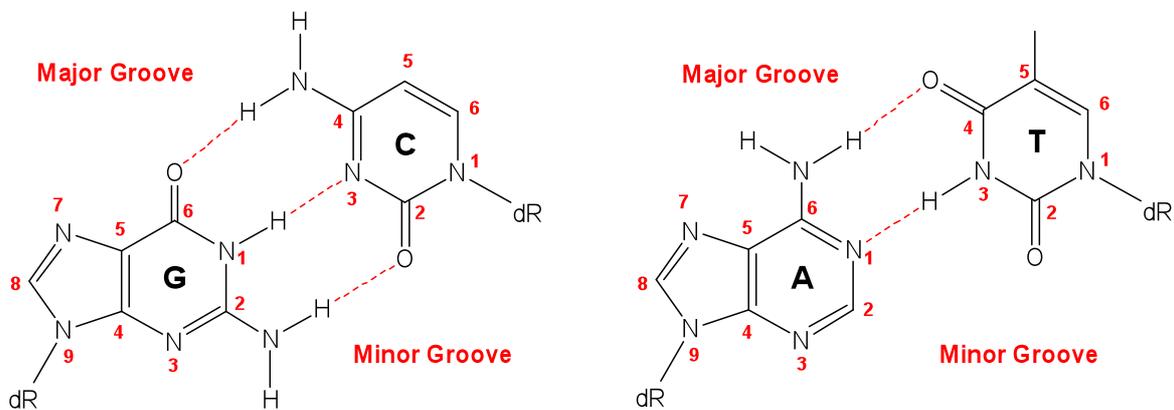


Figure 1.2: Watson-Crick base-pairing of guanine and cytosine (top) and adenine and thymine (bottom). dR is 2-deoxyribose

The DNA double helix can exist in three forms: the most predominant form at physiological conditions is right-handed B-DNA, with a major groove 12Å wide and a minor groove 6Å wide. Other structures include A-DNA, which is also a right-handed helix but differs from B-DNA by the width and depth of the grooves (due to the puckering conformation of the ribose sugars in the backbone), and much less common, left-handed form known as Z-DNA.

Whilst Watson-Crick base pairs are primarily responsible for the formation of the double helix, additional stability is provided by hydrophobic π - π interactions between the aromatic rings of the bases that are stacked up perpendicular to the helical axis. One of the DNA strands runs in a 5'→3' direction and the other runs anti-parallel to it (3'→5'). The length of one full turn of the B-DNA helix is approximately 34Å which corresponds to 10 nucleotides on each strand. The distance of the Watson-Crick hydrogen bonds is 2.80-2.95Å and since the base pairs always consist of a pyrimidine and a purine the width of the helix is constant. This is a distance of approximately 10.6Å from the C1' atom on one ribose sugar to the C1' atom of the ribose in the nucleotide opposite.

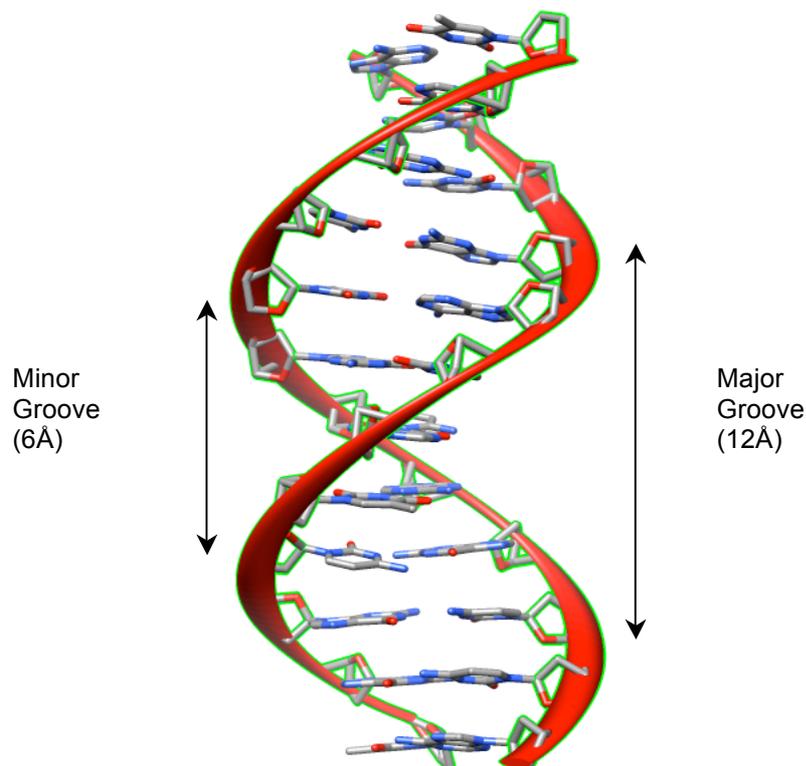


Figure 1.3: The DNA double helix. Note the bases paired up opposite each other and the major and minor grooves

1.2 DNA Replication

Every time a cell divides, a copy of the genetic code must be passed on to each of the nascent cells. This is achieved by DNA replication, which relies on the specificity of base pairing to maintain the integrity of the code. C always pairs with G, and A with T, meaning that each strand of the double helix can act as a template to allow specific enzymes known as DNA polymerases to generate complementary copies, thus giving rise to two new identical double helices. This process is said to be semi-conservative as each new DNA double helix is comprised of one 'new' strand and one 'old' one. For example, a sequence 5'-GATCGTCA will have the complementary sequence 3'-CTAGCAGT.

The first stage of DNA replication is the unwinding of the strands by enzymes known as helicases to create a replication fork.(2) This is followed by binding of single-stranded binding proteins which stabilise the separated DNA strands and allow access by a DNA polymerase. The DNA polymerase synthesises the new strands by incorporating 'free' deoxyribonucleotide triphosphates (dNTPs) into the growing chain using the opposite strand as a template. Replication can only proceed in a 5'→3' direction owing to the requirement of DNA polymerase for a free ribose 3'-OH to join the nucleotides together. The process of DNA replication is shown in figure 1.4 and highlights the key features: unwinding of the helix, replication fork formation, and the two distinct modes of DNA synthesis. Leading strand synthesis, performed by Pol δ , is continuous as it can advance in a 5'→3' direction towards the fork. However, due to the direction of the template sequence as the fork opens, the

lagging strand must be synthesised discontinuously in a series of short stretches of nucleotides, known as Okazaki fragments.(3) Before lagging strand synthesis can occur, a short piece of RNA known as a primer must be added to each segment by DNA primase. As a result, after synthesis the Okazaki fragments are left with an 5'-single-stranded RNA flap that must be cleaved by a structure-specific flap-endonuclease (FEN-1 in humans).(4) Finally, the resolved fragments are sealed together by DNA ligase to complete production of the new strand.

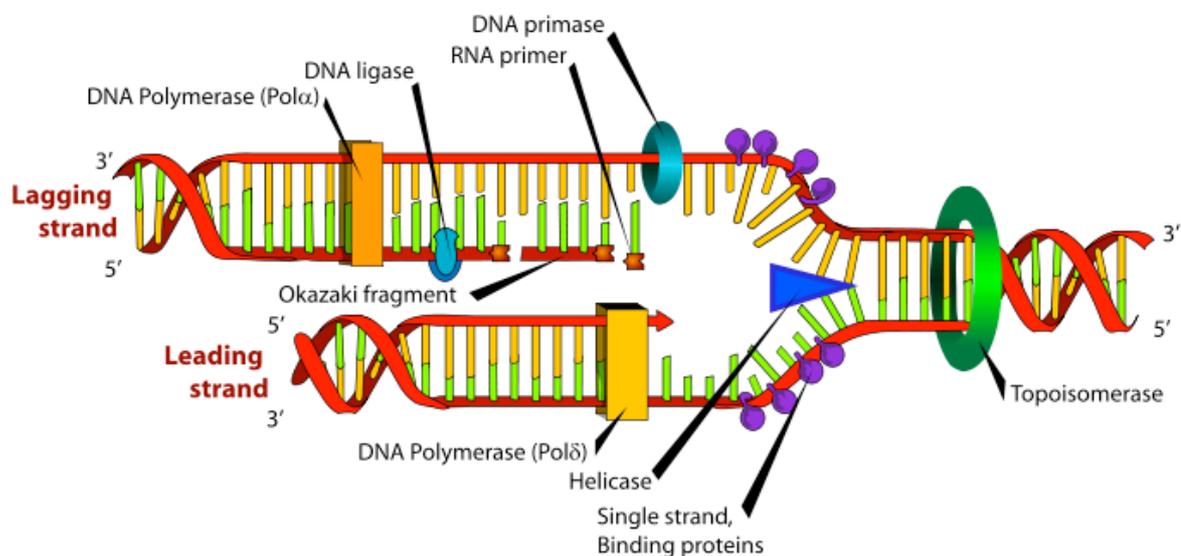


Figure 1.4: A schematic showing the process of DNA replication (taken from replicationfork.com)

1.3 Transcription and Translation

The central dogma of molecular biology, which explains the movement of information in biological systems, is shown in figure 1.5.(5) As mentioned previously, more copies of DNA are produced by replication. The process of transcription creates RNA molecules from the DNA template, whilst translation of the RNA completes the synthesis of proteins from the genetic code.

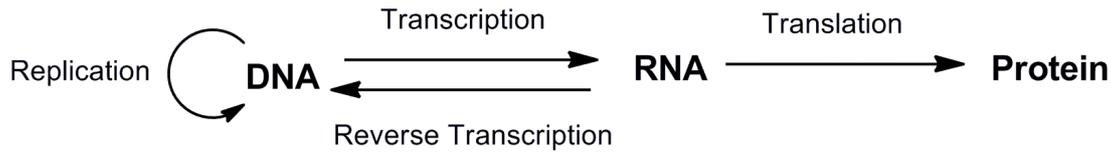


Figure 1.5: The central dogma of molecular biology

Transcription produces a strand of RNA that is complementary to the sequence of the DNA template. The process is carried out by RNA polymerase, which unzips the DNA double helix, synthesises the new RNA strand and then releases it. If the DNA sequence being transcribed codes for a protein (i.e. it is a gene) then messenger RNA (mRNA) is produced. This mRNA is transported to the ribosome, which is a complex structure composed of hundreds of sub-units of both RNA and proteins. The ribosome is essentially the cellular 'factory' of protein biosynthesis. The mRNA is then used by the ribosome as a template for protein synthesis, with each three bases in the sequence forming a codon that specifies an amino acid. These amino acids are brought to the ribosome by transfer RNA (tRNA) molecules and joined together by the ribosomal machinery to form the long polypeptide chains that subsequently fold around themselves to become active proteins. Proteins are the structural and functional building blocks of all living things and as such are responsible for a multitude of essential biological roles.

Suppose that a chemical change occurs to one of the DNA bases such that it no longer only interacts with the complementary base in the double helix, for example if guanine is alkylated at the O^6 -position such that it hydrogen bonds not exclusively with cytosine but also thymine. When replication or transcription takes place, the incorrect complementary dNTP is

likely to be incorporated by DNA polymerase into the growing strand of nucleotides so that the new DNA or RNA produced will not be the same as the original template DNA. This leads to mutations in the genetic code and leads ultimately to changes in the amino acid sequences of the proteins made. This can impair the function of the proteins and affect the regulation of protein synthesis and other biochemical pathways within the cell.

1.4 Protein Structure

Proteins are long polypeptide chains composed of amino acids joined together by peptide (amide) bonds. It is the identity of these amino acids that affect the structure and hence the function of the protein. Individual amino acids have different side chains which in turn gives them different properties (hydrophobicity, charge, chemical reactivity etc.). The twenty one naturally occurring amino acids are shown with their abbreviated names, letter codes and side chains in figure 1.6. The primary structure of a protein is the sequence of amino acids it is composed of, and proteins have two termini: an N-terminus (which is the 'start' of the protein during biosynthesis and consists of an amino group) and a C-terminus (which is the end that contains a carboxylic acid group). The amino acid chains form characteristic, local secondary structures such as α -helices and β -sheets by donating and accepting hydrogen bonds. α -helices are right-handed coil conformations formed when every amino group (N-H) of a polypeptide chain donates a hydrogen bond to the carbonyl group (C=O) of the amino acid four residues earlier in the sequence. β -sheet are formed when N-H to C=O hydrogen bonds are donated between different polypeptide strands which causes them

form a biologically active protein (figure 1.7). Protein folding is incredibly complex and is mediated by factors such as hydrophobicity, formation of H-bonds and disulphide linkages, and packing of amino acid side chains.

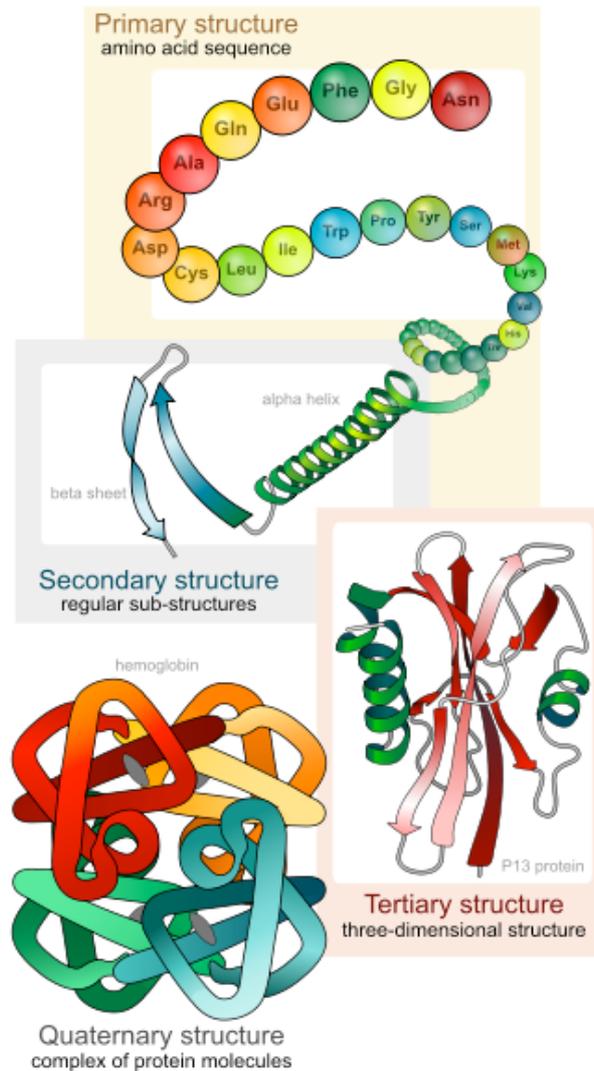


Figure 1.7: Increasing complexity of protein structure (taken from [http:// en.wikipedia.org/wiki/](http://en.wikipedia.org/wiki/File:mainproteinstructurelevels.en.svg)

File:mainproteinstructurelevels.en.svg)

Proteins that catalyse chemical reactions are known as enzymes and promote catalysis by binding substrate molecules in a distinct region called the active site. The nature of the amino acid residues in this active site affects the recognition of substrate molecules and is usually highly specific for a

certain enzyme. Proteins also have structural regions that are known as domains: they are often named for the biological function that they perform and are often found in many different proteins, e.g. a DNA-binding domain (DBD). A related concept is the motif, a distinctive sequence or structural fold found in proteins, e.g. the helix-turn-helix (HTH) motif, a structural feature which is implicated in DNA recognition. Often function can be elucidated or predicted from consideration of these structural elements and they are frequently common to many proteins. Finally, the quaternary structure of a protein describes the amalgamation of proteins or polypeptides into higher order structures. If a protein engages in interactions with other proteins to form a complex, it becomes referred to as a sub-unit and the group of proteins a complex.

1.5 DNA Damage and Mutation

There are various ways in which DNA can become damaged. It can be caused by UV light, radiation and both endogenous and exogenous chemical agents. DNA damage is a physical abnormality in the DNA, which can usually be repaired by specific enzymes that recognise that form of lesion. If not repaired, DNA damage can lead to mutations by causing errors during replication. Mutations are changes in the base sequence of the DNA and hence a corruption of the genetic code. Once incorporated, such changes are replicated into new cells and cannot be repaired.

Some types of DNA damage will eventually result in cell death, known as apoptosis. Apoptosis is a method of regulation by which cells that pose a risk to the survival of the organism as a whole can be eliminated. If the cell does not perish and the damage causes mutations in the DNA code, there are

a variety of effects that it can have on the cell. Some mutations are neutral, that is they cause no ostensible problems for the cell. The majority of mutations have a detrimental effect on the cell concerned, causing these cells to undergo apoptosis and thus be lost from the organism. However, occasional mutations confer a survival advantage on the cell such that it expands at the expense of neighbouring cells. Changes that result in these cells undergoing rapid and unregulated cell division can cause cancer.

1.5.1 DNA Alkylation Damage

One of the more biologically significant types of DNA damage is alkylation. This chemical modification can take place at a number of sites on the purine and pyrimidine nucleobases and at the phosphates of the DNA backbone, occurring with varying frequencies in the cell (figure 1.8). Alkylation can be caused by S_N1 agents (such as methylnitrosourea (MNU)) that react via a carbocation, or S_N2 agents (such as methane methylsulfonate (MMS))

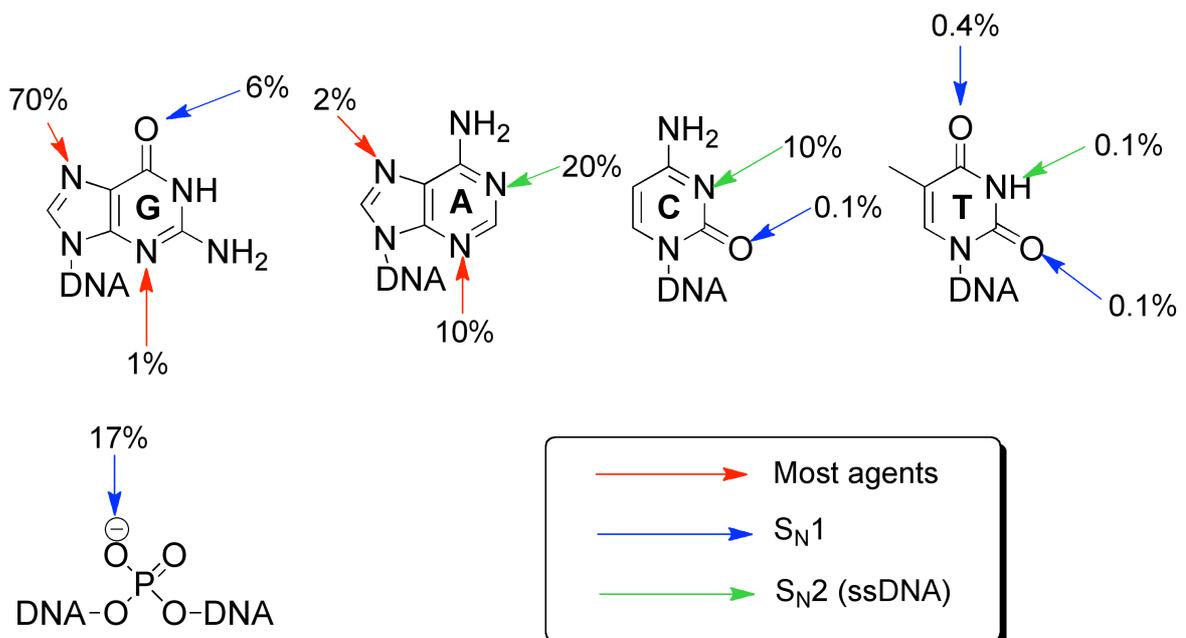


Figure 1.8: Sites and frequency of base alkylation damage by S_N1 and S_N2 agents (6)

which react in a bimolecular mechanism. Different alkylation products have a range of biological consequences, for example N7-alkylation of guanine (which is the most frequently occurring form of damage) is not directly mutagenic, although it can lead to spontaneous depurination, creating an apurinic (AP) site which must be repaired to prevent mutations occurring. In contrast, N¹-alkylpurines are considered extremely cytotoxic lesions due to their ability to block DNA replication. Alkylation of the O⁶-position of guanine occurs relatively frequently (approximately 5000 events per cell per day (7)) and is a highly problematic form of damage. When this chemical modification occurs, it locks the modified guanine base in the enol-tautomeric form, rather than the keto-form. This effectively means that rather than the exclusive base pairing that is observed between guanine and cytosine, O⁶-methylguanine (O⁶-MeG) can mispair with thymine and form a wobble base-pair with cytosine (figure 1.9). O⁶-MeG residues fail to block replication by various DNA

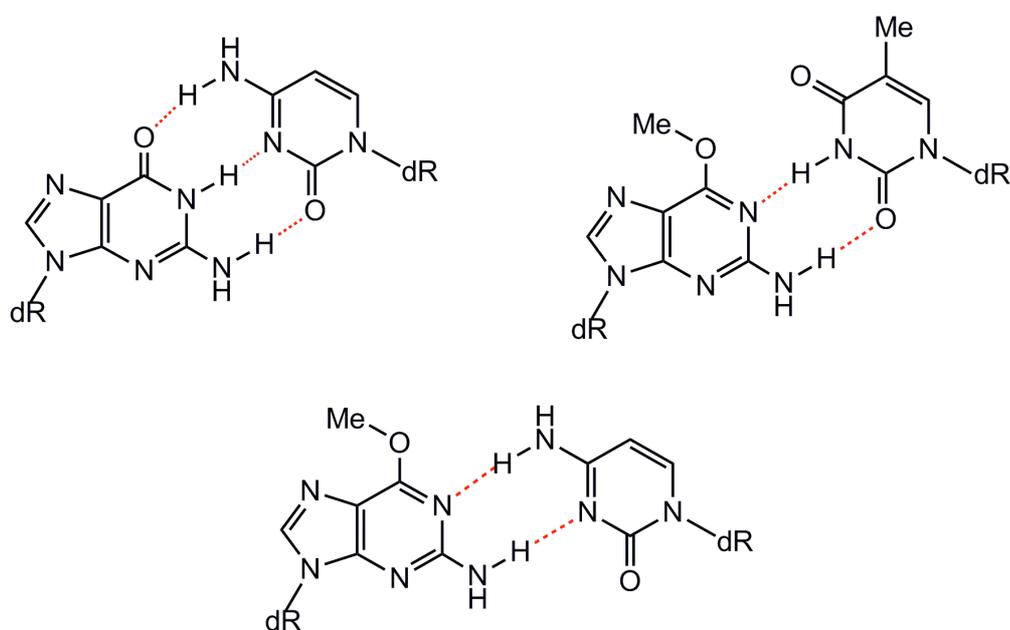


Figure 1.9: G·C base-pair (left), O⁶-MeG·T base-mispair (right), O⁶-MeG·C wobble base-pair (bottom)

polymerases, and since O^6 -MeG·T base-mispairs have a more Watson-Crick-like geometry than O^6 -MeG·C wobble base-pairs, thymine becomes incorporated preferentially in place of cytosine during DNA replication.(8) This preference leads to a GC→AT (or G→A) transition (figure 1.10) and O^6 -alkylguanines are hence known as mutagenic lesions.(9) These lesions are also highly cytotoxic (10) for reasons that will be explained in section 1.6.3. Whilst O^6 -alkylguanine adducts can occur endogenously, they can also be specifically induced by the action of S_N1 alkylating agents, such as alkyl nitrosureas (e.g. MNU, ENU, PNU, BzNU) and *N*-alkyl *N'*-nitro-*N*-nitrosoguanidines (e.g. MNNG, ENNG), some of which are used in cancer chemotherapy to deliberately induce cell death in tumours.

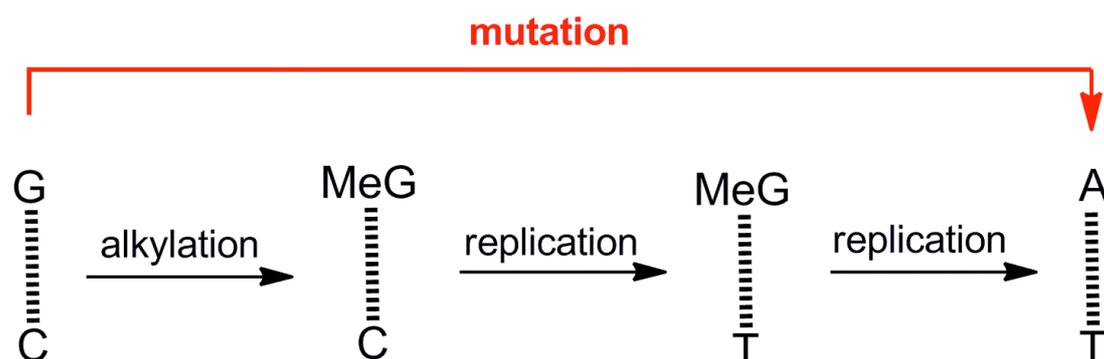


Figure 1.10: A schematic showing a GC→AT transition

1.6 DNA Repair

The presence of damaged DNA can have extremely serious consequences for living cells and therefore various mechanisms and repair systems have evolved to process DNA lesions. For the many types of DNA damage that have adverse biological effects, different enzymatic pathways

exist to repair this damage.(11) Figure 1.11 gives a summary of DNA damage in cells and the pathways by which they are repaired.

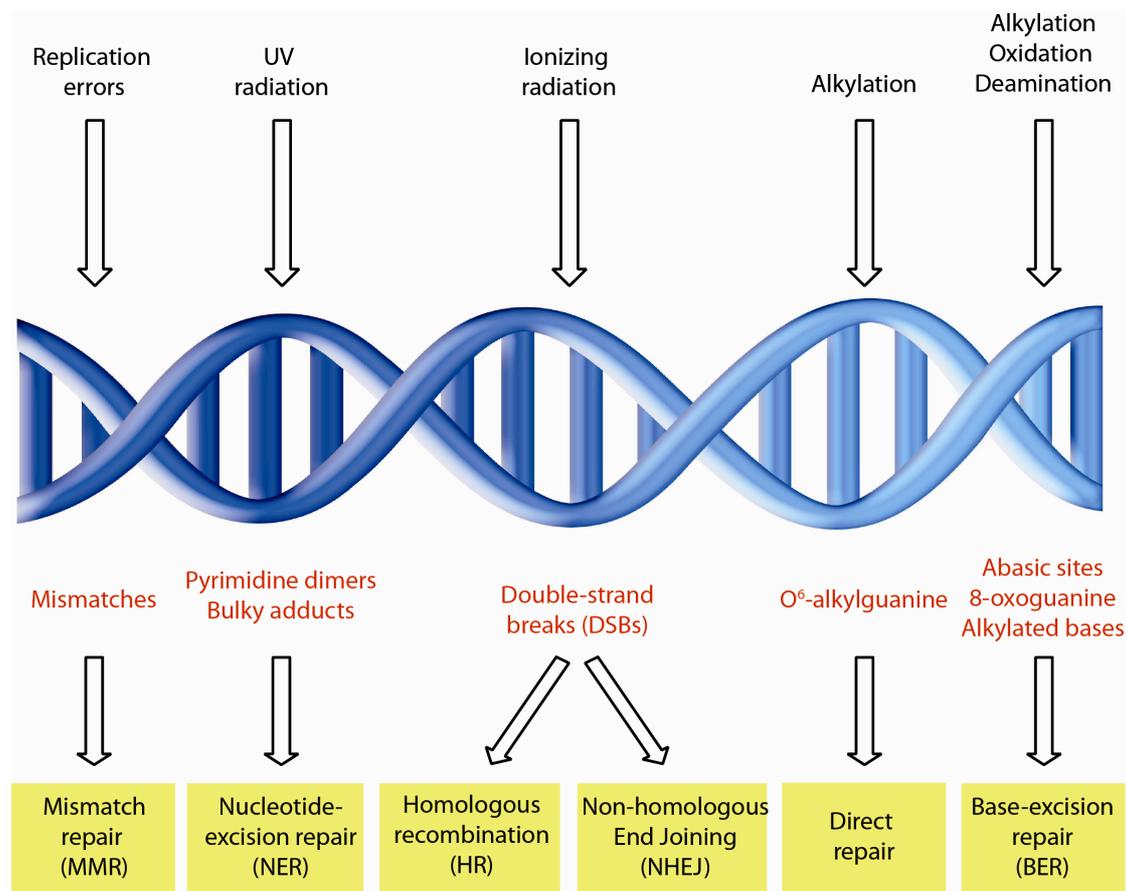


Figure 1.11: Causes, types and repair of DNA damage in cells

1.6.1 Homologous Recombination (HR) and Non-homologous End Joining (NHEJ)

Double-strand breaks (DSBs), which are particularly hazardous to the cell as they can cause rearrangements of the genome and/or cell death, are repaired by two different mechanisms: homologous recombination or non-homologous end joining.(12) DSBs can be caused by X-ray radiation, or by the collapse of replication forks, and are extremely serious in biological terms. In homologous recombination repair (HR), which is considered to be an error-

free mechanism, an almost identical enzyme machinery to that responsible for chromosomal crossover during cell meiosis uses the DNA template from a sister chromatid to repair the break. As such, HR can only take place during the S and G2 phases of the cell cycle when these chromatids are present. In contrast, non-homologous end joining (NHEJ) can take place at any time in the cell cycle but is considered an error-prone mechanism. In NHEJ, a specialised DNA ligase directly joins the ends of the strands back together.

1.6.2 Direct Repair (Damage Reversal)

Direct repair (or damage reversal) is a mechanism used to dealkylate bases that have undergone chemical damage. Although the most abundant alkylated DNA lesion is N^7 -methylguanine, it is not directly mutagenic or cytotoxic (13) and is not repaired by direct transfer (N^7 -MeG is repaired by base-excision repair along with N^3 -MeA, see section 1.6.4). However, O^6 -alkylguanine lesions are highly mutagenic, recombinogenic and cytotoxic and as such are repaired by O^6 -alkylguanine DNA-alkyltransferase (AGT) proteins: those best studied are the Ada and Ogt proteins in *E.coli* and O^6 -methylguanine-DNA methyltransferase (MGMT) in humans.(14) AGTs repair damage by transferring the alkyl group from the modified base to an active site cysteine residue. Ogt and MGMT also have the ability to repair O^4 -methylthymine residues (15), though these lesions are much less abundant and mutagenic than O^6 -methylguanine. The structure and function of MGMT are described in detail in section 1.7. Direct damage reversal is also the mechanism of repair of N^1 -methyladenine and N^3 -methylcytosine which are highly cytotoxic lesions owing to their ability to inhibit or block DNA replication.

These adducts are repaired by 1-methyladenine-DNA dioxygenases, such as AlkB in *E.coli*, and ABH2 and ABH3 in humans, using a reaction known as oxidative dealkylation.

AGT proteins are effectively the first line of defence against O^6 -alkylguanine damage; if the damage persists and leads down the path toward a mutation (e.g. a GC-AT transition mutation) then the resolution of the problem becomes the responsibility of the post-replication mismatch repair (MMR) system.(10)

1.6.3 Mismatch Repair (MMR)

It is the role of mismatch repair (MMR) to detect when incorrect bases are paired with each other in the double helix, usually as a result of bases being inserted or deleted during DNA replication but also due to mutagenic events such as O^6 -alkylguanine: thymine mismatches. It is a highly conserved process from prokaryotes to eukaryotes, as loss of MMR results in greatly increased rates of spontaneous mutation.(16) In *E.coli*, a homodimer of the protein MutS recognises the mismatched base, and then in concert with MutL recruits the endonuclease MutH, which makes an incision in the newly synthesised DNA strand containing the error. Aided by MutL, helicase II finds the nick and unwinds the helix which exposes the single-strand to digestion by exonucleases (such as ExoI but there are also many others). This leaves a gap of between 1000-3000 nucleotides which is then repaired by DNA Pol III and DNA ligase to restore the DNA to the correct sequence.(17)

In eukaryotes, the pathway is largely the same with some small differences. Rather than being homodimers, MutS and MutL are heterodimers

made up of two subunits (e.g. MutS α is composed of MSH2 and MSH6, MutS β of MSH2 and MSH3). The selection of these subunits changes the substrate specificity and cellular function of the complexes allowing greater flexibility.(16) In addition, the eukaryotic MMR system is not as well characterised as that of prokaryotes.

In fact, it is the MMR pathway that is responsible for the cytotoxicity of O^6 -alkylguanine lesions (figure 1.12). Whilst the MMR machinery has the ability to recognise the O^6 -alkylguanine: thymine mismatch and remove thymine from the newly replicated strand, it cannot remove the O^6 -alkylguanine lesion itself. This leads to repeated cycles of MMR recognition,

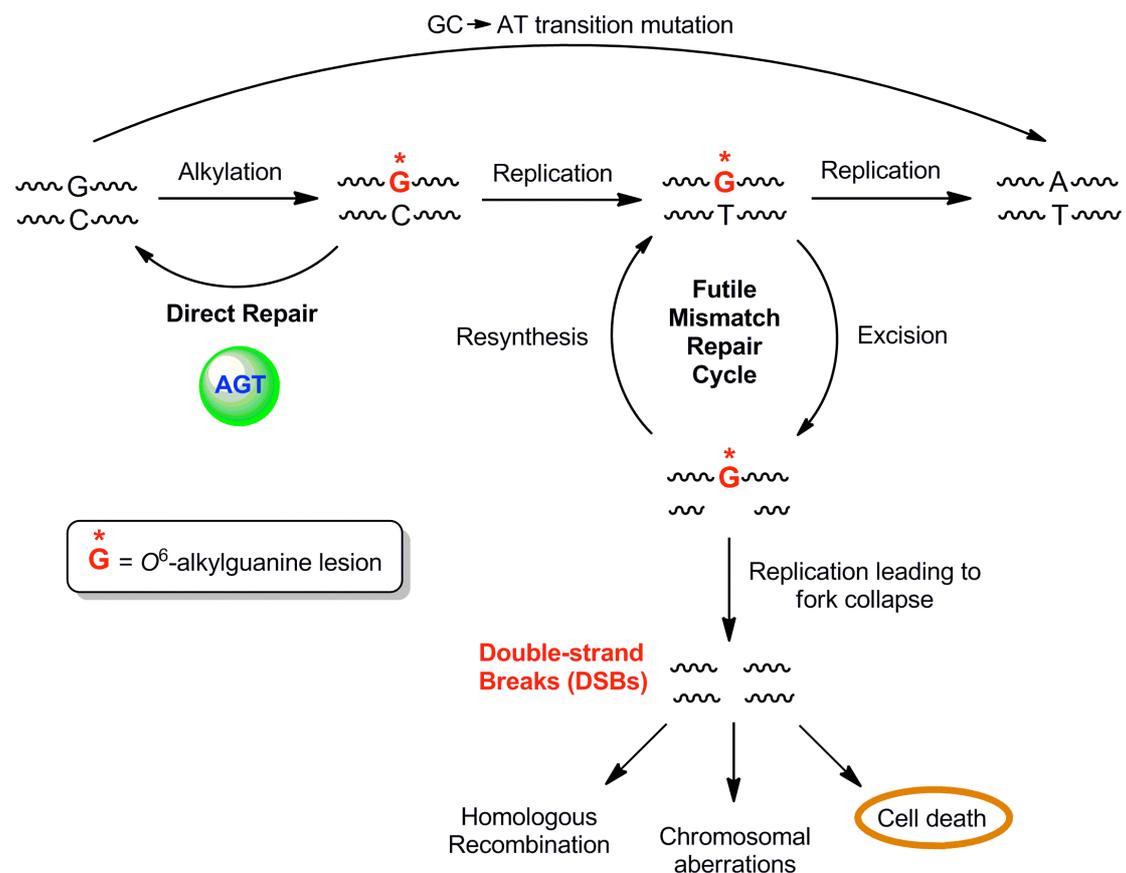


Figure 1.12: Toxic and mutagenic effects of O^6 -alkylguanine residues in DNA

excision and DNA resynthesis, a process referred to as the futile repair pathway.(10,18,19) In fact, the requirement of the MMR system for the toxicity of these lesions has been demonstrated in *S.cerevisiae*, where inactivating MMR genes reduced toxicity of the cells to MNNU (an agent which induces O⁶-methylguanine lesions in DNA).(20)

1.6.4 Base-Excision Repair (BER)

The base-excision repair (BER) pathway mainly repairs base damage caused by deamination, oxidation (such as 8-oxoguanine) and alkylation (such as N¹-methyladenine and N³-methylcytosine).(21) There are two types of BER: short patch, where just one nucleotide is removed during repair, and long patch, where a single-stranded section of between 2-13 nucleotides is removed. The process is initiated by a DNA glycosylase, which recognises the damaged base and then cuts the N-glycosidic bond (between the base and the sugar-phosphate backbone) to generate an apurinic/apyridinimic (AP) site. The AP site, which is itself a form of DNA damage that can be generated spontaneously, is then cleaved by an AP endonuclease to leave a gap of varying length. For short patch repair, the missing nucleotide is filled in by DNA polymerase β , which also removes the 5'-moiety so that the remaining nick can be sealed by DNA ligase. For long patch repair, removal of the strand containing the AP site generates a flap structure, which must be cleaved by the endonuclease FEN-1 before ligation.

The recognition factors in BER, DNA glycosylases, are a large and disparate family of proteins that, although structurally diverse have a single functional role: binding certain damaged bases and cleaving the N-glycosidic

bond to create an AP site and thus initiate BER. In common with AGTs, they flip damaged bases out of the helical stack into their active site in order to bring the base in close enough proximity to the residue/s that perform the cleavage reaction (transfer of the alkyl group for AGTs, glycosidic bond cleavage for DNA glycosylases). Examples of DNA glycosylases include AlkA in *E.coli* which repairs a wide variety of nucleobase alkylation products (N^7 -MeG, N^7 -MeA, N^3 -MeG, N^3 -MeA, O^2 -MeC and O^2 -MeT lesions),(22) AlkD in *B.cereus* which repairs N^7 -MeA and N^3 -MeG (23) and hoGG1 in humans which repairs 8-oxoG residues.(24)

1.6.5 Nucleotide-Excision Repair (NER)

It is the responsibility of the NER system to deal with bulky, helix distorting types of DNA damage, such as cyclobutane pyrimidine dimers (CPD) and 6,4-photoproducts (6,4-PP) that are induced by exposure to UV radiation. There are two partially overlapping pathways of NER that remove UV-induced photolesions: global-genome repair (GGR) and transcription-coupled repair (TCR). In GGR, enzyme complexes constantly scan the genome looking for signs of damage, whereas in TCR recognition is instigated by the stalling of RNA polymerase during transcription on reaching a bulky lesion.(25-27)

NER in *E.coli* is carried out by the UvrABC endonuclease enzyme complex. Briefly, the UvrA-UvrB complex scans the DNA and recognises any distortions, after which UvrA 'loads' UvrB onto the damaged site and subsequently dissociates from the complex, processes that are regulated by ATP.(28) UvrB remains bound to the damage and recruits the endonuclease

UvrC which cleaves the phosphodiester backbone either side of the damage.(29,30) Once incision has taken place UvrD removes both the damage-containing oligonucleotide patch and UvrC, whilst UvrB remains attached to the non-damaged strand and recruits DNA polymerase I to synthesise a new complementary strand. Repair is completed when DNA ligase seals the nick, restoring the DNA to its original sequence.(31) TCR in *E.coli* is identical except that the initial step is the recognition of stalled RNA polymerase III by the Transcription-Repair Coupling factor Mfd, which displaces the polymerase and recruits UvrA.(32)

The situation in eukaryotic organisms is rather more complex, although the general features of repair are maintained. Figure 1.13 (Hoejmakers *et al.* (33)) shows the complexity of repair in human cells. This figure shows the human NER machinery with the homologous proteins in *S.pombe* being labelled in pink. As there is convincing evidence to suggest a link between ATL proteins and the NER system (and part of this project will be to investigate this in *S.pombe*) it is useful to be aware of them. Unfortunately, NER has not been well characterised in *S.pombe*, although it has been extensively studied in *S.cerevisiae* and humans. NER is of fundamental importance to organisms, as evidenced by the three known human syndromes that arise from a non or partially functional NER pathway: Cockayne's syndrome (CS), Trichothiodystrophy (TTD) and Xeroderma Pigmentosum (XP), the symptoms of which include hypersensitivity to UV light and increased cancer risk.(33)

In eukaryotes, GGR and TCR differ in their initial recognition steps before proceeding along the same repair mechanism. In GGR in humans, the

UV-DDB complex binds to damaged DNA and induces repair by recruiting the XPC-hHR23b complex (Rad4-Rad23 in *S.cerevisiae*). It has been suggested that UV-DDB recognises structural abnormalities in the DNA induced by photolesions rather than the damage itself, acting as a sensor for conformational changes in DNA.(34) This could explain how, in addition to CPD and 6,4-PP lesions, UV-DDB can recognise apurinic sites and 2-3 base pair mismatches.(34) It has also been shown that large enough helical distortions (e.g. 6,4-PP) can be recognised by the XPC complex without the

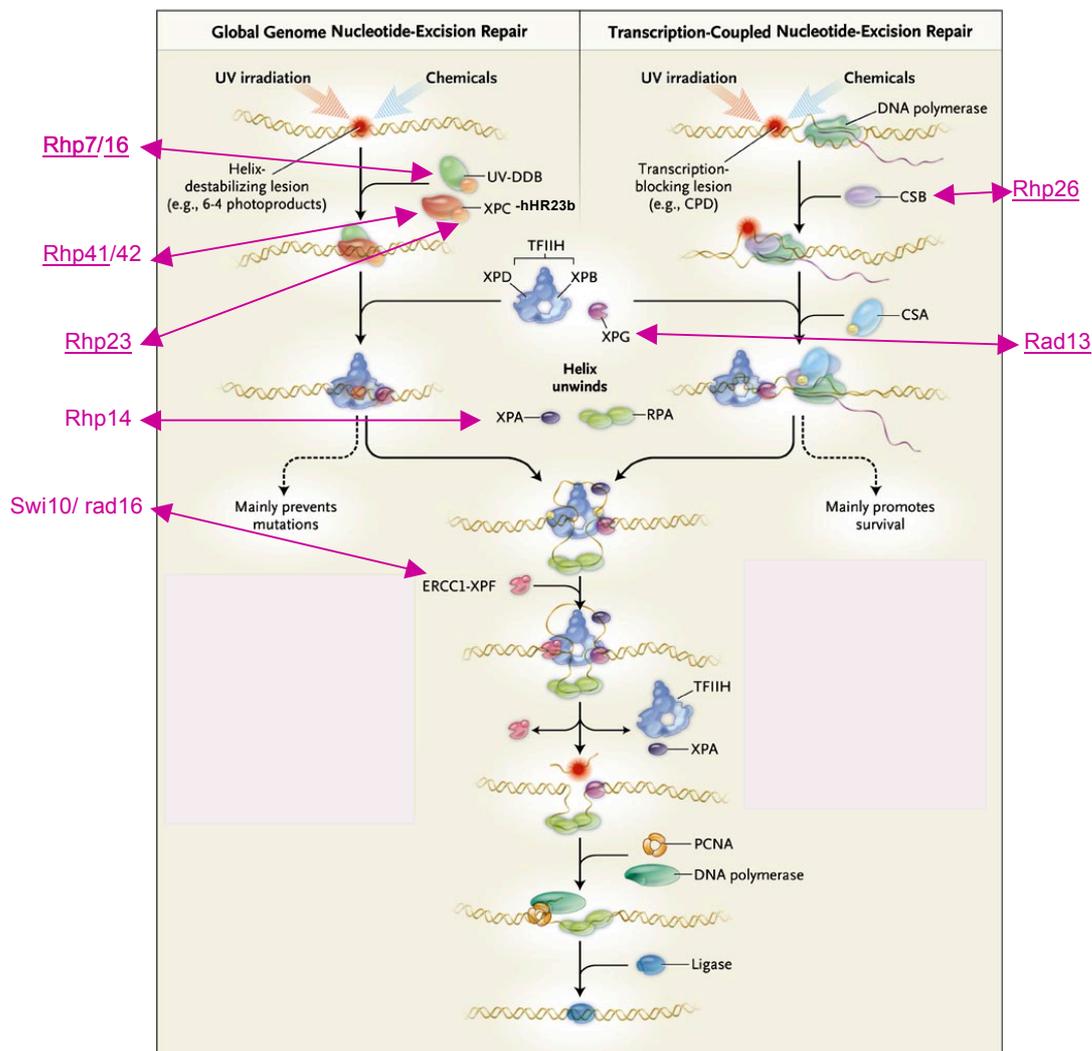


Figure 1.13: NER repair in humans (with *S.pombe* protein homologues shown in pink), taken from Hoeijmakers *et al.* (33)

involvement of UV-DDB, (35) although the UV-DDB complex is thought to mediate efficient targeting of the XPC complex and other NER factors to sites of DNA damage, and to accelerate repair.(36,37) Recent structural studies of Rad4 (the XPC orthologue in *S.cerevisiae*) show that it binds to CPD-containing DNA via the undamaged strand opposite the lesion, and uses a β -hairpin to flip the damaged bases out of the helix.(38)

In TCR, the first step is recognition of the stalled RNA polymerase by the protein CSB. CSB (functional homologues Rad26 in *S.cerevisiae* and Rhp26 in *S.pombe*) has been shown to transiently associate with RNA polymerase III whilst transcription takes place.(39) However, when RNA Pol III reaches a bulky lesion and stalls, CSB increases its association (40) and recruits CSA and the 9 sub-unit complex Transcription Factor II H (TFIIH).(41) It is not fully understood what happens to RNA Pol III after binding of TFIIH and CSA: whilst some studies have suggested that it is degraded before repair continues,(42) other evidence would indicate that the TCR process can continue without the removal of RNA Pol III.(43,44)

Once these damage recognition steps have taken place in GGR and TCR, the helicases XPB and XPD (which are part of the TFIIH complex) unwind the helix in either direction adjacent to the damage. XPA is then recruited which binds to the lesion and keeps the DNA unwound, along with RPA which binds to single-stranded DNA on the damaged and undamaged strands. Further recruitment of 3'-endonuclease XPG causes an incision to be made 2-8 nucleotides away from the lesion, after which the 5'-endonuclease XPF-ERCC1 cuts the DNA 15-24 nucleotides 3' of the lesion.(26) This leads to excision of a patch of 24-32 nucleotides including the lesion. Subsequent

recruitment of PCNA by XPA and RPA targets DNA polymerase and allows synthesis of the new fragment to take place, after which DNA ligase completes the repair by sealing the nick.(45) This restores the DNA to its undamaged, original sequence and maintains the integrity of the genome.

1.7 O⁶-Methylguanine-DNA Methyltransferase (MGMT)

O⁶-Methylguanine-DNA methyltransferase (MGMT) is a protein that removes alkyl groups from the O⁶-position of guanine bases by direct transfer.(46) It is the human form of a representative family of direct transfer DNA repair proteins, the alkylguanine transferases or AGTs, which are ubiquitous across all three domains of life (Prokaryotes, Eukaryotes and Archaea). Despite having markedly low primary sequence homology, crystal structures show that the topology and overall structure of alkylguanine transferases has been remarkably conserved, including a helix-turn-helix (HTH) motif involved in DNA binding coupled by an asparagine-hinge to a cysteine-containing active site.(47) AGT proteins irreversibly transfer the O⁶-alkyl group from the guanine base to an active site cysteine residue in a stoichiometric reaction. Once this transfer has taken place, the alkylated protein is rendered inactive and is subsequently destroyed by proteolysis in the cell. The process is therefore non-enzymatic and AGTs are thus sometimes referred to as suicide proteins.

E.coli has two AGTs, the constitutive Ogt protein and the inducible Ada protein.(48) They differ in their substrate specificities, in that O⁶-MeG is repaired most efficiently by the C-terminal domain of Ada, and O⁴-MeT and

larger O^6 -alkylguanine adducts by Ogt.(49) The N-terminal domain of Ada is also able to repair methylphosphotriester residues.(50)

The mechanism of alkyl transfer by MGMT from O^6 -alkylguanine, based on data from the Tainer group, is shown in figure 1.14.(47) It is thought that an H-bond network increases the reactivity of the cysteine residue, with His146 acting as a general base which deprotonates a water molecule and in turn Cys145. This Cys145 then acts as a nucleophile, removing the alkyl group from the O^6 -position of guanine. It could be that the reaction is also promoted by an H-bond from Tyr114 to N3 of guanine, which may reduce the negative charge on the repaired guanine base.(51)

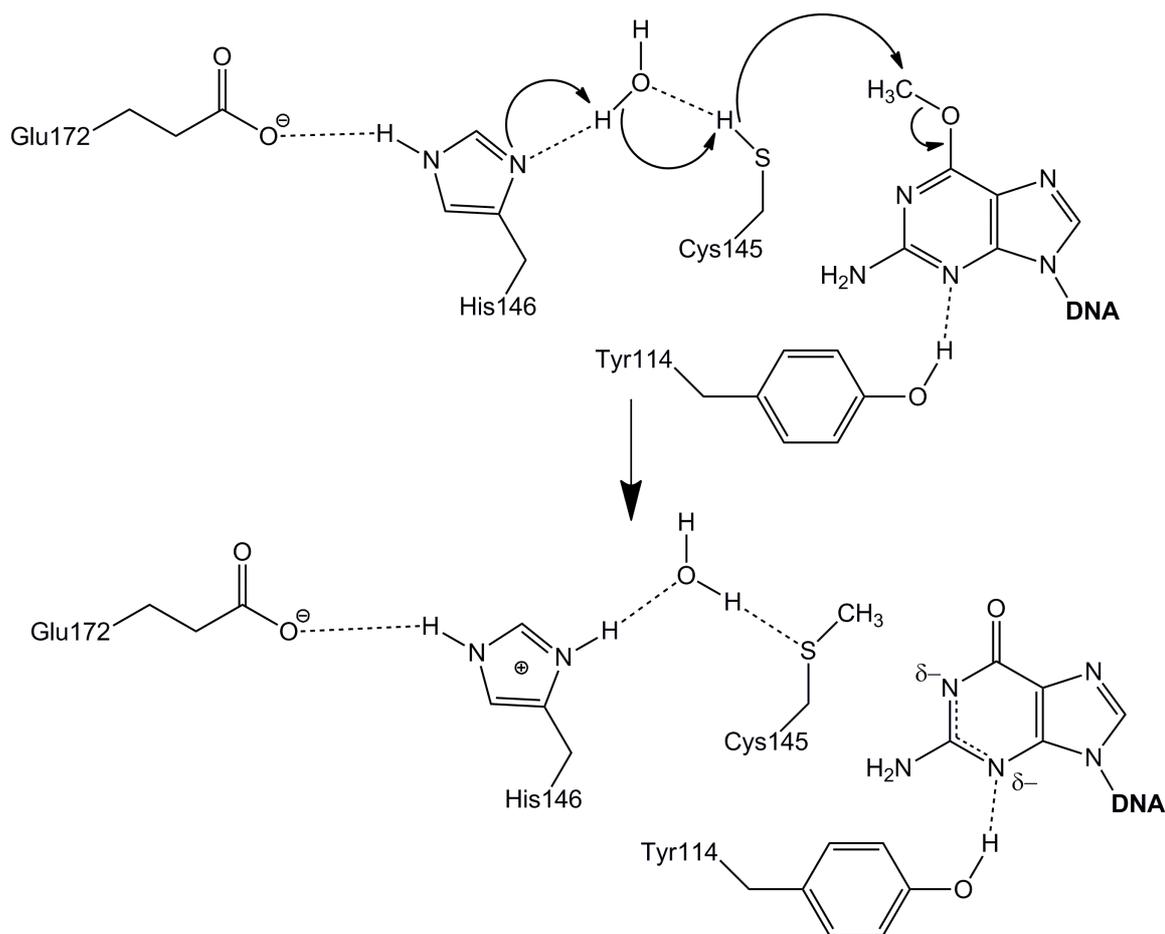


Figure 1.14: Mechanism of alkyl transfer in the MGMT active site

The crystal structure of the MGMT C145S mutant bound to its DNA substrate, published by Daniels *et al.* (figure 1.15), gives profound insights into the structure and function of this protein.(51) The DNA-binding helix-turn-helix motif is located on the C-terminal domain of the two domain α/β -fold, adjacent to which is the active site of sequence PCHR (containing the cysteine nucleophile). The helix-turn helix motif is itself composed of a first helix (Tyr114-Ala121), which interacts with the phosphate backbone, and a second helix (Ala127-Gly136), known as the 'recognition' helix, which binds deep within the minor groove of the DNA. This is unprecedented as previously HTH motifs have only been observed to bind through the major groove.(52) The binding of the recognition helix widens the minor groove by $>3\text{\AA}$ relative

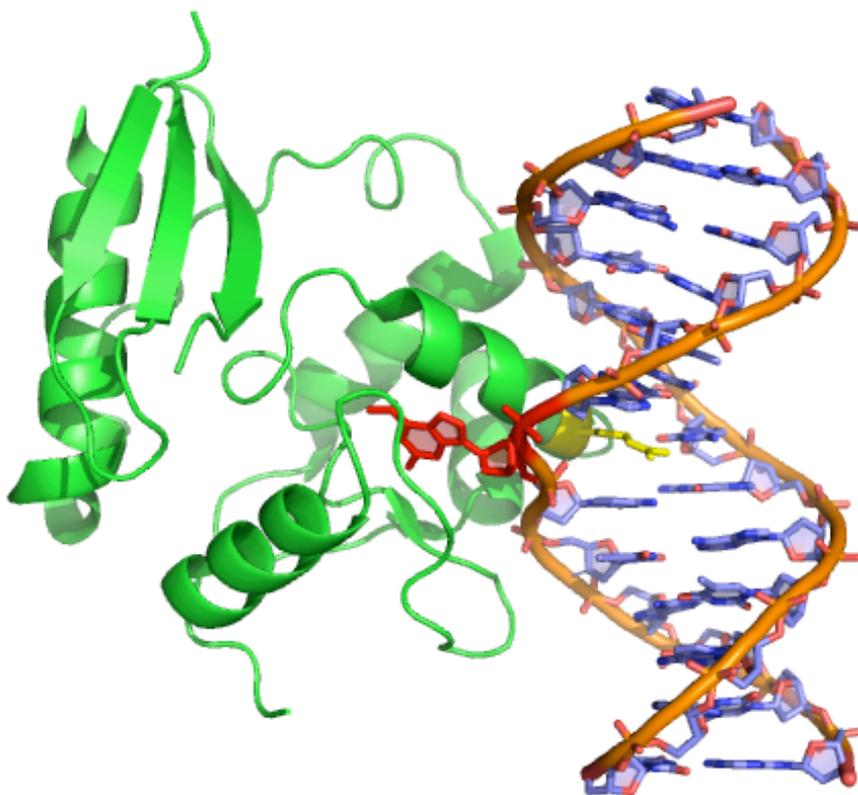


Figure 1.15: Crystal structure of MGMT C145S bound to O⁶-methylguanine containing DNA

to typical B-DNA and also bends the DNA $\sim 15^\circ$ away from the protein. In order to repair the damaged guanine, it is necessary for the protein to flip the nucleotide out of the base stack. The recognition helix also contains Arg128 (the so-called 'arginine finger') which intercalates via the widened minor groove and is thought to push out the nucleotide whilst also stabilising the extra-helical DNA conformation by forming a charged H-bond with the remaining unpaired cytosine. In addition, the Tyr114 residue induces rotation of the 3'-phosphate into the centre of the DNA helix by both charge and steric repulsions.(51) In the crystal structure shown above (figure 1.15) the results of this base-flipping mechanism can clearly be seen, along with the arginine finger protruding into the DNA double helix.

The other domain of the protein, the N-terminal domain, consists of two helices and a three-stranded anti-parallel β -sheet separated by a disordered loop. The human protein (MGMT) also has a zinc ion binding site which is absent in the bacterial and hyperthermophilic proteins.(47) The role of the N-terminal domain appears to be structural in the sense that it maintains the C-terminal domain in an active configuration. Thus, truncation of the protein to leave only the C-terminal domain results in a loss of alkyltransferase activity.(53) In addition, it has been found that substitution of the cysteine residue in the active site also inactivates the protein, as would be expected given the mechanism of repair.(54)

The repair of O^6 -alkylguanine lesions by MGMT is crucial for the maintenance of healthy cells in humans. However, this type of DNA damage is also deliberately induced by the administration of alkylating agents during cancer chemotherapy. The therapeutic aim is to cause significant damage to

the tumour cells, preferentially over healthy cells, and hence destroy the cancer. MGMT reverses the clinical benefits of these alkylating agents, and as such increased expression of MGMT can confer resistance of tumours to chemotherapy.(14) The selective inhibition of MGMT in tumour cells might therefore improve the effectiveness of alkylative anti-cancer treatments, and indeed there is much interest in developing agents which inhibit or inactivate MGMT.(55)

Although AGTs have been found to occur in organisms in all three domains of life there are some notable absences. No AGT genes have been identified in plants, *Schizosaccharomyces pombe* or *Deinococcus radiodurans* amongst others.(53) This fact is all the more intriguing given that distinct homologues of AGT, namely alkyltransferase-like (ATL) proteins, have been found in several organisms (including *E.coli* and *S.pombe*). These proteins possess a similar active site to that of AGT proteins but the sequence PWHR (and occasionally PAHR) is found, rather than the conserved PCHR in AGTs. (14,56) Some organisms such as *E.coli* have genes for both AGT and ATL, some just for AGT and others only for ATL.(14)

1.8 Alkyltransferase-Like (ATL) proteins

Alkyltransferase-like (ATL) proteins were first described by Margison *et al.* in a manuscript reviewing the known genes in different organisms that encode alkyltransferases (AGTs).(14) *In silico* analysis of genetic databases indicated that many prokaryotes and lower eukaryotes would have this homologue (figure 1.16). *E.coli* has two genes encoding AGTs (*ogt* and *ada*)

```

Homo_sapiens-AGT VKFGEVTSQQLAALAGH-----PKAFVAVGGAMRQNPV-----ILIPCHRVVCCSSGAVGVNS-----GG--LAVKEWLLAHEGHRKPGKGGSSGLAGAWLKAGATSG
Schizosaccharomyces_pombe-ATL IPYKQVSTGEIARYVGMPS-----YAQVQVQAKMLHPET-----HVPWHRVINSRGTISKR-----DISAGEQRQKDRLEEGVEIYQTSLG-----EYKLNLPYMKWP
Escherichia_coli-ATL IPEGVTVYGDVAKLAGSPR-----ARQVQGVLRKLPPEG-----STLPWHRVVRNRHGTISLT-----GPDLQRQKQALLAEGVQVSGSG-----QIDLQRVRYWNY
Deinococcus_radiodurans-ATL P-YGRVTSYQALGRELG-----LSEBAVGAALRACRPPF-----LLVPAHRVIHADGRLLGGFQ-----GQE--GLKWLMLFEGAL-----
Nematostella_vectensis-ATL IPYGRVTSYGAIAKYLGAAR-----SARVWGVAMNNSHQ-----SEEVPAHRVVRNRKGLLTGKH-----HFSGTNLMQQLLESSEKRVVDN-----QIQDFENFVWNP
Nanoarchaeum_equitans-ATL IPKGYVTVKELAKALGD-----EATKTFVAMYKKAP-----HWYRVVSSNLIVSPM-----QRALLEKE-----
Ca._Korarchaeum_cryptofilum-ATL VPYGEVTVKTIABALGDGR-----ALAVRNECHRLSKMDPD-----IPWHRVVSCKLEIPG-----AEKLEREG-----
Deinococcus_radiodurans-ATL IPPGRVTVYQALLAGQPG-----ARQVQGVLRKLPPEG-----D-IPWQRVINAQGRVSTY-----KVGLGEVQEGLLRAEGIEFFDAG-----RCDSLRYQWNP
Vibrio_parahaemolyticus-ATL IPKGRVTSYGAIAKMGVYFG-----YARVQKALGNLPEG-----SKLPWFRVINSQKISLQ-----GRDLDRQKQKLEAEGIEVSEIG-----KIALRKYQWP
Vibrio_cholerae-ATL IVPGRVTSYGAIAKMGVYFG-----YARVQKALGNLPEG-----SQLPWFRVINSQKISLQ-----GEDFVRQKQALLAEGIEVSDTG-----KISLRKYQWP
Salmonella_enterica-ATL IPEGVTVYGDVARLAGSPQ-----AARVQVGVLRKLPPEG-----STLPWHRVVRNRHGTISLT-----GPDLQRQKQALLAEGVQVSGSG-----QIDLQRVRYWNY
Pseudomonas_aeruginosa-ATL VPPGRVTSYQALAEHAGLGR-----AARVQVGRTLGQLPADTRL-----PWHRVLAGGGRISLPA-----DSPGGREQRTRLEEGVLLHNRVD-----IRRHWLT
Pseudomonas_fluorescens-ATL VPEGVTVYQALAEHAGLGR-----AARVQVGRTLGQLPADTRL-----PWHRVLAGGGRISLPA-----GSPGDEQRARLRAGEVSI LNNRVD-----IQRHWLP
Pseudomonas_putida-ATL VPAGKVVSYQALAEHAGLGR-----AARVQVGRTLGQLPADTRL-----PWHRVLAGGGRISLPA-----GTPSGDEQRARLRAGEVSI LNNRVD-----MTRHWRP
Synechococcus_elongatus-ATL IPEGVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Thermosynechococcus_elongatus-ATL IPIGRVTVYQALAEHAGLGR-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Prochlorococcus_marinus-ATL IPPGQVTVYQALAEHAGLGR-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Streptomyces_coelicolor-ATL IPPGRVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Neurospora_crassa-ATL IPHGKVTSYGHIKLVGTPQ-----REQVGVCLKHLIPDSARFNHNVVWQRVINAKGVI SPR-----SQPEGSRQATALEAEGIVTGTALG-----ELMVDFAEYGFPP
Gibberella_zeae-ATL IPHGKVTSYGHIKLVGTPQ-----REQVGVCLKHLIPDSARFNHNVVWQRVINAKGVI SPR-----SQPEGSRQATALEAEGIVTGTALG-----ELMVDFAEYGFPP
Chaetomium_globosum-ATL IPPGRVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Flavobacterium_johnsoniae-ATL IPYKQVSTGEIARYVGMPS-----YAQVQVQAKMLHPET-----HVPWHRVINSRGTISKR-----DISAGEQRQKDRLEEGVEIYQTSLG-----EYKLNLPYMKWP
Flavobacterium_psychrophilum-ATL IPCGKVTSYGAIARVLTAR-----SARVWGVAMNNSHQ-----SEEVPAHRVVRNRKGLLTGKH-----HFSGTNLMQQLLESSEKRVVDN-----QIQDFEKHFQWP
Chromobacterium_violaceum-ATL VPPGRVTVYQALAEHAGLGR-----AARVQVGRTLGQLPADTRL-----PWHRVLAGGGRISLPA-----GSPGDEQRARLRAGEVSI LNNRVD-----IQRHWLP
Stigmatella_aurantiaca-ATL VPSGKVTSYGHIKLVGTPQ-----REQVGVCLKHLIPDSARFNHNVVWQRVINAKGVI SPR-----SQPEGSRQATALEAEGIVTGTALG-----ELMVDFAEYGFPP
Chromohalobacter_salexigenus-ATL IPEGVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Citrobacter_koseri-ATL IPEGVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Klebsiella_pneumoniae-ATL IPEGVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Enterobacter_sakazakii-ATL IPEGVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Cytophaga_hutchinsonii-ATL IPRGRVTSYGAIAKYLGMKS-----SARVWGVAMNNSHQ-----SEEVPAHRVVRNRKGLLTGKH-----HFSGTNLMQQLLESSEKRVVDN-----QIQDFEKHFQWP
Yersinia_enterocolitica-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Yersinia_pestis-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Photobacterium_profundum-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Bacillus_clausii-ATL IPRGEVTSYQALAEHAGLGR-----AARVQVGRTLGQLPADTRL-----PWHRVLAGGGRISLPA-----GSPGDEQRARLRAGEVSI LNNRVD-----IQRHWLP
Desulfovibrio_desulfuricans-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Pseudoalteromonas_tunicata-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Natronomonas_pharaonis-ATL VPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Frankia_sp._Cc13-ATL IPPGRVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Frankia_sp._Eul1c-ATL LPAGRVTYGDVAQVAVT-----KAPLQGHQIATCPSC-----QNRVRLDADGTVAADFKWP-----DPSRADTPQVLESEGITFTNGTADP-----SVRLSLTDVAN
Haloarcula_marismortui-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Halobacterium_sp.-ATL VPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Anabaena_variabilis-ATL IPAGKVTYGHILEALGVDRS-----YIAPITPKKASDD-----NYPVHRILDSQVLYKYS-----PNQKNLSESG-IINNLVSSNSFKYFYDYSKYLYQD
Nostoc_punctiforme-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Xanthomonas_oryzae-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Myxococcus_xanthus-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Azotobacter_vinelandii-ATL VPPGVVSYQALAEHAGLGR-----AARVQVGRTLGQLPADTRL-----PWHRVLAGGGRISLPA-----GSPGDEQRARLRAGEVSI LNNRVD-----IQRHWLP
Lactococcus_lactis-ATL LKEGVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Magnaporthe_grisea-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Chloroflexus_aurantiacus-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP
Blastopirellula_marina-ATL IPIGQVTVYGDVARLAGSPQ-----AARVQVGVLYRVAPGL-----TDSIPWHRVVRNRHGTISLT-----RHGSDYL-QRHLLEAEGIEFFDAG-----RISLHRYWNP

```

Figure 1.16: *In silico* analysis of genetic databases discovered a number of ATL protein sequences in many different organisms

and one encoding an ATL product. It was subsequently found that the ATL protein in *E.coli* (eAtl, though also known as Ybaz), which has a tryptophan replacing the cysteine in the putative active site, has no alkyltransferase activity despite binding to substrate DNA containing O^6 -methylguanine (O^6 -MeG).(56) Site-directed mutagenesis of the active site tryptophan into a cysteine residue (W83C) does not restore the ability to transfer the methyl

group from O^6 -MeG, as perhaps would be expected given that this mutation appears to return the active site motif (PCHR) to the same sequence as that of the AGT proteins (*Ogt*, *Ada* and human MGMT). The W83A mutant also has no alkyltransferase activity but is still observed to bind substrate DNA containing O^6 -MeG. It was also shown that eAtl has no glycosylase or endonuclease activity, suggesting that this protein has no direct role in the classical base-excision repair (BER) or nucleotide-excision repair (NER) pathways. In addition, it was shown by electrophoretic mobility shift assay (EMSA) that eATL binds to short single- and double-stranded oligodeoxyribonucleotides (ODNs) containing O^6 -MeG, but not other types of damaged DNA (i.e. DNA containing O^4 -methylthymine, 8-oxoguanine, ethenoadenine or 5-hydroxycytosine). ATL was also shown to strongly, but reversibly, inhibit the transfer of methyl groups to MGMT which indicates that these two proteins are in competition for the binding of O^6 -MeG in substrate DNA.(56)

The analogous ATL protein in *Schizosaccharomyces pombe* (*S.pombe*), known as Atl1, also has a tryptophan residue in its active site (W56) and, like eAtl, has no known alkyltransferase, glycosylase or endonuclease activity.(57) Atl1 was shown to bind tightly to single-stranded DNA substrates containing a number of different O^6 -alkylguanine lesions (figure 1.17), such as O^6 -methylguanine (O^6 -MeG), O^6 -benzylguanine (O^6 -BnG), O^6 -(4-bromophenyl)guanine (O^6 -4-BTG) and O^6 -hydroxyethylguanine (O^6 -HOEtG).(57) Interestingly, the latter adduct is poorly repaired by MGMT. Atl1 is also observed to inhibit the transferase activity of MGMT in the same manner as the *E.coli* protein.(57)

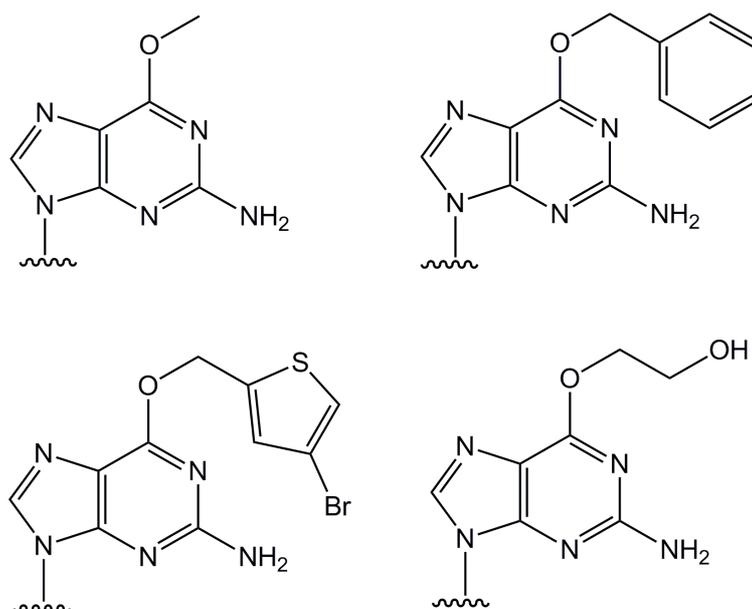


Figure 1.17: Alkylguanine adducts; (left to right, top to bottom) O⁶-MeG, O⁶-BnG, O⁶-4-BTG and O⁶-HOEtG

The *S.pombe* deletant strain $\Delta atl1$ lacks the gene that encodes AtI1 but grows at a similar rate to wild-type, which would indicate that the *atl1* gene is non-essential. In addition, the crude cell extracts from this deletant strain have an inability to bind ODNs containing O⁶-MeG, suggesting that the ability to bind this adduct is due to the presence of AtI1 in the cell. The *S.pombe* $\Delta atl1$ mutants, which also lack any indigenous AGT gene and hence have no alkyltransferase protein, were shown to have a much greater susceptibility to the toxic effects of methylating and other alkylating agents. The results of a cell survival assay are shown in figure 1.18 and clearly demonstrate that cells of the deletant *S.pombe* strain $\Delta atl1$ are much more prone to cell death than wild-type when exposed to increasing amounts of MNU, an S_N1 alkylating agent that induces O⁶-methylguanine lesions.⁽⁵⁷⁾ This would indicate that AtI1 is protecting *S.pombe* cells from the deleterious effects of the O⁶-alkylguanine lesions created by these agents despite its lack of transferase activity. These

observations have led to the suggestion that AtI1 may act by binding O^6 -alkylguanine lesions and signalling them to be processed by other DNA repair pathways.(57)

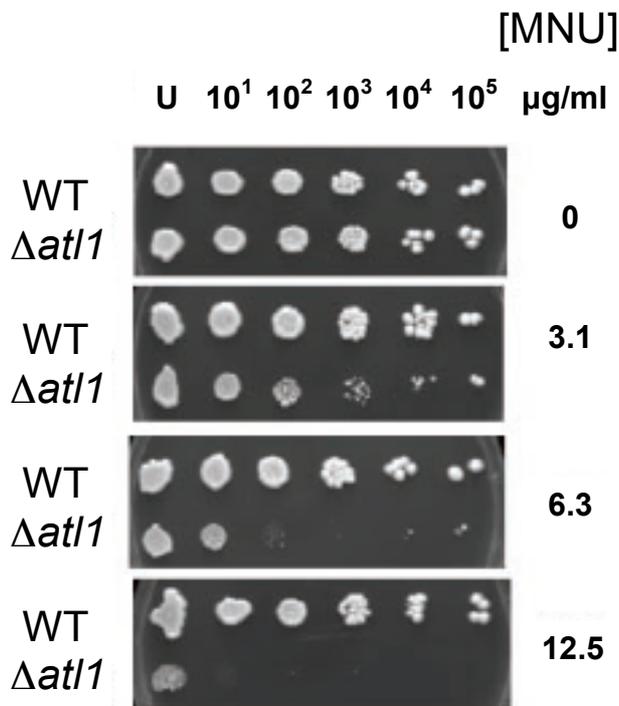


Figure 1.18: Results of survival assays with methylating agent MNU and *S.pombe* cells

1.9 Structure of AtI1 and AtI1-DNA Complex

The publication of crystal structures of AtI1-DNA complexes by Tubbs *et al.* in 2009 have contributed hugely to the understanding of AtI1. Structures were determined for pure AtI1 protein, and in complex with short 13-mer ODN substrates containing O^6 -methylguanine (O^6 -MeG) or O^6 -[4-(3-pyridyl)-4-oxobutyl]guanine (O^6 -PobG) (figure 1.19).(58) O^6 -PobG is a biologically important and toxicologically relevant adduct as it is a product of the nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), a

compound present in tobacco.(59) ODNs containing O^6 -PobG are known to be substrates for MGMT, but are repaired at a much slower rate by the protein than ODNs containing O^6 -MeG.(60) The elucidation of crystal structures has established that At1 shares the MGMT catalytic domain fold (figure 1.20). Almost all residues that are required for alkyltransferase activity and DNA-binding are conserved. Those that promote phosphate rotation and nucleotide flipping of the damaged base, Tyr25 and Arg39 are present. One significant

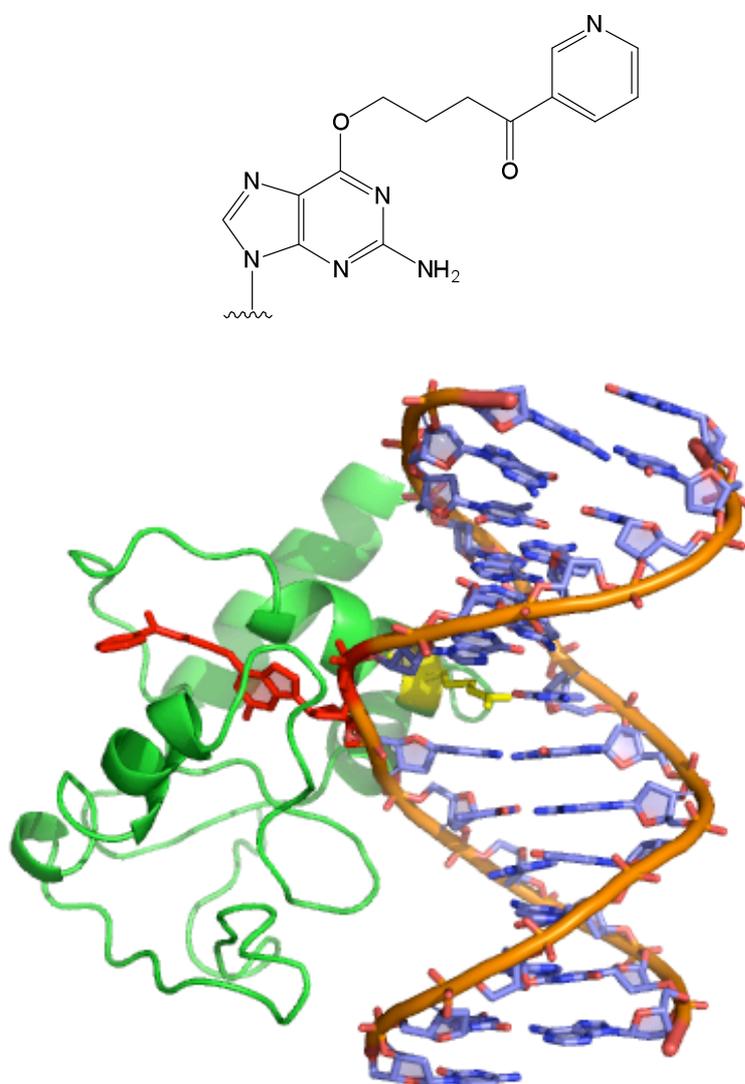


Figure 1.19: Structure of O^6 -4-(3-pyridyl)-4-oxobutylguanine (O^6 -PobG) (above) and At1 in complex with O^6 -PobG-containing DNA substrate (below)

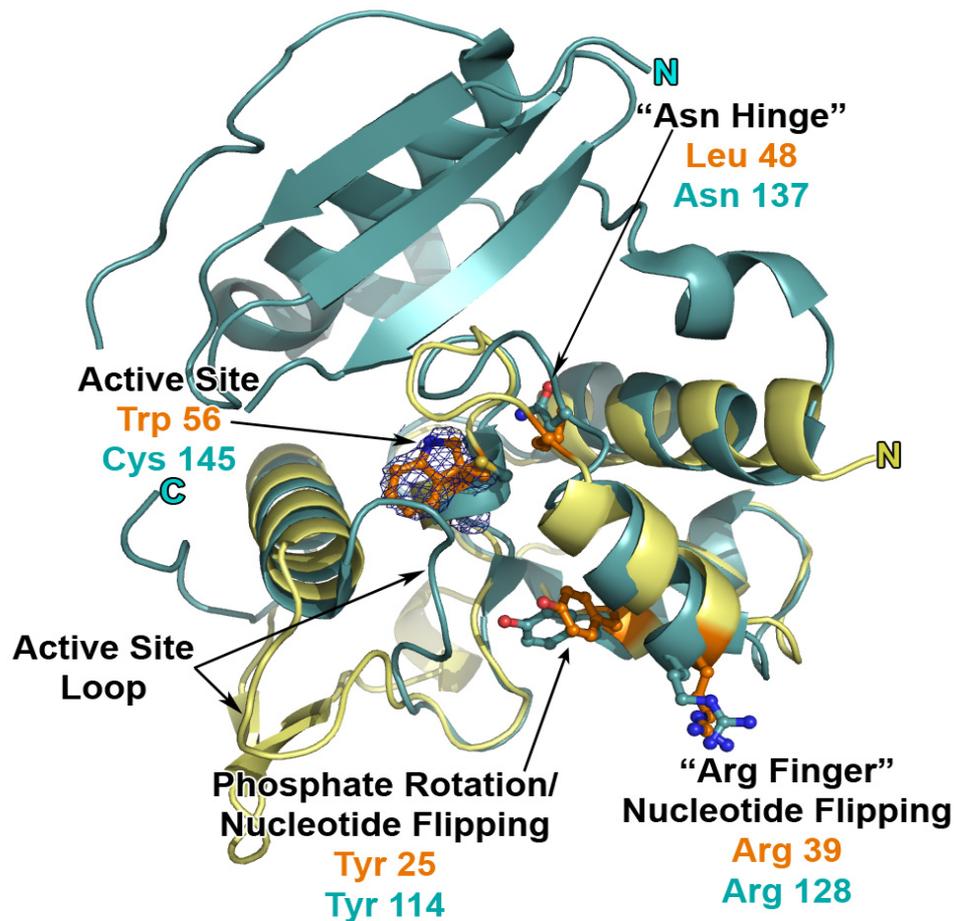


Figure 1.20: Crystal structure overlay of At1 (yellow) and MGMT (cyan) without DNA: the active site and DNA binding motifs are highly conserved.

difference between the structures of MGMT and At1 is that the latter lacks the nucleophilic cysteine residue, which is replaced with tryptophan in the active site fold. The asparagine hinge that joins the helix-turn-helix (HTH) motif to the active site is also not present in At1. In addition, whereas MGMT has a large N-terminal domain consisting of two helices and a three-stranded anti-parallel β -sheet separated by a disordered loop, At1 has a comparatively small N-terminal domain that consists of one helix and as such has a lower molecular

weight.(58) However, the N-terminal domain helix of Atl1 is a few residues longer than the corresponding helix in MGMT.

From the crystal structures of Atl1 with DNA substrate, it can be seen that Atl1 flips the damaged base into the active site in a similar fashion to MGMT. Arg39 intercalates into the DNA base stack via the minor groove, flipping out the base and stabilising the extrahelical conformation by hydrogen bonding with the estranged cytosine, whilst Tyr25 induces the rotation of the phosphate to facilitate base-flipping.(58) In fact, the importance of these residues for binding to O^6 -methylguanine-containing DNA has been shown for another ATL protein, vpATL from *Vibrio parahaemolyticus*, whereby site-directed mutagenesis of the corresponding Tyr23 and Arg37 abolished the ability of the protein to inhibit binding of MGMT to damaged DNA.(61) Rather unexpectedly, SDM of these highly conserved residues (Tyr88 and Arg100) in the *T.thermophilus* ATL (TTHA1564) did not seem to impair the ability of the protein to repair MNNG-induced mutations *in vivo*.(62) However, this result warrants further investigation and explanation, as it was previously shown that TTHA1564 flips O^6 -methylguanine residues upon binding,(63) and these conserved Tyr and Arg residues appear to be extremely important for the nucleotide-flipping mechanism in other ATL proteins. Unfortunately, no structural studies have ever been performed with TTHA1564 so the implications of these results are not clear.

A closer look at the active site pocket highlights some of the important interactions between the protein and the damaged bases (figure 1.21). The tryptophan residue (Trp56) acts in a hydrophobic packing interaction with the alkyl group while the side-chain of the arginine residue (Arg69) stacks against

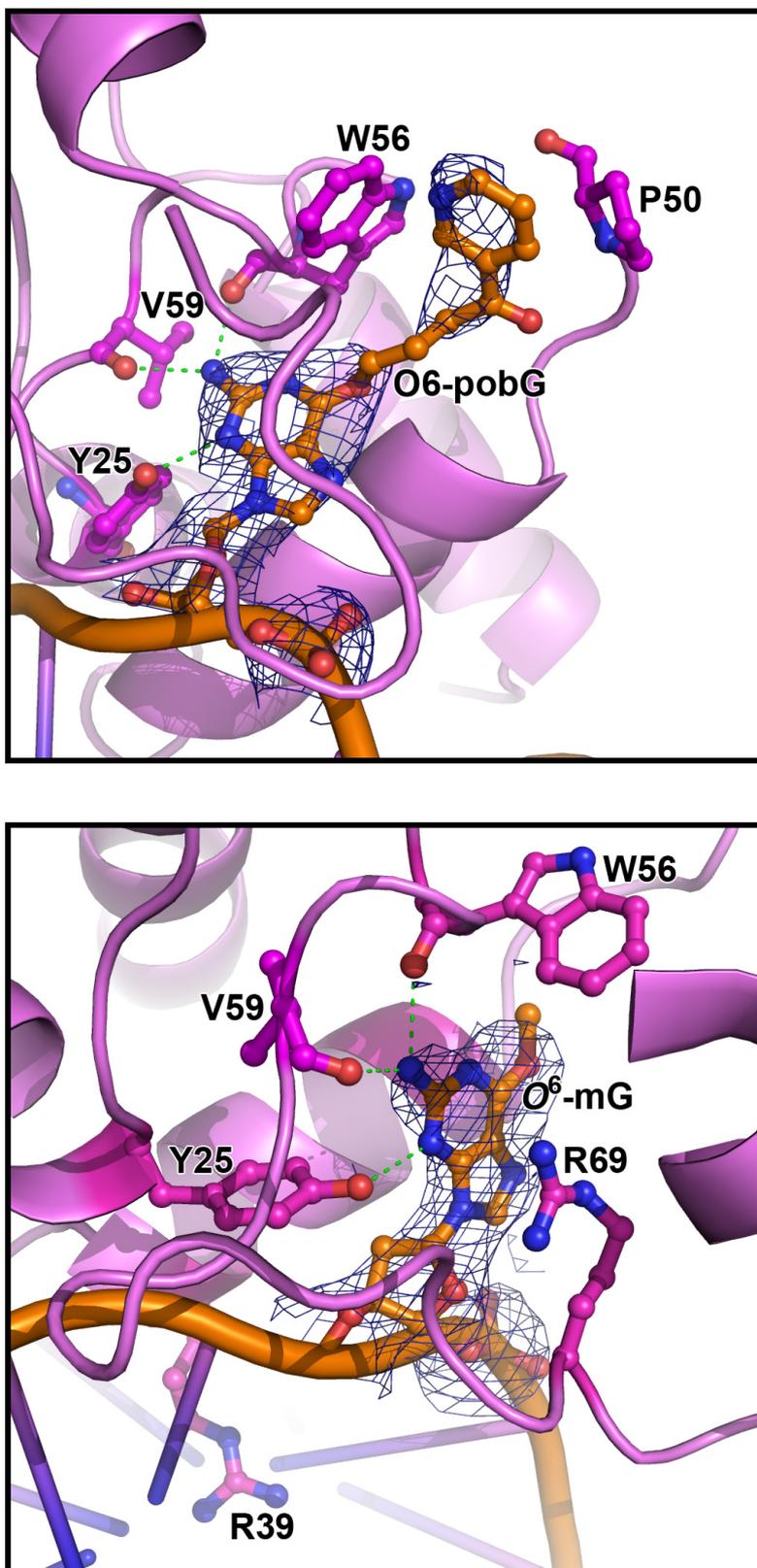


Figure 1.21: Close-up of At1 active site with DNA substrate: containing O^6 -PobG (above) and O^6 -MeG (below)

the guanine base in a cation- π interaction. These interactions are not found in MGMT and so appear to be unique to At11. However, the other active site interactions are conserved from MGMT to At11: the tyrosine (Tyr25) hydroxyl group hydrogen bonds with the alkylguanine N3, whilst the alkylguanine N2 amino group donates hydrogen bonds to the main chain carbonyl oxygen atoms of Trp56 and Val59.(58)

The At11 active site is approximately three times larger than that of MGMT. This is, in part, due to the fact that the residues 65 to 73 in the loop which constitute one wall of the pocket are further from the protein core than the corresponding residues in MGMT. In addition, the proline (Pro50) residue in At11 is further out than the comparable proline (Pro140) residue in MGMT, which is important for interacting with larger alkyl groups. Any potentially unfavourable steric clashes with Trp56 are also reduced by the substitution of tyrosine (Tyr158) of MGMT with the smaller glycine (Gly68) in the active site loop of At11.(58) Furthermore, NMR structural studies of vpATL revealed that this alkyltransferase-like protein has a recognition cavity that is smaller than that of At11 but larger than that of MGMT, and in addition exhibits a significant degree of conformational flexibility that may facilitate the binding of larger O⁶-alkylguanine lesions.(61)

Like MGMT, At11 binds DNA via the minor groove using a helix-turn-helix (HTH) motif. However, for At11 there are some additional contacts with DNA from residues in the active site loop (Ser67 and Lys70) and another loop (Thr92 and Ser93) that are not found in MGMT. It has been calculated that there is a more extensive binding interaction between At11 and DNA than that

of MGMT, with DNA binding accounting for 1050 Å² and 788 Å² of the buried surface area respectively.(58)

Other structural features of the At1-DNA complex provide insights into the proposed role of At1. The helix that composes the At1 N-terminal domain is a few residues longer than the corresponding helix in MGMT and this elongation gives the N-terminus of At1 the ability to push against the phosphate backbone of the strand opposite the flipped nucleotide in the bound DNA.(58) In addition, the active site loop of the protein is able to switch between an 'open' and 'closed' conformation depending on whether DNA is bound. The gating action of the active site loop of At1, which switches between an open or closed conformation depending on the presence of DNA substrate, can be seen in figure 1.22. This mechanism was proposed for MGMT in computer simulations of binding (64) but never seen in crystal structures.(51,65) These structural changes cause the DNA to become bent by a large degree upon At1 binding (approximately 45° in relation to the helical axis of B-DNA) (58) whilst for MGMT the bound DNA is only bent by about 30°. It follows that a greater distortion of DNA by At1 could have the effect of aiding recognition of the damage by a pathway such as NER, which is designed to find bulky, distorted regions of DNA.

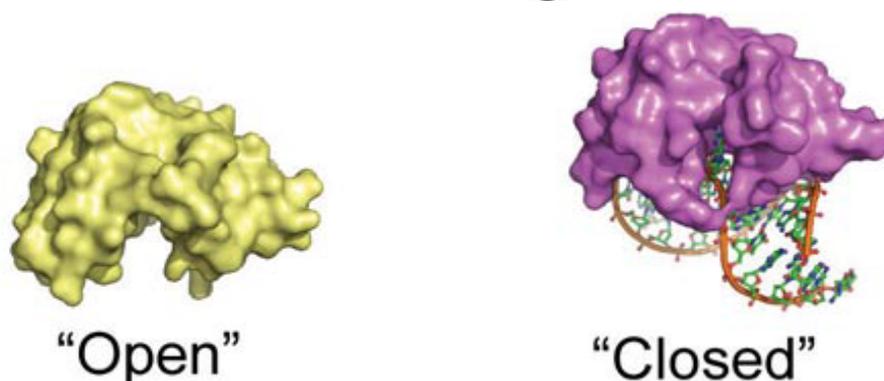


Figure 1.22: The 'gating' action of At1 upon DNA binding

Whilst this conformational switch of the AtI1 active site loop works in concert with the extended N-terminal helix to cause significant bending of the bound DNA, it appears that this movement also exposes the C-terminal loop by moving arginine and isoleucine side-chains that would otherwise be covering it.(58) It is an intriguing possibility that after the switch, the unveiled loop may become free for possible intermolecular interactions and could play a role in signalling to other proteins.

1.10 Substrate Specificity of ATL Proteins

It has been demonstrated in studies with MGMT that as the size of the alkyl group on the O^6 -position of guanine increases, the ability of MGMT to remove it decreases: that is, the repair is less efficient with larger adducts.(66) It was thought that a larger alkyl group creates a bulkier DNA lesion which could be recognised and processed preferentially by the nucleotide-excision repair (NER) system rather than by direct repair. However, some non-bulky lesions are not effectively repaired by MGMT. For example, O^6 -hydroxyethylguanine (O^6 -HOEtG) is a very poor substrate for MGMT (66-68) and O^6 -carboxymethylguanine (O^6 -CMG) is not repaired by MGMT at all (69), although there is evidence that it is repaired by the NER pathway.(70)

In contrast, AtI1 has been shown to bind to oligodeoxyribonucleotides (ODNs) containing a number of different O^6 -alkylguanines.(71) The alkyl groups vary in size, polarity and charge and it seems that AtI1 has the ability to bind a wider range of O^6 -alkylguanine lesions in DNA and hence has less substrate specificity than MGMT. It would appear that the reason for this may be the presence of the enlarged active site cavity which is able to recognise

and accommodate a broad range of adducts. The same protein structure, active site side-chains and DNA conformation are observed for At11 whether binding DNA containing O^6 -MeG or the significantly larger O^6 -PobG.(58) The observation that O^6 -PobG is repaired at a much slower rate by MGMT than O^6 -MeG may be explained by the fact that O^6 -PobG does not fit so well into the MGMT active site due to its larger size. This does not seem to be the case for At11, where it has been shown that O^6 -PobG is tightly bound by the protein.(58)

The substrate specificity of At11 has been examined previously by the Margison group. The binding constants for At11 with DNA substrates containing various adducts were measured by surface plasmon resonance (SPR) and enzyme-linked immunosorbent assay (ELISA). SPR involves immobilising the ODN of interest onto a gold sensor chip and then exposing the reverse of the chip to polarised light from a laser. As protein binds to the ODN, the accumulation of protein on the surface of the chip results in a change in the angle of the reflected light, and this can be measured over the whole time course of the reaction. Data derived from SPR analysis can be used to calculate the association (k_{ass}) and dissociation (k_{diss}) rates of the DNA-protein complex, which can in turn be used to generate a value for the dissociation constant (K_D). Using this technique, it has been determined that At11 binds single-stranded (ssDNA) and double-stranded (dsDNA) containing O^6 -MeG with similar affinity, though with a slight preference for dsDNA (1.8nM and 0.91nM respectively). This is in contrast to MGMT, which has been shown to bind single-stranded DNA with more affinity than double-stranded DNA but to transfer the alkyl group more rapidly from dsDNA than ssDNA.(72)

It was also shown by electrophoretic mobility shift assay (EMSA) that AtI1 binds to double-stranded and hairpin ODNs containing all of the lesions shown in figure 1.23 ((57) and unpublished data). As noted, these substrates vary in the size, polarity and charge of the alkyl group and include O^6 -alkylguanine lesions that are refractory to or poorly repaired by MGMT.

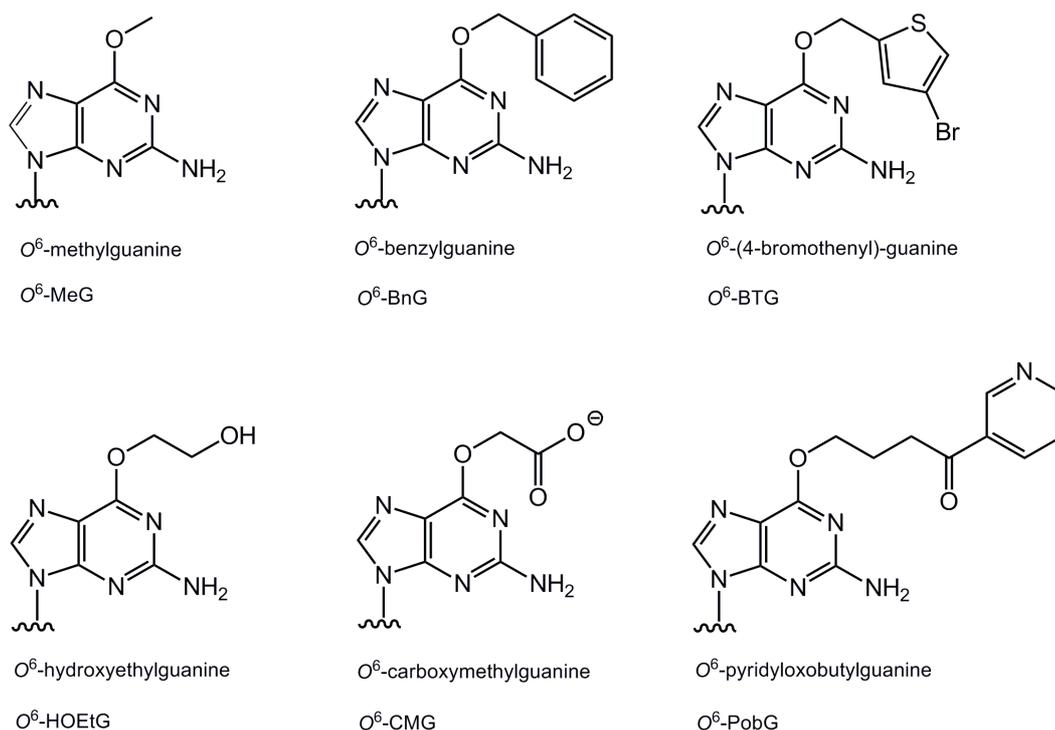


Figure 1.23: Alkyl adducts known to be bound by AtI1 (from *S.pombe*)

The affinity of binding for each adduct has also been quantified using direct ELISA and competition ELISA. The former technique involves immobilising the various adduct-containing ODNs (of identical sequence) on the surface of the wells of a microtiter plate. Incrementally increasing concentrations of AtI1 protein are added to each well, and the amount of bound protein is quantified by measuring the chemiluminescence produced by addition of a reagent that interacts with the antibodies that have bound

specifically to the protein (figure 1.24). The latter method is similar, but this time AtI1 protein is pre-incubated with the ODN of interest, and then added to plates with O^6 -methylguanine-containing ODNs already bound. The binding is quantified in the same manner; this time giving a measure of how much each adduct is inhibiting the binding of AtI1 to the O^6 -MeG on the plate.

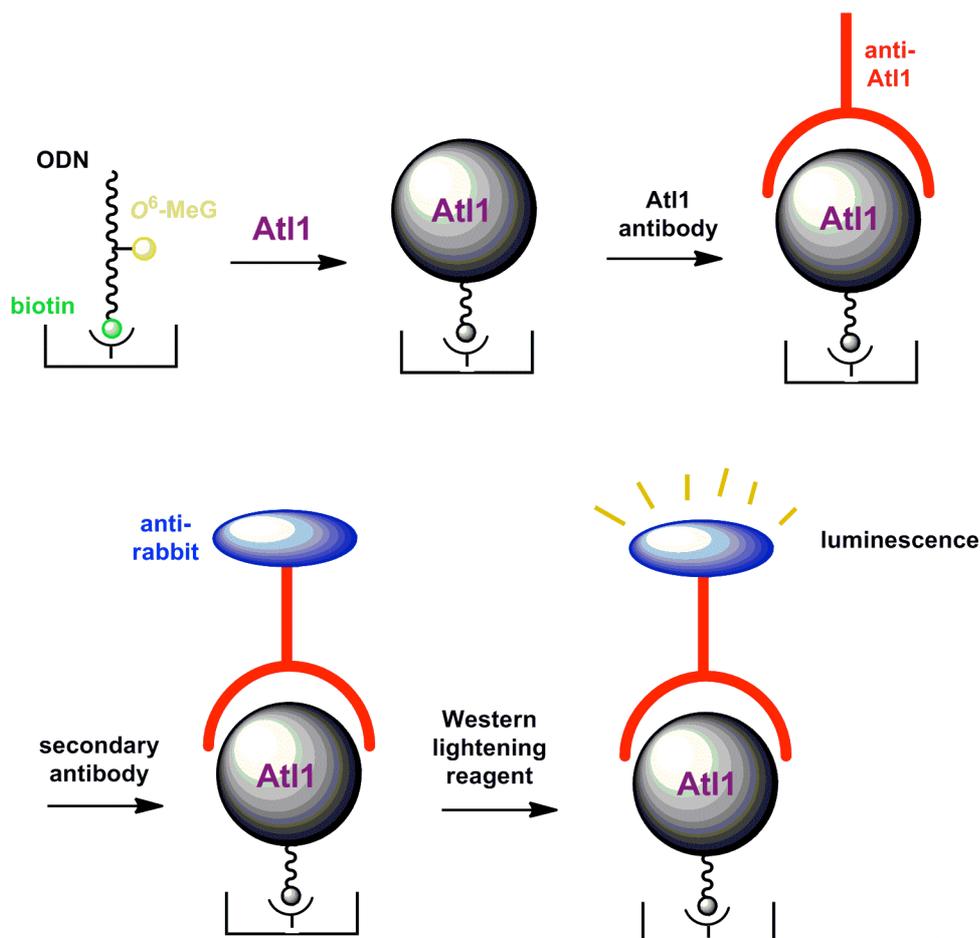


Figure 1.24: Principles of a direct ELISA.

The assays produced largely the same results (table 1.1) and demonstrated that AtI1 binds tightly to ODNs containing a range of O^6 -alkylguanine adducts. Whereas MGMT has difficulty repairing O^6 -PobG (59) and O^6 -HOEtG (66), and will not repair O^6 -CMG (69,70), it is clear from these results that AtI1 recognises these lesions and in fact appears to bind them

with more affinity than smaller lesions such as O^6 -MeG. This lack of substrate discrimination and possible preference for lesions not readily repaired by MGMT, combined with its lack of alkyltransferase activity, indicates that AtI1 may have a role in signalling to other DNA repair pathways. It has been suggested that AtI1 binds a wide range of adducts with high affinity to form bulky, relatively stable AtI1-DNA complexes that are suitable for recognition by nucleotide-excision repair (NER) proteins. In the NER pathway, repair is initiated by the identification of large distortions in DNA caused by bulky adducts. Therefore, the role of AtI1 may be to recognise DNA containing a wide range of O^6 -alkylguanine lesions to form a common, more substantial structure that could be more easily located by the NER apparatus than O^6 -alkylguanine residues alone and hence increase the efficiency of repair.

Assay	Rank Order of Binding
Competition ELISA	O^6 -PobG > O^6 -BnG > O^6 -BTG > O^6 -HOEtG > O^6 -MeG ~ O^6 -CMG
Direct ELISA	O^6 -PobG ~ O^6 -BnG ~ O^6 -BTG ~ O^6 -HOEtG > O^6 -CMG ~ O^6 -MeG

Table 1.1: Binding data for various O^6 -alkylguanine adducts in duplex DNA with AtI1
(Margison laboratory, unpublished data)

eATL, the ATL protein from *E.coli* which is also known as Ybaz (after the gene which encodes it), has been shown to bind double-stranded DNA containing an abasic site.(73) This finding is intriguing as it may indicate a link

between the *E.coli* ATL protein and the base-excision repair (BER) pathway. Furthermore, Ybaz has been found to interact with HelD, a type IV helicase in *E.coli*, which is known to be involved in the repair of methylation-based DNA damage.(74) However, it was recently shown that Atl1 does not recognise abasic sites in DNA and so is unlikely to be involved in BER.(58)

1.11 ATL Proteins and Nucleotide-Excision Repair (NER)

TTHA1564, the ATL homologue from *Thermus thermophilus* TTHB8, possesses a PAHR active site motif rather than the PWHR sequence seen in most ATL proteins. In common with eAtl and Atl1, TTHA1564 binds O⁶-MeG-containing DNA by flipping the damaged nucleotide but has no alkyltransferase activity.(63) In addition, mutation of the active site to the sequence PCHR does not confer alkyltransferase activity to the protein.(62) Deletion of the TTHA1564 gene results in mutants that grow at the same rate as the wild type (WT) strain but display a significantly increased spontaneous mutation rate, which is exacerbated 5-to-8 fold by the presence of methylating agents (MNNG). Thus, ATL proteins clearly seem to have a role in protecting DNA from damage and the deleterious effect this has on cells.(56,57,63) It has been suggested that these proteins, with their ability to bind but not remove lesions in DNA, must be signalling to a known repair pathway. The current hypothesis is that ATL proteins are involved in the NER pathway, and there is an increasing amount of evidence to vindicate this.

Pulldown assays have demonstrated the interaction of TTHA1564 with several proteins, the most significant being with UvrA and RNA polymerase. UvrA is a nucleotide-excision repair (NER) protein, responsible for damage

recognition (along with UvrB). The observation that TTHA1564 interacts with UvrA would seem to provide tentative evidence that ATL proteins are capable of communicating with the NER system and signalling damaged DNA for repair. Interestingly, the additional interaction of TTHA1564 with RNA polymerase may indicate that it is involved in the transcription-coupled repair (TCR) pathway. In TCR, in which repair is initiated by the stalling of RNA polymerase at lesions in DNA, transcription-coupled repair factor (TCRF, or Mfd) acts to remove the stalled RNA polymerase and recruit UvrA.(32) The fact that both TTHA1564 and TCRF are seen to interact with RNA polymerase and UvrA (63,75) may indicate that the ATL protein recruits UvrA to the site of DNA damage and initiates NER in a similar manner to TCRF. Alternatively, TTHA1564 may cause RNA polymerase to stall at the site where it is bound to DNA and thus initiate repair of a lesion that would otherwise be bypassed by the enzyme. Direct interaction of TTHA1564 with purified RNA polymerase holoenzyme has been demonstrated very recently, which would strengthen the argument for a role for TTHA1564 in TCR.(76)

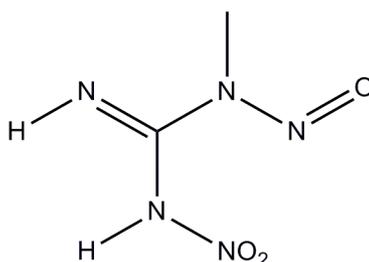


Figure 1.25: 1-Methyl-3-nitro-1-nitrosoguanidine (MNNG), an S_N1 alkylating agent

Further evidence for the involvement of ATL proteins in the NER pathway is provided by experiments with *S.pombe* deletants (figure 1.26).(58)

Inactivation of Atl1 ($\Delta atl1$) causes a nine-fold increase in MNNG-induced mutation rates, with a similar increase being seen for cells missing Rad13 ($\Delta rad13$). Rad13 is the *S.pombe* homologue of mammalian XPG, an NER pathway endonuclease that cleaves at the site 3' of the DNA lesion.(77) MNNG is a S_N1 methylating agent (figure 1.25) known to produce a significant number of O^6 -MeG lesions in DNA.(78)

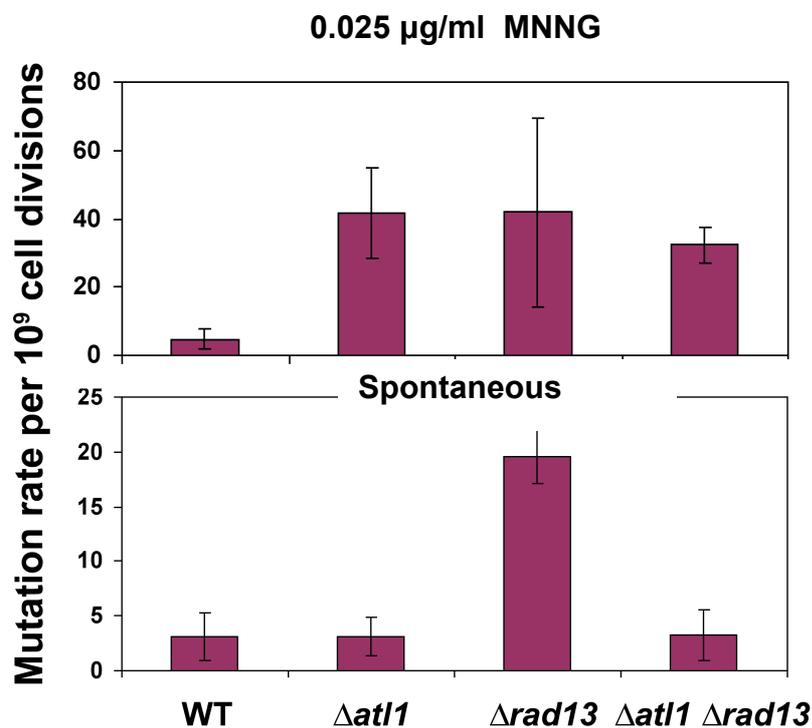


Figure 1.26: Effect of gene deletion in *S.pombe* on MNNG-induced (top) and spontaneous (bottom) mutation rates

An interesting observation is that the double deletant strain ($\Delta atl1 \Delta rad13$) does not show any increased mutation rate when exposed to MNNG than the single deletants. This demonstrates that there is an epistatic relationship between Atl1 and Rad13 (i.e. knocking out one gene disrupts the

repair system as much as removing both genes) and provides evidence that AtI1 is involved in the same DNA repair pathway as Rad13 (i.e. NER).

The effects of inactivation of AtI1 and Rad13 on spontaneous mutation rates are also extremely interesting. The wild-type (WT), AtI1 deletant ($\Delta atI1$) and the double deletant ($\Delta atI1\Delta rad13$) have similar rates of spontaneous mutation, with the Rad13 deletant ($\Delta rad13$) showing a significantly higher rate. It is possible that there is a build-up in the cell of stable DNA-AtI1 complexes that are hard to repair in the absence of Rad13 (i.e. NER cannot take place). The observation that the double deletant returns to wild-type levels of mutation could indicate that when there is no AtI1 in the cell, the stable complex does not form and the damaged DNA can be repaired by another pathway.

Similar experiments in *T.thermophilus* have demonstrated that ATL proteins and NER are mutually involved in the repair of O^6 -methylguanine lesions.(62) Compared to wild-type, the $\Delta uvrA$ mutant and the $\Delta uvrA \Delta atI$ double mutant were equally mutable to MNNG whilst the ΔatI mutant showed mutation rates that were intermediate between WT and the aforementioned mutants. This would seem to indicate that NER is still capable of repair of O^6 -methylguanine lesions in the absence of ATL, but the repair is much more efficient when ATL is present to assist in the recruitment of UvrA. In addition, the increased spontaneous mutation frequency observed in the $\Delta uvrA$ mutant was reduced back to wild-type levels in the double deletant $\Delta uvrA \Delta atI$. This is similar to the result observed in *S.pombe*, and it has been suggested that the in the absence of ATL or NER, responsibility for recognition and repair of O^6 -

methylguanine lesions may fall to MutS of the mismatch repair system, which has been shown in *E.coli* to recognise MeG:T mispairs.(79)

The role of eATL (also known as Ybaz) in the repair of O^6 -alkylguanine adducts has been studied in *E.coli* by Mazon *et al.*. It was demonstrated, using various AGT, eATL and NER deficient mutants, that there is a division of labour whereby smaller lesions (such as O^6 -methylguanine and to an extent O^6 -hydroxyethylguanine) were repaired efficiently by alkyltransferase proteins (Ada and Ogt) whilst the repair of larger alkyl groups (such as O^6 -1-hydroxy-propylguanine and O^6 -2-hydroxypropylguanine) was the responsibility of the NER pathway.(80) Furthermore, it was shown that repair by NER of these larger adducts was enhanced significantly by the presence of eATL in the cells, suggesting that eATL facilitates recognition of the lesion by the NER machinery. Pulldown assays showed an interaction with UvrA (the NER protein that is responsible for recognition of bulky, distorting lesions in DNA), which is in agreement with the finding by Morita *et al.* that TTHA1564 interacts with UvrA in *T.thermophilus* and further implicates ATL proteins in the role of damage sensors for NER. These results support the suggestion that eATL has a similar role to Mfd, recruiting UvrA to the site of damage and hence enhancing the efficiency of NER. This enhancement would be especially important for lesions that are otherwise poor substrates for NER or for alkyltransferase (AGT) proteins, such as O^6 -1-hydroxy-propylguanine and O^6 -2-hydroxypropylguanine. They could be considered to be 'stuck in the middle': too large for efficient repair by AGTs, but too small to be recognised with high affinity by NER factors or to cause stalling of RNA polymerase. eATL could

thus be providing crosstalk between distinct repair pathways, as is suggested for Atl1 in *S.pombe*.(58)

The proposed role of ATL proteins is shown in figure 1.27, whereby they facilitate in the recruitment of NER factors and increase the efficiency of repair of relatively small lesions such as O^6 -methylguanine. This function would seemingly be even more important in organisms such as *S.pombe* and *T.thermophilus* which cannot process these lesions by direct damage repair. However, the exact function of ATL proteins and the mechanisms by which they recruit NER factors or increase the repair efficiency of smaller O^6 -alkylguanine lesions is presently unknown and warrants further investigation.

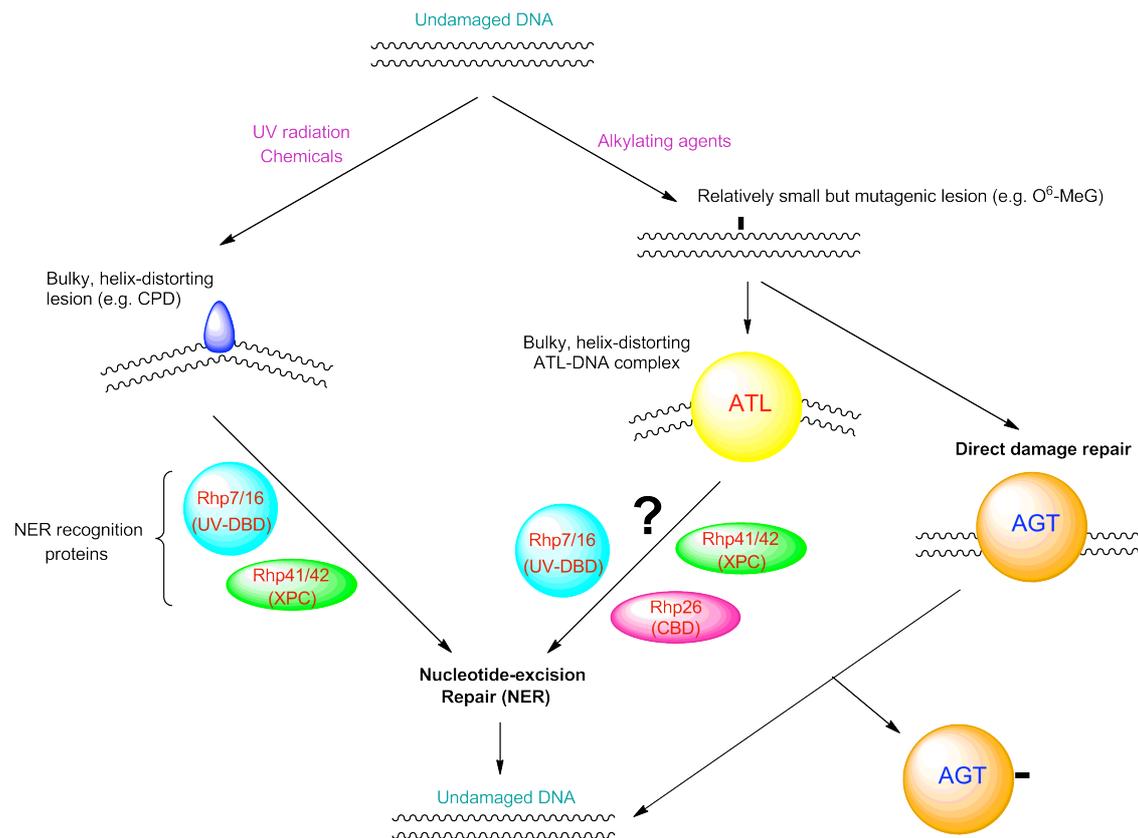


Figure 1.27: Proposed role of ATL proteins in repair of O^6 -alkylguanine lesions. The NER factors shown are labelled with the names of the *S.pombe* and human homologues

In addition, eATL has also been found to prevent the mismatch repair-mediated toxicity induced by certain O^6 -alkylguanine residues in *E.coli*.⁽⁸¹⁾ Using strains with a deficient AGT (Δogf , Δada) and NER ($\Delta uvrA$) background, Mazon *et al.* confirmed that the toxicity of O^6 -alkylguanine (O^6 -MeG) residues was due to activity of the mismatch repair (MMR) system. It was shown that repeated recognition of O^6 -MeG: T mispairs by MMR in the so-called futile cycle (see 1.6.3) led to toxicity of O^6 -alkylguanine adducts, and that this toxicity was strongly dependant upon the function of both MutS and MthH (i.e. it involved steps beyond MutS binding). Cells containing O^6 -MeG, O^6 -HOEtG, O^6 -1-HOPrG and O^6 -2-HOPrG lesions had an \approx 80% survival rate compared to WT in the absence of MMR ($\Delta mutS$) whereas when the MMR system was restored in the absence of eATL ($\Delta ybaz$) the survival rates for all adducts dropped to <10%. Whilst eATL had little effect on survival rates of cells containing O^6 -MeG (it remained <10%), for the larger lesions there was a significant increase in survival rates (50-60%) showing that eATL prevented these adducts from causing MMR-mediated toxicity. Furthermore, the authors demonstrated by EMSA that eATL and MutS compete for binding of O^6 -MeG: T mispairs in double-stranded ODNs, and that O^6 -MeG was only very weakly bound by eATL compared to the other O^6 -alkylguanine adducts. These results suggest an intriguing and hitherto unsuspected role for eATL in *E.coli*. It will be interesting to see if other ATL proteins perform this cellular function in their respective organisms, especially in the cases where there is no AGT protein present (e.g. *S.pombe* and *T.thermophilus*).

1.12 Evolution of Alkyltransferases

It is interesting that a group of ATL proteins exist that have extensive homology and topology to alkyltransferases (AGTs), and that recognise the same or similar lesions in DNA but lack the ability to repair them. It is also intriguing that *E.coli* has two AGTs and an ATL protein, *S.cerevisiae* and mammals only an AGT, and *S.pombe* only an ATL protein.(82) It would seem to suggest that the presence of AGT (and hence the direct repair of alkyl adducts in DNA) is not essential for the survival of living organisms. The observation that mice lacking AGT are phenotypically normal would also seem to suggest this.(83)

It may be that in divergent organisms, a different range of DNA lesions are generated *in vivo*, and that these substrates are processed most effectively either by a direct repair protein (AGT) or by a protein that signals to other DNA repair pathways (ATL). Even if the spectrum of lesions in the cells of different organisms is the same, it could be that some species survived by having a single protein with a broad substrate specificity range, and others by developing a number of proteins, each with their own niche, that worked together to remove genotoxic damage.

It is possible that AGTs evolved from proteins that had the ability to bind O⁶-alkylguanine lesions but were unable to remove the alkyl group. Mutations in these proteins, including the generation of a cysteine residue in the active site, could have led to the acquisition of alkyltransferase activity. It is conceivable that this new pathway then became the dominant pathway for the repair of DNA alkylation damage in some organisms. On the other hand, ATL proteins may be the result of substitutions in an AGT that caused the loss

of alkyltransferase activity but left the ability to bind O^6 -alkylguanine adducts. It is difficult to exclude one or the other of these two possibilities without further evidence.

An ATL sequence has recently been found in the genome of the starlet sea anemone *Nematostella vectensis*.(58) The *N.vectensis* genome is surprisingly more similar to vertebrates than to nematodes or fruit flies.(84) This discovery of an animal ATL sequence could give an indication that there is a possibility of finding ATL proteins in other eukaryotes, including humans. It is also possible that in higher eukaryotes and mammalian systems, proteins exist that perform the same role as ATL proteins, for example it has been suggested that UV-DDB, the NER DNA damage binding complex may act as a functional homologue of ATL in humans.(81)

There are still many unanswered questions about the role of alkyltransferase-like proteins. It is reasonable to suppose that the study of ATL proteins will allow us to gain insight into the mechanisms of O^6 -alkylguanine lesion recognition, to identify any interacting protein partners and to discover if there are connections between direct repair and other seemingly distinct repair pathways.

2.0 Research Aims

The general aim of this project is to investigate the function of ATL proteins, with particular emphasis on AtI1 from *S.pombe* and TTHA1564 from *T.thermophilus*. *S.pombe* (budding yeast) and *T.thermophilus* (extremophile bacterium) do not have an alkyltransferase (AGT) protein and therefore these organisms are especially interesting models for attempting to elucidate the function of ATL proteins.

2.1 DNA Recognition by ATL Proteins

A systematic study of the range of DNA substrates recognised by AtI1 and TTHA1564, including detailed binding affinities, will be carried out using fluorescence anisotropy or similar fluorescence-based methods. This will be the first study to measure a series of accurate K_D values in solution and under true equilibrium conditions for ODNs containing a broad range of O^6 -alkylguanine adducts, and will be complementary to and extend upon the studies already conducted using ELISA and SPR. In addition, these measurements will allow direct comparisons to be made between ATL proteins from markedly different organisms (the prokaryote *T.thermophilus* and eukaryote *S.pombe*) and with different active site motifs (PAHR and PWHR respectively). If insight is gained into the preference of ATL proteins for binding certain types of modified bases, it should facilitate in the elucidation of the mechanism of lesion recognition and may give some clues as to the biological function of ATLs.

The ODN substrates containing a wide variety of O^6 -alkylguanine and related analogues will be prepared using post-DNA synthesis chemistry using synthetic ODNs containing the reactive base 2-amino-6-methylsulfonyl-purine.

2.2 ATL Proteins and NER

It has been proposed that the relatively stable At11-DNA complex acts as a molecular signal to proteins in other repair pathways. Pull-down assays have been used with other ATL proteins (TTHA1564 and eATL) to find interacting proteins, such as UvrA which links ATLs to nucleotide-excision repair (NER).(63,80) Therefore, we will attempt to isolate At11 in complex with any interacting proteins which we hope will give more information about its specific cellular function. Mass spectrometry-based proteomics will be used to identify any bound proteins and thereby give an indication of the possible repair pathways with which At11 is involved.

Although the At11-DNA complex is relatively stable the binding is reversible and therefore attempts shall be made to obtain a covalently cross-linked At11-DNA complex as it will be more robust and possibly more suitable for use in this assay. To this end, a chemical moiety that is capable of reacting specifically with the active site cysteine (in the At11 W56C mutant) will be synthesised and introduced into an ODN substrate. Isolation of At11 with its binding partners may provide further evidence of whether the At11-DNA complex is involved in the NER pathway as proposed or an alternative repair mechanism. The stable At11-DNA complex appears to be suitable to block RNA polymerase during transcription and also large enough to be recognised as a bulky lesion by the NER machinery and therefore these experiments may

clarify whether At11 is involved in transcription-coupled repair (TCR), global genome repair (GGR) or both sub-pathways.

3.0 Synthesis and Modification of O⁶-alkylguanine-containing Oligodeoxyribonucleotides (ODNs)

In order to investigate the properties and behaviour of AGT and ATL proteins, it was necessary to synthesise and chemically modify various oligodeoxyribonucleotides (ODNs) for use in biochemical assays and structural studies. This chapter describes the synthesis and characterisation of the ODNs that were made for this project.

3.1 Introduction to ODN Synthesis and Modification

3.1.1 Synthesis of Oligodeoxyribonucleotides

Oligodeoxyribonucleotides (ODNs) are short pieces of single-stranded DNA, where each nucleoside unit is joined to the next by a phosphodiester linkage to comprise a strand that is typically less than one hundred and fifty nucleotides in length. Oligodeoxyribonucleotides are usually chemically synthesised in order to generate a DNA strand of specific sequence; the method used most commonly utilises phosphite triester (or phosphoramidite) chemistry, which was developed by Caruthers into an automated process.⁽⁸⁵⁾ At the heart of the synthesis is a highly efficient (greater than 99%) coupling reaction between the 5'-hydroxyl group of a 2'-deoxyribonucleoside bound to a solid support and a 5'-DMT-(*N*-protected)-2'-deoxyribonucleoside 3'-*O*-(*N,N*-diisopropyl) *O*-alkyl phosphoramidite.

The solid-phase synthesis of oligodeoxyribonucleotides is generally carried out in the 3'→5' direction, the reverse of the biosynthetic pathway during DNA replication. It is extremely efficient due to the fact that a large excess of the soluble phosphoramidite can be used to drive the reaction to high yield. In addition, the heterogenous nature of the reaction allows for a simple washing step to remove excess reagents and byproducts from the solid support-bound polymer at each step. The solid support is made of controlled pore glass (CPG) beads which are rigid, non-swelling and inert. When the synthesis is complete, the ODN can be cleaved from the beads by a thirty minute treatment with concentrated aqueous ammonia.

The synthesis itself consists of a cycle of chemical reactions that extend the nucleotide chain by one residue at a time. The cycle (figure 3.1) is repeated until the specific sequence of nucleotides is incorporated into the ODN of desired length. The basic steps of this cycle are as follows:

1. Detritylation (removal of the 5'-DMT group);
2. Activation of the appropriate phosphoramidite by a coupling agent;
3. Addition of the activated phosphoramidite to the growing chain;
4. Capping to block the chains that have not reacted during the coupling;
reaction
5. Oxidation of the phosphite to a phosphate triester.

Initial removal of the trityl group on the CPG-bound protected nucleotide is carried out by reaction with trichloroacetic acid in methylene chloride (i.e. acid deprotection). The dimethoxytrityl cation (DMT⁺) that is

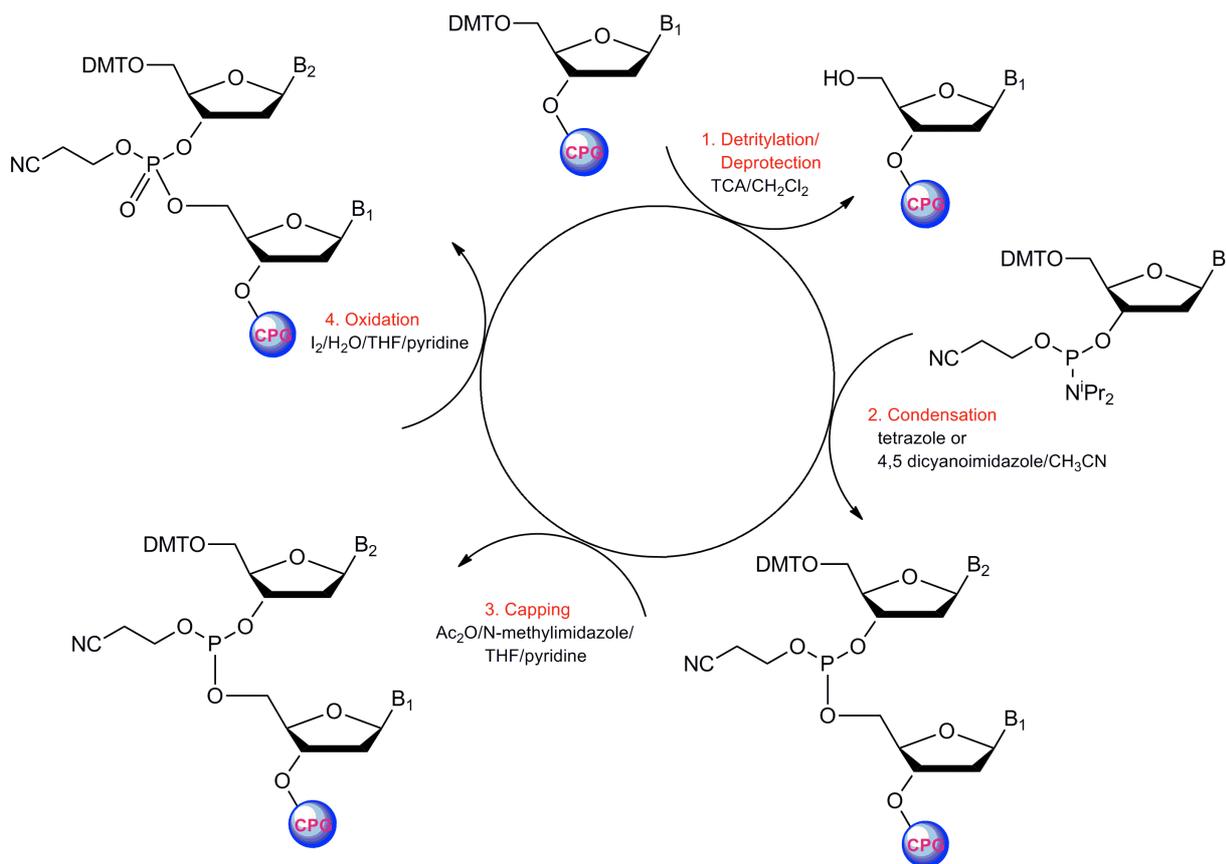


Figure 3.1: The cycle of steps in phosphoramidite synthesis of oligonucleotides

released into solution has a bright orange colour: comparison of the intensity of this colour with that of the previous cycle can be used to calculate the coupling efficiency. The phosphoramidite to be added to the growing nucleotide chain is activated by mixing with 5-ethylthiotetrazole in acetonitrile solution: the tetrazole protonates the diisopropylamine group of the phosphoramidite which then reacts to give an activated tetrazolophosphoramidite. This then reacts with the 5'-OH group of the support-bound nucleotide resulting in the formation of a new P-O bond and the expulsion of the NⁱPr₂ leaving group. The efficiency of coupling is excellent (>99%) and the only side reaction is phosphitylation of the O⁶-position of

guanine, which is completely reversed by the conditions in the capping step. Capping involves treatment with acetic anhydride and *N*-methylimidazole in THF to block any chains that have not reacted during coupling, in order that they cannot react any further and lead to failure sequences (i.e. ODNs one, two etc. nucleotides shorter than the desired full length sequence). Finally, the intermediate phosphite is oxidised to the phosphate triester by treatment with iodine and 2,6-collidine in aqueous THF. The hydrogen iodide that is generated during the oxidation reaction is neutralised by pyridine. After washing, the cycle can start again which introduces the next nucleotide to the growing chain.

Once the synthesis is complete, and the required ODN cleaved from the CPG beads, the base and phosphate protecting groups are removed by treatment for 6 hours with aqueous ammonia at 50°C. The oligonucleotides can then be purified by polyacrylamide gel electrophoresis (PAGE) or reversed-phase high performance liquid chromatography (RP-HPLC). Reversed-phase HPLC, which separates according to hydrophobicity, is a common technique and useful for purifying ODNs with and without the 5'-terminal DMT group left on. The advantage of leaving the 5'-DMT protecting group on is that the impurities (e.g. shorter length ODNs) lacking this hydrophobic group have much shorter retention times by RP-HPLC and can therefore be easily removed. The purified oligonucleotide can then be treated with acetic acid in order to remove the 5'-DMT and subsequently de-salted using a gel filtration column.

In addition to being able to synthesise a specific sequence of single-stranded DNA by the phosphoramidite method, it is also possible to use this

approach to introduce a variety of reporter groups at the terminal end of the chain.(86) Reporter groups are chemical entities that functionalise the oligomer so that it can be used for certain applications. For example, a fluorophore can be introduced, such as fluorescein or hexachlorofluorescein (figure 3.2), which effectively labels the DNA in order that it can be used for fluorescence studies (e.g. fluorescence anisotropy).(87) Also useful is the introduction of a biotin molecule onto the 5'-end of an ODN by use of a phosphoramidite reagent. The presence of the biotin group allows the ODN to be separated from other molecules by a highly selective affinity step, exploiting the extremely high affinity interaction of biotin with streptavidin, which itself can be immobilised on beads.

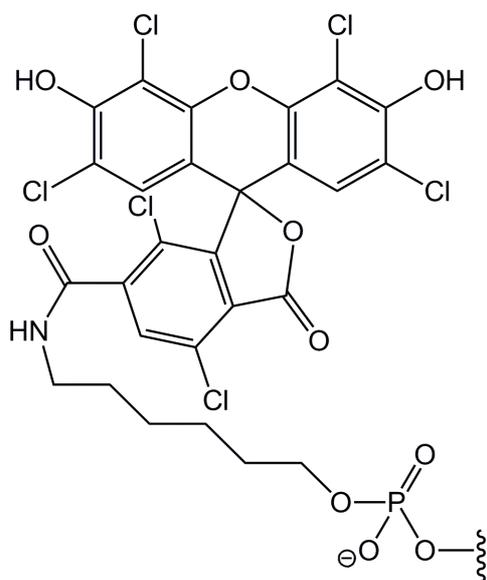


Figure 3.2: Hexachlorofluorescein, a fluorescent reporter group

ODNs are useful substrates for proteins that bind DNA. Using reporter groups such as those mentioned above, they can be used as probes in many types of assay. Whilst they are usually in the single-stranded form after DNA synthesis, two complementary ODNs can be annealed together to form

double-stranded DNA duplexes. It is also possible to deliberately incorporate modified bases into ODNs, using specific phosphoramidites, in order that they can be used as substrates for proteins that recognise those modifications. For example, the phosphoramidite reagent is commercially available that can be used to place an O^6 -methylguanine-containing nucleotide residue at a specific site in the synthesised ODN. These ODNs could then be used in assays with alkyltransferase (AGT) and alkyltransferase-like (ATL) proteins that recognise DNA containing this modified base.

3.1.2 Post-DNA Synthesis Modification of ODNs

As mentioned, ODNs with specific modifications such as O^6 -methylguanine can be synthesised using specific base-modified phosphoramidite reagents. One of the methods that can be used to introduce these modifications into ODNs (other than generating a specific phosphoramidite to place it into the oligonucleotide during synthesis) is that of post-synthesis modification. This strategy involves first synthesising an ODN with a reactive base that is stable to the conditions of DNA synthesis but which can be subsequently modified in a simple chemical transformation following this process.

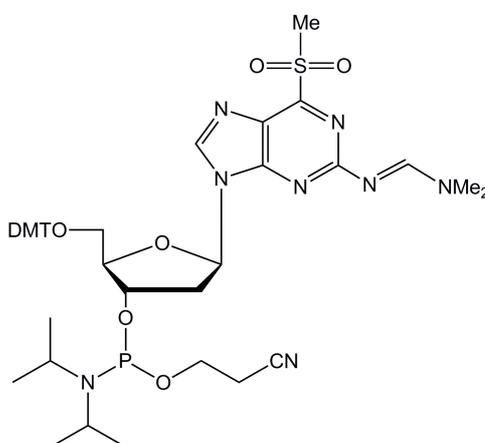


Figure 3.3: Sulfone phosphoramidite

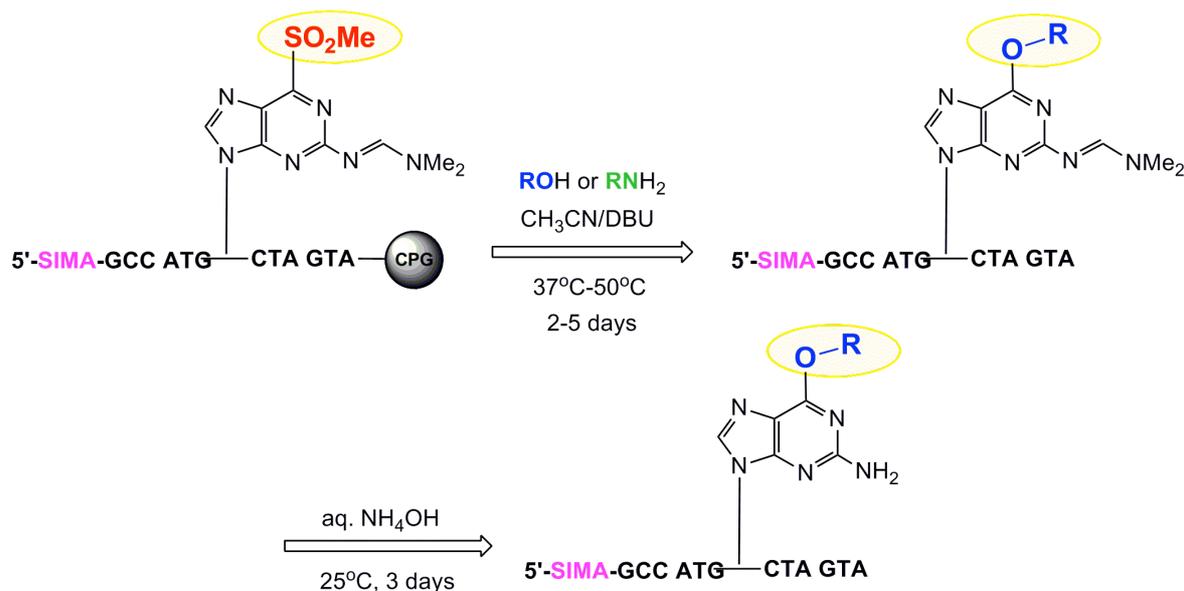


Figure 3.4: Scheme for post-DNA synthesis modification of oligonucleotides, as used for the preparation of SIMA-labelled 13-mer ODNs (see section 3.2) (68)

Whilst a number of strategies have been described (88-90) the Williams group have developed a simple and efficient route to the preparation of ODNs containing any O^6 -alkylguanine modification. In this procedure, 2-amino-6-methylsulfonyl-purine is introduced to an ODN during DNA synthesis using a unique phosphoramidite (figure 3.3).(68) Following DNA synthesis, treatment with an alcohol under basic conditions can be used to yield a variety of modified O^6 -alkylguanines. Complete cleavage of the ODN from the CPG beads and base deprotection is achieved by treatment with aqueous ammonia to yield an ODN with an O^6 -alkylguanine base (figure 3.4). Alternatively, an amine can be used in the displacement step to yield an N^6 -alkylpurine base. It is worth mentioning that cleavage of the ODN from the CPG and removal of the phosphate diester protecting groups is likely to occur during the first step of this post-DNA synthesis chemistry. The natural bases are likely to be deprotected after approximately one day in aqueous ammonia, although the

modified base forms an N^2 -formylamino analogue which requires three days for full deprotection.(68)

This route has been used previously to make ODNs containing O^6 -methylguanine (O^6 -MeG), O^6 -benzylguanine (O^6 -BnG), O^6 -(2-hydroxyethyl)guanine (O^6 -HOEtG), O^6 -(4-bromophenyl)guanine (O^6 -BTG) (68) and more recently O^6 -carboxymethylguanine (O^6 -CMG) and O^6 -[4-oxo-4-(3-pyridyl)butyl]guanine (O^6 -PobG).(71) These ODNs have been used to probe the substrate specificity of At11, an alkyltransferase-like (ATL) protein from *S.pombe*. There is scope for the method to be used to create a whole range of modified ODNs that could be used in experiments to probe the mechanisms of DNA repair involving ATL proteins.

3.2 ODN Substrates for ATL-DNA Binding Assays

It was our intention to conduct the first comprehensive, quantitative and comparative study of the substrate specificity of two ATL proteins, At11 from *S.pombe* and TTHA1564 from *T.thermophilus*. It was decided that determining dissociation constants (K_D) by titration using fluorescence-based methods, such as fluorescence anisotropy or total fluorescent emission intensity, would be desirable (section 4.2). To this end, we required a series of ODNs, labelled at the 5'-terminus with a fluorescent reporter group and of identical sequence but containing different O^6 -alkylguanine residues at a specific position. This would allow us to make direct comparisons of the binding affinities of ATL proteins with ODNs containing various interesting and relevant lesions.

3.2.1 Selection of Fluorescent Dye

As previously mentioned (section 3.1.1), hexachlorofluoroscein (HEX) is a useful reporter group that allows dissociation constants in the low nM range to be measured accurately using fluorescence-based methods. Indeed, we were able to show that we could quantitatively measure the interaction of MBP-AtI1 with an ODN containing O^6 -methylguanine using changes in total fluorescent emission intensity (section 4.2). The problem when planning the study as a whole was the chemical stability of the HEX label. In order to produce the range of substrates necessary for these experiments, post-synthesis displacement chemistry must be utilised. Having considered the data available from Glen Research (the manufacturer of HEX), as well as a recent manuscript by Chuvilin *et al.* (91) regarding the instability of HEX under basic conditions, it was not certain whether this tag would be robust enough to survive the treatment with aqueous ammonia for 3 days. Recently, Glen Research have produced a new fluorescein-based dye, SIMA (HEX) (figure 3.5), that is available as a suitable phosphoramidite reagent for 5'-labelling of ODNs. This fluorescent reporter group has very similar fluorescent properties to HEX whilst displaying a much greater stability at high pH. For example, a HEX-labelled ODN deprotected with aqueous ammonia at 55°C overnight showed considerable degradation, while the SIMA (HEX) equivalent showed no decomposition (<http://www.glenresearch.com//GlenReports/GR21-110.html>). Although the deprotection step used for our reactions is milder than these conditions, it seemed prudent to opt for the fluorescent label that is more chemically robust. The standard deprotection conditions for HEX are treatment for 24h in aqueous ammonia at room temperature and it follows

that SIMA is likely to be more suitable for the 3 day deprotection that is required in the post-synthetic modification (PSM) chemistry. It was therefore decided that the SIMA tag would be used in our experiments after we confirmed that, like HEX, we could use ODNs labelled with it to accurately measure dissociation constants (see section 4.2).

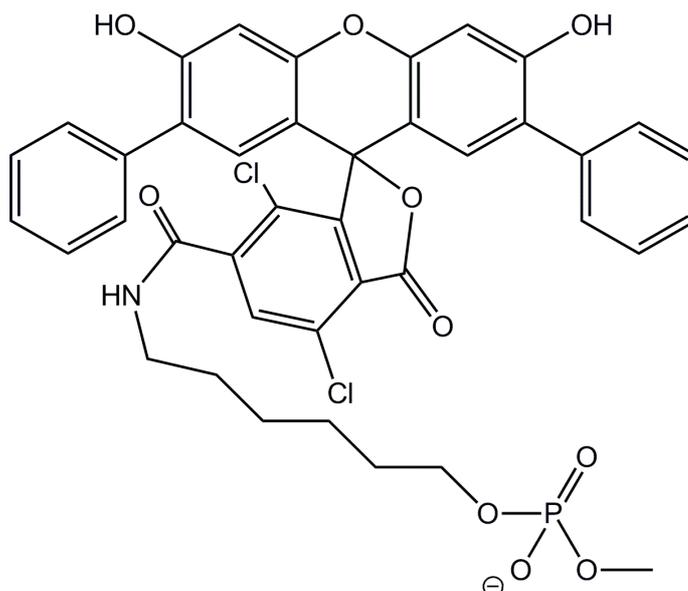


Figure 3.5: Structure of SIMA (HEX) fluorescent label

3.2.2 Synthesis of 5'-SIMA-labelled ODNs



Figure 3.6: Single-stranded 13-mer ODN for use in binding assays

A short sequence of DNA (13-mer ODN) was selected for modification and subsequent use as a substrate for studying the binding of ATL proteins (figure 3.6). This sequence was chosen as it had been shown to form 1:1

complexes with AtI1 by sedimentation analysis, and in addition is the sequence of the substrate used in the published crystal structures of the AtI1-DNA complex that were solved by Tubbs *et al.* (58). The 5'-SIMA-labelled ODN containing *O*⁶-methylguanine was synthesised by DNA Technology (by the standard phosphoramidite method), and once it was established that accurate and repeatable dissociation constants could be measured using substrates bearing this probe, ODNs containing a full range of modifications at the same position (X) in the sequence were prepared using the post-synthesis modification (PSM) chemistry developed by Shibata *et al.* (92). This allowed us to incorporate a wide range of adducts at the 6-position of guanine using both alcohols and amines (figure 3.4).

For liquid alcohols, a ratio of 9: 9: 2 (alcohol: acetonitrile: DBU) was used, whilst for solid alcohols the displacement was performed with a 5M solution in acetonitrile in a 9:1 ratio with DBU (section 9.3). For some of the bulkier alcohols (where the deprotonated nucleophile will be more sterically hindered) the two day incubation period was increased in length, or the reactions performed at elevated temperatures, the full details of which are discussed in greater detail later in this section. For reactions using fluorescently-labelled ODNs, the incubations were carried out in the dark to protect the SIMA dye, though in the case of reactions with unlabelled ODNs this was not necessary.

It is extremely important that the reaction is dry during this displacement step so that no hydroxide ions compete with the alkoxide ions as the displacing nucleophile (which would result in guanine as the product). Therefore, all reagents used in these reactions were dried over molecular

sieves, and the tips used to transfer the reagents dried in a vacuum desiccator for one hour prior to use. In addition, the acetonitrile was dried and purified using Grubbs apparatus and the reaction vessel was flushed with nitrogen prior to incubation in order to drive out air and moisture. After this displacement step, treatment in aqueous ammonia (33%) for 3 days at room temperature resulted in full deprotection of all the bases of the ODN. The oligonucleotides were evaporated to remove the ammonia, redissolved in water, extracted with ether, and pre-purified by NAP-10 column if necessary. The ODNs were subsequently purified by reverse phase HPLC (RP-HPLC) in TEAB buffer (see section 9.4)

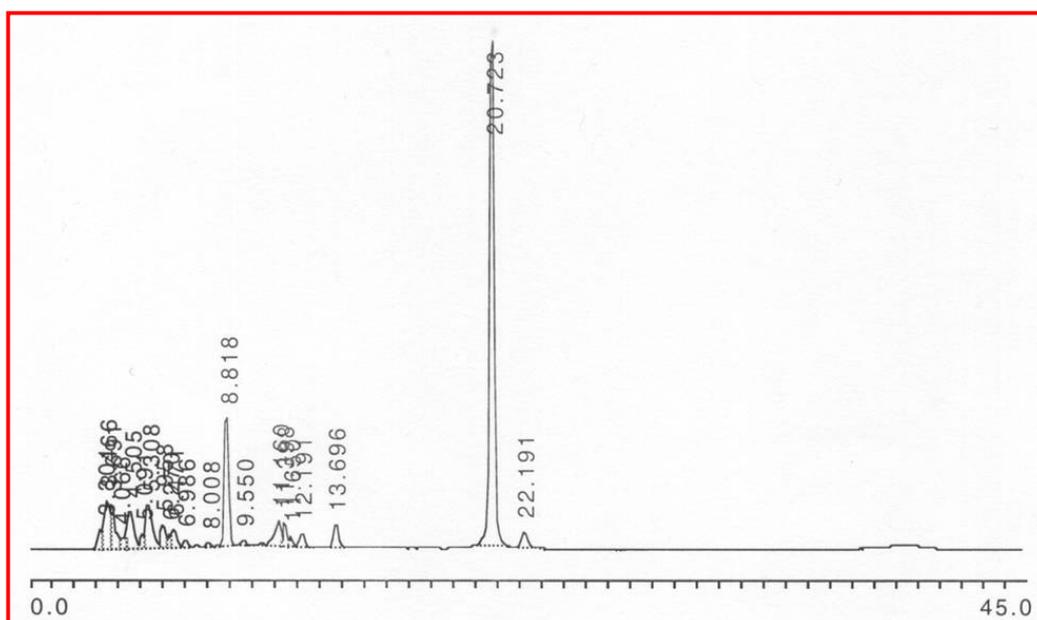


Figure 3.7: An RP-HPLC trace of reaction mixture from the synthesis of OW18 (O^6 -CMG).

The main peak at retention time $T = 20.7$ min is the product.

A typical RP-HPLC trace is shown in figure 3.7. The SIMA-labelled ODNs were purified on a 0-40% gradient (buffer A= 0.1M TEAB, 5% MeCN, buffer B= 100% MeCN) over 30 min (section 9.4). This is a steeper gradient

than for unlabelled ODNs which were typically purified on a 0-10% or 0-15% gradient. This modified gradient is required due to the hydrophobic nature of the fluorescent tag (which increases the affinity of the ODN for the column). All ODNs purified had similar retention times (approximately 19-21 min) despite having different guanine adducts (section 9.4). The 5'-SIMA label gives the ODNs a characteristic reddish-pink colouration. The ODNs were kept in the dark (as far as possible) throughout the purification process to minimise any potential damage to the dye.

After removal of the buffer salts by evaporation, the purified ODNs were redissolved in distilled water and stored at -20°C . The identity of the adduct-containing ODNs were confirmed by mass spectrometry (figure 3.9) and the DNA concentrations calculated by measuring the absorbance at 260nm (typically ODN concentration was in the range 100-300 μM).

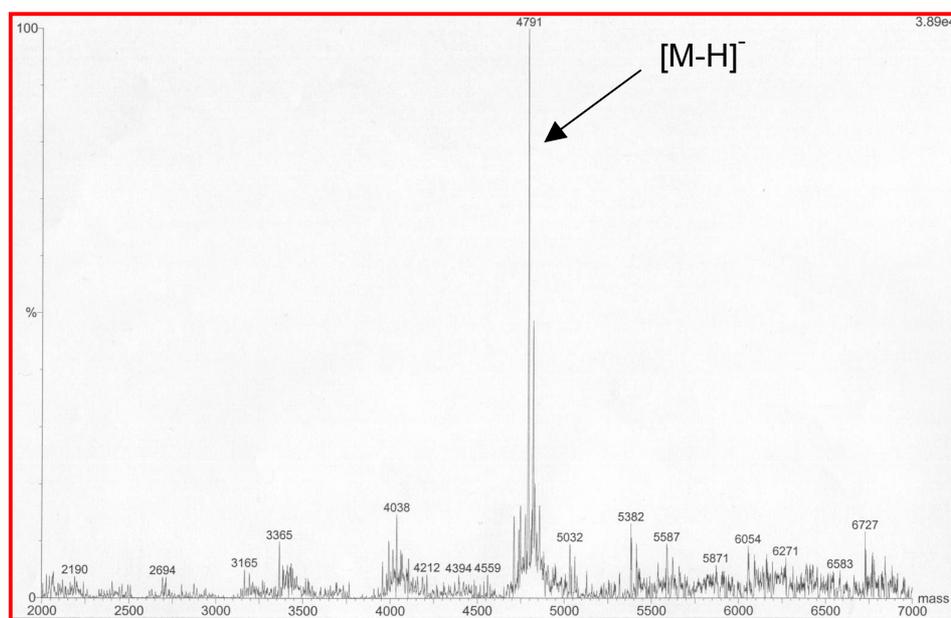


Figure 3.9: Electrospray TOF mass spectrograph of purified OW18 (O^6 -CMG).

The mass ion at 4791 is the product.

Using different alcohols and amines as reactants in the displacements, a broad range of O^6 -alkylguanine and other purine lesions that vary in size, charge, hydrophobicity and polarity were introduced into ODNs. Many were selected that are biologically relevant and others were chosen as they are known to be poorly repaired by MGMT (or not repaired at all). The modifications incorporated into ODNs will be examined in some detail in the pages to follow (all the modified bases prepared are shown in figure 5.9).

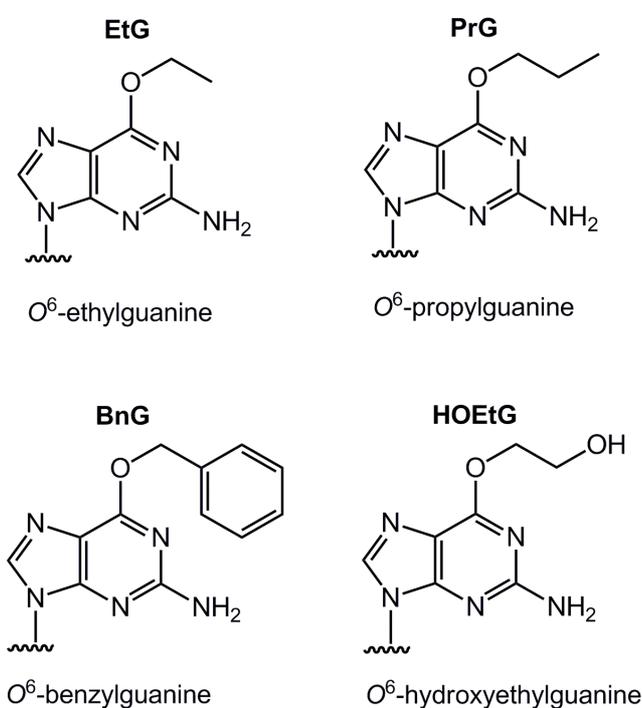


Figure 3.10: Lesions introduced to ODNs in the first batch of PSM reactions

The first ODNs to be made were those shown in figure 3.10. The synthesis was successful in all cases, as demonstrated by HPLC analysis and subsequent identification of the modified ODNs by mass spectrometry. They are all biologically relevant lesions: O^6 -ethyl, O^6 -propyl and O^6 -benzylguanine adducts are common products arising from endogenous and exogenous alkylating agents (including those used in cancer chemotherapy), whilst O^6 -

hydroxyethylguanine lesions are generated by exposure to ethylene oxide (93) and also produced endogenously by certain metabolic processes such as oxidation of methionine and lipid peroxidation.(94)

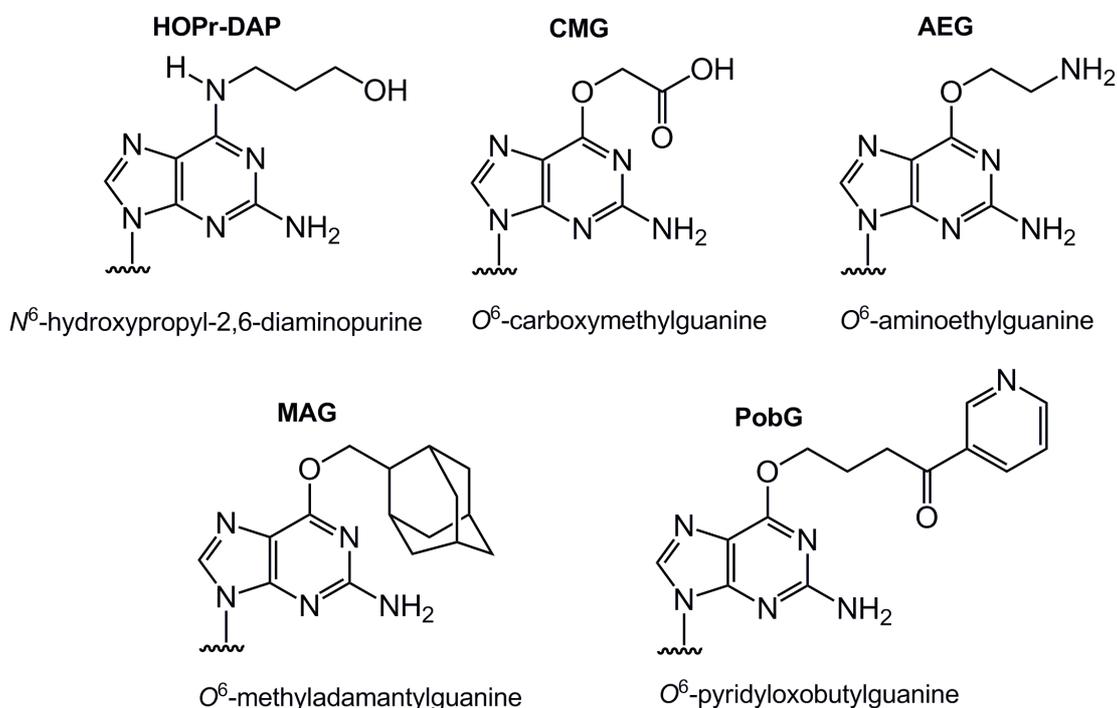


Figure 3.11: Lesions introduced to ODNs in the second batch of PSM reactions

The next reactions performed were a little more adventurous: the modified bases that were introduced are shown in figure 3.11. The syntheses of *O*⁶-pyridyloxobutylguanine (*O*⁶-PobG), *O*⁶-carboxymethylguanine (*O*⁶-CMG), *O*⁶-aminoethylguanine (*O*⁶-AEG) and *N*⁶-hydroxypropylguanine (HOPr-DAP) all worked well to give the expected products. For preparing *O*⁶-AEG, trifluoroacetyl (TFA) *N*-protected ethanolamine was used to ensure that the desired product would be formed (i.e. displacement of the methylsulfonyl group by the less nucleophilic alcohol function). The TFA-protecting group is base labile and was therefore removed under the standard deprotection conditions (three days in aqueous ammonia). In order to form the HOPr-DAP,

unprotected aminopropanol was used which reacts via the amino group. The synthesis of O^6 -methyladamantylguanine (O^6 -MAG) required the duration of the displacement step to be increased (5 days at 37°C instead of the standard 2 days), which is likely to be due to its large size and hence greater steric hindrance when acting as a nucleophile. The synthesis of O^6 -carboxymethylguanine (O^6 -CMG) by reaction of sulfone-modified ODN with methyl glycolate also requires a slight modification to the procedure. Thus, after the initial displacement step, the product must be treated with 1M NaOH for one day *before* the addition of aqueous ammonia for 3 days. This step is necessary to convert the methyl ester into a carboxylate moiety as direct treatment with ammonia would produce the O^6 -carboxamido analogue. Deprotection of the nucleobases then proceeds as usual.

O^6 -CMG and O^6 -PobG were prepared as they are both highly relevant types of biological damage. These lesions have been shown to occur by the action of exogenous chemical agents on DNA: by nitrosamines present in red and preserved meats, and by NNK present in tobacco smoke, respectively. O^6 -MAG was made due to its large size which it was thought may prevent it from being bound efficiently by a base-flipping protein such as AtI1. To examine the effects of charge on recognition, O^6 -AEG and O^6 -CMG were prepared: the O^6 -AEG residue will be positively charged at pH 7.5, whilst O^6 -CMG will carry a negative charge. For HOPr-DAP, the substitution of nitrogen for oxygen at the 6-position of guanine, whilst still having an alkyl chain, made it an interesting analogue to incorporate into the study.

Subsequently, attempts were made to synthesise ODNs containing the analogues shown in figure 3.12. Unfortunately there were some difficulties

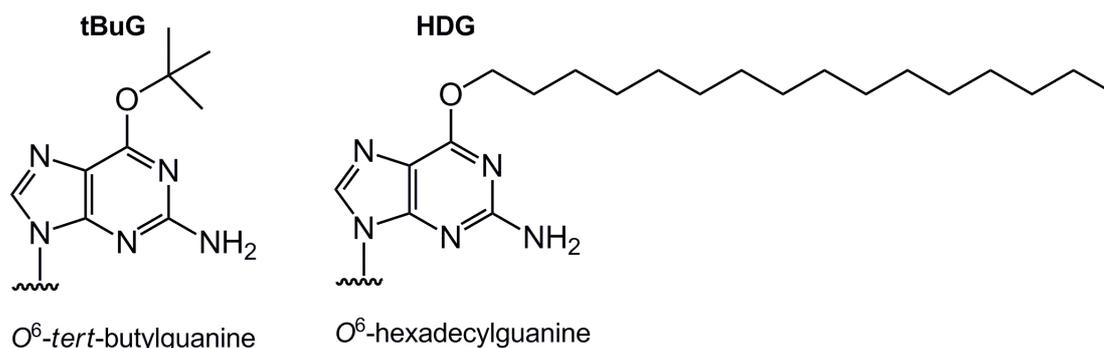


Figure 3.12: O^6 -alkylguanine adducts that failed to be introduced to ODNs using PSM chemistry with *t*-butanol and hexadecanol

with the synthesis of both O^6 -hexadecylguanine (O^6 -HDG) and O^6 -*tert*-butylguanine (O^6 -tBuG) and the desired products could not be formed. Hexadecanol is poorly soluble in acetonitrile, and after addition of all the reagents there was clearly undissolved solid alcohol in the reaction mixture. After the first attempt using the standard reaction conditions failed, more aggressive conditions were used (7 days at 50°C) which improved the solubility of the alcohol but still did not give the required product. These conditions were the same used for the other reaction shown, of sulfone-modified ODN with *tert*-butanol, as this latter reaction suffers from using a poorly nucleophilic tertiary alcohol. The HPLC traces and mass spectra of the product mixtures from these reactions are shown in figure 3.13. The HPLC traces show more peaks than would be expected to be observed for a clean and successful synthesis, with a cluster of peaks occurring close to the expected retention time of the product (approximately 19-21 mins). The calculated masses of the ODN products are 4789 Da for the ODN containing O^6 -(*t*-butyl)guanine and 4957 Da for ODN containing O^6 -(hexadecyl)guanine. There are no peaks in the spectra corresponding to these masses or even any

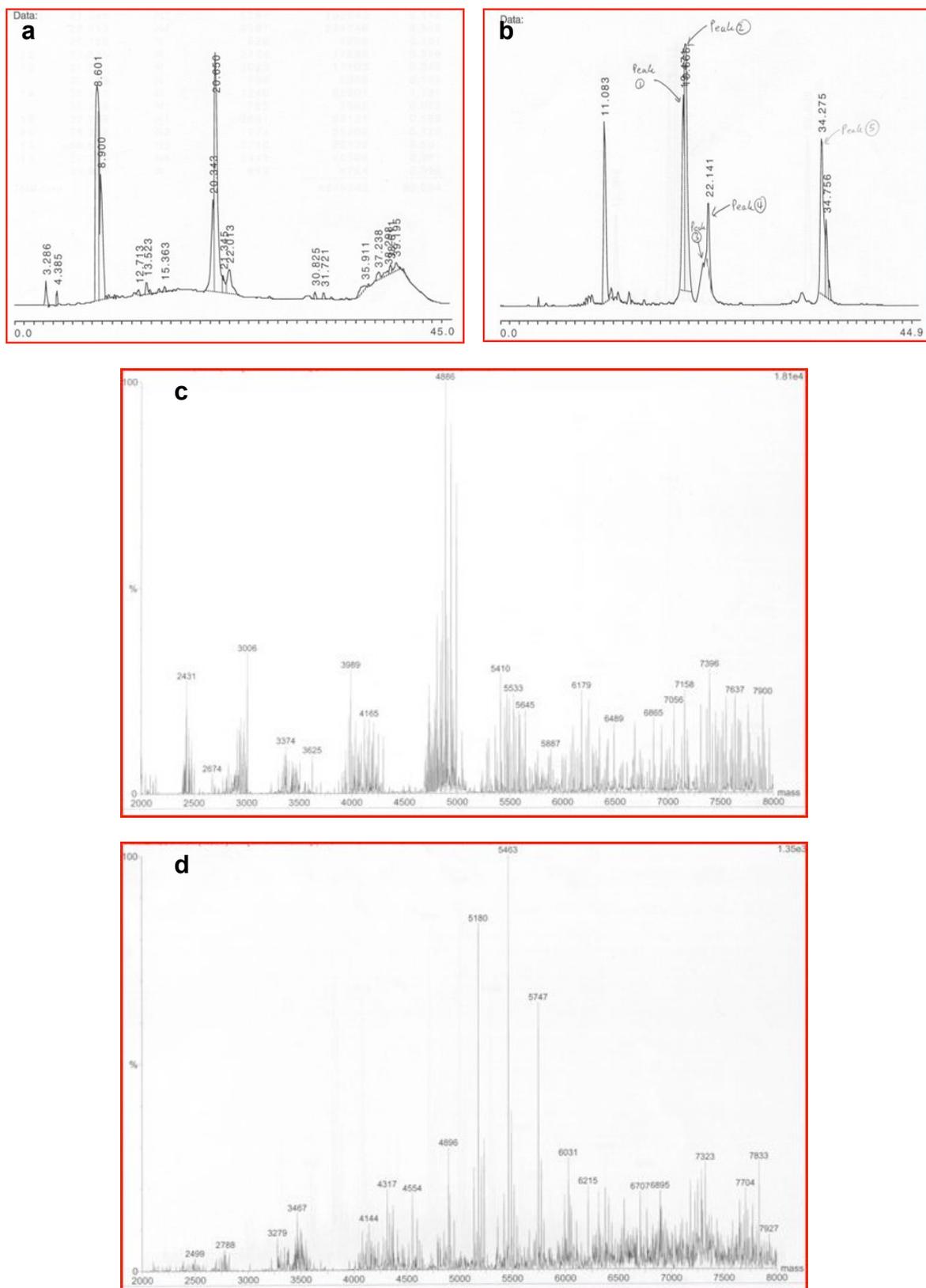


Figure 3.13: HPLC traces and ESI mass spectra of the product mixtures formed from PSM reactions of ODN with t-butanol (a, c) and hexadecanol (b, d). Calculated mass of ODN containing O^6 -tBuG = 4789 Da (c) and of ODN containing O^6 -HDG = 4957 Da (d)

that are close to these values.

For the synthesis of ODNs containing 2,6-diaminopurine (figure 3.14), ammonia gas was bubbled through dry acetonitrile for one hour, and the resulting solution (containing no DBU) added to the solid support-bound sulfone modified ODN with the reaction then incubated at 37°C for 2 days. Following this, aqueous ammonia treatment was performed as normal (3 days at r.t.). This provided the desired product as confirmed by mass spectrometry. It was of interest to look at ODNs containing 2,6-DAP as substrates for ATL proteins due to the presence of a 6-amino group, rather than a methoxy (in O^6 -methylguanine) or a carbonyl in the case of guanine.

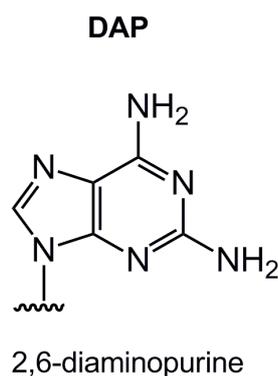


Figure 3.14: Lesion introduced to ODNs by PSM reaction with ammonia in acetonitrile

In addition, an ODN containing O^6 -methylhypoxanthine (O^6 -MHx) was prepared by post-DNA synthesis displacement using a commercially available 6-chloropurine containing ODN (figure 3.15). The modified ODN was treated for 4 days at 37°C with MeOH and DBU in acetonitrile instead of the 2 days used for sulfone modified DNA. After treatment with aqueous ammonia for three days to remove the fast/mild deprotection groups on the other bases in the sequence, we recovered our desired product, which was purified by RP-HPLC in the same manner as for other ODNs. The reason we were interested

in using an ODN containing this modified base was to observe how important the N^2 -amino group is for recognition of O^6 -alkylguanines by AtI1.

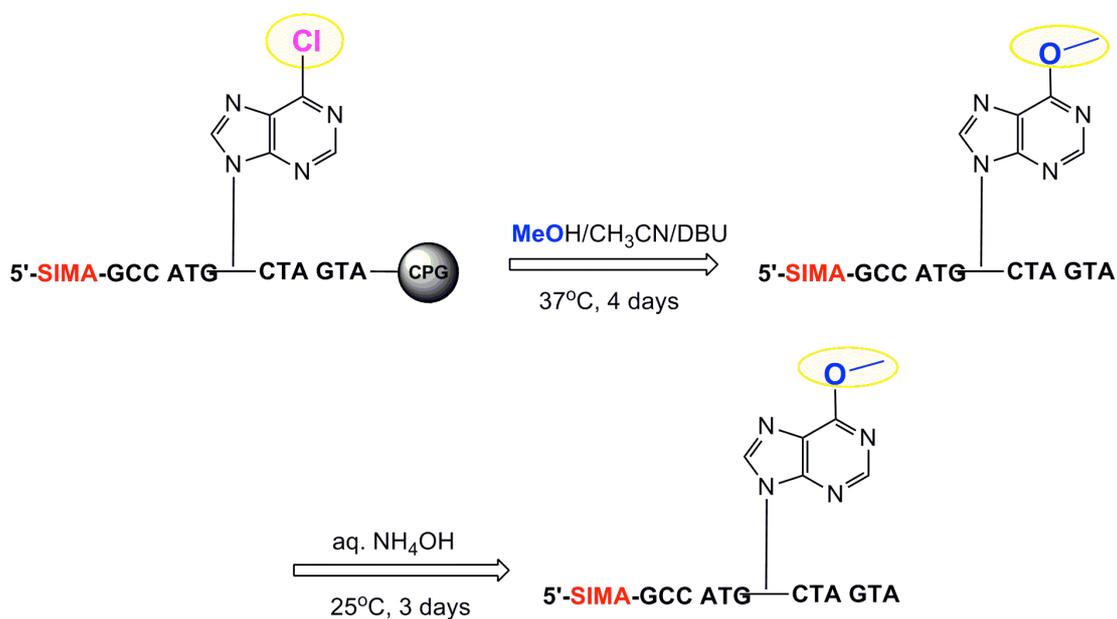


Figure 3.15: PSM reaction using a reactive 6-chloropurine base and methanol to prepare O^6 -methylhypoxanthine (O^6 -MHx)

The other ODNs that were used in the substrate recognition study of ATIs were made by the standard method, i.e. by incorporation of the appropriate phosphoramidite into the chosen position in the sequence during DNA synthesis (section 3.1). The lesions introduced into ODNs in this manner are shown in figure 3.16. ODNs containing guanine (G), O^6 -methylguanine (O^6 -MeG), S^6 -thioguanine (ThioG) and 2-aminopurine (2-AP) were obtained from DNA Technology A/S. The ODN containing guanine at the same position as the modified base is the control sequence. ODN containing ThioG was prepared to investigate the effect of substitution of the oxygen by sulfur (a larger and less electron dense atom) on recognition by AtI1. The thiocarbonyl group of ThioG forms weaker H-bonds than that of a carbonyl group (due to

the reduced polarity of the C=S bond) and this could probe whether the protein recognises guanine by donation of an H-bond to O6. Similarly, the ODN containing 2-AP allowed the investigation of the role of the O6-alkyl group in recognition.

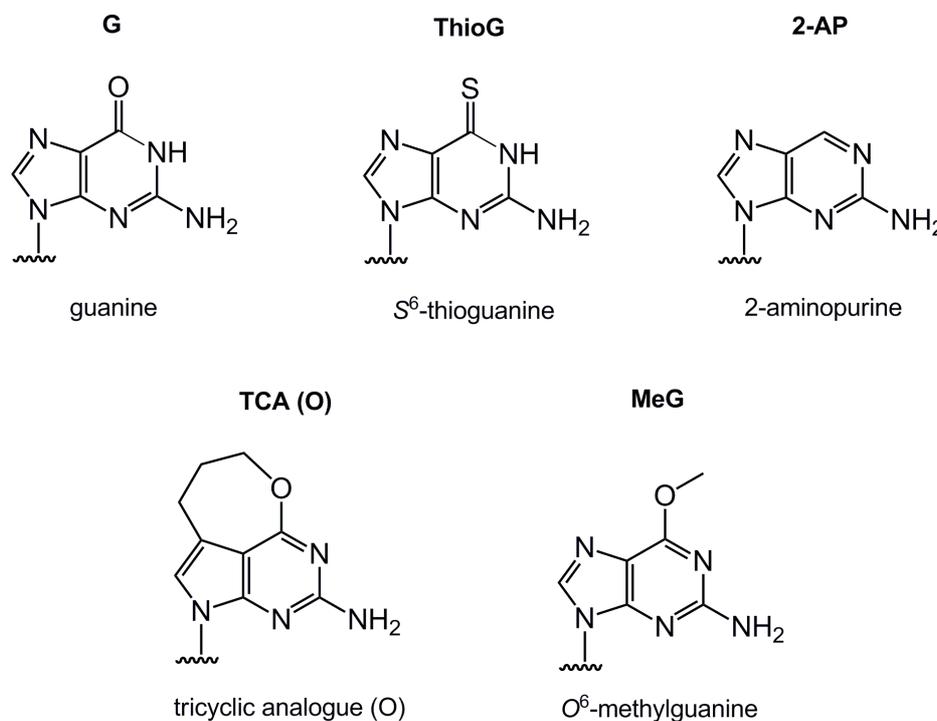


Figure 3.16: Modified bases incorporated into ODNs by standard phosphoramidite method

The ODNs containing the tricyclic guanine analogue were made in Sheffield using a phosphoramidite that was synthesised by Kabir Abdu of the Williams group as part of his PhD project. This modified base was originally conceived as a possible crosslinker for MGMT (the human AGT protein) but was also included in the current study. It is an interesting analogue as it is 'locked' in the *anti*- rather than the *syn*-conformation (relative to the purine ring and the alkyl group). It has been previously reported that alkyl groups in the *anti*-conformation are repaired very slowly (or not at all) by MGMT (95) and

thus it was of interest to examine whether a substrate with an alkyl group locked in the *anti* conformation would be recognised by At1.

5'-GAA CT X CAG CTC CGT GCT GGC CC

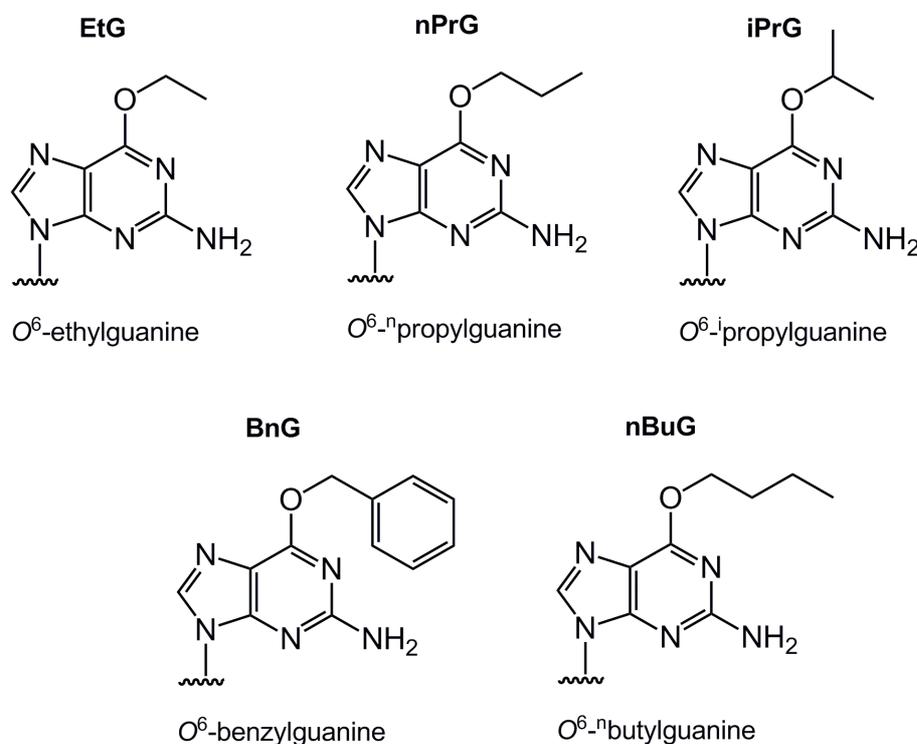


Figure 3.17: Modified bases incorporated into ODNs for SPR-based binding studies with At1

As part of a collaboration with the Margison group, ODNs containing O^6 -ethyl, O^6 -*n*-propyl, O^6 -benzyl and O^6 -*n*-butylguanine were prepared for use as substrates for At1 in SPR-based binding studies (figure 3.17). Whilst most of the reactions worked well as expected, the reaction with isopropanol was not successful under the standard conditions. HPLC analysis of the reaction mixture indicated that the expected product peak appeared fine when the DMT group was intact. However, after purification of the main peak and subsequent deprotection to remove DMT, the ODN product showed multiple peaks in the HPLC trace which made it impossible to purify any further (see

figure 3.18). As a result, the reaction was repeated but the duration of the displacement step was increased from 2 to 7 days at 37°C. However, even under these conditions the reaction was still unsuccessful and resulted in HPLC data that was almost identical. Therefore, another variation was attempted which involved addition of DABCO as well as DBU in the reaction (at the same concentration, section 9.1 for details) in the hope that this would

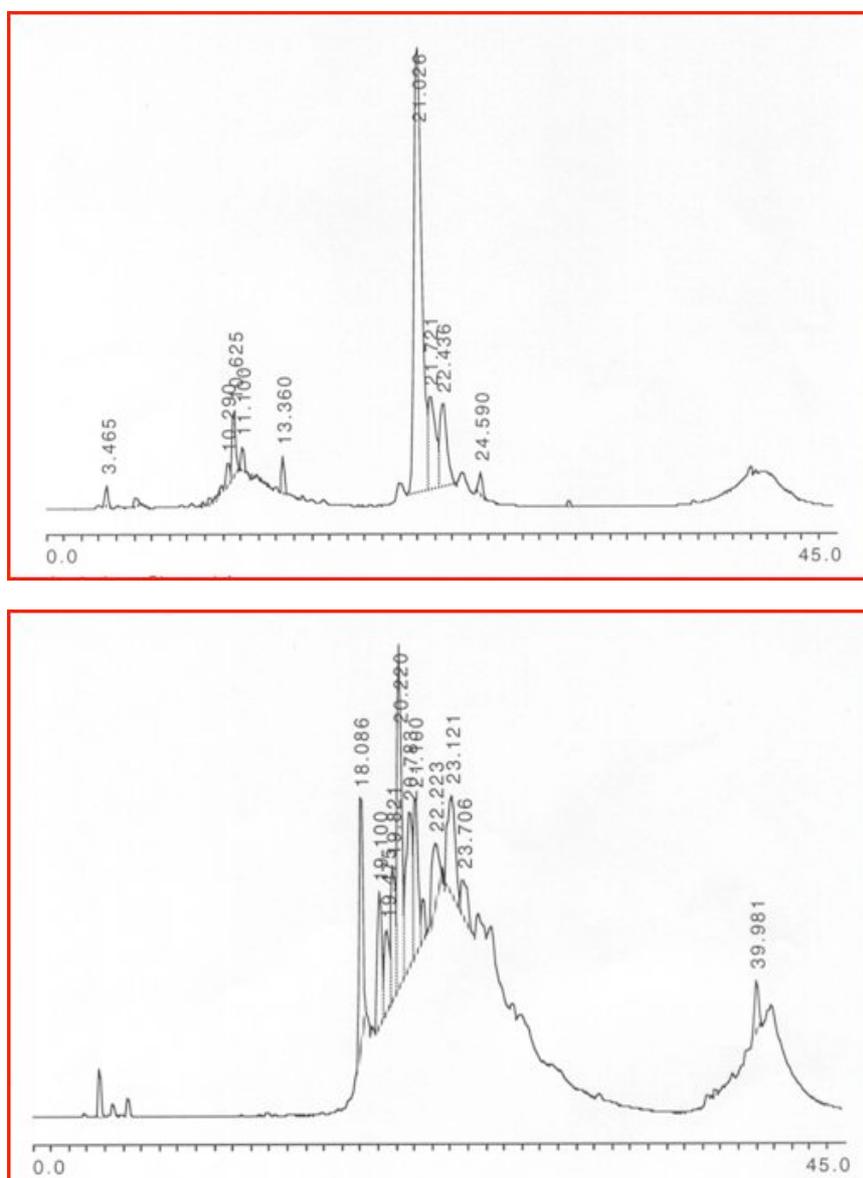


Figure 3.18: HPLC traces of the product from the reaction of ODN with isopropanol, DMT on (top) and DMT off (bottom)

increase the efficiency of the displacement. DABCO would be expected to initially displace the methylsulfonyl group on the purine base to produce an intermediate that is more reactive towards alcohols, due to DABCO being a better leaving group than the methylsulfonyl group (see figure 3.19). However, despite the longer incubation time and higher temperature this reaction also failed to give the desired product and produced an HPLC trace that displayed a multitude of unresolvable peaks akin to those from the previous reactions.

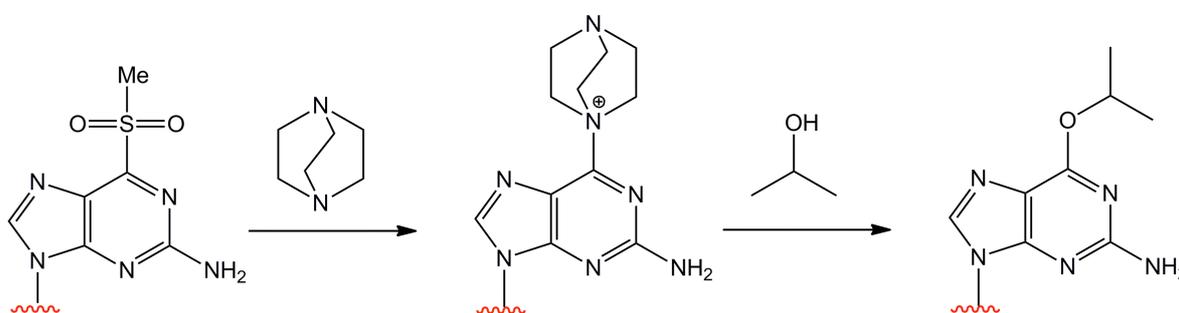


Figure 3.19: Displacement reaction with isopropanol using DABCO

3.3 ODN Substrates for *S.pombe* Affinity Purification Assays

A number of ODN substrates were prepared that could be used as 'bait' in pulldown assays. The intention was to attempt to isolate AtI1, along with other interacting proteins, from whole-cell extracts of *S.pombe* in order to elucidate more information about the role of AtI1 in the NER pathway. The pulldown experiments are described in detail in section 6.3, whilst the preparation of the ODNs is discussed here.

It is known in both humans and *S.cerevisiae* that during NER a patch of between 25-30 nucleotides including the damage is excised from the DNA.(26,96) Consequently it was considered that substrates significantly

longer than this would be required in order to have a good chance of detecting any protein-DNA and protein-protein interactions related to NER. If the ODNs were too short it could hamper the formation of any nascent repair complexes on the DNA. Traditional DNA synthesis by the standard phosphoramidite method is limited to making ODNs of no more than around 100-150 nucleotides in length, as the inefficiencies of the process with each progressive step are incorporated into the final yield. In addition, with this method one is restricted to incorporating base lesions for which there are commercially available phosphoramidites. This effectively means only O^6 -methylguanine can be used with this approach: indeed, two of the 102-mer ODN substrates (those containing O^6 -methylguanine and guanine) were made this way. However, different strategies were required to make the others, as described below.

3.3.1 Preparation of 219-mer ODN Substrates by Primer Extension

Due to the limitations mentioned above, it was decided to make two 219-mer ODNs by the following approach: standard phosphoramidite synthesis of two 120-mer ODNs with short terminal regions complementary to each other, hybridisation of these regions to form an overlapping 'template', and then primer extension by a DNA polymerase I Klenow fragment to give the full-length double-stranded ODN (figure 3.20). Full details of the procedure are given in section 9.6. This approach successfully produced two 219-mer ODNs that were identical except for one key difference: one (OW61) had an O^6 -methylguanine residue in the overlapping region, and the other (OW60) simply had a guanine residue at the same position in the sequence. The products

were analysed on a PAGE gel, which showed bands of the correct size for 219-mer ODNs and demonstrated that the process had been successful (figure 3.21). In addition, duplex 102-mer ODNs OW62 and OW63 that were synthesised commercially and annealed to their complementary ODN (section 3.3.2 and figure 3.22) were analysed on the same gel, and an EMSA was also performed to show that AtI1 would bind to OW61 (containing O^6 -MeG) but not to OW60 (the control ODN) (figure 3.21).

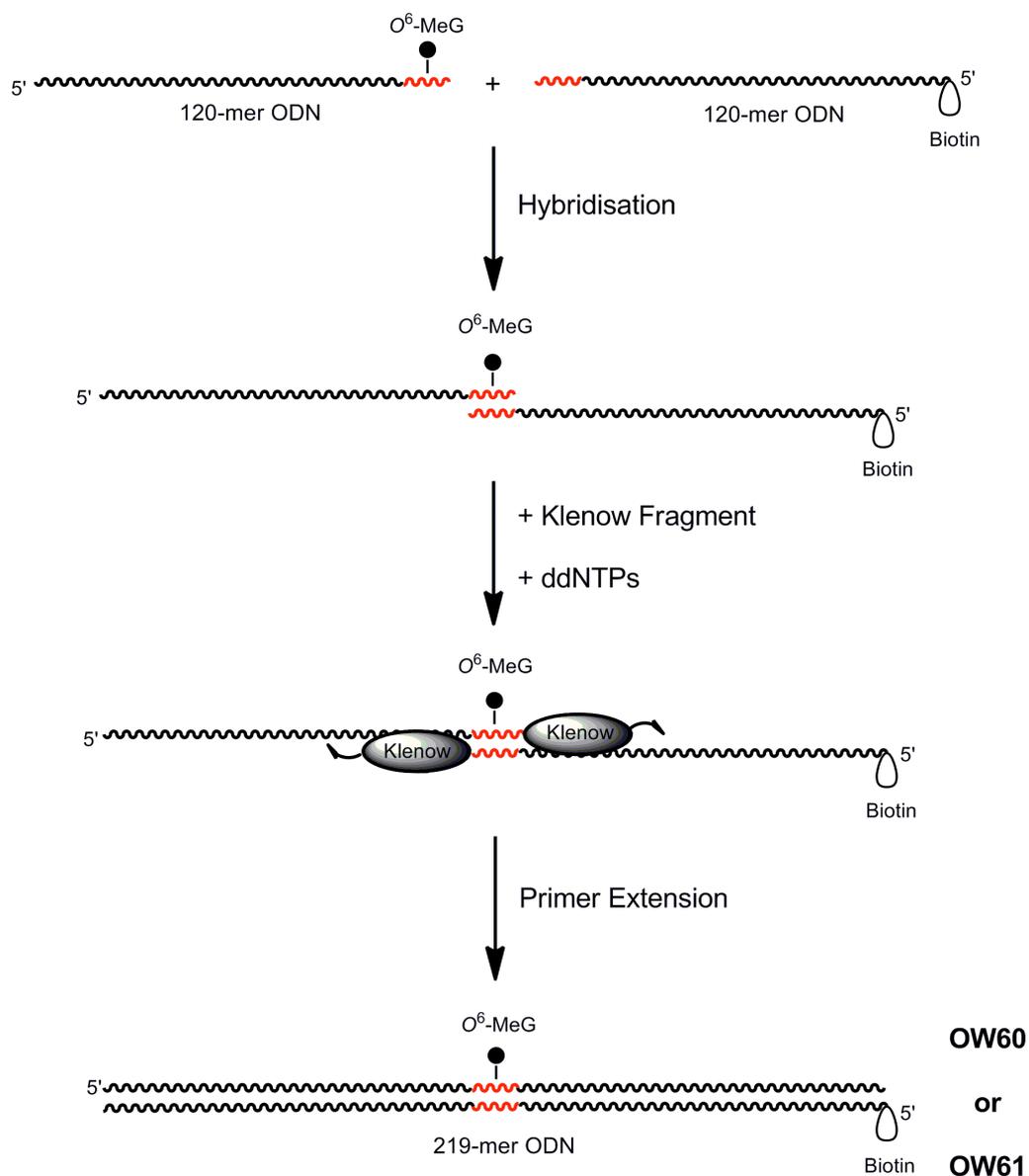


Figure 3.20: Synthesis of double-stranded 200-mer ODNs by primer extension

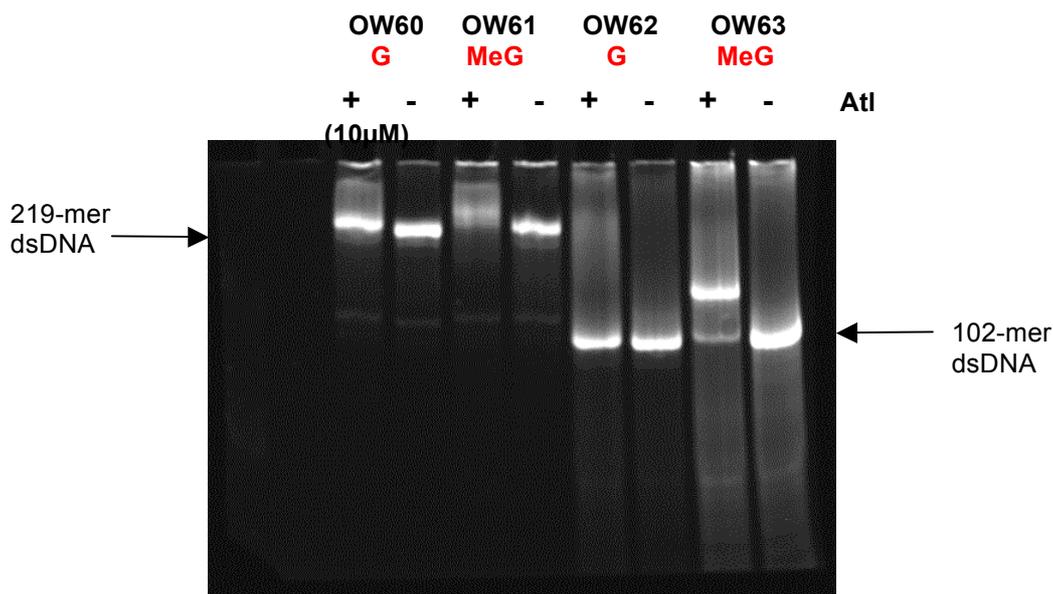


Figure 3.21: PAGE analysis of double-stranded 200-mer ODNs prepared by primer extension

3.3.2 Preparation of 102-mer ODN Substrates by Annealing and Ligation

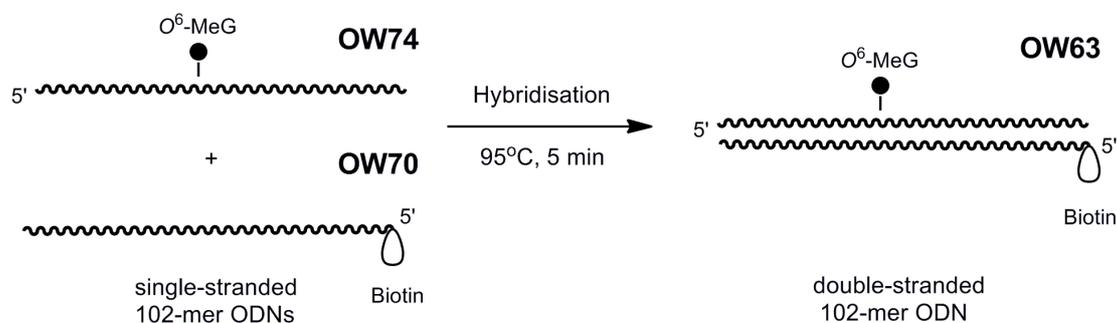


Figure 3.22: Annealing of single-stranded ODNs made by standard DNA synthesis to form duplex DNA substrates (sequences of ODNs and code names are in section 9.2)

The single-stranded 102-mer ODNs containing O^6 -methylguanine (OW74) and guanine (OW73) were synthesised using the standard phosphoramidite method, along with the 5'-biotinylated 102-mer ODN

complement (OW70). These ODNs could simply be annealed together to form the double-stranded substrates for use in the pulldown assays (figure 3.22, OW73+OW70 = OW62 and OW74+OW70 = OW63). However, an analogous double-stranded substrate containing an O^6 -benzylguanine residue was also required. It was not possible to have this ODN made commercially (a phosphoramidite for incorporating O^6 -benzylguanine is not available), and so

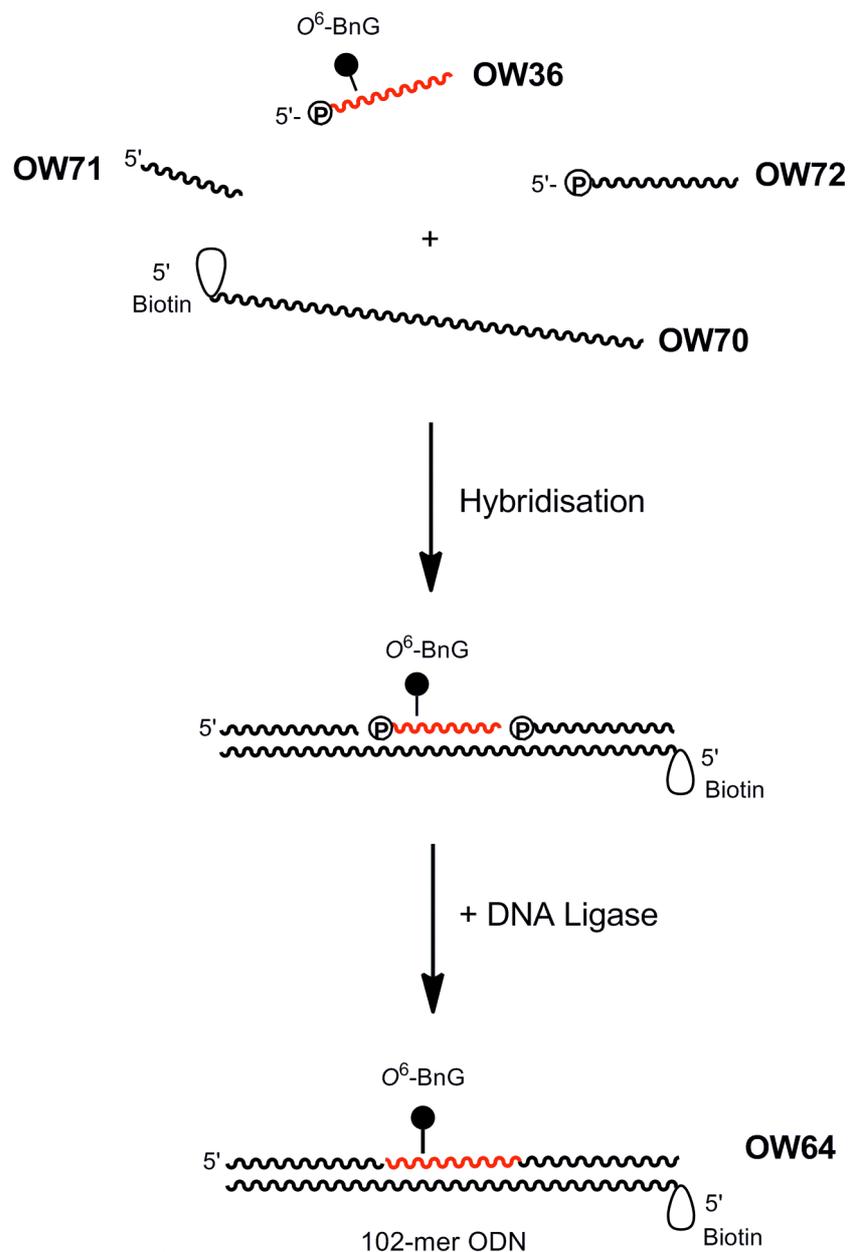


Figure 3.23: Synthesis of double-stranded 102-mer ODN by annealing followed by ligation

it was decided to prepare it by ligation: by joining shorter, phosphorylated ODN fragments together with DNA ligase, using the 102-mer complement as a template. This approach is shown in figure 3.23. Three short ODNs (OW72, OW36 and OW73) that were complementary to sequential regions of the biotinylated 102-mer (OW70) were synthesised. For the ligation to be successful, it was necessary that OW36 and OW72 contained a phosphate group on the 5'-terminus, which is required by DNA ligase to form the phosphodiester linkage of the DNA backbone and thus make a continuous strand. After mixing together, the four single-stranded ODNs were hybridised together by heating to 90°C for 3 min in 50mM NaCl and being allowed to cool slowly. After the efficiency of annealing had been checked by PAGE, DNA ligase was added and the reaction incubated at 16°C for 16h. This produced a double-stranded 102-mer ODN product (OW64), as confirmed by PAGE analysis (by comparison to OW63, the dsODN containing O^6 -methylguanine which was made by a standard annealing process). EMSA also demonstrated that OW64 (and also OW36 annealed to its complement) were bound by AtI1 (figure 3.24). Of course, to produce the modified, double-stranded DNA substrate described above it was necessary to make the appropriate ODN containing an O^6 -benzylguanine residue and a 5'-phosphate group (OW36, shown in red in figure 3.23). In order to do this, an ODN containing the 2-amino-6-methylsulfonyl-purine phosphoramidite was prepared, also using a CPRII phosphoramidite (Glen Research) at the 5'-terminus. This CPRII group was later converted into a 5'-phosphate (figure 3.25). The displacement chemistry is described in detail earlier in this chapter and proceeded here with no difficulty to give the expected ODN product containing O^6 -benzylguanine

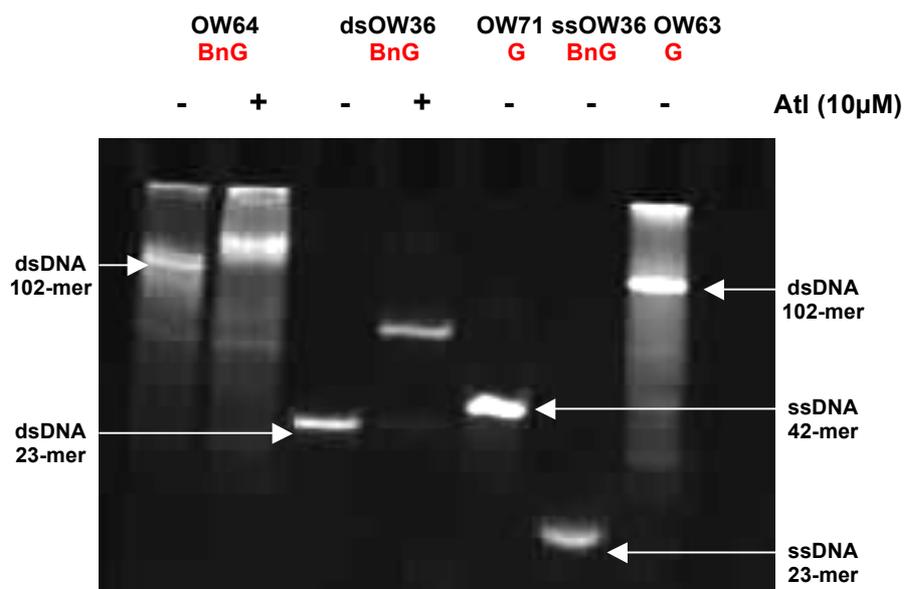


Figure 3.24: PAGE analysis of the ODN product (OW64) of the ligation reaction

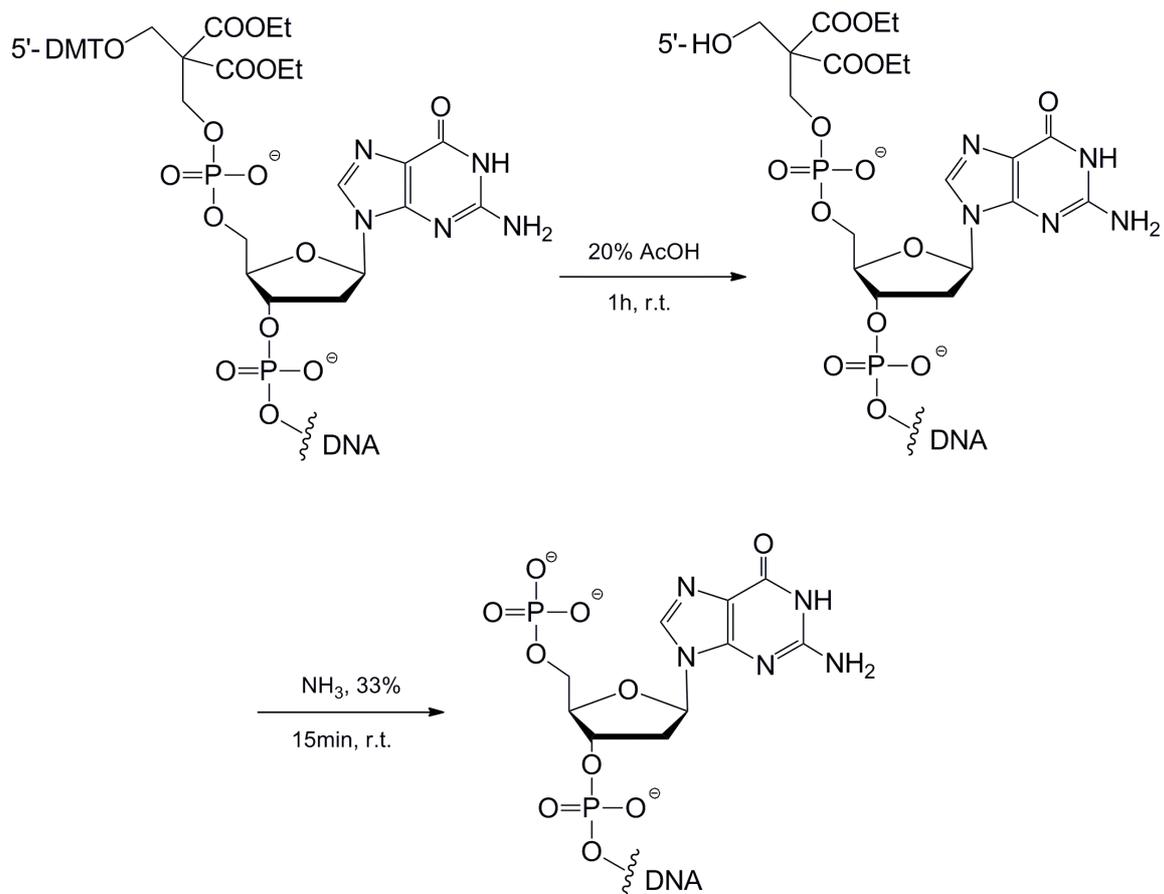


Figure 3.25: Deprotection and conversion of the CPRII reagent to a yield a 5'-phosphate

and the CPRII group with the 5'-DMT still intact. This was purified by HPLC (DMT-ON), the DMT protecting group removed by treatment with 20% (v/v) acetic acid for 1h and the DNA dried to a pellet. In order to eliminate the remainder of the 5'-CPRII group the pellet was simply treated with aqueous NH_3 (33%) for 15 min at room temperature. The product ODN (OW36, figure 3.26) was then purified by RP-HPLC and characterised by ES-MS.

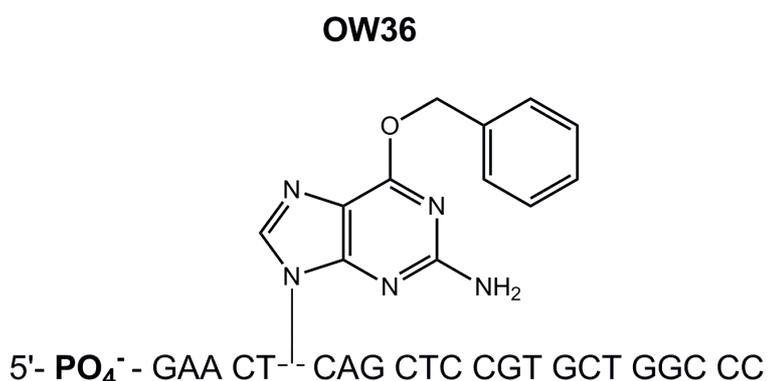


Figure 3.26: 5'-phosphorylated 23-mer ODN containing O^6 -benzylguanine (OW36)

3.4 ODNs for Fluorescence-based MGMT Activity Assay

In a different part of the project, we wished to develop a non-radioactive, fluorescence-based assay for measuring the inhibition of MGMT alkyltransferase activity by various modified ODN substrates. To this end, an assay was devised (section 4.5) using probes that exploit the properties of ODNs known as molecular beacons (section 4.1.4). The molecular beacon ODNs OW31 and OW39 are shown in figure 3.27. These were synthesised using a 3'-BHQ-label (Black Hole Quencher 2, Glen Research) on the CPG. BHQ-2 is compatible with ammonia deprotection and exhibits an excellent coupling efficiency (<http://www.glenresearch.com/GlenReports/GR17-14.html>). The O^6 -methylguanine residue (in OW31) is placed in the sequence

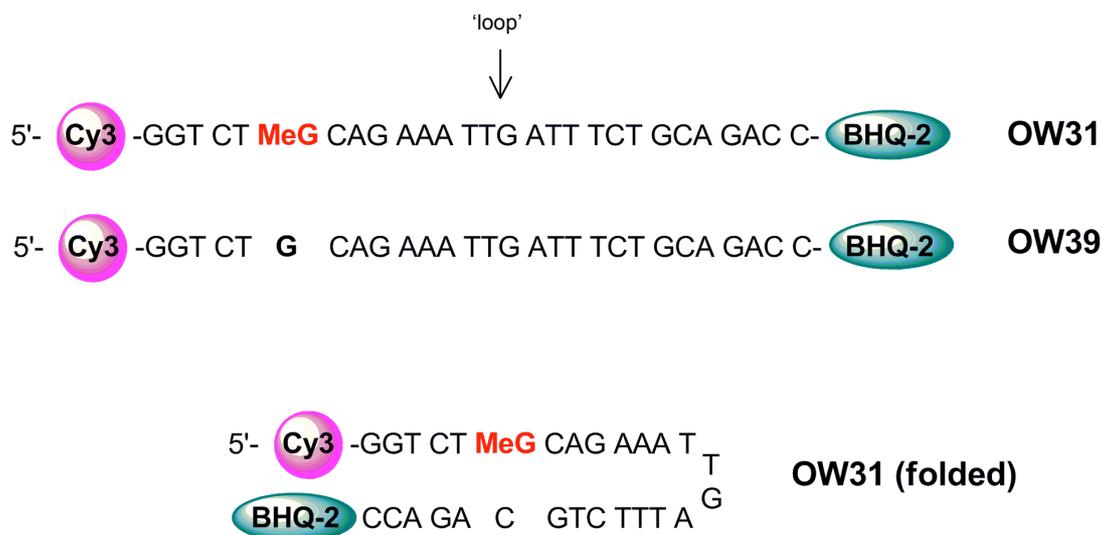


Figure 3.27: Molecular beacon ODNs for use in fluorescence-based MGMT activity assays

using the appropriate phosphoramidite; in the control sequence (OW39) this is simply replaced with guanine. The ODNs were also labelled at the 5'-terminus with a Cy3 fluorescent dye. These sequences have been deliberately designed so that the ODNs will self-hybridise into hairpin stems (i.e. fold back on themselves so that single-stranded structure becomes double stranded). The formation of this secondary structure will also bring the 5'-Cy3 and 3'-BHQ-2 labels into close proximity.

The principles of molecular beacons are described in 4.1.4. Briefly, BHQ-2 is a dark quencher molecule produced by Biosearch Technologies Inc. that is completely non-fluorescent (figure 3.28). It produces its quenching effect by having a large absorbance maximum (and correspondingly high extinction coefficient) at a similar wavelength to the emission maximum of various fluorophore dyes, including Cy3 (BHQ-2 $A_{\max} = 579\text{nm}$, Cy3 $E_{\max} = 563\text{nm}$). In essence, these ODNs were designed based on the properties of molecular beacons so that when the terminal labels are close to each other,

BHQ-2 will quench the fluorescent signal of Cy3. In fact, this contact quenching is 93% efficient for this dye-quencher pair.(97) When they become separated from each other, for example if the DNA were cut with a nuclease and the strands dissociated from each other, this quenching effect would no longer take place and the fluorescent emission signal of Cy3 should increase accordingly.

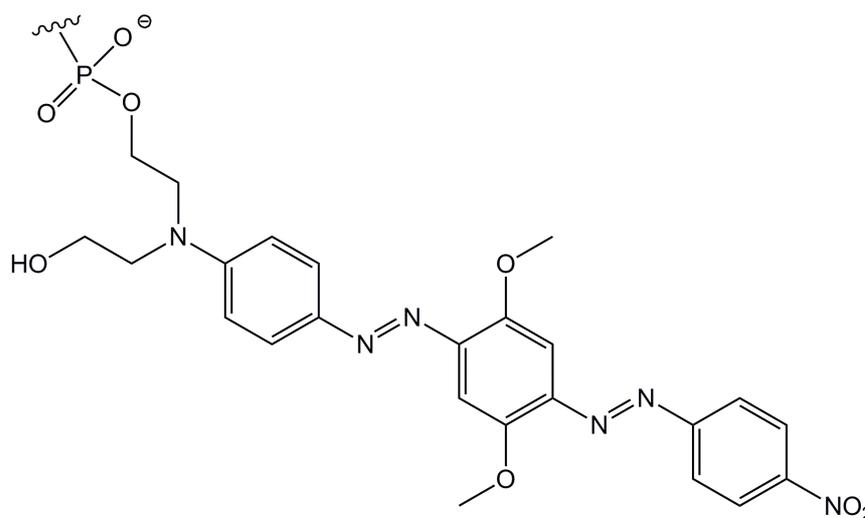


Figure 3.28: Structure of dark quencher molecule BHQ-2 at the 3'-terminus of an ODN

3.5 ODNs for Structural Studies of O^6 -carboxymethylguanine-Containing DNA

O^6 -carboxymethylguanine (O^6 -CMG) lesions in DNA are caused by the endogenous intestinal *N*-nitrosation of glycine and are associated with increased risk of gastrointestinal cancer.(69) Therefore, it was desirable to investigate the pairing geometry of duplexes containing O^6 -CMG. In collaboration with Professor Akio Takenaka and co-workers of Iwaki Meisei University, Japan, two ODNs that were suitable for use in structural studies of double-stranded DNA containing O^6 -CMG were prepared. The sequences

used are the well-characterised Dickerson dodecamers: they are 12-mer ODNs that will hybridize to form self-complementary duplexes.(98-100) ODNs containing the convertible base 2-amino-6-methylsulfonylpurine were synthesised by the standard phosphoramidite method, and were then modified using the displacement chemistry described in 3.2 to form ODNs containing (O^6 -CMG) (figure 3.29). They were purified by RP-HPLC and characterised by ES MS as described previously. The ODNs were then sent to Japan so Prof. Takenaka and co-workers could crystallise them and solve the X-ray structures.



where $X = O^6$ -CMG

Figure 3.29: Modified Dickerson dodecamers containing O^6 -CMG for use in structural studies

Both modified dodecamers formed a right-handed double helix with B form conformation similar to those of unmodified duplexes. The O^6 -carboxymethyl group on the damaged nucleotide protrudes into the major groove and as such does not disrupt the DNA conformation. In OW29, O^6 -CMG (X) is wobble-base paired with cytosine (C) as expected, with two hydrogen bonds present between $N^1(X)\dots N^4(C)$ and $N^2(X)\dots N^3(C)$.

Interestingly, there is also an additional hydrogen bond between the carbonyl of the X group and N⁴ of cytosine which stabilises this base pair (figure 3.30 and 3.31). In OW30, X forms a pair with thymine (T) (with hydrogen bonds between N²(X)...O²(T) and N³(X)...N³(T)) that has a Watson-Crick type geometry, with propeller twisting to release the unfavourable O⁶(X)-O⁴(T) interaction. In addition, there is an interaction between the hydroxyl hydrogen of X and the O⁴ of T (figure 3.30).

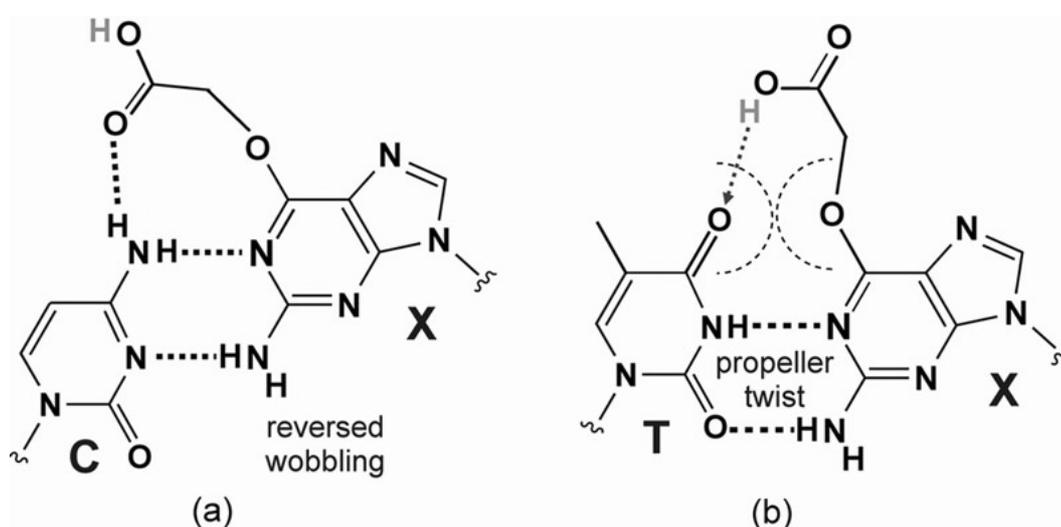


Figure 3.30: Base-pair interactions between O⁶-carboxymethylguanine (O⁶-CMG) and cytosine (a) and thymine (b) observed in crystal structures

When DNA polymerase was modelled in complex with OW29 and OW30, it suggested that both pairing modes would be acceptable to the enzyme. Hence, the presence of O⁶-CMG in a DNA template is capable of directing the incorporation of both dCTP and dTTP into the newly synthesised DNA and leads to GC→AT transition mutations. This corroborates existing evidence of the mutagenic effects of these lesions and is likely to be, in part, the origin of increased risk of gastrointestinal cancer.

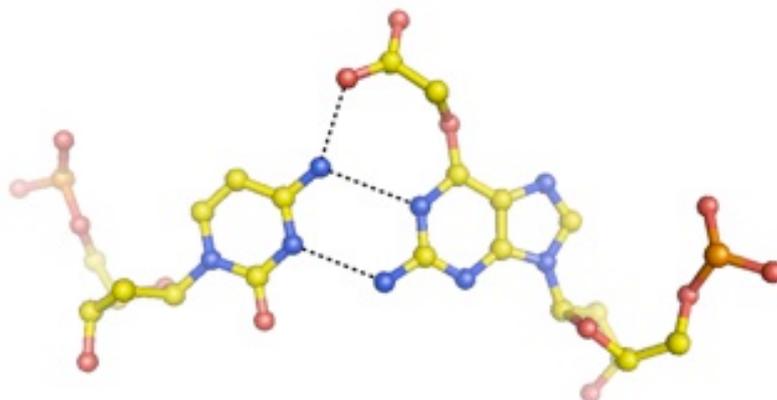


Figure 3.31: Stabilisation of the O^6 -CMG: T base-pair by an H-bonding interaction between the carboxymethyl carbonyl group and the thymine N4 atom

4.0 Fluorescence-based Methods for Studying ATL and AGT Proteins

There are many techniques for studying and quantifying protein-DNA interactions. It was decided to primarily use fluorescence-based methods, such as fluorescence anisotropy (FA), total fluorescent emission intensity (TFEI) and Förster resonance energy transfer (FRET) for this project. All these methods involve the labelling of an ODN with a fluorescent tag, which can then be monitored using a fluorimeter: changes in its signal provide information about the environment or properties of the label. In terms of protein-DNA interactions, fluorescent assays such as FA and TFEI are solution-based, real time, true-equilibrium techniques and as such may give them certain advantages over other methods of calculating dissociation constants (such as electrophoretic mobility shift assays (EMSA), surface plasmon resonance (SPR) or enzyme-linked immunosorbent assay (ELISA)).(101) Compared to SPR or ELISA, where typically one of the partners in the interaction is immobilised, steric or entropic factors that may affect the value of the K_D will be less of an issue and in this sense assays of this type may give more accurate and relevant binding data. The theoretical aspects of these fluorescence-based methods and experimental design are examined in more detail in this chapter.

4.1 Introduction to Fluorescence-based Methods

4.1.1 Fluorescence Anisotropy (FA)

Fluorescence anisotropy is a technique often used to monitor protein-DNA interactions, having been first utilised in this context by Heyduk and Lee in 1990.⁽¹⁰²⁾ It has since been used to evaluate a variety of interactions between proteins including EcoKI methyltransferase, EcoRV restriction endonuclease and Archaeal family B DNA polymerases and their DNA substrates.^(87,103,104) The principles of the method are based on the fact that fluorescent molecules have an excitation and an emission dipole, and that there is a short time lapse (the fluorescent lifetime) between the absorbance of a photon that excites the fluorophore and the subsequent emission of a fluorescent photon. Polarised light will only excite molecules in the solution that have the correct orientation; that is, their dipole is aligned with the incident light. Fluorophores will continuously be tumbling into alignment with the polarised light, being excited, and then emitting photons at an angle based on their position after the fluorescent lifetime has elapsed. If a molecule is rotating faster, this angle will be greater relative to the original position of the fluorophore. When a number of molecules are moving quickly the emitted light is depolarised more effectively which leads to a larger difference in the angle between the incident and emitted light. This leads to a low anisotropy measurement as less of the emitted light is still polarised. When molecules are moving at a slower rate, there is more retention of emission polarisation, and the anisotropy value will be correspondingly higher (figure 4.1). The spectrometer allows measurement of how much polarised

light is emitted in the vertical and horizontal plane and then the anisotropy is calculated from these measurements. Anisotropy is defined by:

$$\text{anisotropy} = (I_{VV} - I_{VH}) / (I_{VV} + 2I_{VH})$$

where I_{VV} and I_{VH} are the intensities of the vertical and horizontal components of the emitted light using vertical polarised excitation.

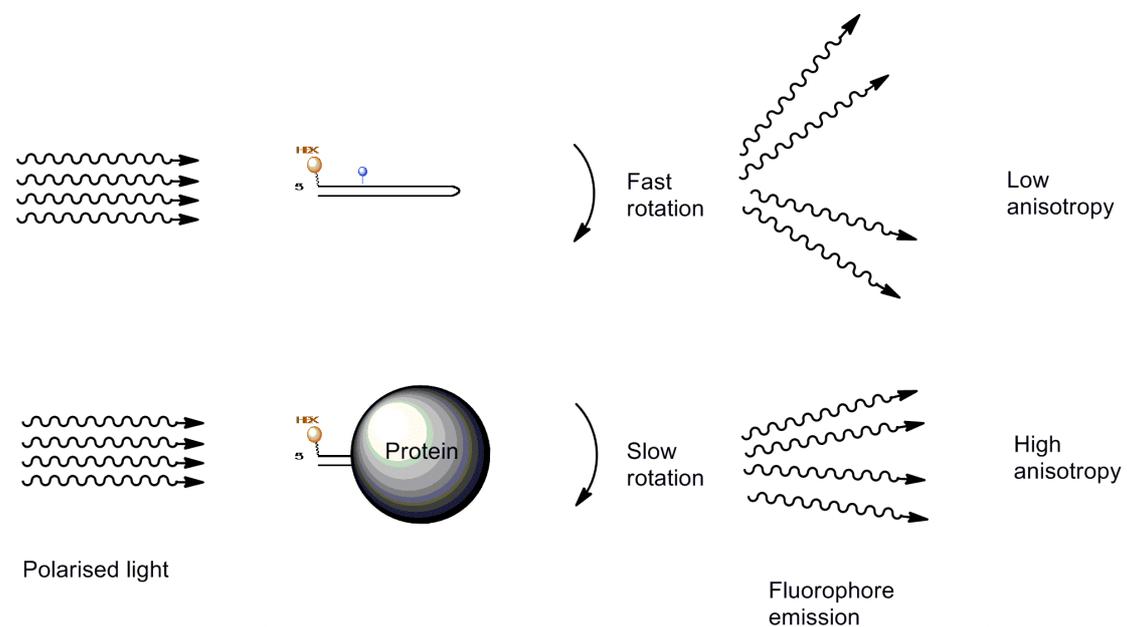


Figure 4.1: Principles of fluorescence anisotropy: free fluorescently-labelled DNA in solution has low anisotropy, which increases as DNA becomes bound by protein

Fluorescence anisotropy can be used to measure protein-DNA interactions due to the fact that a DNA-protein complex is larger and bulkier than a free DNA molecule and hence will tumble more slowly in solution. This means when all the fluorescently-labelled DNA is unbound (and therefore tumbling quickly) the anisotropy value will be relatively low, and as the DNA is bound by protein the value will become higher as the complex tumbles more

slowly and a larger proportion of the emitted light remains polarised or anisotropic. The change in anisotropy thus gives a measure of the fraction of the DNA that is bound, and can be used to calculate the dissociation constant (K_D).

Oligodeoxyribonucleotides (ODNs) can be labelled with fluorescent labels in order to be used for this purpose. Hexachlorofluorescein (HEX) is a useful dye as it is available commercially as a phosphoramidite for labelling the 5'-terminus of the ODN and has a sensitivity that allows K_D values in the low nanomolar (nM) range to be measured.⁽⁸⁷⁾ A range of ODNs containing both a HEX tag and one of the many O^6 -alkylguanine lesions would allow dissociation constants of various different DNA-AtI1 complexes to be measured. This would give information about the binding preferences of AtI1 for various types of O^6 -alkylguanine adducts in DNA, allowing a systematic and quantitative study of AtI1 substrate recognition.

Dissociation constants (K_D) have been reported for AtI1 bound to ODNs containing various O^6 -alkylguanine lesions using enzyme-linked immunosorbent assay (ELISA) and surface plasmon resonance (SPR). As described earlier, these methods involve immobilising the ODN of interest onto a surface before binding the protein and measuring data from which the K_D can be derived. Fluorescence anisotropy has the advantage of being a solution-based equilibrium technique which therefore should provide more reliable values for the dissociation constants for the binding of AtI1 to DNA.

4.1.2 Total Fluorescent Emission Intensity (TFEI)

A related fluorescence-based technique that can also be used to quantify molecular interactions involves measurement of changes in total

fluorescent emission intensity upon titrating protein into a solution containing the labelled ODN. It is important to note that this method can only be used if the specific interaction between the protein and DNA substrate causes an effect on the fluorescent label, either quenching or enhancing its fluorescent signal. If, for instance, addition of protein into a solution containing fluorescent ODN causes a concentration-dependent change in total fluorescent emission intensity (TFEI) then this type of titration may be used rather than fluorescence anisotropy (figure 4.2). Typically this change will involve a quenching effect caused by the change in the local environment of the fluorophore, from a hydrophilic environment when unbound in solution to a more hydrophobic one when bound by and therefore in close proximity to the protein.

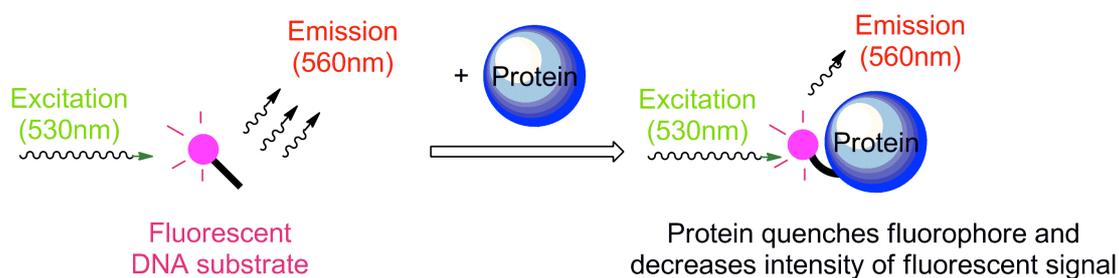


Figure 4.2: Principles of using total fluorescent emission intensity to measure protein binding to ODN substrates.

The advantage of this method is that lower concentrations of ODN may be used as the corresponding fluorescent signal is greater than that of FA, which is useful if the K_D values that are being measured are in the low nanomolar-to-picomolar range. The titrations can also be performed on any fluorimeter, rather than one that is equipped to permit FA measurements to be

taken: hence the experiments are somewhat more versatile. Finally, the experiments themselves require less time to complete.

4.1.3 Förster Resonance Energy Transfer (FRET)

Förster Resonance Energy Transfer, or FRET, is a process of nonradiative energy transfer between two chromophore molecules. During this process, one chromophore acts as the donor and transfers excitation to the other (the acceptor) without the emission of a photon. The theory was developed by Theodore Förster in the late 1940s, and it has since become widely used for biological applications. The efficiency of energy transfer is dependant on distance ($1 / r^6$, where r = distance between donor and acceptor) which effectively means that it operates over about 10-100 Å, of similar dimensions to many biological macromolecules. This makes it a useful tool for studying, for example, protein-protein, protein-DNA interactions and RNA structure. The efficiency of FRET is determined by the following factors:

- 1) the proximity of donor and acceptor (these need to be in 10-100 Å distance range)
- 2) the overlap of the absorption spectrum of the acceptor with the emission spectrum of the donor (figure 4.3)
- 3) the relative orientation of the transition dipoles of donor and acceptor

If these conditions are correctly satisfied, then FRET occurs: that is, the fluorescent excitation of the donor is transferred to the acceptor, quenching the fluorescence of the donor whilst increasing the fluorescent emission

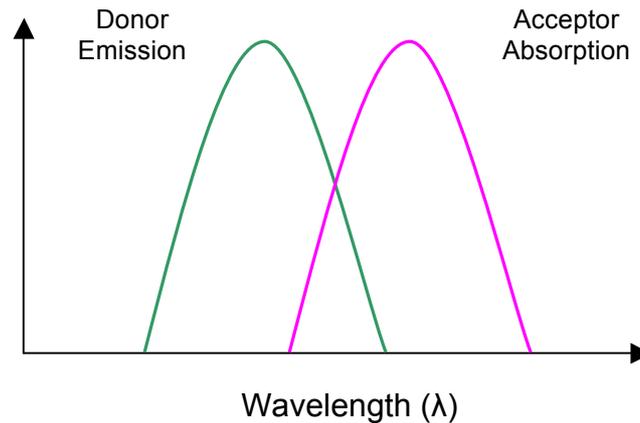


Figure 4.3: Spectral overlap required for FRET to occur between a donor-acceptor pair

intensity of the acceptor. The closer the donor and acceptor are to each other, the greater the transfer, and vice versa. This is given by the equation:

$$E = \frac{1}{1 + (r / R_0)^6}$$

E = FRET efficiency

r = distance between donor and acceptor

R₀ = Förster distance (the distance where the FRET efficiency is 50% for a specific donor-acceptor pair)

In practice, when using FRET to study biomolecules, the donor and acceptor pair are usually fluorophore dyes, such as fluorescein, cyanine etc. There are many such commercially available dyes which have the required overlapping emission/excitation spectra to be suitable for FRET studies. When covalently attached to proteins or DNA they can be used to detect

conformational changes or bimolecular interactions, due to the fact that changes in their proximity will cause relatively predictable changes in the FRET signal. For example, if two complementary ODNs of a suitable length (i.e. less than 30 base pairs) have appropriate dye molecules at their 5'-terminii, then annealing them together should cause them to become close enough to each other to observe FRET.

4.1.4 Molecular Beacons

There are ways other than FRET that a fluorophore can be quenched of its fluorescent emission energy. For example, static quenching involves the formation of a non-fluorescent ground state complex between a dye molecule (such as FAM, Cy3 etc.) and a so-called dark quencher chromophore (such as DABCYL, BHQ-2 etc.). The complex is stabilised by induced dipole and hydrophobic interactions and as such is efficient only when the dye-quencher pair are in close proximity. This static quenching effect is utilised in functionalised ODNs known as molecular beacons.

Molecular beacons were first reported by Tyagi *et al.* in 1996.⁽¹⁰⁵⁾ They designed nucleic acid probes that recognised and reported the presence of specific nucleic acid sequences in solution. The molecular beacons are ODN hairpin stem structures with a 5'-fluorophore (EDANS) and a 3'-quencher (DABCYL) (figure 4.4). They also contain a single-stranded loop region, which contains a sequence that is complementary to the target nucleic acid. When in the hairpin conformation, which is held together by 5-8 base pairs forming a stem structure, the fluorophore and quencher are close enough together for the emission energy of EDANS to be transferred to DABCYL and dissipated as heat, therefore generating no fluorescence.

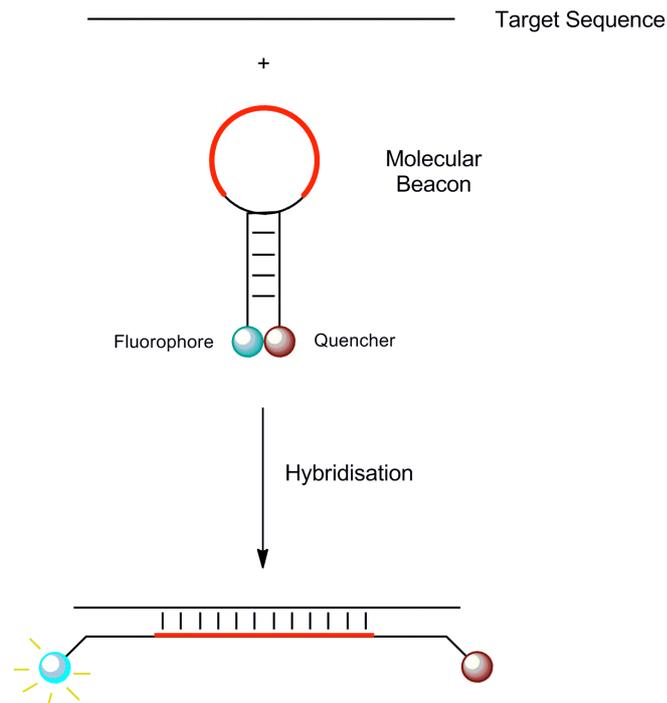


Figure 4.4: Molecular beacon ODNs used as probes for sequence detection by Tyagi *et al.*

However, once the molecular beacon encounters the complementary target sequence the loop region hybridises to it, driven by the thermodynamic forces inherent in the formation of many more base pairs than are present in the stem. Consequently, EDANS and DABCYL become estranged and this leads to removal of the quenching effect and a corresponding increase in the fluorescent emission signal of EDANS. Quantifying this increase allows the degree of hybridisation to be monitored easily and accurately.

There are now many dye-quencher pairs available (97) and molecular beacons are used in a huge range of biological applications.(106) For example, molecular beacons have subsequently been used to detect enzymatic activity *in vitro*. Li *et al.* developed a highly sensitive and convenient fluorescence-based assay to monitor the cleavage of single-stranded ODN substrates by single-strand-specific DNA nucleases (107). In

addition, a molecular beacon assay has been developed by Maksimenko *et al.* to measure the base-excision repair activities of DNA glycosylase and AP endonuclease enzymes.(108) Both these assays have in common that cleavage of a molecular beacon ODN substrate causes the fluorophore-quencher pair to become separated in space and hence lead to a corresponding rise in the fluorescent signal of the dye. In this way, the enzymatic reactions can be monitored in real time.

It was decided to exploit some of the unique properties of molecular beacons to develop an assay that could assess the alkyltransferase activity of MGMT (section 4.5).

4.2 Measurement of ATL-DNA Dissociation Constants (K_D values)

Fluorescence anisotropy has been used successfully by several groups to study protein-DNA interactions and accurately measure dissociation constants (K_D values) (87,103,104). The Connolly group have used oligonucleotides (ODNs) labelled with hexachlorofluorescein (HEX) to measure K_D values in the low nanomolar range by direct titration. This is due to the sensitivity of HEX which allows ODN concentrations of as low as 1 nM to be used in the assay (87), in contrast to fluorescein (FAM) where higher ODN concentrations must be used. Consideration of the available data (based on SPR experiments conducted by the Margison group) for At11-DNA interactions led us to initially select HEX as the label of choice as it was expected that the dissociation constants would be in the low nanomolar range.

The binding curves were generated by plotting protein concentration against anisotropy (see section 1.16 for definition), and the data was fit to the following equation:

$$A = A_{\min} + \left[\frac{(D+E+K_D) - \left((D+E+K_D)^2 - 4DE \right)^{0.5}}{2D} \right] (A_{\max} - A_{\min})$$

A = the anisotropy measured at a specific concentration of enzyme (E)

D = the oligonucleotide concentration

A_{min} = the lowest measured anisotropy (i.e. when no protein is added)

A_{max} = the highest measured anisotropy (i.e. when the binding is saturated)

K_D = the dissociation constant

Initially, problems were encountered when using this technique to measure AtI1-DNA interactions. In the first experiments MBP-AtI1 fusion protein was used with a single-stranded HEX-labelled 23-mer ODN containing O⁶-methylguanine, O⁶-MeG (OW55, table 9.1 has full details of all ODN names and sequences). Fusion protein was initially used in the titration since it was reasoned that a larger protein (AtI1 = 12,600 Da, MBP-AtI1 = 54,600 Da) would result in a complex with a slower rotation in solution when bound to the DNA substrate and hence cause a greater change in the anisotropy. The 23-mer ODN was the same sequence as that used in the SPR binding assays by the Margison group. However, when we conducted the assay the data points (i.e. anisotropy values) were extremely erratic and deviated markedly from the fitted curve (figure 4.5). This was the case when using both single-stranded (ss) and double-stranded (ds) ODNs in the titrations, and also when

using wild-type AtI1 rather than MBP-AtI1 fusion protein. In addition, the value of K_D calculated from this plot (88nM) was significantly higher than that determined by SPR measurements (1.8nM). Consequently a number of modifications were made to the experiments.

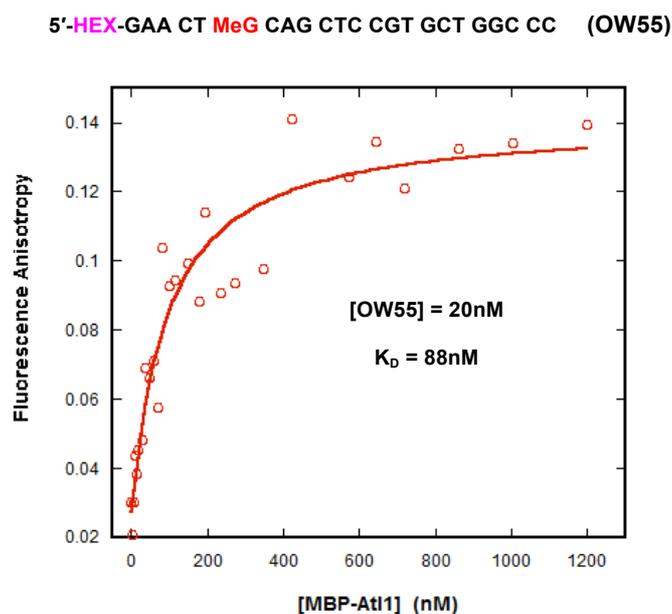


Figure 4.5: Plot to show titration of MBP-AtI1 protein with 5'-HEX-labelled 23-mer ODN containing O^6 -methylguanine

A shorter, 13-mer ODN (OW53) was used to see if this would make any difference. The reasoning here was similar to that for using the fusion protein: a shorter ODN would have greater movement in solution when unbound and would therefore be affected more by protein binding, leading to greater changes in the anisotropy values. The assays performed with OW53 generated slightly less erratic plots, but they were still not acceptable and continued to give comparatively large values for the K_D (≈ 100 nM). It may be that this was due to the limitations of the fluorimeter that the assays were performed on (LS-50B Perkin Elmer), which does not possess a photon-

counting detector and could therefore be considered too insensitive for measurements at low DNA concentrations (in the initial experiments an ODN concentration of 20nM was used which is actually reasonably high). As a result, experiments were undertaken in Newcastle using a fluorimeter (SLM-Aminco 8100) in the group of Professor Bernard Connolly. This instrument was much more sensitive and calibrated specifically for assays of this type which allowed an ODN concentration of 5nM to be used (a lower DNA concentration is preferable in experiments of this kind so that any error in DNA concentration is not extrapolated into the final K_D value) and the plots generated with the data were clearly an improvement on the previous ones (figure 4.6).

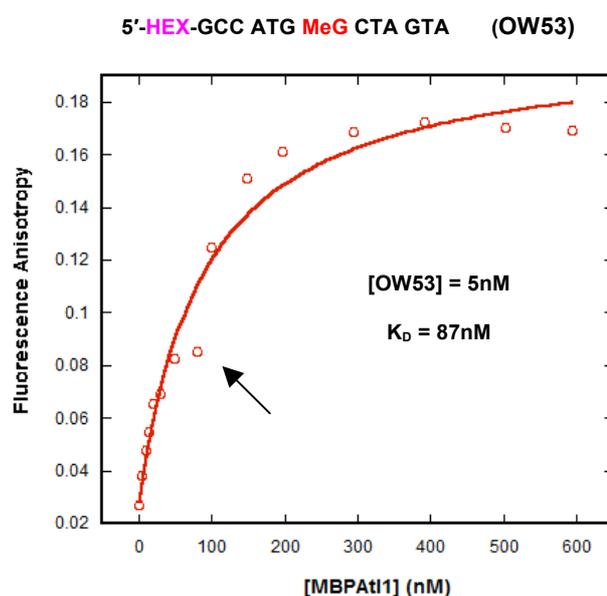


Figure 4.6: Plot to show titration of MBP-At1 with 5'-HEX-labelled 13-mer ODN containing O^6 -methylguanine. The arrow indicates the initial levelling out of the binding curve (see text)

However, the value of the dissociation constant was 87 nM, which was still considerably larger than the value determined previously using SPR.

Furthermore, there appeared to be an initial levelling out of the curve (around 50-80 nM, see arrow on figure 4.6) which occurs before a larger change in the anisotropy takes place. This occurred consistently in repeats of the same experiment in Newcastle and was interpreted as initial tight binding of AtI1, followed by subsequent non-specific binding interactions. When a plot was made and the dissociation constant derived from this initial binding event, a K_D of 15nM was obtained. It was decided that if saturation of the protein-DNA complex formed from the specific binding interaction (to reach an end-point in the plot) cannot be achieved without further non-specific binding interactions affecting the anisotropy measurement then it would be difficult to calculate accurate K_D values. Hence, an alternative way of quantifying the protein-DNA interaction was pursued, by using total fluorescent emission intensity (TFEI) measurements rather than fluorescence anisotropy.

TFEI can be used to monitor protein-DNA interactions if the binding causes a change in the fluorescent signal of labelled ODN. When a saturating amount of MBP-AtI1 was added to HEX-labelled 13-mer ODN containing O⁶-methylguanine (OW53) the fluorescent intensity decreased by approximately 20-25%. Presumably this is due to the proximity of the protein to the HEX label which causes a quenching effect on its fluorescent signal. By measuring fluorescent intensity changes as protein concentration increases, a plot can be generated and used to calculate the dissociation constant. The data is plotted and fit in a similar way to the anisotropy measurements, using the following equation:

$$I = I_{\max} + [(D+E+K_D) - ((D+E+K_D)^2 - (4DE))^{0.5}] (I_{\min} - I_{\max}) / 2D$$

I = the intensity measured at a specific concentration of enzyme (E)

D = the oligonucleotide concentration

I_{\min} = the lowest measured intensity (i.e. when the binding is saturated)

I_{\max} = the highest measured intensity (i.e. when no protein is titrated)

K_D = the dissociation constant

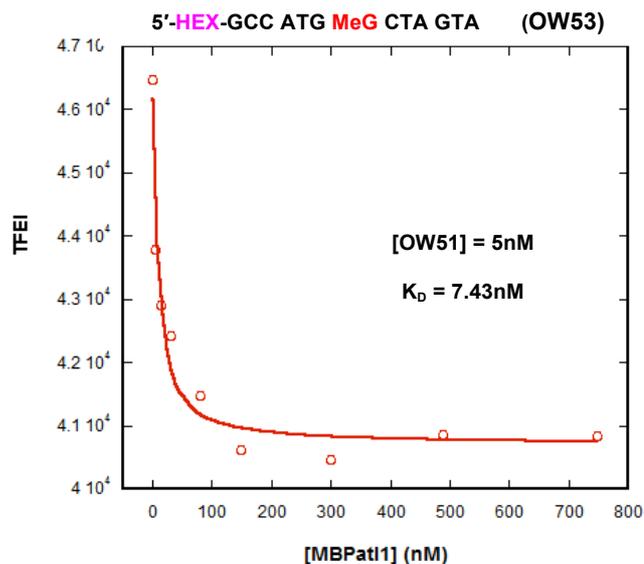


Figure 4.7: Plot to show titration of MBP-AtI1 with 5'-HEX-labelled 13-mer ODN containing O⁶-methylguanine

The plot for the initial experiment with OW53 is shown in figure 4.7. The data points show reasonably good agreement with the fit, and the binding constant was calculated to be 7.4nM, with a very similar value of 6.9nM for double-stranded ODN (i.e. OW51 annealed to its unlabelled complement). Furthermore, the negative control experiment with HEX-labelled but unmodified ODN (i.e. guanine in place of O⁶-methylguanine, OW52)) gave a binding constant that was significantly larger than that for OW53. This was consistent with the specific binding of MBP-AtI1 to ODNs containing alkylguanine lesions and suggested that it would indeed be possible to use

these assays to carry out a study of the DNA substrate specificity of ATL proteins.

From this point on, the assays were performed on a Horiba Jobin-Yvon Fluoro-Max3 fluorimeter, which continued to give us reasonably accurate and repeatable data. However, it was also decided to make a number of improvements to the experiment. Firstly, we changed the fluorescent label from HEX to SIMA(HEX) due to the instability of HEX under basic conditions. This was necessary as an ammonia deprotection step was required in the post-synthetic modification reactions used to produce a range of O^6 -alkylguanine lesions in DNA (section 3.2.1). The ODN labelled with SIMA (OW51) gave a TFEI signal that was similar in peak intensity and wavelength to the HEX-labelled sequence. In addition, At11 (i.e. cleaved from MBP) was used in the experiments. Although controls had been performed to ensure there was no interaction between the substrate ODNs and MBP, it seemed

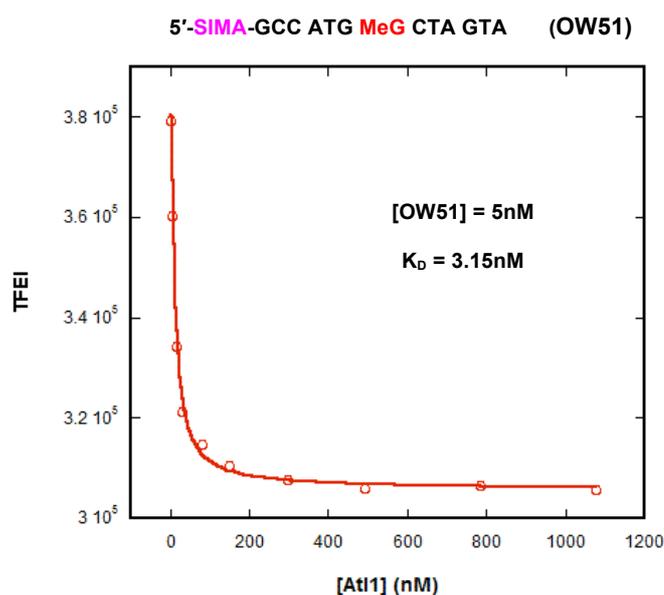


Figure 4.8: Plot to show titration of wild-type At11 with 5'-SIMA-labelled 13-mer ODN containing O^6 -methylguanine

desirable to use native protein in the assays if possible. It was found that titration of wild-type At11 with OW51 resulted in a concentration-dependent reduction of TFEI, of similar magnitude to that observed using MBP-At11 fusion protein (figure 4.8).

Using SIMA label allowed the ODN concentration to be reduced to 1nM in the experiments. This was an advantage as it is desirable to have the DNA concentration below that of the K_D value if possible to ensure true-equilibrium conditions in the assays and prevent stoichiometric binding. In addition, more data points at low protein concentrations were included to improve the accuracy of the assays. The plots of 13-mer ODN containing O^6 -methylguanine (OW51) and guanine (OW50) using these conditions are shown in figure 4.9. In addition, it was decided to compare the binding of ATL proteins to ODNs containing different O^6 -alkylguanine lesions by using ssDNA

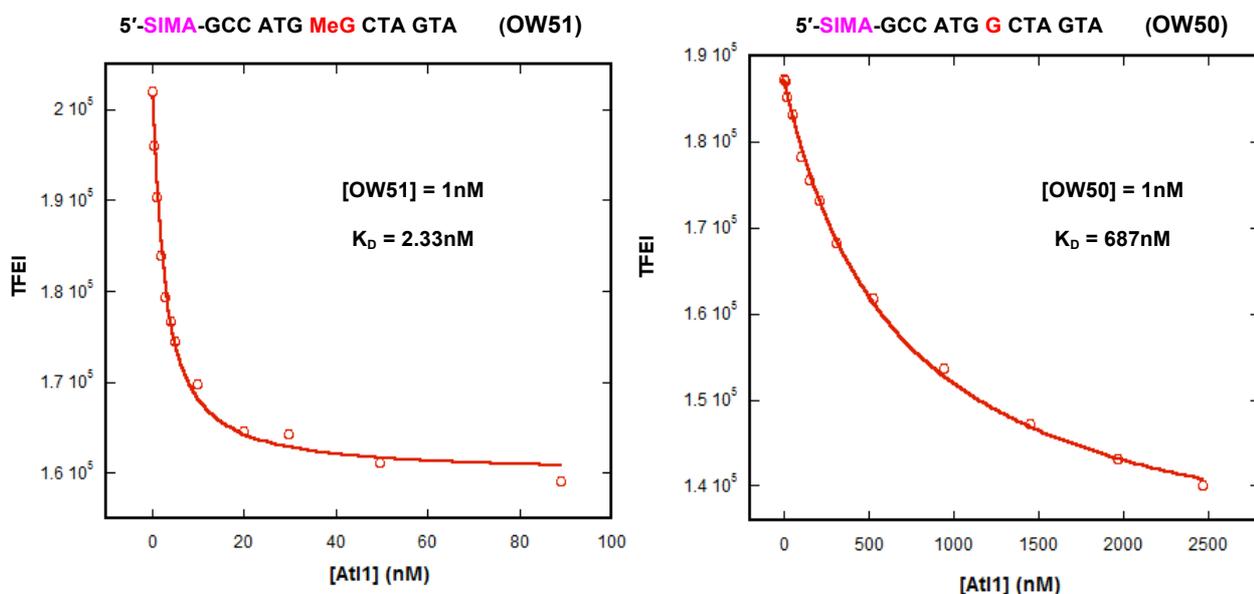


Figure 4.9: Plots to show titration of wild-type At11 with 5'-SIMA-labelled 13-mer ODNs containing O^6 -methylguanine (left) and guanine (right)

in the assays: it meant that there was no need to anneal the substrates to form duplexes, which would introduce the possibility of variable annealing efficiency that could affect the experiment. Another advantage was that for assays using ssDNA rather than dsDNA the fit of the data to the curve was more precise (and hence the value of the error smaller).

Once the titrations had been optimised, ODNs containing a variety of O^6 -alkylguanine and related modifications were synthesised (section 3.2.2). It is worth briefly mentioning a consideration relating to the titrations using the control sequence (OW50). At low protein concentrations, the TFEI of OW50 (the guanine-containing control ODN) changed very little and the fluorescent intensity decreased only at much larger protein concentrations. A consequence of this behaviour was that the value of the K_D was dependent on where the titration was stopped (i.e. the final concentration of protein) as the curve did not flatten out in the same way as for ODNs recognised with high affinity by ATL proteins. It was decided that in order for there to be a fair comparison between the bound substrates and the unbound control, protein would be added until the TFEI had decreased by the same amount as for the bound ODNs (i.e. around 25-30% of the initial value). This gave the value of K_D for the control which was between two and three orders of magnitude larger than that of the bound substrates. The precise detail of recognition by ATL proteins of all the ODN substrates is discussed at length in Chapter 5.

Whilst carrying out the assays it was noticed that the initial fluorescent signal of the dilute DNA solution was changing by a small amount before any protein had been added to the cuvette. As a result, the fluorescent intensity of an ODN substrate was measured over 20 min in the absence of added

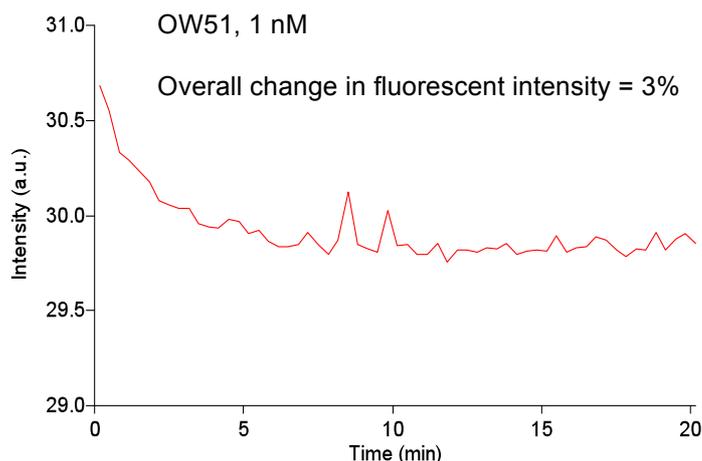


Figure 4.10: Plot to show the change in emission intensity of SIMA-labelled ODN over time before addition of AtI1.

protein. The fluorescent intensity decreased steeply in the first 5 min before settling out after about 10-20 min (figure 4.10). This change in intensity was usually around 2-5% of the initial intensity value. In comparison, AtI1 binding an ODN to saturation typically changes the intensity by 25-35%. It was initially decided to allow a period of 20 min at the start of the experiment for the ODN to equilibrate, so to have as little effect on the fluorescent intensity reading during protein binding as possible. At first it was thought that this effect could be due to temperature (as the solution in the cuvette equilibrates in the thermostatically-controlled chamber) or that it may be caused by changes in the conformation of the dye-labelled DNA: the SIMA fluorophore contains biphenyl groups that may be prone to conformational preferences which may in turn depend on ODN concentration. It was also suggested that the effect may be caused by the presence of gases in the buffer which slowly diffuse out after initial mixing at the beginning of the titration. It was eventually found that by pumping the ODN solution gently in and out of a Gilson pipette (P200)

once the cuvette was in the fluorimeter, this initial decrease in TFEI could be made to occur much faster than if simply left alone (i.e. as shown in figure 4.10). This makes sense if the buffer is being actively degassed by pipetting rather than the gases being allowed to diffuse out slowly (i.e. over a 20min period). Henceforth, each titration was begun by mixing the ODN solution with a pipette three times and checking the TFEI reading after each mix. In practice, there would be a comparatively large drop in TFEI after the first mix, followed by little difference in the readings between the second and third mix. In this way we ensured the TFEI signal of the ODN solution was stable and unchanging before titrating protein and thus causing a deliberate decrease in fluorescent intensity.

Finally, to confirm that the decrease in TFEI during the titration was due exclusively to binding of AtI1 protein to ODN, SDS was added to a final concentration of 0.1% v/v at the end of the titration (i.e. to saturated AtI1-DNA complex) to denature the AtI1 in the cuvette. This restored the fluorescent intensity to its initial value before any protein was added, and shows that the decrease in TFEI is due to the specific interaction of AtI1 with the ODNs.

4.3 FRET Analysis of AtI1-DNA Binding

It has been determined by structural analysis of the AtI1-DNA complex that the DNA is bent by an angle of approximately 45° compared to free B-DNA.(58) It is currently thought that this bulky, helically distorted complex is important for subsequent recognition by the NER machinery. It was therefore considered interesting to attempt to provide further evidence of this DNA distortion in free solution and under equilibrium conditions using FRET. It was hoped that changes in FRET between two dyes located at opposite ends of a

DNA duplex would allow us to observe the binding of At1 and subsequent bending of the DNA helix. The experiment is shown in figure 4.11.

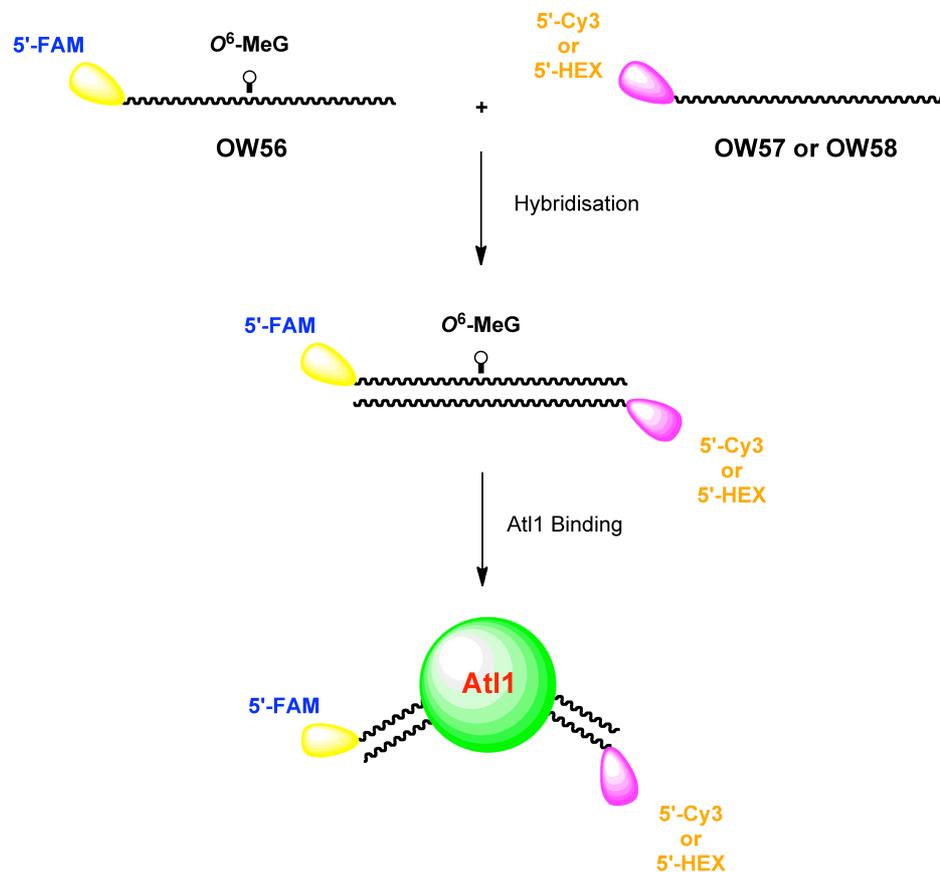


Figure 4.11: FRET-based experiment to measure DNA bending by At1 upon binding

Briefly, two complementary 13-mer ODNs each labelled with a 5'-dye are annealed together to form a DNA duplex. In this conformation, the two chromophores should be a defined distance apart, and able to interact through FRET (as the length of the 13-mer ODN is approximately 45Å). As At1 is added and binds to the DNA, it should cause the helix to bend and hence bring the dyes in closer proximity to each other. As FRET efficiency is strongly affected by the distance between the donor and acceptor, this should cause a change in the FRET signal (section 4.1.3). The ODN OW56 is

labelled with a 5'-FAM dye and the sequence also contains an O^6 -methylguanine residue which will be recognised by AtI1. Additionally, two ODNs complementary to OW56 were synthesised: OW57 which is labelled with 5'-Cy3, and OW58 is labelled with 5'-HEX (figure 4.11). The FAM fluorophore should engage in FRET with both Cy3 and HEX due to the overlap of its emission spectra with their absorption spectra (table 4.1).

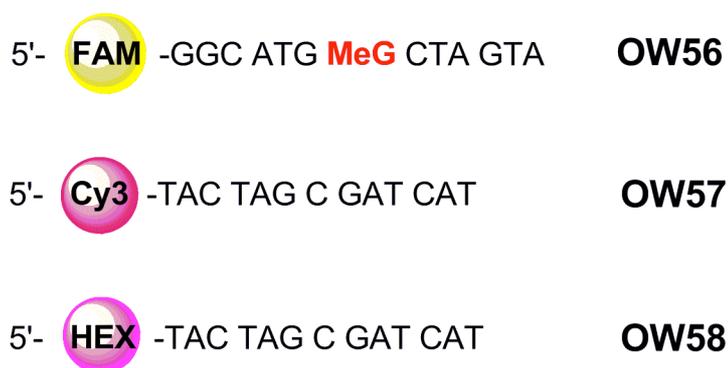


Figure 4.11: ODNs used in FRET-based experiments of DNA bending by AtI1

	FAM	Cy3	HEX
Excitation λ_{\max}	495nm	547nm	537nm
Emission λ_{\max}	521nm	563nm	556nm

Table 4.1: Excitation and emission maxima for selected fluorescent dye molecules

Initially, the experiment was attempted with ODNs OW56 (FAM) and OW57 (Cy3). The first step was to check that when the ODNs annealed together that FRET was occurring between the dye molecules at either end of the duplex (figure 4.12). A 50nM solution of OW56 (FAM) was placed in the cuvette and the FAM label excited at a wavelength of 490nm. This wavelength

was used because it is close to the absorption maximum of FAM but should cause minimal direct excitation of the Cy3 dye (table 4.1). The emission spectrum for the experiment is shown in figure 4.12 with the initial FAM

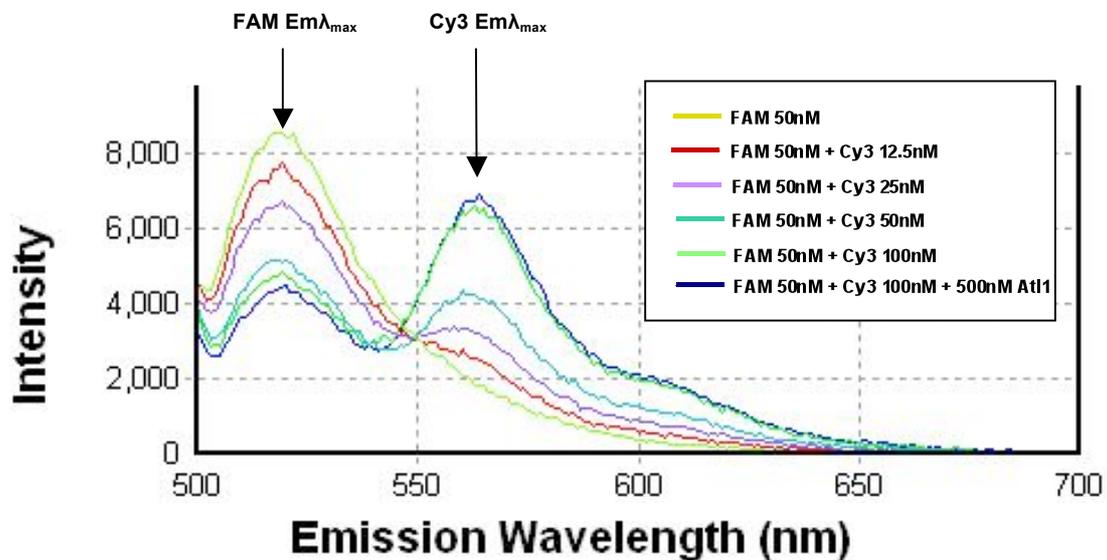


Figure 4.12: Plot to show occurrence of FRET between FAM and Cy3 dyes as annealing takes place in solution to form duplex DNA

emission peak shown as a yellow trace. Increasing amounts of OW57 (Cy3) was then added so that the ODNs would hybridise in the cuvette and cause FRET to occur from FAM to Cy3. As the concentration of OW57 increases a clear quenching of the FAM emission peak can be observed with a corresponding increase in the Cy3 emission peak, as FAM transfers its excitation energy to Cy3 by FRET. Initially 2 equivalents of the complementary sequence OW57 were added to make sure all the OW56 was in a duplex (as this ODN contains the O^6 -methylguanine residue that will be bound by AtI1). A 10-fold molar excess of AtI1 protein was then added, a concentration of protein which had previously been shown to form a saturated AtI1-DNA complex. If the binding of AtI1 to the DNA involved it becoming bent by 45° ,

this was expected to cause the labels to become closer to each other and hence cause an additional change in FRET (i.e. the FAM emission peak being quenched further and an increase in the intensity of the Cy3 emission peak). Unfortunately, as can be observed in figure 4.12, there is a slight change in peak intensities after addition of AtI1 but additional FRET corresponding to a change in proximity of the dye molecules did not occur to any significant degree.

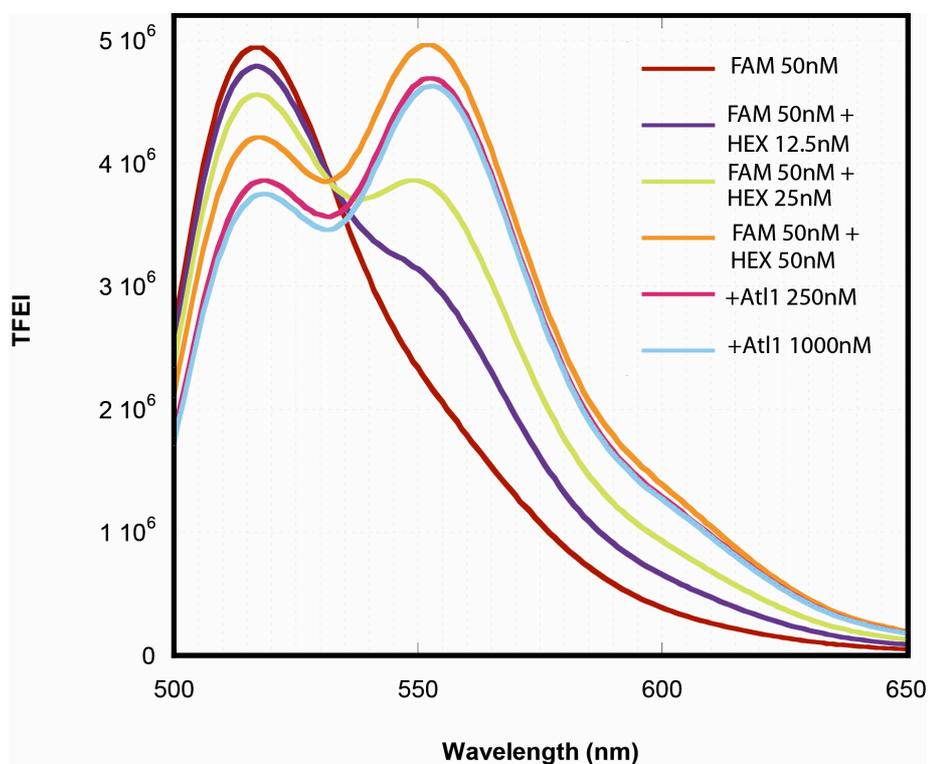


Figure 4.13: FRET-based analysis of DNA bending by AtI1 protein

It was decided to repeat the experiment, this time using OW56 (FAM) and OW58 which has the same sequence as OW57 but is labelled with HEX. The experiment was conducted in the same manner, adding increasing amounts of OW58 to see the FRET developing between the dye molecules at either end of the duplex as it hybridises. When a 1:1 ratio of OW56: OW58 at 50nM concentration had formed, increasing amounts of AtI1 were then added

(figure 4.13). In common with the previous experiment, no change in FRET between the dye molecules was observed upon addition of protein. There is an equal decrease in fluorescent intensity of both the FAM and HEX signals that occurs after addition of AtI1; however this is almost certainly caused by quenching of the fluorescent dyes upon AtI1 binding (which is the effect observed in the binding assays discussed previously, i.e. the concentration-dependent decrease in TFEI of a SIMA-labelled ODN by AtI1).

The reason that the distortion of DNA could not be observed by FRET may be that binding of AtI1 is not causing the DNA to bend, which seems unlikely given that this conformation is clearly observed in crystal structures of AtI1 with DNA containing various O^6 -alkylguanine residues (and of the same sequence as our probe). It is more likely that either there is not enough of a change in distance when this bending occurs to be observed by FRET, or that the protein interacting with one of the labels changes its orientation and prevents a change in FRET occurring.

4.4 FRET Analysis of ODN Hybridisation

Although the majority of the binding assays to investigate the interaction of O^6 -alkylguanine-containing ODNs with ATL proteins were performed with ssDNA, the recognition of selected O^6 -alkylguanine residues using dsDNA was also studied (section 5.2.3). The ODNs used in the assays are relatively short and as such the duplexes were calculated to have a fairly low melting temperature (T_m). For example, the 13-mer duplex of OW51 + OW6 in 50mM NaCl has a calculated T_m of approximately 40°C (figure 4.14).



Figure 4.14: Duplex DNA used in the fluorescent binding assays

The binding assays were carried out at 25°C which is reasonably close to the T_m and could lead to a significant proportion of the duplex being in the single-stranded form even after annealing. The standard method to assess whether DNA is single-stranded or double-stranded at a certain temperature is to perform a DNA melt (figure 4.15) where UV absorbance at 260nm is plotted as a function of temperature. The natural UV absorbance of the nucleobases is quenched when they are base-paired (i.e. in a duplex) and so as the strands unwind the absorbance at 260nm increases. Plotting absorbance against temperature gives an empirically-determined value of the T_m . The DNA duplex can be observed denaturing into single strands as the temperature increases and hybridising back to dsDNA as the temperature decreases. The T_m is the point at which 50% of the DNA is double-stranded.

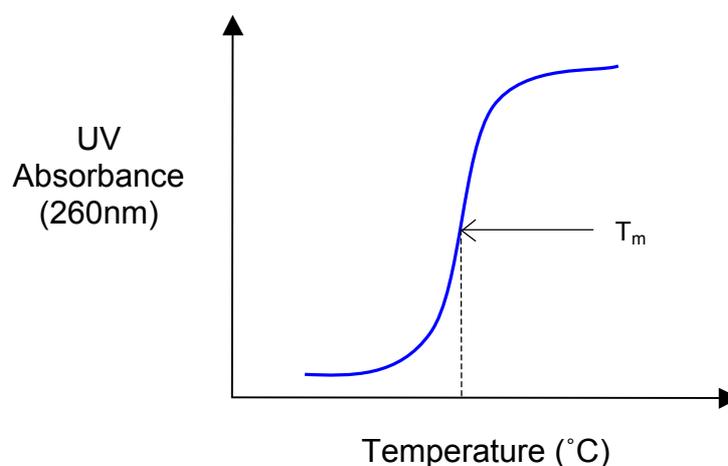


Figure 4.15: A typical DNA melting curve. T_m is the temperature at which 50% is duplex DNA

The problem was that our ODN concentration in the assays is very low (1nM). The DNA concentration required to give an accurate DNA melting curve is much higher than this in order that the Abs_{260} values are large enough to be measured. So whilst it was likely most of the ODNs were duplexes at high concentration at 25°C, we also wanted to confirm that our ODN substrates were double-stranded after dilution to 1nM in the cuvette.

Previously it had been shown that when two complementary FAM- and HEX-labelled ODNs (OW56 and OW58) were hybridised, FRET was observed between the dye molecules (section 4.3). Hence attempts were made to demonstrate by FRET that duplexes were formed under the experimental conditions used in the binding assays (1nM ODN concentration, 50mM NaCl, 25°C). The experiment involved measuring the emission spectra of OW56 and OW58 using an excitation wavelength of 490nm. As can be observed in figure 4.16, this causes excitation of the FAM label and, to a lesser degree, the HEX label. The signals from these ODNs alone are shown, as well as those of the same HEX- and FAM-labelled ODNs mixed together but treated under different conditions. In one reaction, OW56 and OW58 were annealed (at 1 μ M ODN concentration and in 50mM NaCl) by heating to 80°C for 5 min and being allowed to cool slowly (~1h). This dsDNA was then diluted to 1nM in the cuvette, left to equilibrate for a few minutes and its emission spectrum recorded ($\lambda_{ex} = 490nm$). In the other, OW56 and OW58 were simply mixed together (also at 1 μ M ODN concentration and in 50mM NaCl) at room temperature and then diluted to 1nM. The emission spectrum was then recorded in exactly the same way. In the case of the annealed substrates, there was a decrease in the intensity of the FAM emission peak whilst the

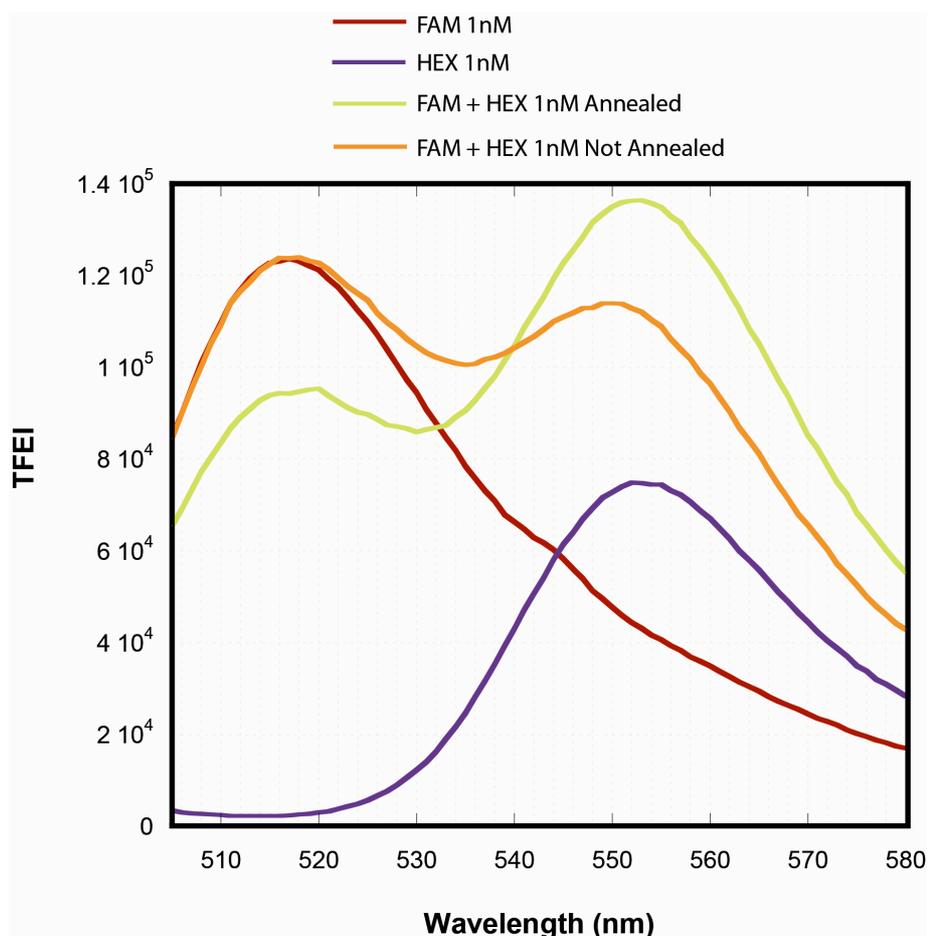


Figure 4.16: FRET analysis of DNA hybridisation of **complementary** strands at 1nM ODN concentration

emission peak of HEX increased in intensity as excitation was transferred by FRET. This was not the case for the mixture of OW56 and OW58 that was not annealed. In fact, the FAM emission peak was identical to that for OW56 alone (no quenching through FRET), and the emission peak at 553nm had the same TFEI as the addition of the signals of the FAM- and HEX-labelled ODNs alone at this wavelength. This would strongly suggest that the ODNs were double-stranded under these conditions (i.e. [ODN] = 1nM, [NaCl] = 50mM, 25°C) after being annealed.

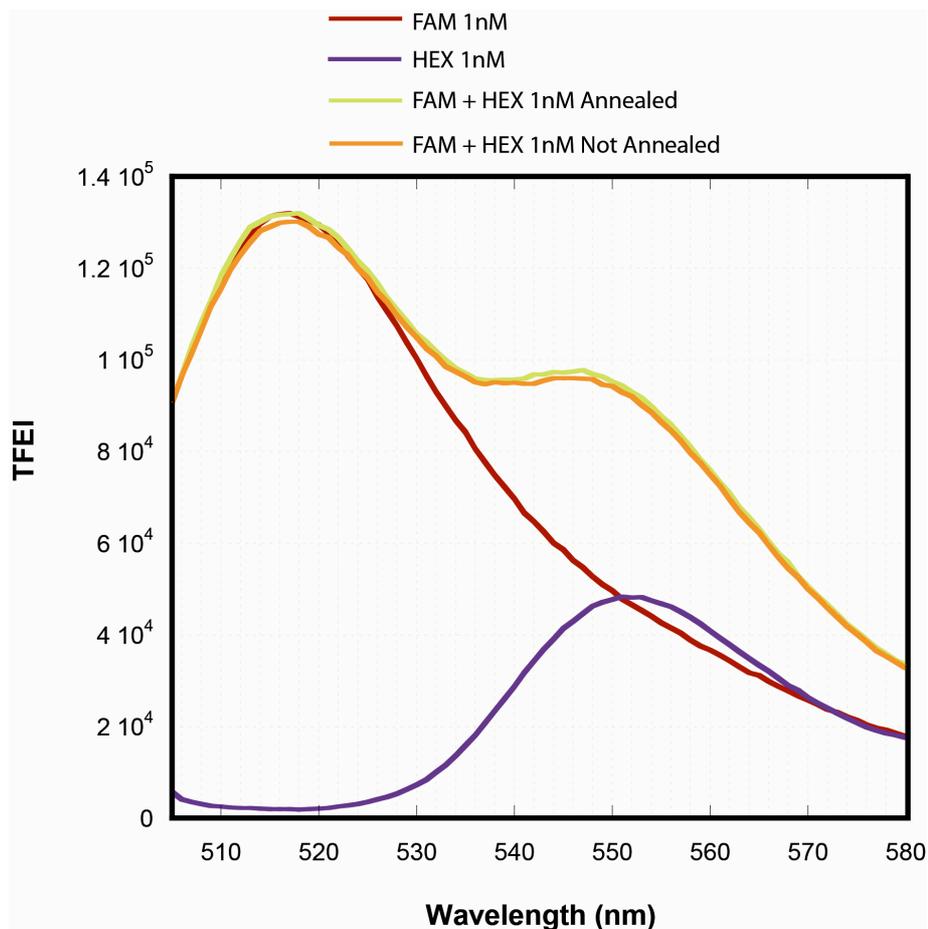


Figure 4.17: FRET analysis of DNA hybridisation of **non-complementary** strands at 1nM ODN concentration

In order to provide further evidence that the ODNs had formed dsDNA, the experiment was set up exactly as before but this time OW56 was mixed with a HEX-labelled ODN with the same sequence as itself (OW53) so there was no possibility of hybridisation. Figure 4.17 demonstrates that the emission spectra are identical regardless of whether the ODNs were annealed together or not. In common with the previous experiment, the emission peak at 553nm is an additive intensity signal from the FAM and HEX emissions at this wavelength. This would suggest that it is the close proximity of FAM and HEX when OW56 and OW58 are in a duplex that causes FRET to occur, and that

this effect is not observed between the terminal dye molecules on non-complementary sequences that are incapable of hybridising.

4.5 Fluorescence-based MGMT Activity Assay

MGMT is a protein that repairs DNA by removing O^6 -alkylguanine lesions in a stoichiometric, irreversible reaction (see 1.7).(47) It is an important drug target for cancer chemotherapy as selective inhibition of MGMT in tumour cells increases their sensitivity to the toxic effects of DNA alkylating agents.(14) It is therefore necessary to have a method for quantifying the effects of inhibitors on MGMT. Previously, MGMT inhibition by various O^6 -alkylguanine-containing ODN substrates had been measured by a radioactive assay using [3H]-methylurea-methylated calf thymus substrate DNA.(68) MGMT was incubated for 1h at 37°C with an ODN substrate, and then residual alkyltransferase activity was measured by relative levels of transfer of tritium from the radiolabelled substrate to the protein. IC_{50} values (representing a 50% inhibition of MGMT activity) were then derived from the dose-response curves for each ODN. However, a similar assay capable of quantifying MGMT inhibition by modified ODNs but without having to use any radioactive reagents would be advantageous as the assays would be cleaner, safer and easier to carry out. It was decided to design a fluorescent probe that would allow the monitoring of MGMT activity by exploiting the properties of a dye-quencher interaction using a molecular beacon (see 4.1.4).

This ODN (OW31, figure 4.18) contains a 5'-Cy3 fluorescent label, a 3'-BHQ-2 quencher molecule, and was self-complementary so that it folds into a hairpin secondary structure which brings the dye and quencher into close

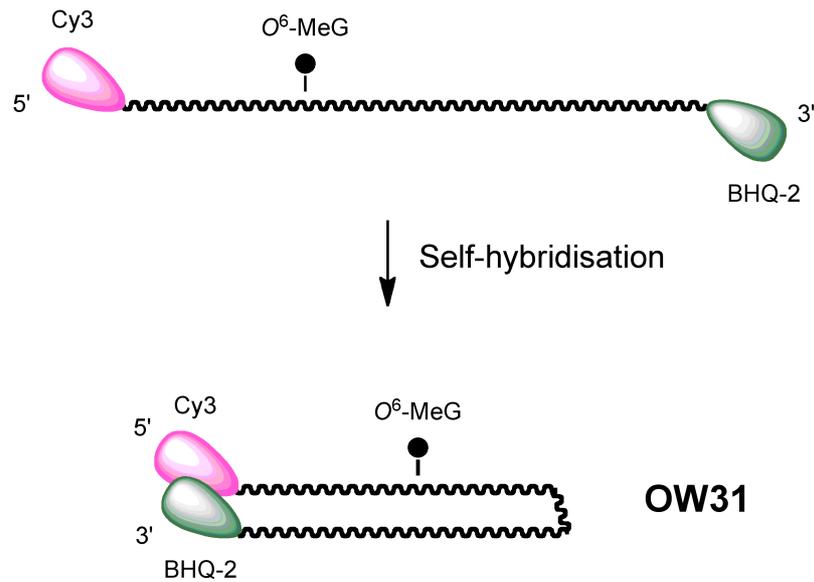


Figure 4.18: The molecular beacon ODN OW31

proximity. It has been shown that in this arrangement BHQ-2 will quench the fluorescence of Cy3 very efficiently.(97) The outline of the assay is shown in figure 4.19. Firstly, the molecular beacon (OW31) is denatured by heating and re-folds into its most stable conformation (i.e. a hairpin stem with Cy3 and BHQ-2 close together). This ODN also contains an O^6 -methylguanine residue in the middle of a PstI restriction site, which will effectively block the sequence from digestion by the PstI enzyme. It has previously been demonstrated that MGMT repairs DNA containing O^6 -methylguanine (109) and therefore when the alkyl transfer reaction takes place to produce guanine, the PstI site will be restored. If PstI is added to the molecular beacon ODN after treatment with MGMT, then any ODN that has been repaired will be digested by the restriction enzyme. This will cause the ODN to be excised and produce two fragments, one containing the Cy3 dye, and the other the BHQ-2 quencher. As a consequence, the strands will dissociate due to their lower T_m , and these

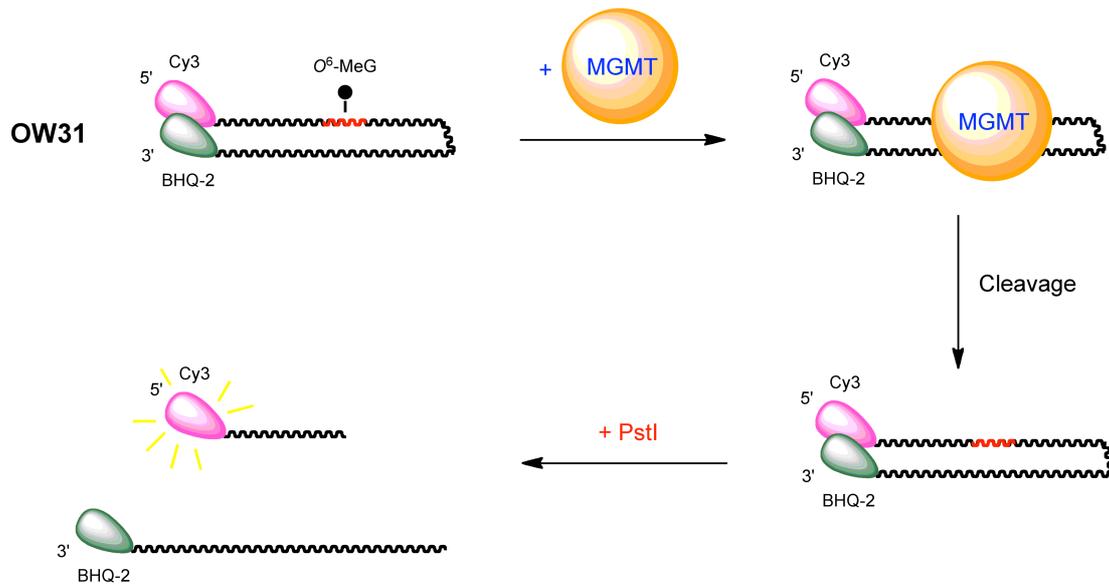


Figure 4.19: Outline of the fluorescent MGMT activity assay using a molecular beacon probe

molecules will no longer be in close proximity. Correspondingly an increase in the fluorescence of Cy3 should result as it ceases to be quenched. Measuring the emission intensity of Cy3 throughout the assay should allow MGMT activity to be determined (as the amount of 'free' Cy3 at the end of the assay is dependant on the amount of repaired ODN substrate and hence upon the activity of MGMT). When MGMT is used in the assay without any pre-treatment, it would be considered to have 100% activity. Therefore, if MGMT is pre-treated with an ODN containing a specific O^6 -alkylguanine lesion, this assay should allow the determination of an IC_{50} value (i.e. the concentration of inhibitor/inactivator at which MGMT activity is reduced to 50%).

Initially a control experiment was performed using the ODN OW39 which is identical to OW31 but contains guanine at the same position as O^6 -methylguanine, i.e. has an unhindered PstI restriction site. The ODN substrate was heated to 95°C for 1 min and then cooled on ice for 5 min to ensure it was

folded correctly, and then diluted to 2nM and placed in the fluorescence cuvette. This was incubated at 25°C for 10min, 20 units of PstI were added and the reaction continued to be incubated at 25°C for a further 30min. The Cy3 dye was excited at a wavelength of 547nm (its excitation λ_{\max}) and the emission intensity measured at 563nm (its emission λ_{\max}) as a function of

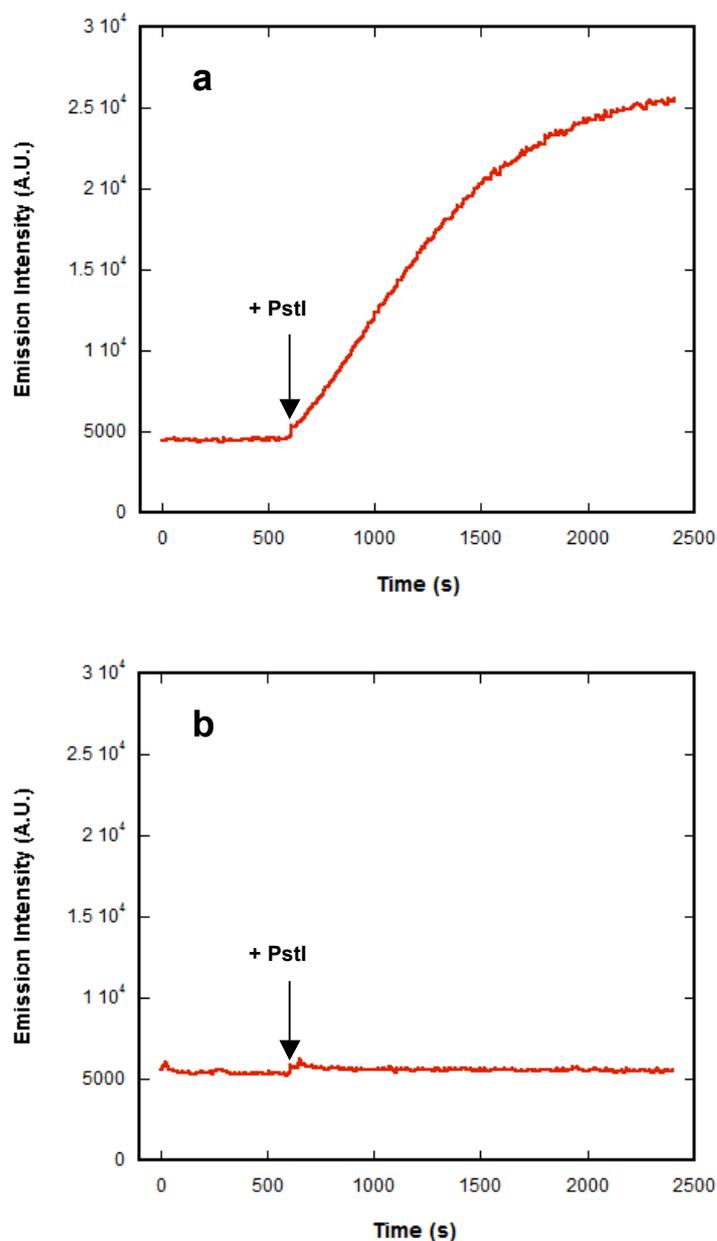


Figure 4.20: Plots to show digestion of molecular beacon ODNs (2nM) by PstI (20 units); with OW39 which contains guanine so has an unhindered restriction site (a), and with OW31 which contains O^6 -MeG and therefore the same restriction site is blocked (b).

time. The addition of PstI and incubation over 30 min caused a 4.8-fold increase in emission intensity of Cy3 which is a large and significant change (figure 4.20) which was consistent and reproducible. When the same experiment was undertaken using OW39 (which contains O^6 -MeG so the PstI site is blocked) in place of OW31, very little change in emission intensity was observed. Thus it seemed from the initial experiments that this molecular beacon may be useful for assaying MGMT activity.

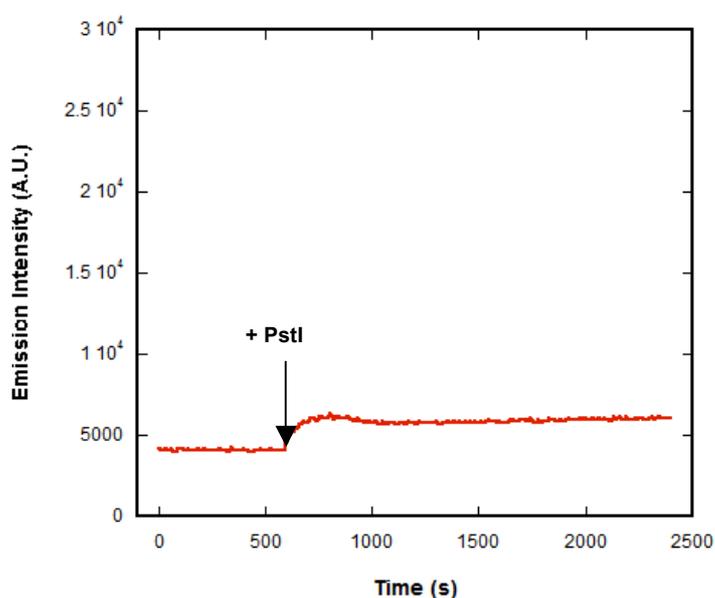


Figure 4.21: Plot to show digestion of OW31 (2nM) after treatment with MGMT (4nM) and PstI (20 units)

In the next experiment, OW31 was folded and diluted to 2nM as before, and its emission intensity measured over time for 10 min. MGMT was then added to a final concentration of 4nM (2 molar equivalents) and this was incubated in the cuvette for 30 min at 25°C. This caused no significant change in the emission intensity of the molecular beacon. Finally, 20 units of PstI was added as before and incubated for a further 30min. The emission intensity of Cy3 was measured throughout the experiment. Unfortunately, the emission

intensity only increased by a small amount (1.2-fold) in this experiment (figure 4.21).

It was reasoned that the issue may have been the MGMT concentration being too low in the experiment. As an analogy, when At11 binds O^6 -methylguanine-containing ODN at 1nM concentration, the binding becomes saturated at around 50nM. The experiment was therefore repeated but this time using an MGMT concentration of 100nM (50 molar equivalents). This was considered a sufficient amount of protein to demethylate a significant proportion of the molecular beacon ODN. The experiment was successful, with cleavage by MGMT and subsequent digestion by PstI causing a 3.2-fold increase of fluorescent signal (figure 4.22).

Unfortunately due to time constraints this work was not able to be continued, however these preliminary results are promising and therefore this assay should be developed in the future.

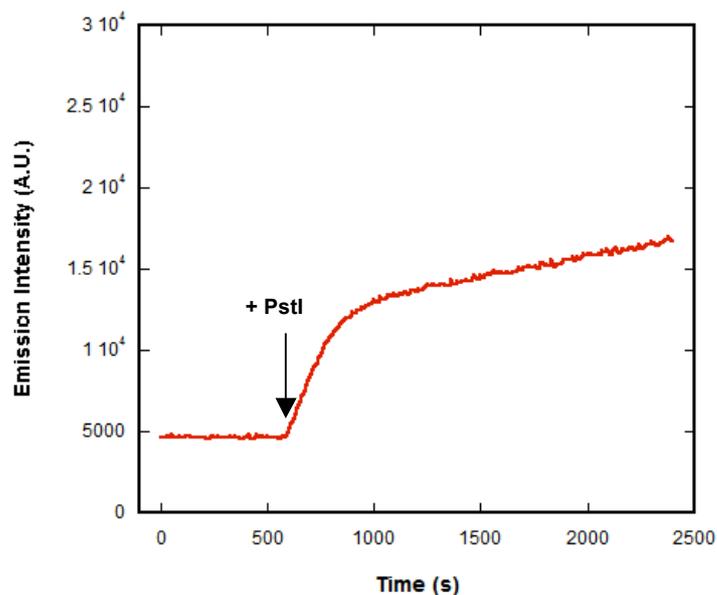


Figure 4.22: Plot to show digestion of OW31 (2nM) after treatment with MGMT (100nM) and PstI (20 units)

5.0 Damaged DNA Recognition by ATL Proteins

5.1 Protein Recognition of DNA Damage

Certain DNA binding proteins, such as alkyltransferases, alkyltransferase-like proteins and DNA glycosylases, process lesions by flipping damaged bases into an active site pocket, usually in order to perform the specific chemical reaction required for activity (but not in the case of ATL proteins). It is a unique mechanistic solution for damaged base recognition and repair, but the way in which damaged bases within in the helical stack are detected is still not fully understood. However, once the base has been flipped out into the protein active site, recognition is mediated by a number of interactions between residues in the active site fold and the damaged base. These interactions can take a number of forms, and bestow the protein with the ability to discriminate between the specific damaged base that it repairs, and undamaged bases or those with other types of damage. This introduction will briefly review the current understanding of damaged base detection, and examine the active site interactions of a selection of nucleotide-flipping proteins, based on structural studies.

5.1.1 Detection of Damaged Bases by Nucleotide-flipping Proteins

Detection of damaged bases has been best studied in a large family of structurally diverse enzymes, the DNA glycosylases, which initiate base-excision repair. Although DNA glycosylases fall into different structural classes and are specialised for the removal of different types of damaged bases, they

engage in a single mechanistic solution for damaged base recognition and excision: flipping of the damaged base from the base stack into the protein active site (by rotation of the sugar-phosphate backbone) to facilitate cleavage of the N-glycosidic bond.(110) This is an improbable process in energetic terms: the considerable penalties that must be overcome by breaking the Watson-Crick base pairs and straining the sugar-phosphate backbone must be paid for by the binding energy of the protein. This can, for example, involve insertion of a protein residue into the DNA base stack to stabilise the extrahelical conformation of the base, as is seen for the DNA glycosylases UDG (111), AAG,(112) and hOGG1,(24) and also for At1 (58) and MGMT.(47) In addition, the ability to pick out damaged bases against a vast background of undamaged bases in the genome is also intriguing.

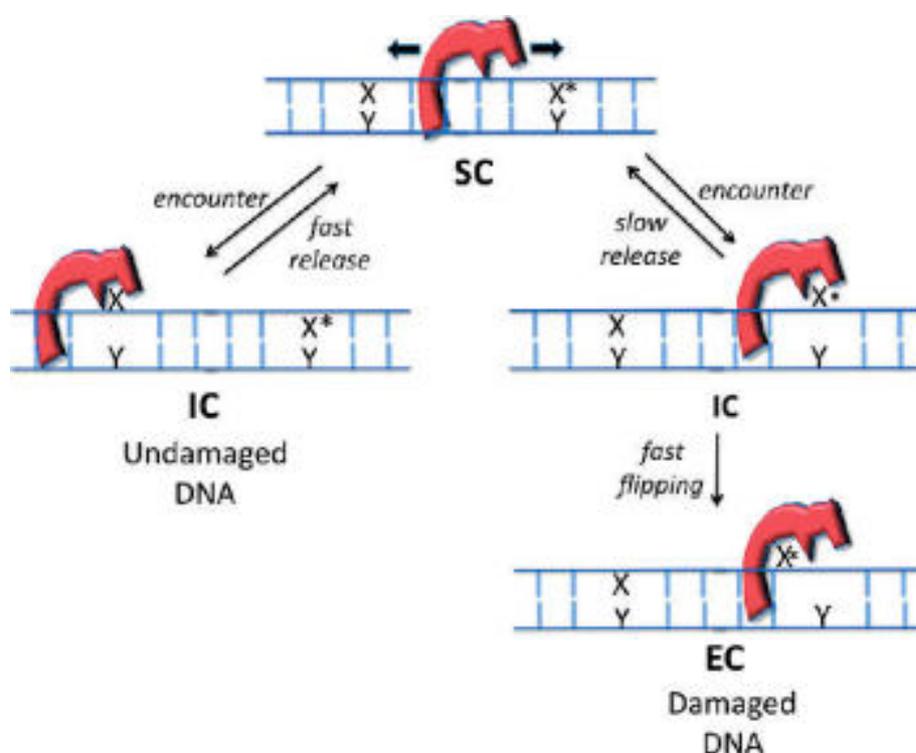


Figure 5.1: Model for recognition of damaged bases by DNA glycosylases, taken from Friedman and Stivers (113)

Recent studies have elucidated the mechanism of detection of damaged bases by DNA glycosylases (figure 5.1).(113) The first stage is the formation of a 'search complex' (SC), a distinct conformation of the protein which slides along the DNA chain using non-specific binding interactions. This sliding is frequently interrupted by the formation of a transient 'interrogation complex' (IC) in which the enzyme extrahelically inspects both damaged and non-damaged bases. When the enzyme encounters a normal base, the IC rapidly falls back to the SC, whereas a damaged base will be efficiently inserted into the active site to form the catalytically competent excision complex (EC). Here, the specific active-site interactions are crucial as they effectively hold the damaged base in the active site and are responsible for the discrimination between damaged and non-damaged bases.

Damaged base detection by the human AGT protein, MGMT (or hAGT), has also been studied in some detail. Meyer *et al.*, after analysis of kinetic data, suggested a plausible mechanism for base recognition by MGMT. In this model, the protein binds non-specifically to DNA, subsequently binds the lesion and flips it out of the base stack before finally instigating repair in the active site pocket buried deep in the protein. This is described in the equation shown below, where k_s is the rate of base flipping and k_{inact} is the rate of repair.(114) This model is supported by kinetic data from Coulter *et al.* who also found that the nucleophilic displacement reaction (i.e. alkyl transfer) is the rate-limiting step.(95)



E = MGMT, **S** = O^6 -alkylguanine-containing DNA, **ES** = MGMT-DNA complex

X = Repaired DNA, **D** = Alkylated MGMT, k_s = rate of base flipping (binding)

k_{inact} = rate of repair (alkyl transfer)

It has been proposed by Duigaid *et al.* that MGMT does not detect DNA lesions by searching for the adduct itself but rather by looking for small distortions or weaknesses in the duplex caused by the damaged base. They also suggest that, due to weakened base-pairing between a damaged base and its helical partner, the damaged base will sample the extrahelical conformation more frequently which may also enhance its detection.(65) They also suggest that the *E. coli* alkyltransferase, C-Ada, does not actively flip out every base for damage searching but locates damaged bases by simply capturing a lesion that is sampling the extrahelical conformation transiently due to its significantly weakened base-pairing ability. It is proposed MGMT is able to detect damaged bases that form unstable base pairs whilst also being able to extrude base lesions that are stabilized intrahelically in the DNA structure in a less efficient process. If both capturing an extrahelical base lesion and detecting the unstable nature of a damaged base pair followed by base-flipping can be used to locate damaged bases for repair it means that actively flipping a base out for damage searching may not be necessary in many cases.(115)

More recently, Hu *et al.* have used computational methods to examine the recognition of damaged bases by MGMT. Nucleotide flipping is shown to proceed in a two-step mechanism, involving a metastable transition state.(64) The protein initially binds the DNA through a helix-turn-helix (HTH) motif, and

buries a second 'recognition' helix (Ala127-Gly136) deep in the minor groove. The O^6 -methylguanine and cytosine bases, whilst maintaining the orientation of the planes containing their aromatic rings, are fluctuating apart by shifting their pattern of hydrogen bonding. The arginine finger (Arg128) utilises this transient motion to form a structure in which it interacts comparably with both bases (this is the transition state). Arg128 then shifts its interaction to the cytosine residue which allows O^6 -MeG to diffuse into solution and form the extrahelical intermediate. (64) It is also suggested that Tyr114, rather than interacting sterically with the 3'-phosphate, acts electrostatically and affects repair in three ways: by stabilising binding before flipping, by providing hydrogen bonds during flipping, and by influencing the reactivity of the damaged base.(64)

Free energy calculations have shown that in common with other DNA repair proteins, MGMT significantly stabilises the base separated state relative to the intrahelical conformation. Crucially, the free-energy barrier to flipping guanine rather than O^6 -methylguanine is much higher which suggests a 'gate-keeping' strategy for lesion discrimination. Both G and O^6 -MeG would flip to the extrahelical state, but only the latter would have a significant chance to continue to the active site before flipping back. In common with that proposed for hOGG1,(116) this gate-keeping mechanism appears to be kinetic rather than thermodynamic in nature: there is only a 3-fold preference for O^6 -MeG over G once in the active site pocket of AGT (117). The rate for partial flipping for O^6 -MeG and G is comparable ($\sim 10^{-3}$ to 10^{-2} ps $^{-1}$) but whereas O^6 -MeG proceeds rapidly to the active site ($\sim 10^{-1}$ ps $^{-1}$), G is much slower ($\sim 10^{-6}$ ps $^{-1}$).

Ada, a bacterial homologue of AGT from *E.coli*, has been shown by single-molecule tracking studies (DNS) to exhibit a helical sliding rate comparable to that of hOGG1, a human glycosylase (116). hOGG1 has been observed by high-speed imaging sliding along DNA at an extremely fast rate which would seem to exclude the possibility of flipping out every single base in the search for damaged bases. Instead, it is proposed that hOGG1 extrudes 8-oxoG lesions into the active site at a higher rate than undamaged bases: thus the recognition is under kinetic control. For C-Ada and hOGG1, comparison of the rates for sliding with those of base flipping suggests that the kinetic gate-keeping mechanism is feasible, and would allow the protein to scan the sequence rapidly compared with one requiring full flipping.

5.1.2 Active-site Interactions of Nucleotide-flipping Proteins

hOGG1 is a DNA glycosylase of HhH superfamily responsible for the recognition of 8-oxoguanine DNA lesions in humans. It has been shown in structural studies that discrimination between 8-oxoguanine and guanine (which differ by only two atoms) is a consequence of specific interactions between the active site and the base (figure 5.2). (118) (24) The Lys249 (NH_3^+) / Cys253 (S-) salt bridge forms a dipole which runs antiparallel to the local dipole of 8-oxoguanine and parallel to that of guanine (this is opposite due to charge inversion at positions 7 and 8). Thus, this interaction stabilises the interaction of 8-oxoguanine with the active site pocket whilst destabilising that of guanine. However, the crucial stabilising interaction in the active site is the formation of a hydrogen bond between the carbonyl oxygen of Gly42 and

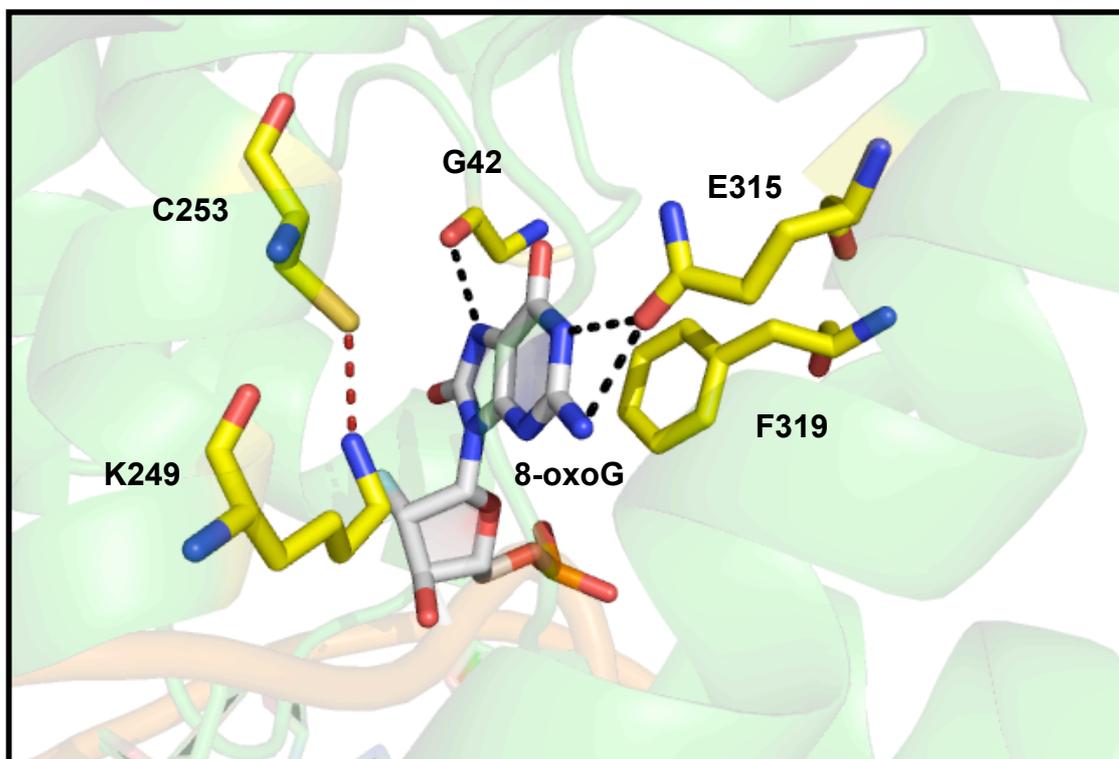


Figure 5.2: Active site of hOGG1 bound to ODN containing 8-oxoG with interactions shown

the N7 hydrogen of 8-oxoguanine. This protonated N7 atom is not present in the natural base, and therefore the corresponding interaction with guanine is that of repulsion between the lone pairs, though this destabilising interaction is partly relieved by a local reorientation of Gly42 due to some conformational flexibility in the protein. The specificity of the protein for guanine over the other canonical bases is also mediated by active site interactions. Phe319 interacts with the π -face of the base, and Glu315 takes part in two interactions: its amide group in conjunction with a tightly-bound water molecule recognises O6, and the side-chain carbonyl forms H-bonds with N1 and N2H. These interactions are present with guanine and 8-oxoguanine, but absent for adenine, cytosine and thymine.

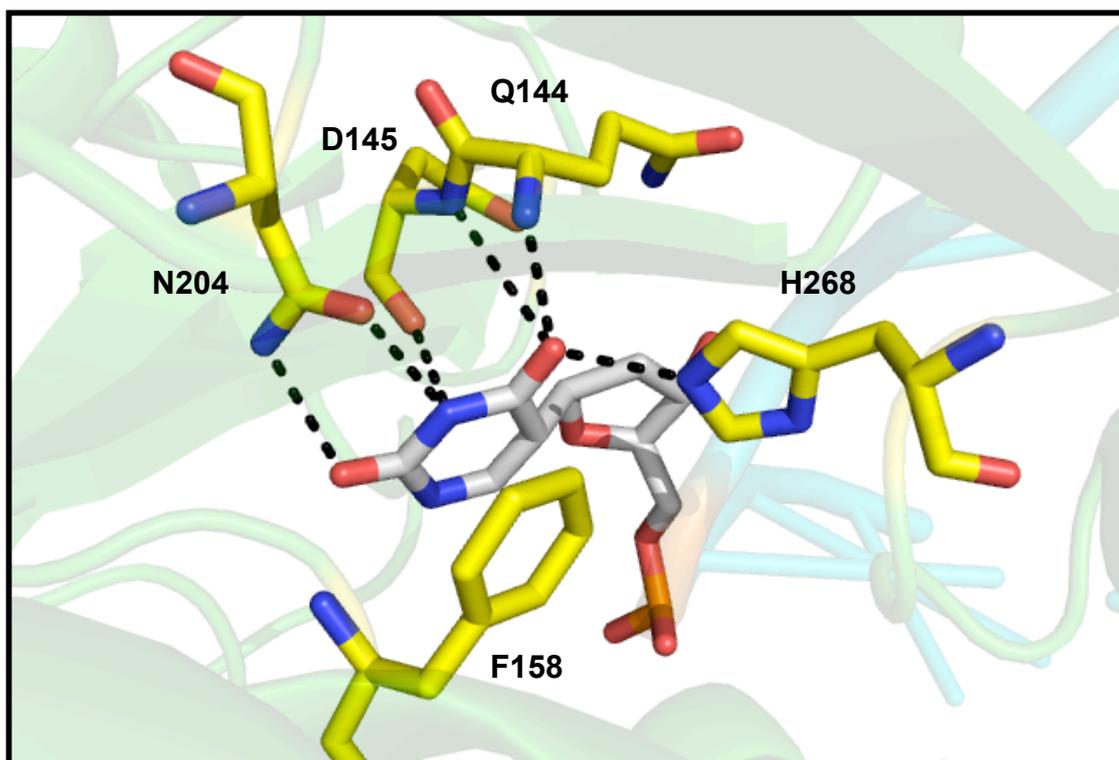


Figure 5.3: Active site of hUDG bound to ODN containing 6-aU showing interactions

Human UDG is a DNA glycosylase of the UNG superfamily that exclusively recognises uracil and 6-aminouracil (6-aU).^{(119), (120)} These bases have highly specific interactions with the active site pocket that result in discrimination between these residues and other bases (figure 5.3). The Asn204 side chain N δ 2 and O δ 1 atoms form hydrogen bonds with the O4 and N3 positions respectively of 6aU. The N3 atom of the base also forms an H-bond with the backbone carbonyl oxygen of Asp-145. Both of the backbone amides of Asp-145 and Gln-144 form hydrogen bonds to the O2 atom of the base, which also forms an H-bond with the N2 atom of His-268 via a bridging water molecule. Thus recognition of 6-aU and U by hUDG is mediated almost entirely by hydrogen bonding interactions, with additional stability being provided by a π - π interaction between Phe158 and the face of the base.

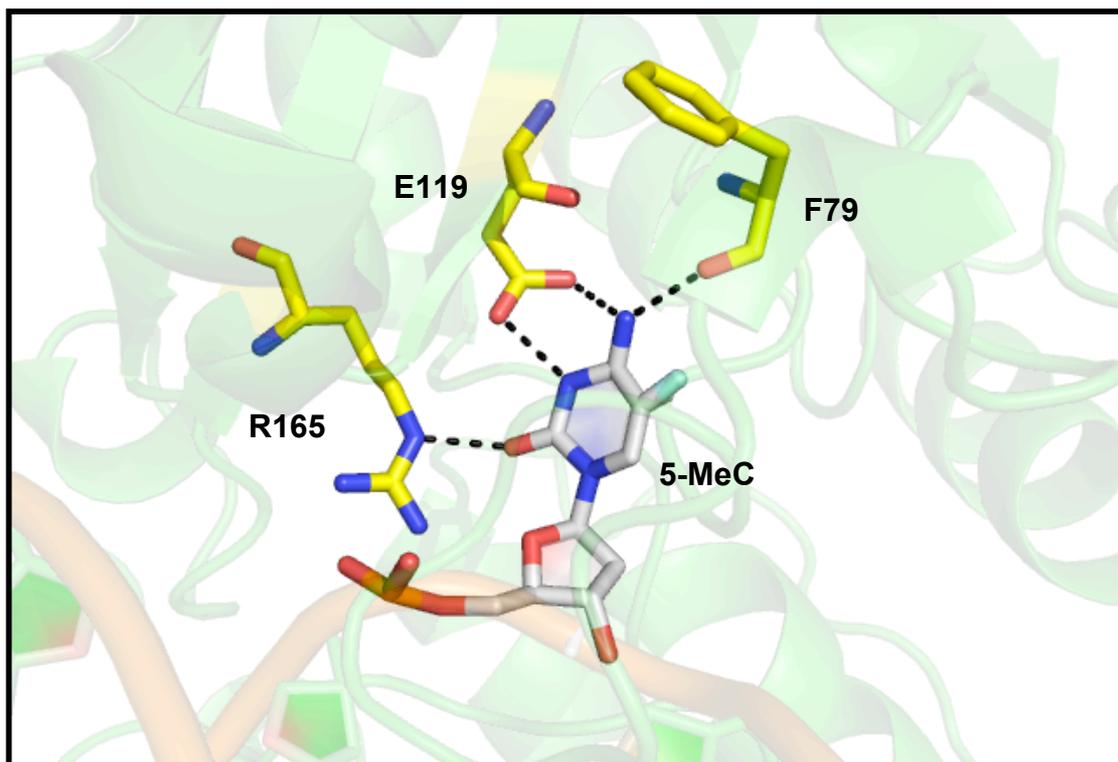


Figure 5.4: Active site of HhaI bound to ODN containing 5-MeC showing interactions

HhaI is a cytosine-5-methyltransferase from *H.haemolyticus*.(121) Extensive interactions exist between the target cytosine and the active site that lead to the specificity of the enzyme (figure 5.4). Arg165's non-terminal nitrogen forms a hydrogen bond with O2, the two side-chain carbonyls of Glu119 H-bond with N3 and N4 and the main-chain carbonyl of Phe79 H-bonds with N4. These residues are invariant across the entire enzyme family. The interactions between Glu119 and the cytosine base are of particular importance for substrate specificity: in two nucleotide methylases with similar mechanisms to HhaI, mutation of an invariant asparagine (with contacts to N3 and O2) to an aspartate or glutamate residue changes the activity of the

enzyme from processing dUMP to dCMP. Thus the nature of this residue and the contacts it makes are crucial for recognition of a certain nucleobase.

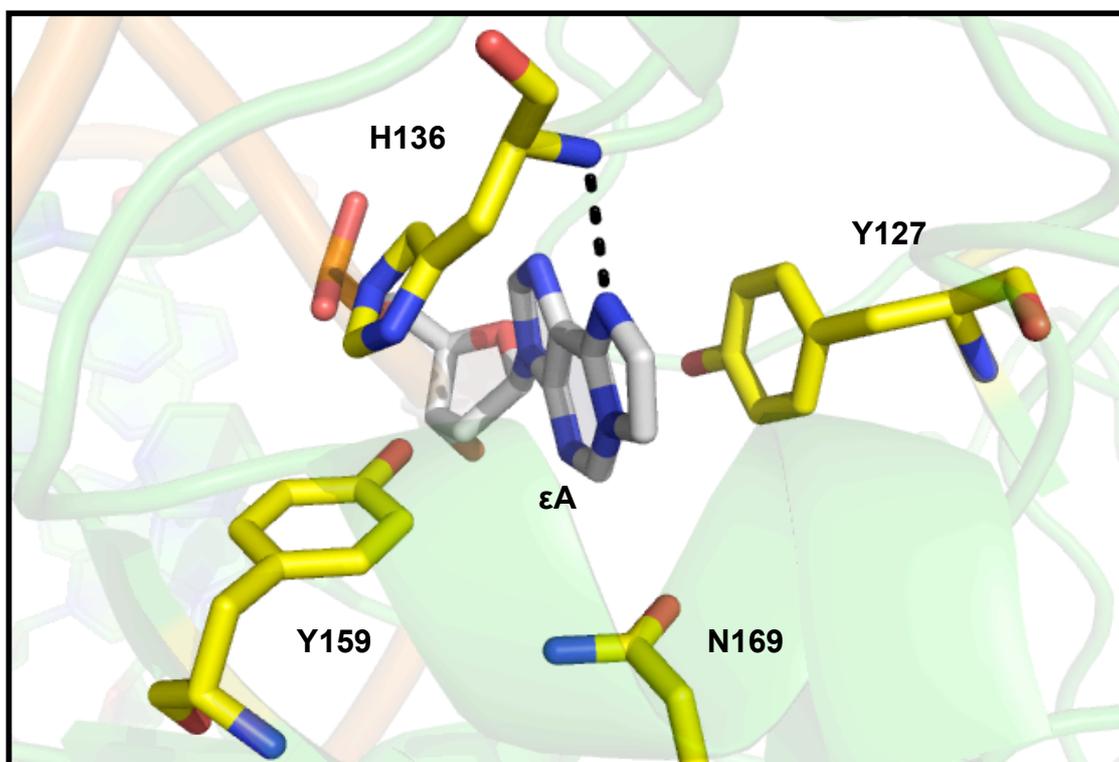


Figure 5.5: Active site of AAG bound to ODN containing ϵ A showing interactions

Human alkyladenine glycosylase (AAG) repairs a variety of alkylated bases, such as 3-methyladenine (3-MeA), 7-methylguanine (7-MeG), hypoxanthine (Hx) and 1- N^6 -ethenoadenine (ϵ A). The structure of AAG bound to DNA containing ϵ A shows that in the active site (figure 5.5) there are multiple interactions with the flipped base: the π -faces of Tyr127 and His136 are orientated towards the alkylated adenine residue, and the -OH group of Tyr159 also contacts the face of the base.(112) The selectivity for positively charged lesions, such as 3-MeA and 7-MeG is enhanced due to the potential for cation- π interactions between Tyr127 and His136 and the alkylated base.

In terms of discrimination between damaged and undamaged bases, the only H-bonding interaction is between the main chain nitrogen of His136 and the N6 of the alkylated base. In adenine, this N6 is an amine rather than an imine and so cannot act as a hydrogen bond acceptor with the main chain amido group of His136. Furthermore, binding of guanine is prevented because the exocyclic N2 amino group of guanine would clash with the highly constrained Asn169 side chain. It is thought that the other enhanced interactions due to the positive charge on 7-MeG must be enough to overcome this energetic barrier when this damaged base is bound.

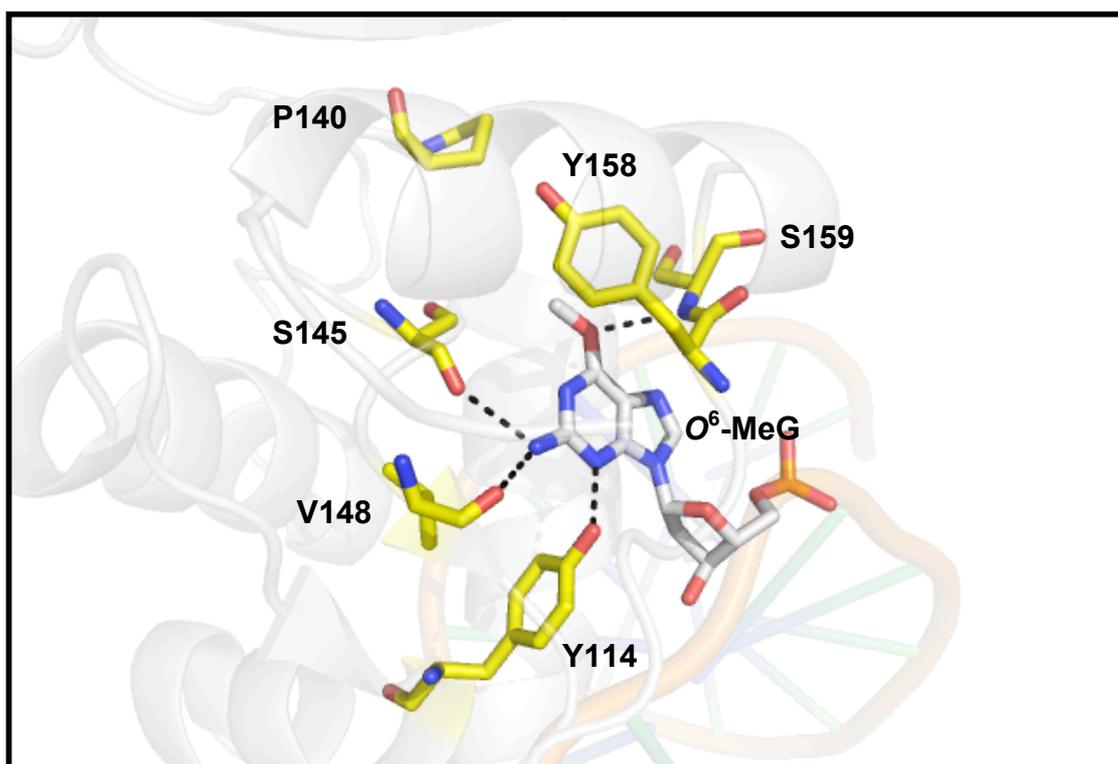


Figure 5.6: Active site of MGMT bound to ODN containing O⁶-MeG showing interactions

MGMT is the human AGT protein which specifically recognises and repairs O⁶-alkylguanine lesions in DNA. In the MGMT active site (figure 5.6)

the following hydrogen bonding interactions occur: the main-chain carbonyl groups from Cys145 and Val148 accept a hydrogen bond from the N2 amino group of guanine, the Tyr114 hydroxyl donates a hydrogen bond to the N3 atom on the base, and the main chain nitrogen from Ser159 donates to the O6 atom.(51) These active site interactions are sufficient for MGMT to discriminate between guanine and the other three canonical bases. However, there is only a slight preference for O^6 -MeG over G (values range from 3-fold (117) to 10-fold (114)) and this would seem to be due to the larger hydrophobic surface that is a consequence of alkylation. The Tyr158 side-chain is seen in the structures to stack against smaller alkyl residues, whereas Pro140 would seem positioned to interact with larger ones. Perhaps surprisingly, AGT does not form an H-bond with the purine N1 atom, which could seemingly be used to discriminate between G and O^6 -MeG due to the difference in protonation state.

Interestingly, it has been suggested that perhaps the inherent difficulty in discriminating between guanine and O^6 -alkylguanine has led to the unique direct damage repair pathway.(51) If DNA glycosylases mistakenly remove undamaged bases during the course of the more commonly occurring base-excision repair, it can be deleterious to the organism.(122,123) However, this would not be the case for MGMT, where any lack of specificity is neither cytotoxic nor mutagenic as the undamaged base will simply go unrepaired and thus be flipped back into the DNA helix.

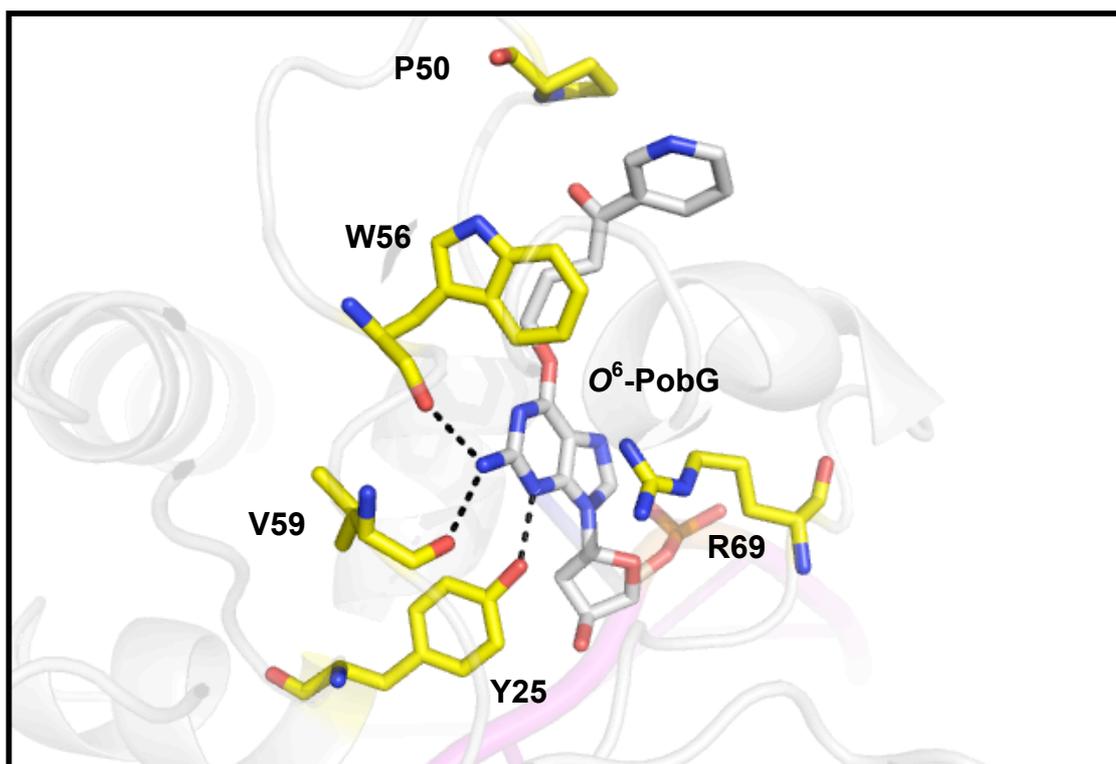


Figure 5.7: Active site of AtI1 bound to ODN containing O^6 -PobG showing interactions

AtI1 is the alkyltransferase-like (ATL) protein homologue from *S.pombe* that also recognises O^6 -alkylguanine lesions. Whilst the active site (figure 5.7) is approximately three times bigger than that of MGMT, the major interactions present in the MGMT active site are conserved. The hydroxyl of Tyr25 forms a hydrogen bond with the N3 atom of guanine and the side-chain carbonyls of Val59 and Trp56 hydrogen bond with N2.(58) Trp56 (and possibly Ala117 in the active site of TTHA1564, although structural data is unavailable for this protein) forms a hydrophobic interaction with the O^6 -alkyl group, and Pro50 is seen to interact with larger O^6 -alkyl residues such as PobG. One of the crucial factors not observed in the MGMT active site is the presence of Arg69 which forms a cation- π interaction with the pyrimidine base (the corresponding residue in TTHA1564 is Phe130).

5.2 Base Recognition by ATL Proteins

In an attempt to understand the mechanisms of base recognition in greater detail, ODN substrates containing a variety of modified purine bases were used to quantify the binding affinity of two ATL proteins, Atl1 from *S.pombe* and TTHA1564 from *T.thermophilus*. The fluorescence-based assays that were performed are described in detail in Chapter 4. The K_D values presented here are an average of those derived from three separate titrations, and the calculated error values are the standard error of the mean (SEM). The experiments were carried out with SIMA-labelled single-stranded (ss) ODNs, duplexes where the modified base was paired opposite cytosine (C), and duplexes where the modified base was paired opposite thymine (T) (figure 5.8).

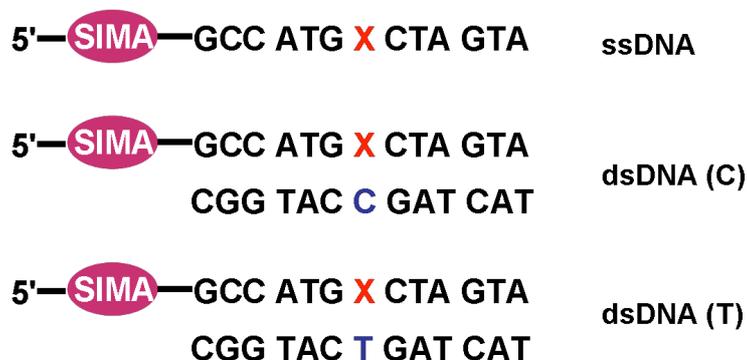


Figure 5.8: Sequences of ODNs used as substrates in the study of base recognition

The modified bases that were incorporated into the study are shown in figure 5.9 and include many O^6 -alkylguanine lesions as well as a number of related purine bases. The O^6 -alkyl groups incorporated vary in size (from the

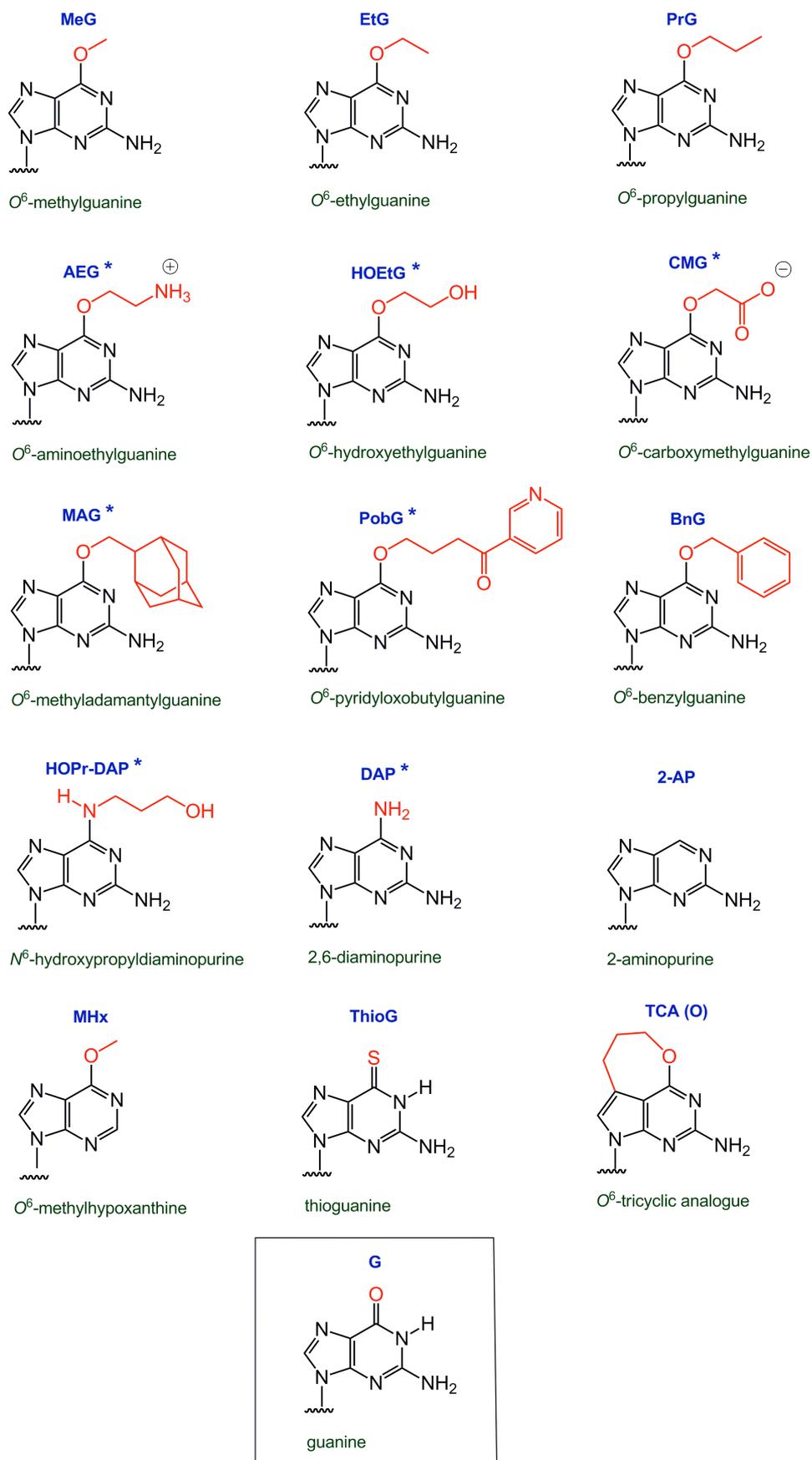


Figure 5.9: Modified bases incorporated into ODNs for the study. The lesions labelled with asterisks are known to be poorly or not repaired by MGMT

smallest, O⁶-MeG, to the very large, bulky residues O⁶-PobG and O⁶-MAG), charge (O⁶-AEG is positively charged and O⁶-CMG negatively charged at pH 7.5), hydrophobicity (which increases with size of the alkyl group) and conformation relative to the O⁶-bond (TCA(O) is locked in the *anti*-conformation whereas a bond in free rotation would be expected to orientate into the *syn*-conformation upon binding to MGMT).(95) In addition, the modified purine residues included in the study were 2-AP, 2,6-DAP, HOPr-DAP and O⁶-MHx. These purine analogues were interesting due to the presence or lack of N2, N6 (or O6) atoms, and alkyl groups at the 6-position of the base. In combination with structural data, these experiments allowed us to discover the relative importance of the various interactions in the protein active site involved in base recognition and provided insights into the mechanism of lesion binding. (Complete tables of binding data with single-stranded and double-stranded ODNs, and the AtI1 mutants, are shown at the rear of this chapter (section 5.7)).

5.2.1 Recognition of Single-Stranded ODN Substrates by AtI1

It was demonstrated that AtI1 specifically recognises ODNs containing all of the O⁶-alkylguanine lesions that were evaluated. The results are shown in figure 5.10 (the result for the control ODN containing guanine in place of the modified base is omitted for clarity: $K_D=741\text{nM}$). AtI1 has a preference for binding bulky, hydrophobic lesions which can be seen in the greater binding affinities (and correspondingly lower values of the dissociation constant, K_D) for O⁶-benzylguanine (O⁶-BnG), O⁶-propylguanine (O⁶-PrG) and O⁶-methyladamantylguanine (O⁶-MAG). This is likely to be in part due to the

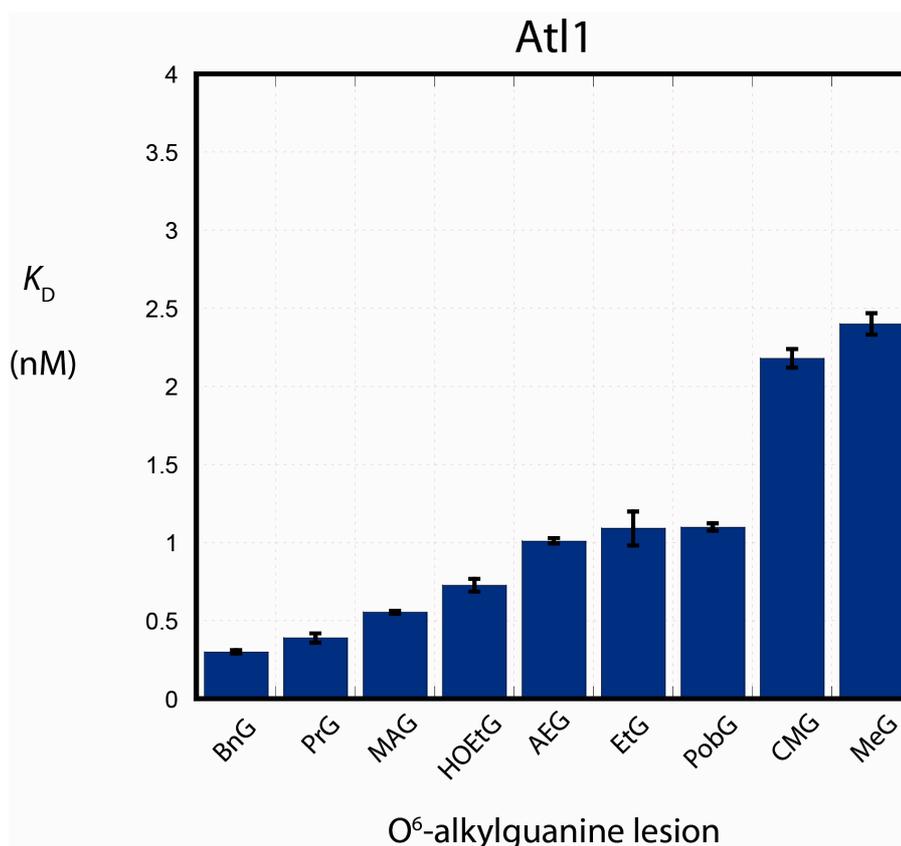


Figure 5.10: Plot to show dissociation constants between Atl1 and single-stranded ODNs containing modified bases

increased hydrophobic interactions with the Atl1 active-site tryptophan residue (W56) which would favour larger hydrophobic groups. The difference in affinity of Atl1 for ODN-containing O⁶-benzylguanine compared with O⁶-methylguanine is almost ten-fold, a significant difference which may have profound implications in terms of which type of repair Atl1 initiates (section 5.6). It has also been shown that ODNs containing O⁶-carboxymethylguanine (along with O⁶-methylguanine) is one of the poorer substrates for Atl1 despite the moderate size of the alkyl group: this may be due to the presence of a negative charge, though ODN containing the positively-charged O⁶-aminoethylguanine lesion is recognised with reasonably high affinity making it difficult to draw any firm conclusions with regard to the effect of charge on

recognition. More interestingly, a significant number of lesions that are poorly repaired or refractory to repair by MGMT are recognised by At1. This includes O^6 -CMG-containing ODNs in addition to those containing O^6 -AEG, O^6 -MAG, O^6 -PobG and O^6 -HOEtG. It is not clear whether these lesions are poorly repaired by MGMT because they are not bound with high affinity, or because of a subsequent slow rate of alkyltransfer reaction after recognition. It is possible the smaller active-site pocket of MGMT does not accommodate the larger lesions such as O^6 -MAG and O^6 -PobG effectively and hence the protein cannot repair them very efficiently. Furthermore, these results would seem to be consistent with the recent findings in *E.coli* that ATL proteins mediate efficient repair of O^6 -alkylguanine lesions that are not processed by AGT proteins.(80)

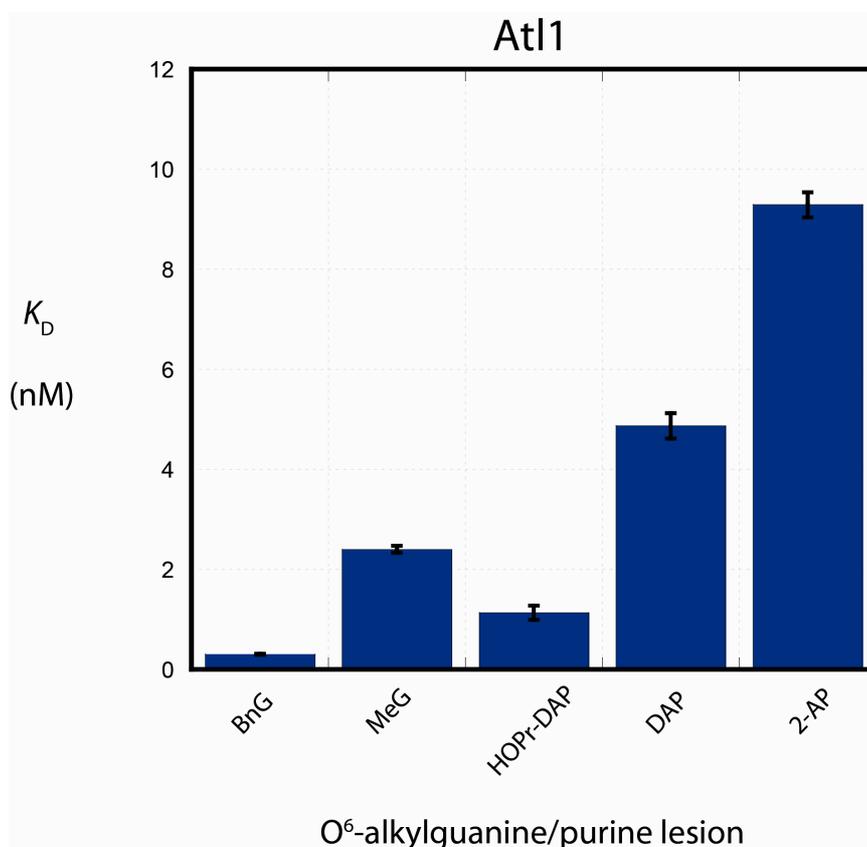


Figure 5.11: Plot to show dissociation constants between At1 and single-stranded ODNs containing modified bases

The results for some of the modified purine bases with O^6 -BnG and O^6 -MeG as references are shown in figure 5.11. Perhaps surprisingly N^6 -hydroxypropyl-2,6-diaminopurine (HOPr-DAP) was bound by AtI1 with high affinity, with a K_D approximately midway between that of O^6 -BnG and O^6 -MeG. This would suggest that the O6 atom itself is not important for recognition (i.e. it can be replaced with N and not cause significant disruption to the interactions). Remarkably, 2,6-diaminopurine and 2-aminopurine are bound with high affinity compared to control ODN (c.f. $K_D = 741\text{nM}$), though with slightly less affinity than any of the O^6 -alkylguanine lesions. It seems clear from these results that the presence of an alkyl group affects recognition: HOPr-DAP is recognised with 5-fold more affinity than 2,6-DAP (where the alkyl group is replaced with an NH_2) and 9-fold more affinity than 2-AP (where there is an absence of any group). The observation that AtI1 recognises these purine residues, along with subsequent structural studies of AtI1 in complex with ODNs containing 2,6-DAP and 2-AP (section 5.5), has led to the proposal of a unique and unprecedented method of base recognition by AtI1. Thus, the active-site residue Arg69 appears to act as a molecular probe by scanning the electrostatic potential of the base (section 5.4).

The dissociation constants of various other modified purine bases are shown in figure 5.12. All of these bases are considered to be poorly bound or not recognised by AtI1. The control ODN containing guanine is also included in the plot and as expected has the largest K_D value of all the substrates. The importance of the N2-amino group for recognition by the protein can be clearly seen: O^6 -methylhypoxanthine (O^6 -MHx) differs from O^6 -methylguanine only by

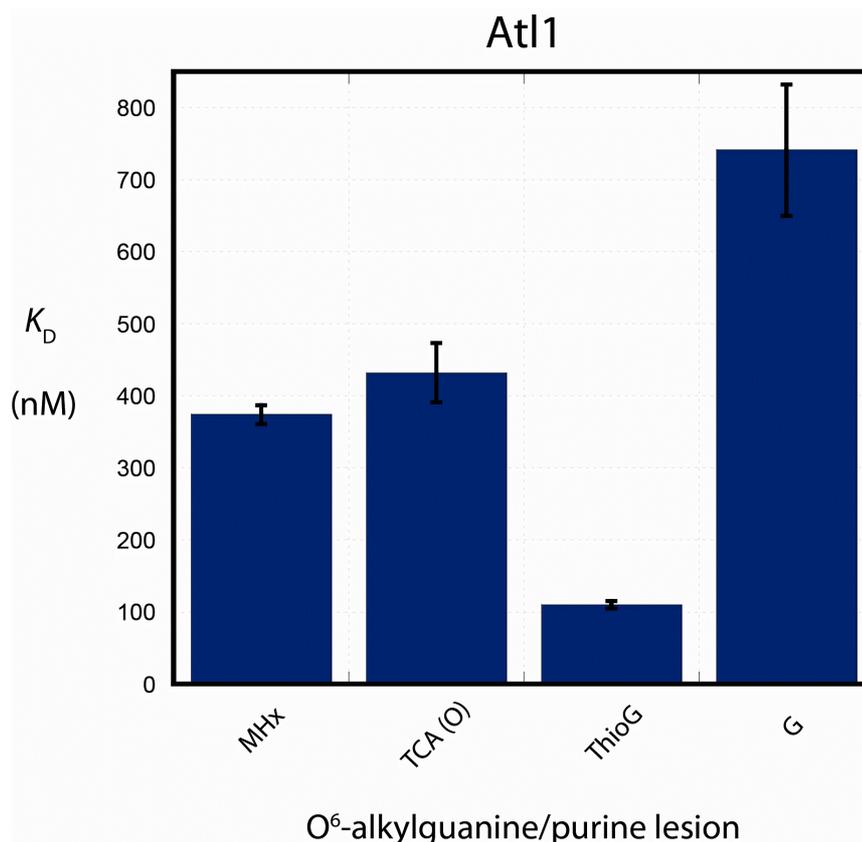


Figure 5.12: Plot to show dissociation constants between Atl1 and single-stranded ODNs containing modified bases

the absence of the 2-amino group, however ODN containing O⁶-MHx is bound with 150-fold less affinity than the ODN containing O⁶-MeG and only 2-fold more affinity than the control ODN. This is almost certainly due to the inability of the Atl1 active-site residues Val59 and Trp56 to hydrogen-bond with the missing N2-amino group. These interactions are conserved between MGMT and Atl1 and are likely to have the role of aiding in discrimination between guanine and the other canonical bases.

Also of interest is the inability of Atl1 to recognise ODNs containing the tricyclic alkylguanine analogue (TCA (O)). As previously mentioned, this modified base is locked in the *anti*-conformation, but apart from this it has the

other features of O^6 -alkylguanine residues. It has been reported that MGMT will not repair O^6 -alkylguanine bases when the alkyl group is in the *anti*-conformation as this orientates it away from the reactive cysteine residue.(95) It is also suggested that MGMT favours alkyl groups in the *syn*-conformation as they are better able to fit into the hydrophobic binding pocket. Perhaps the reason for the lack of recognition of ODN containing TCA (O) by AtI1 is that the locked *anti*-conformation of the alkyl group prevents the initial inclusion of the residue in the active site binding pocket. If this is the case, a consequence of its exclusion from the binding cavity would be that it is unable to be recognised by the specific active-site interactions mentioned previously. Indeed, in all available structures of AtI1 in complex with ODNs containing a wide variety of O^6 -alkylguanine residues, the alkyl group is always orientated in the *syn*-conformation. It would therefore seem likely that this conformation is required for AtI1 binding (and possibly also for MGMT).

The ODN containing 6-thioguanine (ThioG) is also poorly recognised, though intriguingly it would appear to be bound with around 7-fold more affinity than the ODN containing guanine. The reasons for this are unclear at present and warrant further investigation.

5.2.2 Recognition of Single-Stranded ODN Substrates by TTHA1564

In a set of identical experiments, the dissociation constants of TTHA1564 from *T.thermophilus* with various modified DNA substrates were measured. In common with AtI1, this protein recognised all the ODNs containing O^6 -alkylguanine residues that were evaluated. The results are shown in figure 5.13 (as for the previous plot, the result for the control ODN

containing guanine is omitted for clarity: $K_D=103\text{nM}$). Whilst TTHA1564 recognises all these substrates with relatively high affinity compared to the control, it does not appear to show any preference for binding ODNs containing large, bulky O^6 -alkylguanine lesions. In fact, the substrate recognised with the highest affinity is that containing O^6 -methylguanine, which is the poorest bound of the O^6 -alkylguanine substrates for AtI1. It may be that

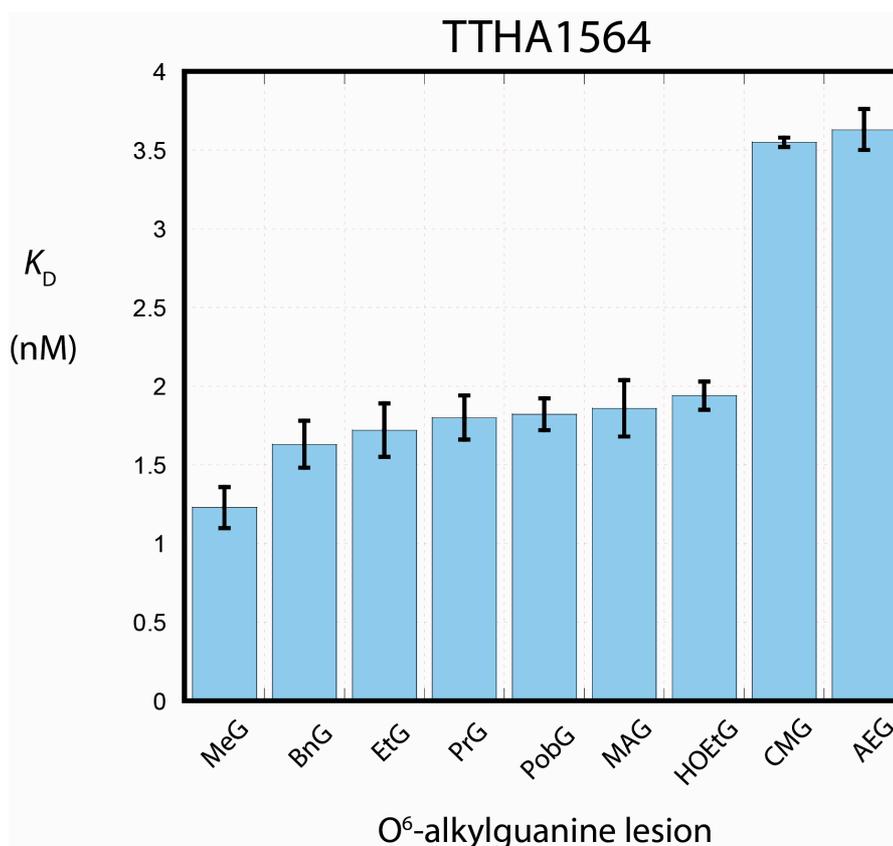


Figure 5.13: Plot to show dissociation constants between TTHA1564 and single-stranded ODNs containing modified bases

the presence of alanine rather than tryptophan in the TTHA1564 active site motif (it has the sequence PAHR, compared to PWHR in AtI1) limits the ability of the protein to engage in such an extensive hydrophobic interaction with larger O^6 -alkyl groups (such as O^6 -benzylguanine and O^6 -propylguanine).

Certainly, there appears to be little variation in the dissociation constants of the majority of the ODN substrates containing O^6 -alkylguanine lesions regardless of the identity of the alkyl group. The exceptions to this are the ODNs containing O^6 -carboxymethylguanine (O^6 -CMG, which was also bound with relatively poor affinity by AtI1) and O^6 -aminoethylguanine (O^6 -AEG). Whether this is due to the charged nature of these modified bases is open to speculation, however O^6 -AEG was bound with moderately high affinity by AtI1.

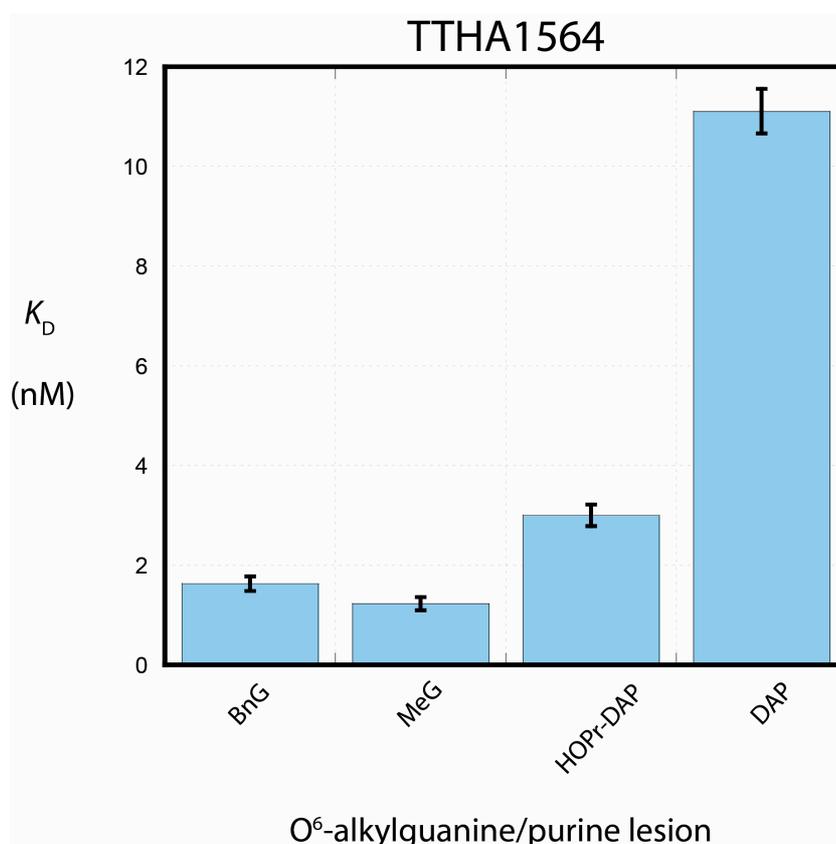


Figure 5.14: Plot to show dissociation constants between TTHA1564 and single-stranded ODNs containing modified bases

It was also demonstrated that TTHA1564 was able to recognise N^6 -hydroxypropyl-2,6-diaminopurine (HOPr-DAP) and 2,6-diaminopurine (DAP) (figure 5.14). Whilst the ODN containing HOPr-DAP was bound with a similar

affinity to the O^6 -alkylguanine-containing substrates, the ODN containing DAP was bound with around 8-to-10 fold less affinity than the O^6 -alkylguanine-containing substrates (for comparison, DAP was recognised with around 5 fold less affinity by AtI1 than most of the O^6 -alkylguanine-containing substrates). These results reinforce the observation that the presence of an alkyl group at the 6-position on the purine base aids recognition by ATL proteins, and also that these proteins are able, seemingly inexplicably, to recognise other modified purines such as DAP (but not adenine). The experiment with an ODN containing 2-AP using TTHA1564 protein was not able to be carried out due to the unavailability of this ODN in Japan, but it would be expected to be recognised with less affinity than DAP (in common with AtI1).

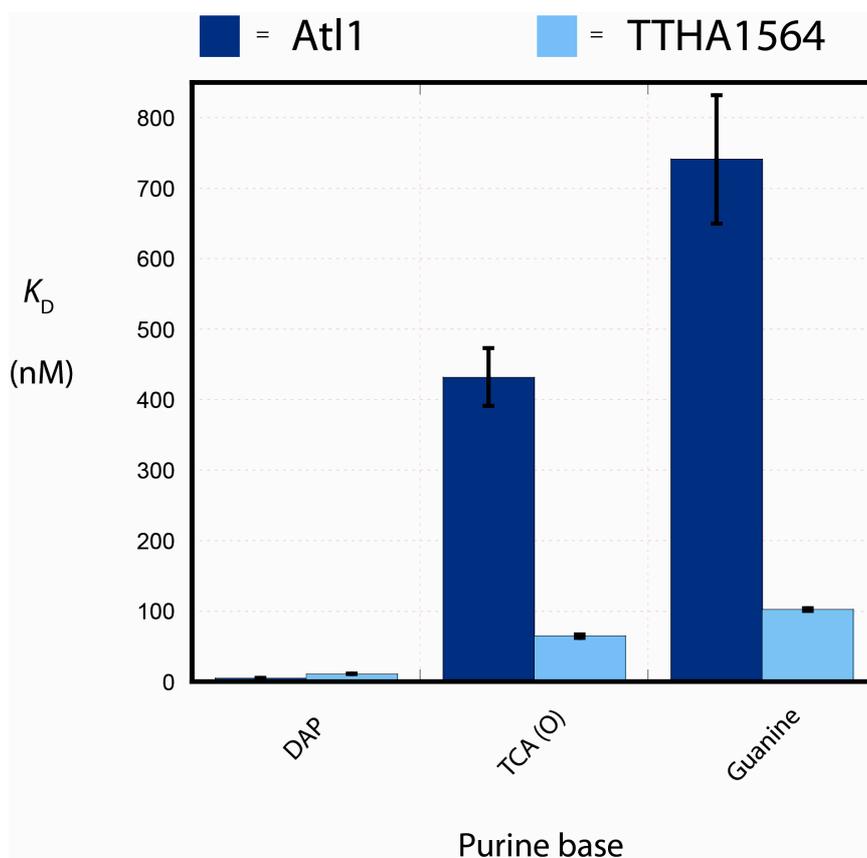


Figure 5.15: Plot to show dissociation constants between AtI1 and TTHA1564 and single-stranded ODNs containing modified bases

For the same reasons mentioned above, it was not possible to carry out assays with O^6 -MHx and thioguanine using TTHA1564. However, the K_D was measured for the tricyclic analogue (TCA (O)) and it found to be only very weakly recognised by TTHA1564. Selected results are shown in figure 5.15.

An observation that becomes immediately apparent from these results is that TTHA1564 has much more affinity for the control sequence containing guanine than Atl1. This means that TTHA1564 appears to have far less ability than Atl1 to discriminate between an ODN containing an O^6 -alkylguanine lesion and one that does not, though obviously TTHA1564 is still capable of specifically recognising O^6 -alkylguanine lesions (see table 5.3). Whilst it is true that both ATL proteins have some affinity for unalkylated DNA, which is likely to be due to the presence of the helix-turn-helix (HTH) binding motif that binds the minor groove, it is unclear why the *T.thermophilus* protein would bind unmodified DNA more tightly than Atl1. It is reasonable to suggest that general recognition of DNA is advantageous for ATL proteins in order to bring them in close contact with potential substrates which are then bound tightly if an O^6 -alkylguanine residue is present, in an analogous manner to DNA glycosylases (section 5.1). It may be that the difference between the high temperatures in which the *T.thermophilus* resides ($\sim 75^\circ\text{C}$) may necessitate a stronger affinity of the protein with DNA *per se* than is needed for Atl1.

The difference in the ability of the two ATL proteins to discriminate between DNA containing O^6 -alkylguanine residues and undamaged DNA can clearly be seen in table 5.1. The comparisons (i.e. 2-fold etc.) are calculated by comparing the K_D of each substrate with the K_D of the ODN bound with the greatest affinity for that protein. For example, for Atl1, the control substrate is

	BnG	MeG	DAP	TCA (O)	G
AtI1	0.3 ± 0.01	2.4 ± 0.08	4.9 ± 0.25	432 ± 41	741 ± 91
TTHA1564	1.63 ± 0.15	1.23 ± 0.13	11.1 ± 0.45	64.7 ± 2.6	102.6 ± 2
AtI1	1-fold	8-fold	16-fold	1450-fold	2500-fold
TTHA1564	1.3-fold	1-fold	9-fold	50-fold	80-fold

Table 5.1: Discrimination between ODNs containing various modified bases and guanine by AtI1 and TTHA1564 compared to the substrate bound with the highest affinity

recognised with 741nM / 0.3nM = 2500-fold less affinity than O^6 -BnG, and for TTHA1564 the control substrate is recognised with 102.6nM / 1.23nM = 80-fold less affinity than O^6 -MeG. The reasons for these differences will be considered in section 5.6.

5.2.3 Recognition of Double-Stranded Substrates

In general, it was found that both AtI1 and TTHA1564 recognised single-stranded (ss) and double-stranded (ds) ODNs containing the same lesions with similar affinity. Rather than evaluating every single substrate as dsDNA, we chose a selection of the SIMA-labelled ODNs containing modified bases and annealed them to complementary sequences to form duplex DNA. As mentioned previously, two forms of duplex were made: one where the modified base (X) was base-paired with cytosine (C) and the other where X was paired with thymine (T) (figure 5.8). The annealing process was analysed by FRET to confirm that the substrates were duplexes at a concentration of 1nM in 50mM NaCl solution and at room temperature, the conditions used in

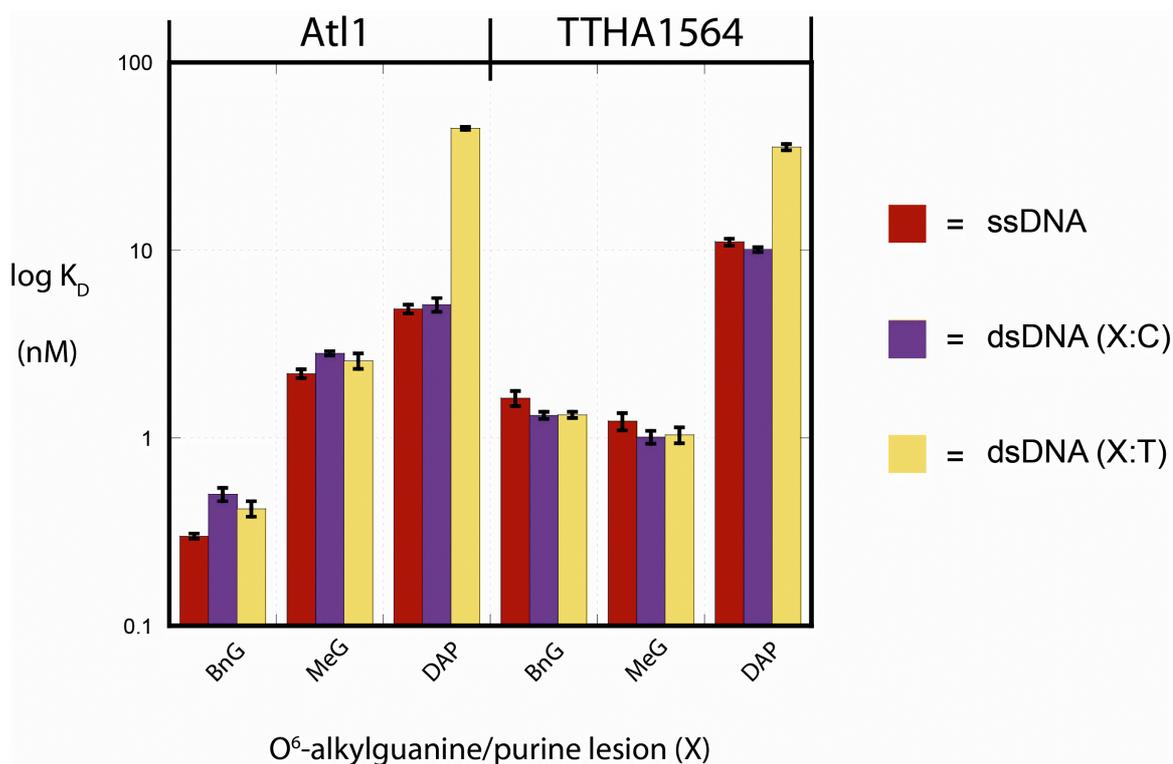
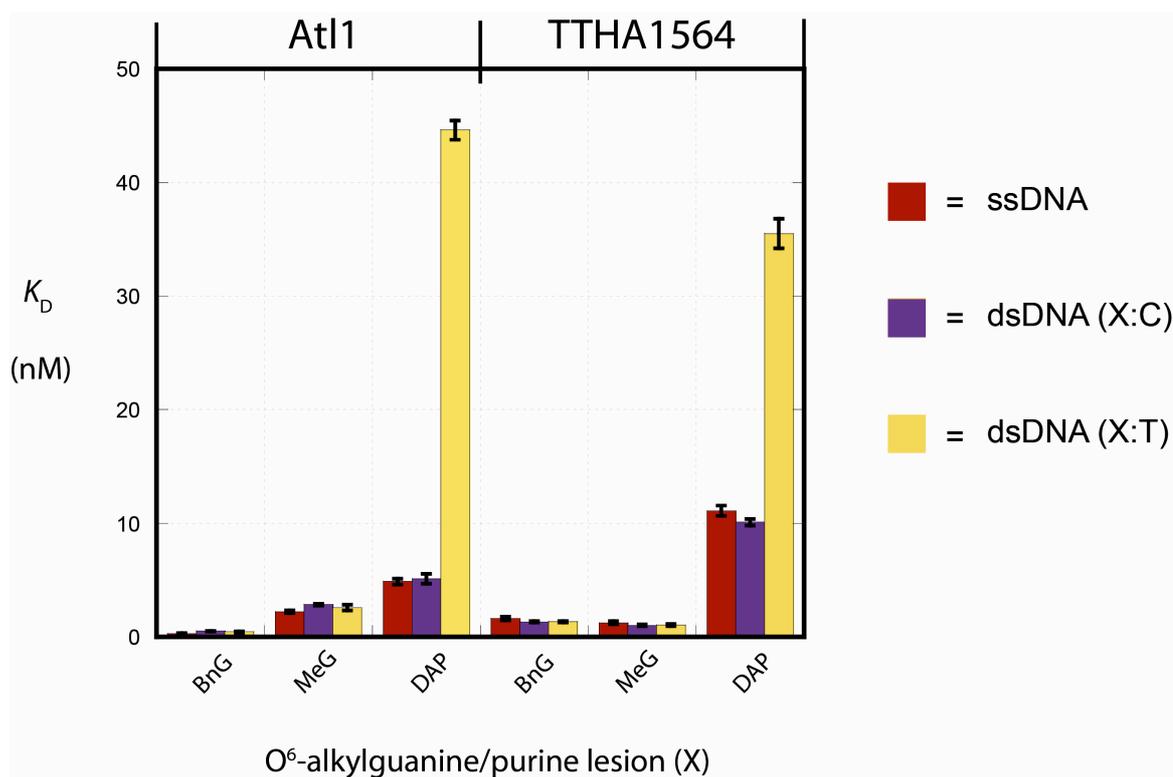


Figure 5.16: Plots to show binding of AtI1 and TTHA1564 to double-stranded ODNs containing various modified bases

the assay (section 4.5). O^6 -benzylguanine (O^6 -BnG), O^6 -methylguanine (O^6 -MeG), 2,6-diaminopurine (DAP), and guanine (G) were evaluated and the results are shown in figure 5.16. For O^6 -BnG, and O^6 -MeG, there is little preference of either ATL protein for single- or double-stranded DNA, and little difference in K_D value whether the modified base (X) is opposite C or T. In the case of DAP, there is a pronounced and significant effect which is seen for binding of both At11 and TTHA1564 proteins. The double-stranded ODN where DAP is base paired with C is bound with approximately the same affinity as the corresponding single-stranded ODN. However, the double-stranded ODN where DAP is base paired with T has a dissociation constant that is far larger than for when DAP is paired with C (9-fold larger for At11, 3.5-fold larger for TTHA1564). This result may be explained by the nature of the base pairing in these duplexes and the mechanism of recognition by ATLs. ATL proteins flip damaged nucleotides into an active site pocket where recognition takes place. When DAP is base paired with cytosine there are two hydrogen bonds between the bases, whereas when it is paired with thymine, there are three (figure 5.17). For O^6 -alkylguanine residues, there are two

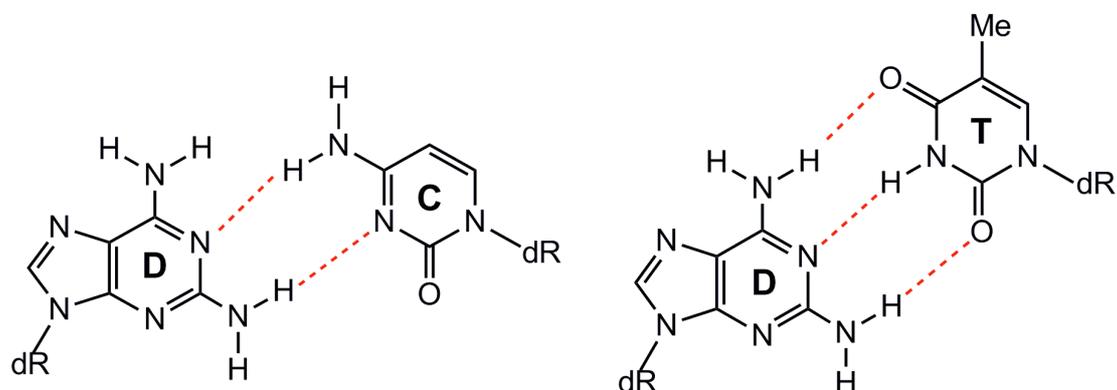


Figure 5.17: Base pairing of 2,6-diaminopurine with cytosine (left) and thymine (right)

hydrogen bonds in the base-pair regardless of whether C or T is the partner. This additional hydrogen bond between DAP and T will stabilise the base pair and is likely to provide an additional thermodynamic barrier to active flipping of the modified base by ATL proteins. Alternatively, the more stable DAP:T base pair is likely to sample the random extra-helical conformation less frequently than DAP:C and so will be less exposed and hence less prone to detection by ATL proteins. In either case, this is reflected in the reduced affinity of AtI1 and TTHA1564 for DAP when it is paired with T in a duplex. These results are consistent with the base-flipping mechanism seen in crystal structures for AtI1 (58) and demonstrated using fluorescent assays for TTHA1564.(63)

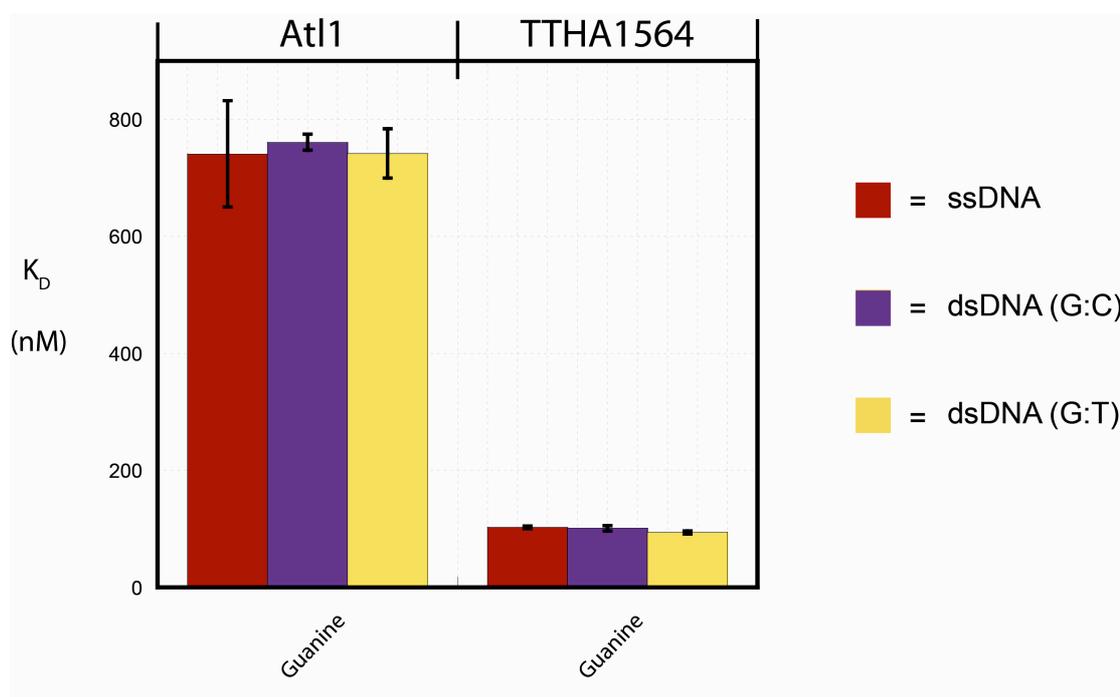


Figure 5.18: Plot to show AtI1 and TTHA1564 weakly binding to control ODNs

Finally, the results with the control ODNs (figure 5.18) showed that ATL proteins do not specifically recognise mismatches. Atl1 and TTHA1564 showed little preference for double-stranded over single-stranded DNA and also did not appear to discriminate between the duplexes whether guanine was paired with cytosine or thymine.

5.3 Mutagenesis Studies of Atl1

It was decided to evaluate the importance of the Arg69 residue in Atl1 for recognition of O^6 -alkylguanine and related purine residues by performing the experiments with mutant Atl1 proteins. In collaboration with the Margison laboratory, Atl1-R69F and Atl1-R69A were prepared by site-directed mutagenesis (section 9.10) and overexpression in *E.coli* (see section 9.9). Identical assays to those with wild-type (WT) Atl1 were conducted to measure the dissociation constants of the mutant proteins with a selection of the ODNs. The results for Atl R69A are shown in figure 5.19 where they are compared with the Atl1 WT values. Here it is demonstrated that substitution of the Arg69 residue for Ala has a profound effect on the ability of Atl1 to recognise ODNs containing O^6 -alkylguanine and related purine lesions. The values of the dissociation constant for the Atl1 R69A mutant compared to WT are on average approximately 50-fold greater for the bound substrates, although surprisingly the control ODN is recognised with 5-fold more affinity by the mutant than WT. It may be that the interaction of Arg69 with the base in the active site is attractive for O^6 -alkylguanine lesions and related purine residues, but repulsive for guanine, and so when Arg69 is not present, the protein

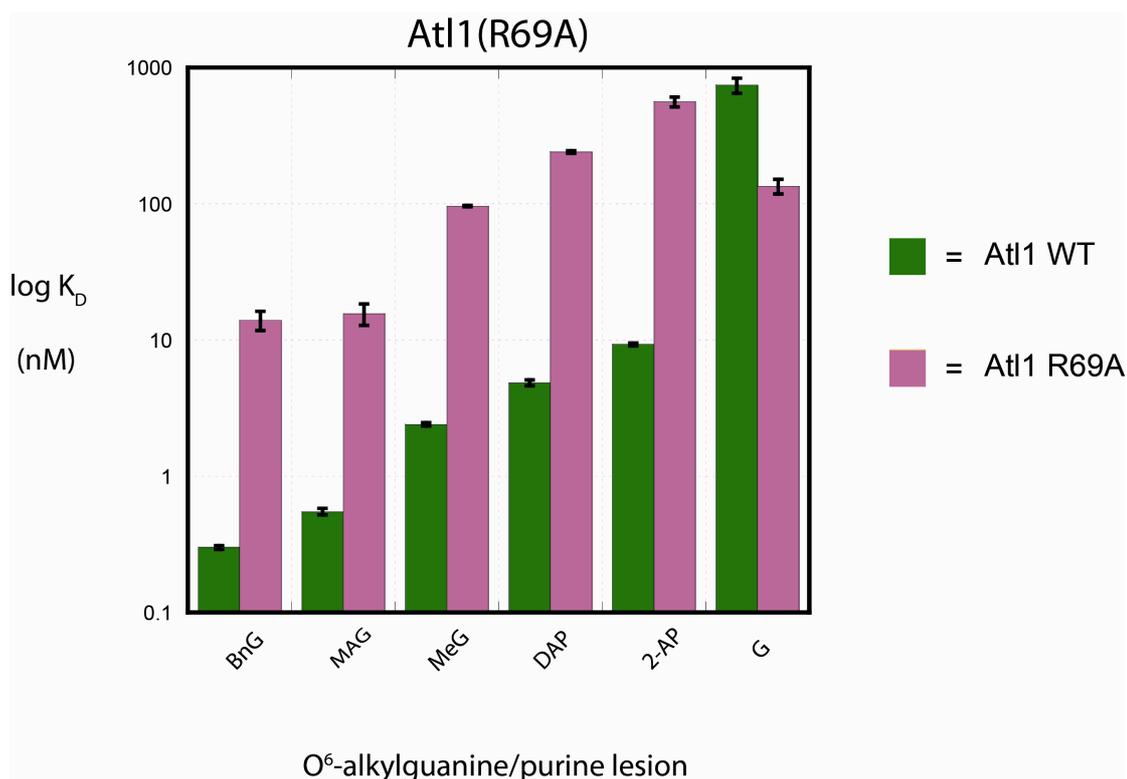


Figure 5.19: Plot to show AtI1 R69A mutant and wild-type AtI1 binding ODNs containing modified bases

cannot adequately discriminate between the damaged and undamaged bases. Clearly, non-conservative mutation of Arg69 to Ala severely disrupts the ability of AtI1 to recognise lesions and shows the importance of this key residue.

In addition, it was decided that the AtI1 R69F mutant would be studied. Comparison of the sequences of AtI1 and TTHA1564 indicated that the residue corresponding to Arg69 in AtI1 would be phenylalanine (Phe130, or F130) in the *T.thermophilus* protein. Whilst this residue is not positively charged like Arg69 (figure 5.20) and so cannot engage with the bound nucleobase through a cation- π interaction, it is hydrophobic in nature and as such could have the ability to form a π - π stacking interaction that may be

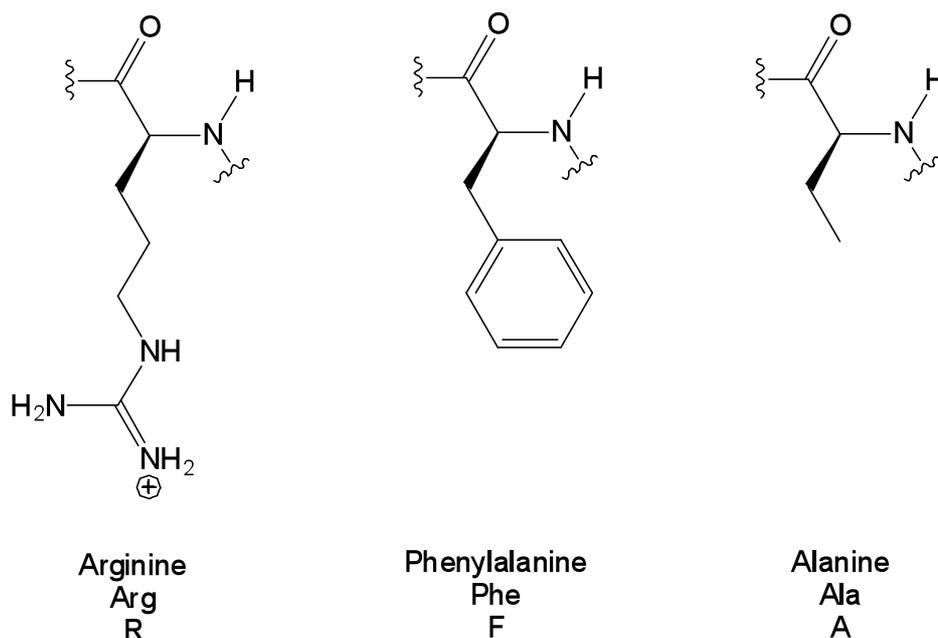


Figure 5.20: Amino acid side chain structures at position 69 of At1 wild-type (Arg) and mutants (Phe and Ala)

involved in recognition, assuming that the TTHA1564 active site loop is folded in a similar manner to that of At1. Of course, without any structural data it is impossible to draw any firm conclusions. However, it has already been demonstrated that TTHA1564 displays less ability to discriminate between O^6 -alkylguanine and guanine residues than At1 and it was therefore interesting to see if this was also the case for the At1 R69F mutant. The results of the binding assays using the At1 R69F mutant are shown in figure 5.21.

Compared to At1 WT, the ODNs containing bulky, hydrophobic O^6 -alkylguanine residues are recognised with around 5-fold less affinity, whereas for ODNs containing O^6 -methylguanine and 2,6-diaminopurine this difference is around 15-fold. Possibly, the contribution from the hydrophobic interaction between the active site Trp56 and the O^6 -alkyl group becomes more

significant when Arg69 is missing. Table 5.2 shows the differences in binding affinity of the mutant At1 proteins compared to wild-type. It can clearly be

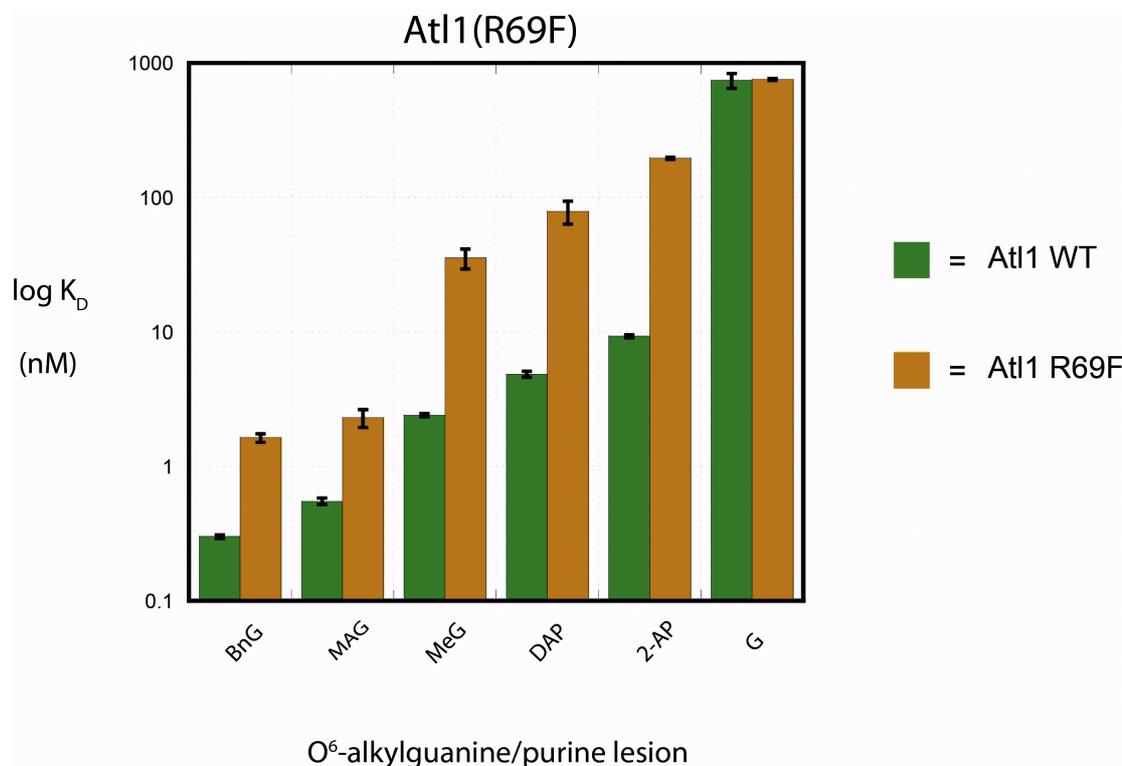


Figure 5.21: Plot to show At1 R69F mutant and wild-type At1 binding ODNs containing modified bases

	BnG	MAG	MeG	DAP	2-AP	G
At1 WT	1-fold	1-fold	1-fold	1-fold	1-fold	1-fold
At1 R69A	47-fold	28-fold	40-fold	50-fold	60-fold	0.2-fold
At1 R69F	5.5-fold	4.2-fold	15-fold	16-fold	21-fold	1-fold
TTHA1564	5.5-fold	3.4-fold	0.5-fold	2.3-fold	N/A	0.15-fold

Table 5.2: Difference in affinity of the mutant proteins and TTHA1564 for ODNs containing selected modified bases compared to wild-type At1

seen that R69A only weakly recognises the O^6 -alkylguanine-containing substrates compared to WT, whilst R69F recognises them with moderate affinity. Therefore it would seem that Phe69 is still able to mediate recognition, albeit with less precision than Arg69. If the data is analysed to calculate the relative affinities of each protein individually for each of the modified substrates, then the discrimination that AtI1 and each mutant has for different modified bases and guanine can be displayed. The data is shown in table 5.3.

	BnG	MAG	MeG	DAP	2-AP	G
AtI1	1-fold	2-fold	8-fold	16-fold	31-fold	2500-fold
AtI1 R69A	1-fold	1.1-fold	7-fold	17-fold	40-fold	10-fold
AtI1 R69F	1-fold	1.4-fold	7-fold	50-fold	120-fold	450-fold
TTHA1564	1.3-fold	1.5-fold	1-fold	9-fold	N/A	80-fold

Table 5.3: Difference in affinity of AtI wild type and mutant proteins and TTHA1564 for ODNs containing selected modified bases **compared to the tightest bound substrate** for each protein (i.e. the **discrimination** of each individual protein between different bases)

Interestingly, the R69A mutant and the R69F show similar levels of discrimination between O^6 -BnG, O^6 -MAG and O^6 -MeG to wild-type AtI1. This may be explained by the fact that mutation of Arg69 does not affect the hydrophobic interaction between the O^6 -alkyl group and Trp56, which would be more extensive for the larger chains and therefore affect the value of K_D . In addition, whilst R69A shows comparatively little discrimination between O^6 -alkylguanine residues and guanine, R69F still has the ability to recognise ODNs containing O^6 -BnG with 450-fold more affinity than ODN containing G,

and ODN containing O^6 -MeG with 65-fold more affinity than that containing G. Intriguingly, this latter discrimination between O^6 -MeG and G is similar to that shown by TTHA1564 (approximately 80-fold).

5.4 Molecular Electrostatic Potentials of Nucleobases

Insights were given into the mechanism of AtI1 base recognition by the calculation of the molecular electrostatic potential (MEP) of different modified and unmodified bases. The MEPs describe the relative attractive and repulsive forces between the atoms in a heterocyclic base and a cation interacting with them. They are calculated by placing a positive charge (the cation, in this case the charged N-terminus of the Arg69 residue) at a certain distance from the base (this distance was calculated from the positions of Arg69 and the base in the crystal structures which was approximately 3.5Å). The MEP calculations allow subtle differences in the electronic nature of the bases to be examined that may help explain the discrimination and selectivity of the AtI1 active site, and provide a mechanistic basis for 'proof-reading' by Arg69 in its proposed role as a molecular probe of electrostatic potential.

In collaboration with Prof. Chris Hunter, the molecular electrostatic potentials for a selection of the modified bases were calculated and are shown in figure 5.22. Comparison of the MEP values for O^6 -methylguanine (which is representative of all O^6 -alkylguanine residues for the purposes of this analysis) and guanine highlights some clear differences between them. Consider the C2 atom of the heterocyclic rings: for O^6 -MeG there is a positive interaction with a relative value of +50, whereas for guanine the interaction is negative with a value of -10. Therefore, according to these calculations, the

presence of Arg69 will cause O^6 -alkylguanine residues to be held in the binding pocket by the attractive force, whilst guanine residues will be repelled. By the same measure, the greater attraction of Arg69 for N1 in O^6 -MeG (MEP values for N1 of +90 for O^6 -MeG, and +20 for G) is highly likely to contribute to the Arg69-mediated selectivity of the active site for O^6 -alkylguanine residues.

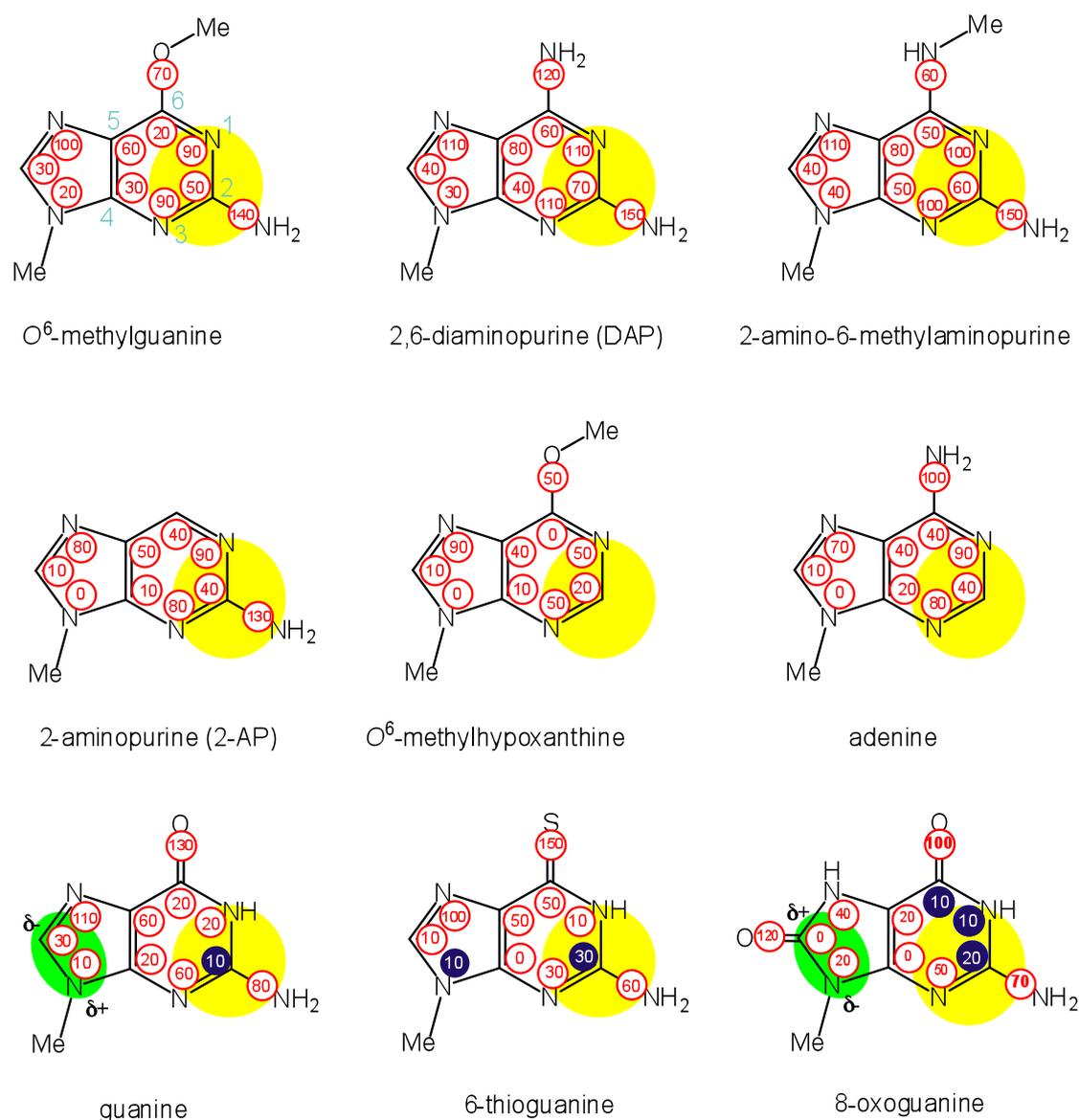


Figure 5.22: Molecular electrostatic potentials of selected modified purine bases. Numbers in red indicate an attractive interaction with a cation, and blue a repulsive one: therefore these values are related to electron density. Larger numbers indicate more extensive interactions.

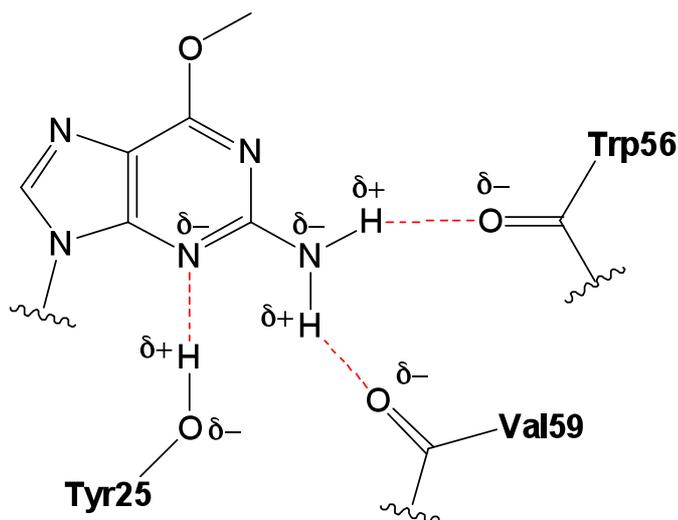


Figure 5.23: Increased hydrogen bonding interactions due to polarisation caused by changes in molecular electrostatic potentials induced by Arg69

In addition, the presence of a cationic charge on Arg69 in proximity to the heterocyclic ring leads to differences in electrostatic potential between some of the other atoms in the bases. For O^6 -MeG, N2 has a value of +140 whilst for G this value is +80. Similarly, the MEP values for N3 between O^6 -MeG and G are +80 and +60 respectively. Given that these MEP values represent the interaction of each atom in the base with a positive charge, a larger value effectively means increased polarisation. This in turn means increased electrostatic charge (δ^-) on this atom which will affect hydrogen bonding. As previously shown in AtI1-DNA structures, the crucial interactions in the active site are H-bonds between the N2 hydrogens and the carbonyl groups of Trp56 and Val59, and an H-bond between N3 and the side chain hydroxyl H of Tyr25 (figure 5.23). It follows that if the δ^- of the N2 and N3 atoms increases, then all of these H-bonding interactions will be strengthened.

For 2,6-diaminopurine (DAP) and 2-aminopurine (2-AP) the MEP values are far more similar to those of O^6 -methylguanine than to guanine (see table 5.4 for MEP values of the more important atoms on the bases). Therefore, the observation that these residues are also recognised by AtI1 can be explained by the similar interactions that Arg69 has with these purine bases. The MEP values for adenine are also similar to those of O^6 -MeG, but crucially adenine lacks the N2 atom. It would seem that the hydrogen bonding interactions between the active site residues Trp56 and Val59 are essential for discrimination between a guanine base (modified or otherwise) and the other canonical bases. Only when this interaction is present are the other aspects of recognition important, i.e. the probing of electrostatic potential of

	N1	C2	N2	N3
O^6-MeG	+90	+50	+140	+90
DAP	+110	+70	+150	+110
2-AP	+90	+40	+130	+80
G	+20	-10	+80	+60
A	+90	+40	Not Present	+80

Table 5.4: Molecular electrostatic potentials of some of the modified bases

the base by Arg69. The lack of recognition of O^6 -methoxyhypoxanthine (O^6 -MHX, which also lacks an N2 atom) also provides evidence that this is the case.

However, it is difficult to explain the observation that ODNs containing 6-thioguanine are recognised with more affinity than ODNs containing

guanine. The MEP calculations suggest that ODNs containing ThioG would be recognised with an affinity similar to that of the natural sequence and this is not entirely consistent with the binding data where the K_D value was around 7-fold less for the ODN containing ThioG than the control. There may be other factors involved that affect the binding of ThioG that are not presently known. However, what would seem to be more important is that ODNs containing ThioG are recognised with low affinity compared to ODNs containing O⁶-alkylguanine, 2,6-diaminopurine and 2-aminopurine residues.

The same MEP calculations were performed for 8-oxoguanine and the findings were found to be in agreement with previous results concerning the recognition of 8-oxoguanine by hOGG1.(118) It was proposed that in the hOGG1 active site a Lys249 (NH₃⁺) / Cys253 (S-) salt bridge forms a dipole which runs antiparallel to the local dipole of 8-oxoguanine and parallel to that of guanine (section 5.1.2). Comparison of the MEP calculations for guanine and 8-oxoguanine (figure 5.22) show that the dipole between C8 and N9 is indeed running in different directions for these residues: the MEP values are C8 (0) → N9 (20) for 8-oxoguanine and N9 (10) → C8 (30) for guanine, where a larger number denotes more electron density and therefore the dipole runs positive to negative from the atom with the smaller value to that with the larger. This, in addition to our own findings, is to our knowledge the only other example of a DNA recognition protein using electrostatic interactions to distinguish between natural and modified bases. However, for hOGG1 the key features of recognition are mediated by hydrogen bonding interactions; for Atl1 there is no recognition of the absent N1 proton whereas for hOGG1 the protein recognises protonation of N7 of 8-oxoG compared to the natural base.

In summary, it has been proposed that there is a two-fold basis for Arg69-mediated recognition of O^6 -alkylguanine and related purine residues, and their discrimination from guanine in the At1 active site: firstly, by interacting with the base, Arg69 holds O^6 -MeG in the binding cavity whilst repelling G, and secondly this cation- π interaction increases the strength of the H-bonds between the active site residues and O^6 -MeG but not G. In this way, Arg69 acts as a molecular probe of electrostatic potential and is crucial for At1 function.

5.5 Structural Studies of At1

In order to provide additional evidence that At1 recognises DAP and 2-AP, and to support our hypothesis that Arg69 acts as a molecular probe of electrostatic potential and is a key part of recognition of O^6 -alkylguanines and related purine bases, At1 was co-crystallised with short ODNs containing DAP and 2-AP residues. This work was undertaken in collaboration with Dr. Julie Tubbs and Prof. John Tainer at the Scripps Institute. It was confirmed that ODNs containing DAP and 2-AP are bound by At1 in the same manner as those containing O^6 -alkylguanines: the modified base is flipped out of the base stack and into the protein active site, with Arg39 protruding into the helix to stabilise the flipped conformation (figure 5.24). The solving of these structures demonstrates that At1 recognises ODNs containing DAP and 2-AP with reasonably high affinity and is consistent with the data generated by our fluorescence-based binding assays. In order to further examine the details of recognition, the active sites are shown in figure 5.25. As expected, the interactions are the same for all the structures shown: At1 binding ODNs

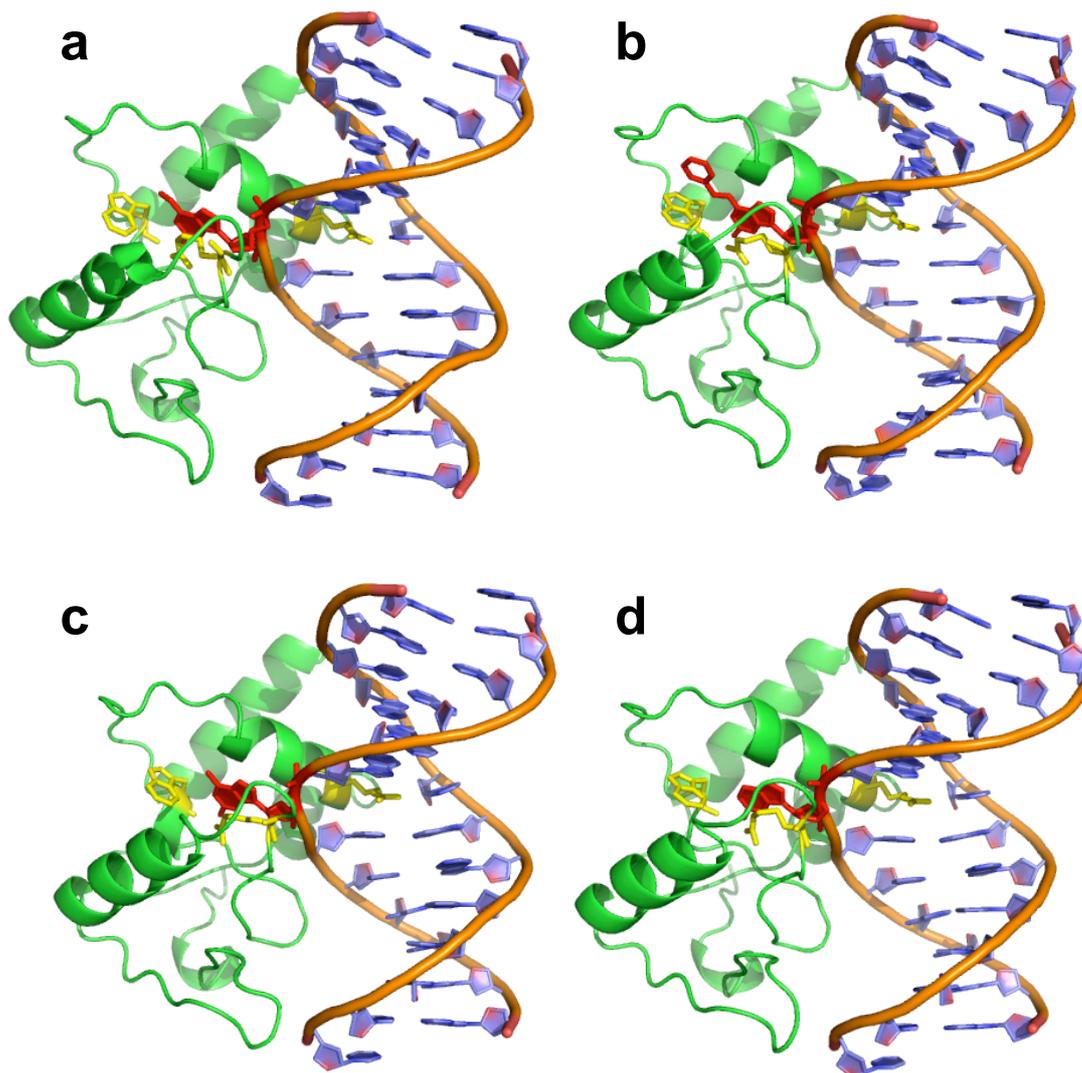


Figure 5.24: Full structures of At11 binding to 13-mer ODNs containing ((a) O^6 -MeG (b) O^6 -BnG (c) DAP (d) 2-AP, showing the Arg39 finger and active site residues Trp56 and Arg69

containing O^6 -MeG, O^6 -BnG, DAP and 2-AP. The importance of the N2 and N3 atoms can clearly be seen, as well as the Arg69 residue which appears to be hovering over the bases in all cases, consistent with its role as a molecular probe that 'reads' the electrostatic potential of the base. It can be seen in the active site structures that the Arg69 residue is tilted at a slightly different angle when DAP and 2-AP are in the active site pocket than when O^6 -MeG and O^6 -

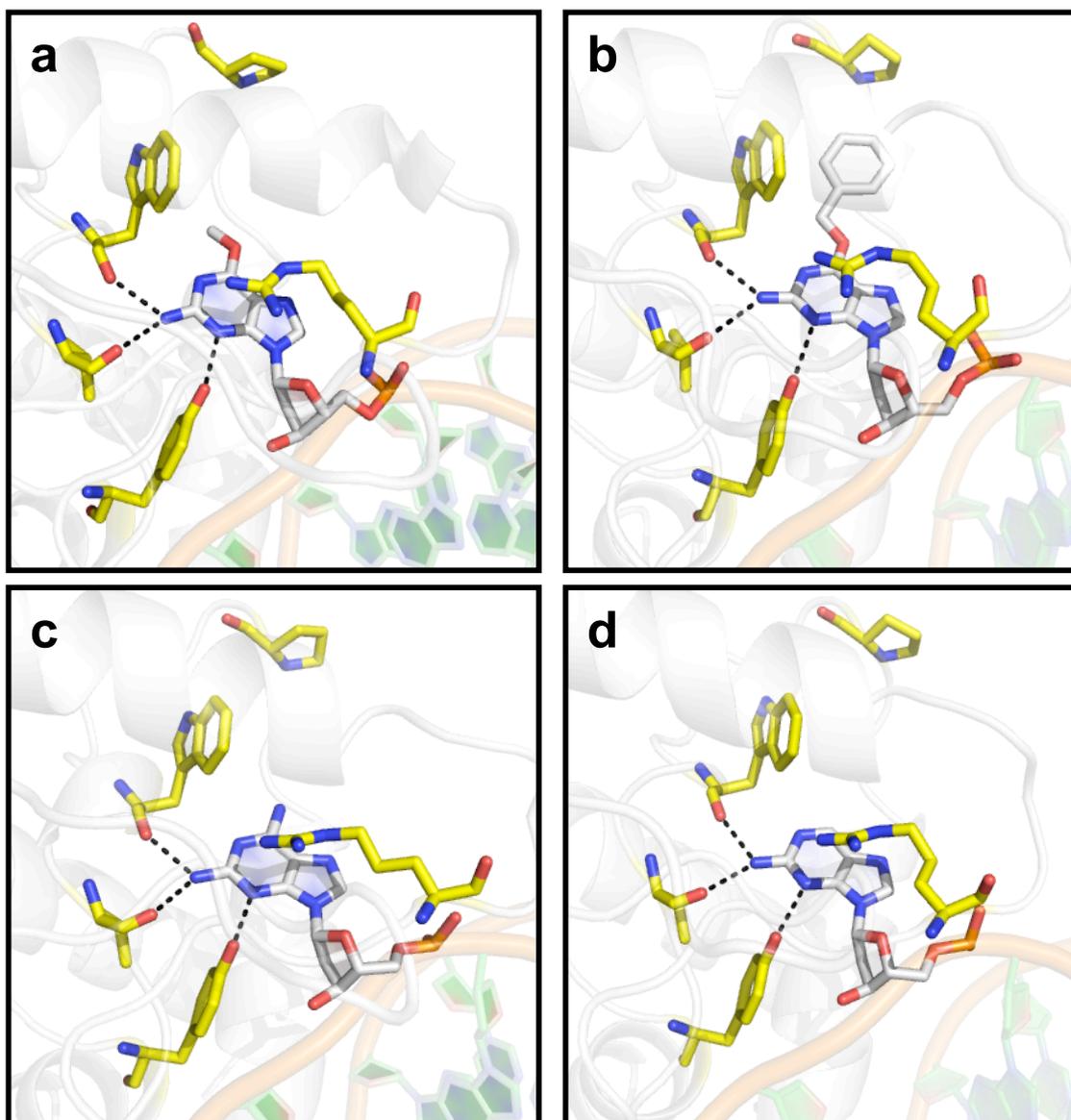


Figure 5.25: At1 active site interactions with selected bases ((a) O^6 -MeG (b) O^6 -BnG (c) DAP (d) 2-AP

BnG are bound (figure 5.25). In addition, Arg69 is in a slightly different position when associated with O^6 -BnG rather than O^6 -MeG, whereby it appears to be marginally closer to N1. Table 5.5 gives the distances between the Arg69 N^+ cation and some selected ring atoms on the base for the different residues, and figure 5.26 gives close up views of the bases in the active site to make the variations in position of the Arg69 residue clearer.

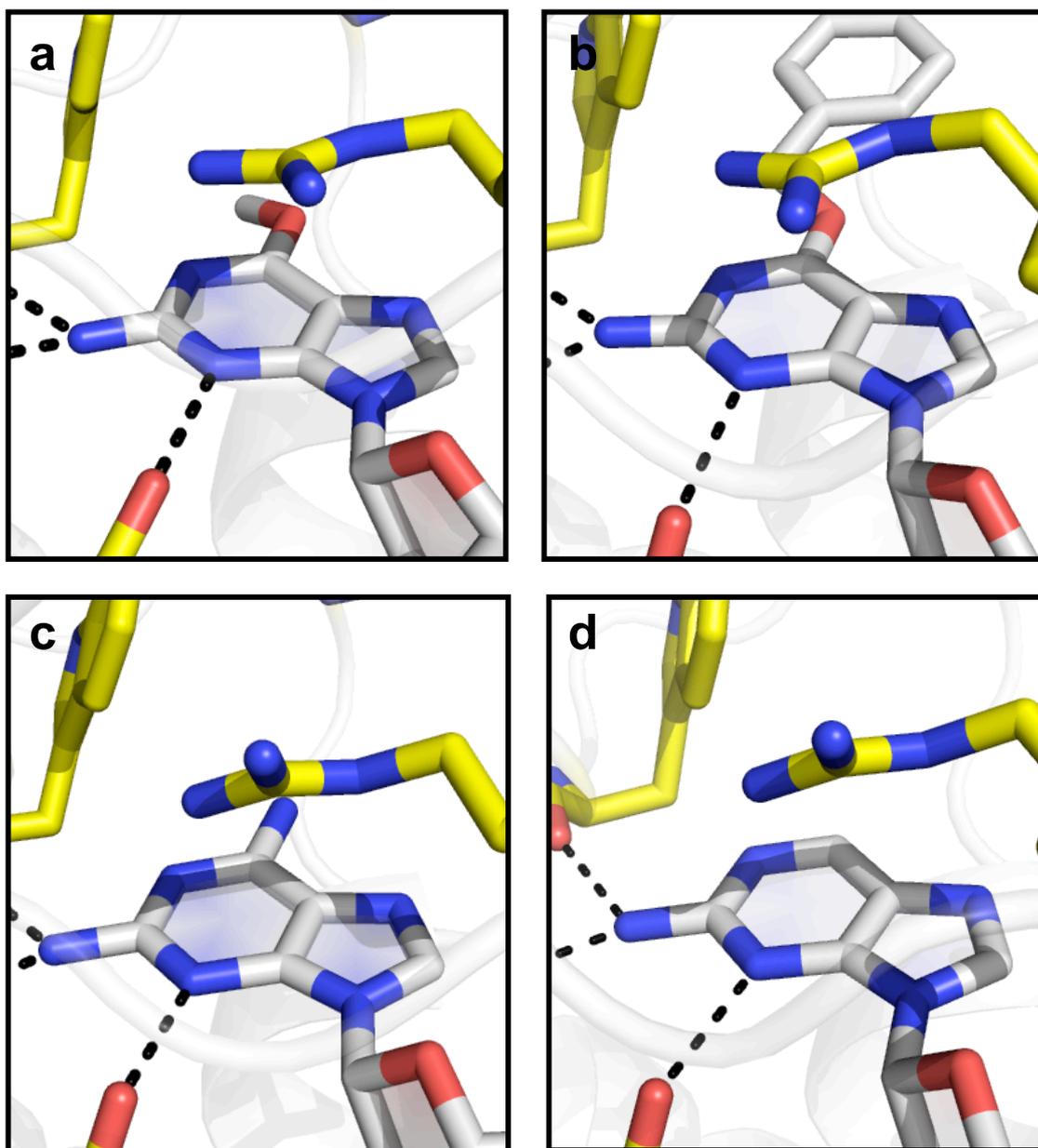


Figure 5.26: Close-up views of the interaction between Arg69 and selected bases ((a) O⁶-MeG (b) O⁶-BnG (c) DAP (d) 2-AP

	N1	C2	N2	N3
O⁶-MeG (a)	3.8 Å	3.5 Å	3.9 Å	3.4 Å
O⁶-BnG (b)	3.7 Å	3.5 Å	4.0 Å	3.4 Å
DAP (c)	3.7 Å	3.4 Å	4.1 Å	3.2 Å
2-AP (d)	4.1 Å	3.8 Å	4.4 Å	3.4 Å

Table 5.5: Inter-atomic distances between the Arg69 N⁺ cation atom and selected atoms of the heterocyclic ring for modified guanine and purine bases

5.6 Mechanism of Base Recognition by ATL Proteins

It has been demonstrated that the substrate specificity of AtI1 and TTHA1564 is relatively broad, with a wide range of O⁶-alkylguanine residues in DNA being recognised with high affinity compared to the control sequence which contains guanine in place of the modified base. Whilst AtI1 clearly displays a preference for ODNs containing large, bulky O⁶-alkylguanine lesions, this is not the case for TTHA1564 which binds most O⁶-alkylguanines with similar affinity regardless of size. Of all the O⁶-alkylguanine lesions, the ODN containing O⁶-carboxymethylguanine (O⁶-CMG) is recognised with the least affinity by both ATL proteins. The ability to discriminate between O⁶-methylguanine and guanine is approximately 300-fold for AtI1 and 80-fold for TTHA1564 and between O⁶-benzylguanine and guanine around 2500-fold for AtI1 and 60-fold for TTHA1564. Interestingly, the ODN containing a tricyclic guanine analogue, in which the O⁶-alkyl group is locked in the *anti*-conformation, was recognised with fairly poor affinity (i.e. similar to control) by

both ATL proteins. This may be evidence of a requirement by ATL proteins to orientate the alkyl group into the *syn*-conformation in order to bind the damaged base, analogous to the lack of repair of O⁶-alkyl groups by MGMT when they are in the *anti*-conformation.(95)

Surprisingly, an ODN containing 2,6-diaminopurine (DAP) was recognised with high affinity by both ATL proteins compared to control ODN but with less affinity than any of the O⁶-alkylguanine residues. In addition, it was also shown that AtI1 recognises ODN containing 2-aminopurine (2-AP) with similar affinity to DAP. This observation was unexpected and demanded an explanation: hence, structural studies of AtI1 bound to ODNs containing DAP and 2-AP were conducted. Thus, it started to become clear that the Arg69 residue of AtI1 was intimately involved in recognition of these bases.

The importance of the active site interactions involved in recognition was investigated by performing the same fluorescence-based binding assays using ODNs containing a number of modified bases. It was previously shown that AtI1 does not specifically recognise adenine or guanine residues (as both these bases are present in the control sequence). Using an ODN containing the modified base O⁶-methylhypoxanthine (O⁶-MHx) which differs from O⁶-MeG only in that it does not possess an N2-amino group, it was shown that AtI1 binds poorly to ODN containing this modified base. Thus, the H-bond interactions in the active site from the main chain carbonyl groups of Trp56 and Val59 to the N2 atom are absolutely essential for recognition. This means that cytosine, thymine and adenine are all unsuitable for recognition in the AtI1 active site. However, both guanine and O⁶-alkylguanine residues have an

N2-amino group and therefore there must be some other basis of discrimination between these residues by AtI1.

It may be that this discrimination is more important for AtI1 than for AGT proteins such as MGMT, which has the same H-bonding interactions in its active site as AtI1 but only a 3- to 10-fold ability to discriminate between O^6 -MeG and G. If MGMT transiently binds to a guanine residue in DNA rather than an O^6 -alkylguanine lesion, the repair reaction cannot be performed and so the undamaged base simply flips back out of the active site. However, if AtI1 were unable to discriminate so effectively between guanine and O^6 -alkylguanine residues (300- to 2500-fold), either these lesions in DNA would not be specifically recognised by AtI1 and go unrepaired, or complexes would be formed between AtI1 and undamaged guanine bases. If these 'false' AtI1 recognition complexes were to instigate repair of DNA by nucleotide-excision repair proteins then this would presumably have deleterious effects, akin to those resulting from the unwanted removal of non-damaged bases by DNA glycosylases during base-excision repair.(122,123) Furthermore, whilst MGMT can perform the repair reaction alone, it is likely that AtI1 needs to form stable, bulky complexes that persist long enough to be recognised by factors from the NER machinery that can continue to process the lesion.

Hence, AtI1 has an Arg69 residue that is not present in MGMT and which can be seen in crystal structures to be interacting with the π -face of O^6 -alkylguanine and certain purine bases when they are in the active site. Arg69 carries a positive charge on its terminal nitrogen atom, and calculation of the molecular electrostatic potentials of the bases when interacting with a cation clearly show that there are more extensive attractive interactions between

Arg69 and O^6 -methylguanine than between Arg69 and guanine. In fact, in addition to the more modest attractive forces between Arg69 and guanine there is also a slight repulsion between Arg69 and the C2 atom of guanine. Therefore, the Arg69 residue is the main basis for discrimination between guanine and O^6 -alkylguanine residues by AtI1, based on differences in the molecular electrostatic potential of the atoms in the damaged and undamaged bases. This is consistent with the results of the binding assays, where a difference in the ability of AtI1 to recognise ODNs containing O^6 -alkylguanine lesions compared to those containing guanine of between 300-fold (for O^6 -MeG) and 2500-fold (for O^6 -BnG) was observed. In addition, site-directed mutagenesis of Arg69 to an alanine residue (AtI1 R69A) disrupted the ability of the protein to recognise O^6 -alkylguanine lesions (around 47-fold for O^6 -BnG compared to AtI1 WT), and almost completely destroys the ability of the protein to discriminate between O^6 -MeG and G (1.5-fold difference) and O^6 -BnG and G (10-fold difference). This lack of discrimination is due in part to the R69A mutant not recognising ODNs containing O^6 -alkylguanines with as much affinity as wild-type, but also due to the control sequence being bound with greater affinity by the mutant protein (for ODN containing G, $K_D = 741\text{nM}$ (AtI1), $K_D = 115\text{nM}$ (AtI1 R69A)). This difference in affinity for control substrate between WT and R69A may be caused by the absence of the repulsive interaction between Arg69 and C2 of guanine in the R69A active site.

Intriguingly, mutation of Arg69 to phenylalanine (AtI1 R69F) does not seem to impair the ability of the protein to recognise O^6 -alkylguanine lesions to the same degree as for the R69A mutant, although it is still disrupted compared to wild-type (around 5.5-fold for O^6 -BnG). Similarly, the ability of

Atl1 R69F to discriminate between O^6 -alkylguanine and guanine is not as disrupted as for R69A, with the difference between O^6 -BnG and G being 450-fold and between O^6 -MeG and G around 65-fold. It is difficult to explain why this is the case: possibly the Phe69 residue is able to take part in a π - π interaction with the face of the base that is more extensive for O^6 -alkylguanine than for guanine residues, due to differences in the electrostatic potentials of the bases. In fact, in the active site of hOGG1 there is an interaction of this kind (between Phe319 and 8-oxoG), although it does not appear to be one of the crucial interactions for discrimination between guanine and 8-oxoguanine.(24) Interestingly, the residue corresponding to Arg69 in TTHA1564 is Phe130 (figure 5.26), though due to a lack of structural data it is hard to draw any conclusions about which interactions may be important for recognition in the active site of the *T.thermophilus* ATL.

It has been suggested by Latypov *et al.* (in a recently submitted manuscript) that the difference in the affinity of Atl1 for large, bulky O^6 -alkylguanine lesions (such as O^6 -benzylguanine) and smaller lesions (such as O^6 -methylguanine) is extremely important and determines the fate of the lesion (i.e. the subsequent repair pathway that processes the damage). Using SPR-based methods (Biocore[®]) to measure dissociation constants, it was found that there was around an 11-fold difference in the affinity of Atl1 for ODNs containing O^6 -BnG ($K_D = 0.08\text{nM}$) and O^6 -MeG ($K_D = 0.91\text{nM}$). This is consistent with our data in which we observed a difference of around 8-fold (O^6 -BnG ($K_D = 0.3\text{nM}$) and O^6 -MeG ($K_D = 2.4\text{nM}$)). It was also shown for all of the ODNs containing O^6 -alkylguanine lesions that were evaluated, that the association rates (k_{ass}) were largely similar, and that it was the differences in

- 1) **Atl1**
- 2) **TTHA1564**

```

1) MRMDEFYTKVYDAVCEIPYGVSTYGEIARYVGMPSYARQVQAMKHLHPETHVPWHRV 59
2) --LSPARLRLYERVRLVPYGRVSYGALGRELGLSP--RAVGAALRACPFLLVPAHRV 120

1) INSRGTISKRDISAGEQRQKDRLEEEGVEIYQTSLGEYKLNLPEYMWKP 108
2) IHADGRLGGEQQQEGLKLWLLRFEGA----- 146

```

Figure 5.27: Sequence alignment of ATL protein homologues Atl1 from *S.pombe* and TTHA1564 from *T.thermophilus* generated using Sequoia.(124) The highlighted amino acids correspond to key residues.

dissociation rates (k_{diss}) that affected the final value of K_D . It is likely that the more extensive hydrophobic interactions between Trp56 and larger O^6 -alkyl groups are responsible for retaining these lesions in the Atl1 active site and increasing the binding affinity for residues such as O^6 -BnG.

The K_D values measured using fluorescence-based assays are on average approximately three times larger than those derived from SPR-based experiments. This may be due to the experimental conditions: for SPR measurements the ODNs are immobilised on a sensor chip, whereas for the fluorescence-based assays the binding takes place in free solution under true-equilibrium conditions. More importantly, the relative differences in affinity between ODNs containing O^6 -alkylguanine residues are broadly similar and so the results by these different methods are complementary and consistent.

In a series of elegant experiments Latypov *et al.* demonstrated *in vivo* that the type of NER repair initiated by Atl1 is dependent on the size of lesion and the relative binding affinity for that lesion. Using various *S.pombe*

deletants, it was shown that repair of smaller lesions such as O^6 -methylguanine was performed by global genome repair (GGR) and that of larger lesions such as O^6 -benzylguanine by transcription-coupled repair (TCR). Therefore, it has been proposed that O^6 -MeG lesions are recognised by AtI1 but that the weaker affinity of binding facilitates dissociation of the protein and allows GGR factors to continue repair, whilst O^6 -BnG lesions are recognised and bound with stronger affinity and therefore cause stalling of RNA polymerase which subsequently initiates repair by the transcription-coupled repair (TCR) pathway. It has thus been suggested that this difference in affinity for various lesions leads to different processing pathways for the DNA damage.

5.7 Tables of Binding Data

Modified Base	AtI1		MBP-TTHA1564	
	K_D (nM)		K_D (nM)	
O ⁶ -benzyl (BnG)	0.3	± 0.01	1.63	± 0.15
O ⁶ -propyl (PrG)	0.389	± 0.02	1.8	± 0.14
O ⁶ -methyladamantyl (MAG)	0.554	± 0.03	1.86	± 0.18
O ⁶ -hydroxyethyl (HOEtG)	0.727	± 0.04	1.94	± 0.09
O ⁶ -aminoethyl (AEG)	1.01	± 0.02	3.61	± 0.13
O ⁶ -ethyl (EtG)	1.09	± 0.12	1.72	± 0.17
O ⁶ -pyridyloxobutyl (PobG)	1.1	± 0.02	1.82	± 0.1
N ⁶ -hydroxypropyl (HOPr-DAP)	1.13	± 0.14	3	± 0.22
O ⁶ -carboxymethyl (CMG)	2.18	± 0.06	3.55	± 0.03
O ⁶ -methyl (MeG)	2.4	± 0.12	1.23	± 0.13
2,6-diaminopurine (DAP)	4.87	± 0.25	11.1	± 0.45
2-aminopurine (2-AP)	9.29	± 0.09	-	-
S ⁶ -thioguanine (ThG)	110	± 4.5	-	-
O ⁶ -methylhypoxanthine (MHx)	374	± 13	-	-
Tricyclic analogue (TCA (O))	432	± 41	64.7	± 2.6
Guanine (G)	741	± 91	102.6	± 2

Table 5.6: Dissociation constants of single-stranded ODNs containing modified bases with AtI1 and TTHA1564 proteins

Modified Base	AtI1		MBP-TTHA1564	
		K_D (nM)		K_D (nM)
O⁶-benzylguanine (BnG)	ssDNA	0.3 ± 0.01	1.63 ± 0.15	
	dsDNA (C)	0.502 ± 0.04	1.32 ± 0.06	
	dsDNA (T)	0.421 ± 0.04	1.33 ± 0.05	
O⁶-methylguanine (MeG)	ssDNA	2.4 ± 0.12	1.23 ± 0.13	
	dsDNA (C)	2.83 ± 0.08	1.01 ± 0.08	
	dsDNA (T)	2.58 ± 0.25	1.04 ± 0.1	
2,6-diaminopurine (DAP)	ssDNA	4.87 ± 0.25	11.1 ± 0.45	
	dsDNA (C)	5.13 ± 0.43	10.1 ± 0.3	
	dsDNA (T)	44.6 ± 0.86	35.5 ± 1.3	
Guanine (G)	ssDNA	741 ± 91	102.6 ± 2	
	dsDNA (C)	762 ± 14	101.1 ± 4.3	
	dsDNA (T)	742 ± 42	94.1 ± 2.7	

Table 5.9: Dissociation constants of double-stranded ODNs containing selected modified bases with AtI1 and TTHA1564 proteins

Modified Base	AtI1	AtI1 R69F	AtI1 R69A
	K_D (nM)	K_D (nM)	K_D (nM)
O⁶-benzyl (BnG)	0.3 ± 0.01	1.64 ± 0.12	14 ± 2.3
O⁶-methyladamantyl (MAG)	0.55 ± 0.03	2.3 ± 0.35	15.6 ± 2.8
O⁶-methyl (MeG)	2.4 ± 0.12	35.4 ± 6.0	96.3 ± 1.0
2,6-diaminopurine (DAP)	4.87 ± 0.25	78.8 ± 15.2	241 ± 4.9
2-aminopurine (2-AP)	9.29 ± 0.10	196 ± 3.3	561 ± 47
Guanine (G)	741 ± 91	755 ± 12.3	135 ± 16

Table 5.10: Dissociation constants of single-stranded ODNs containing modified bases with AtI1 wild-type, and mutant proteins AtI1 R69F and AtI1 R69A

6.0 Isolation and Identification of AtI1 and Interacting Proteins

ATL proteins are thought to be involved in the nucleotide-excision repair (NER) pathway, with their proposed role being damage sensors that signal O^6 -alkylguanine lesions for repair. Whilst there is some genetic and biochemical evidence to suggest that this is the case, their exact function is still unknown. In order to produce evidence that AtI1 (or the AtI1-DNA complex) interacts directly with proteins in the NER machinery, a series of affinity purification (pull-down) assays were carried out that it was hoped would allow the isolation of AtI1 and its binding partners. Mass spectrometry-based proteomic analysis was then used to identify any associated proteins.

6.1 Introduction to Affinity Purification

The study of protein-protein interactions is extremely valuable in molecular biology. In cells, proteins participate in interactions with one another, forming complexes that function in various biological pathways. Identification of the binding partners of a specific protein facilitates the determination of which cellular pathways that protein is involved in, and can assist in the elucidation of its function. The pull-down assay is a technique that is often used to identify the binding partners of a protein of interest. The term pull-down assay can be used to describe two related methods, co-immunoprecipitation and affinity purification, which differ in the strategy for removing the target protein complexes from the protein mixture. Immunoprecipitation exploits the highly specific interaction between a target

protein (antigen) and its antibody. Incubation of an immobilised antibody (the 'bait') with the protein mixture results in removal of the target protein, potentially with other interacting proteins still attached. Affinity purification is very similar but uses another molecule rather than an antibody as the bait.

A pull-down approach that could be utilised in the context of this project involves the protein of interest being expressed as a fusion protein with an attached affinity label (such as GST, MBP, His etc.), then immobilised onto a solid-support using the highly specific interaction between the affinity tag and a ligand on the support. A mixture of proteins is then passed over the immobilised protein of interest (which acts as the bait) in the hope that that any protein binding partners will specifically interact with the bait, and after washing to remove any non-interacting proteins, subsequently co-elute (figure

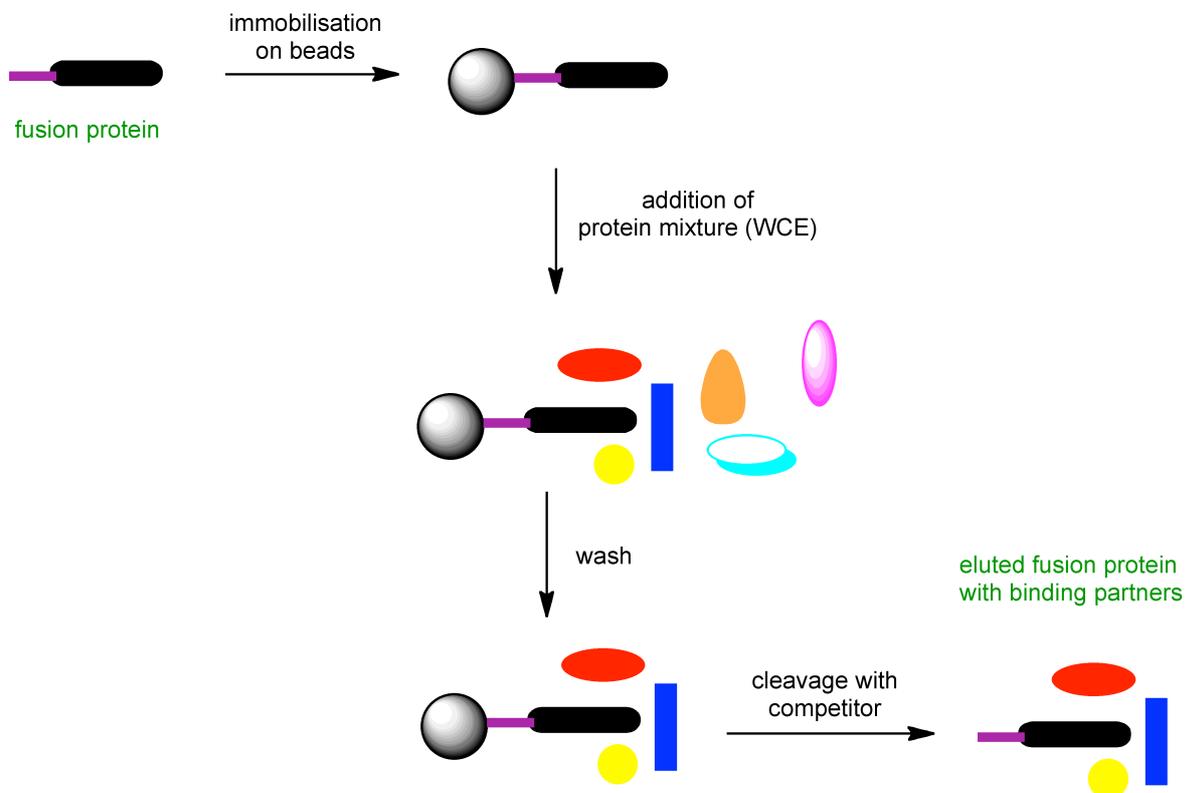


Figure 6.1: Principles of a protein-immobilised pull-down assay

6.1). The intact protein-protein complexes can be cleaved from the solid-support using a competitor. These protein complexes are then separated and analysed by SDS-PAGE (SDS-polyacrylamide gel electrophoresis) and the identity of the proteins present in the sample established using a combination of mass spectrometry and *in silico* analysis.(125)

This method could be used to identify which proteins in *S.pombe* interact with At11. At11 has previously been overexpressed in *E.coli* and purified as a fusion protein tagged with maltose-binding protein (MBP-At11).(57) This fusion protein could subsequently be immobilised on amylose resin and the pull-down assays performed using whole cell extracts (WCEs) of *S.pombe* to attempt to identify the binding partners of At11 *in vitro*. This is a similar approach to that used by Morita *et al.* to identify and demonstrate the interaction between TTHA1564 and UvrA, the NER recognition factor.(63)

Alternatively, the experiment may be more successful using the At11-DNA complex as bait, as NER factors may recognise this complex rather than At11 *per se*. One option would be to use a biotinylated duplex ODN immobilised on streptavidin-coated beads. The ODN would be expected to be recognised by At11 and hence remove it from the WCE, potentially bringing with it any interacting proteins. Another possibility would be to pre-form the At11-DNA complex by treating the ODN-coated beads with purified recombinant At11, and then use this as the bait in a pulldown with *S.pombe* WCE. If it is found that the At11-DNA complex is not stable enough to be used for this purpose, another option would be to use a covalently cross-linked At11-DNA complex (section 7.1).

6.2 Introduction to Protein Identification by Mass Spectrometry

6.2.1 Mass Spectrometry-based Proteomics

Mass spectrometry (MS) is an analytical technique which involves generating ions from a sample and accelerating them in an electric field, to produce a spectrum by measurement of the mass-to-charge (m/z) ratio of the molecules. It is a valuable tool for the characterisation and identification of proteins, either by introduction of an intact protein and subsequent analysis ('top-down' proteomics) or by the introduction of peptides from a pre-digested protein sample that can then be identified ('bottom-up' proteomics). A mass spectrometer by definition consists of three parts; an ion source which creates gaseous phase ionised analytes, a mass analyser which separates the ions based on the m/z ratio, and a detector which interprets the data into mass spectra, including abundance measurements (figure 6.2).(125)



Figure 6.2: Overview of a mass spectrometer

The most commonly used techniques to create ions are matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). MALDI involves using laser pulses to sublime and ionise a sample from a dry crystalline matrix, whereas ESI involves ionising the sample from solution and so can be integrated with an on-line sample fractionation step such as liquid chromatography (LC). The mass analyser is the most integral part of the

experimental set-up as it is responsible for generating mass spectra that are rich in information. There are four types generally used in proteomics: ion trap, time-of-flight (TOF), quadrupole and Fourier transform ion cyclotron (FT-MS), and they vary in their performance in terms of sensitivity, mass accuracy and resolution.(125) MALDI is usually coupled with a time-of-flight mass analyser (MALDI-TOF) which allows the masses of intact peptides to be measured, and matching of this data with theoretical masses in a database allows proteins to be identified in a procedure known as peptide mass fingerprinting (PMF).(126) ESI is more frequently coupled to ion-trap, triple quadrupole or quadrupole-TOF mass analysers which allow tandem mass spectrometry (MS/MS) experiments to take place.(127) Tandem mass spectrometry (MS/MS) has the advantage of revealing more information as it incorporates a fragmentation step before the second analyser measures m/z values of the resulting fragments (figure 6.3).

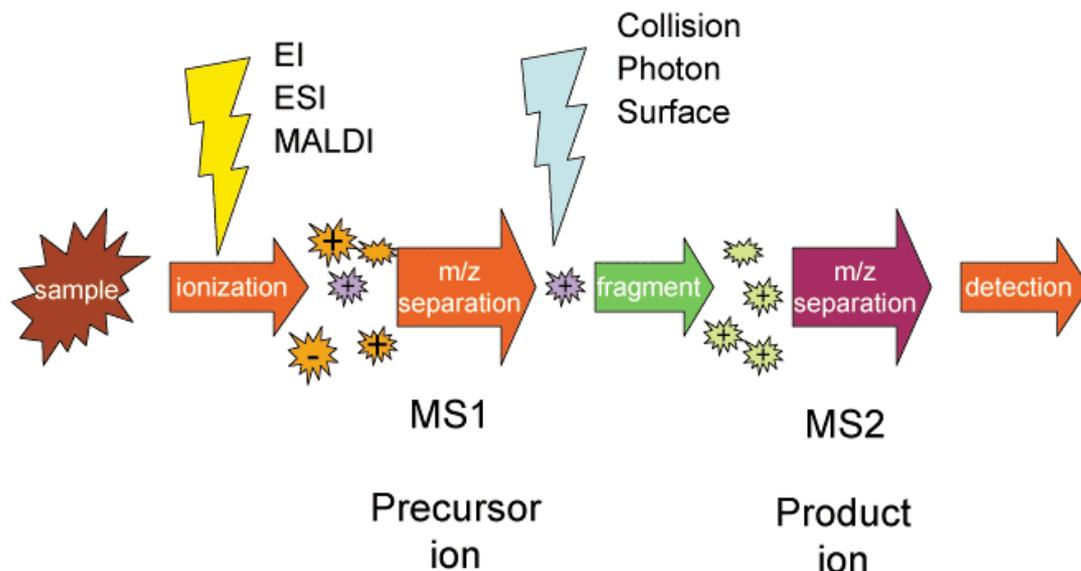


Figure 6.3: Principles of a tandem mass spectrometry (MS/MS) experiment

More specifically, MS/MS experiments are comprised of three stages: the initial MS which separates the peptides on the basis of m/z ratio, and then a second MS step which involves the selection and further fragmentation of selected peptides in the sample. This generates a spectrum in the third stage, the analysis of which allows the masses and sequences of peptides and peptide fragments to be elucidated and hence the proteins in the original sample identified.(128) For the analysis of complex mixtures comprised of peptides from a large number of proteins (e.g. from whole cell extracts) integrated liquid chromatography ESI-MS (or LC-MS) systems coupled with a quadrupole ion trap (Q-TRAP) is an appropriate choice.(129) This is because the LC can fractionate (i.e. simplify) the mixture before ionisation and then allow subsequent MS/MS analysis to take place. The quadrupole ion trap is shown in figure 6.4. The peptide ions are first separated on the basis of m/z ratio, and then precursor ions of a particular m/z are selected and fragmented

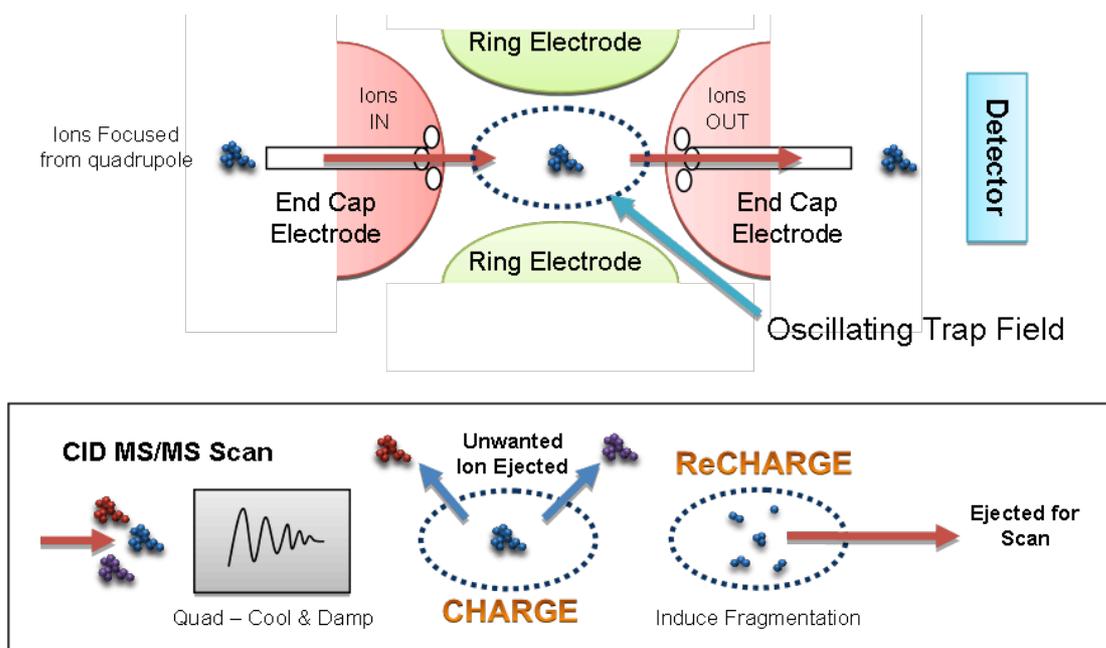


Figure 6.4: Principles of a quadrupole ion trap mass analyser (from Saw Yen Ow, ChELSI)

in the ion trap. The fragments (product ions) are then separated before passing to the detector. This process leads to the creation of collision-induced dissociation (CID) spectra which can be subsequently analysed using peptide databases.

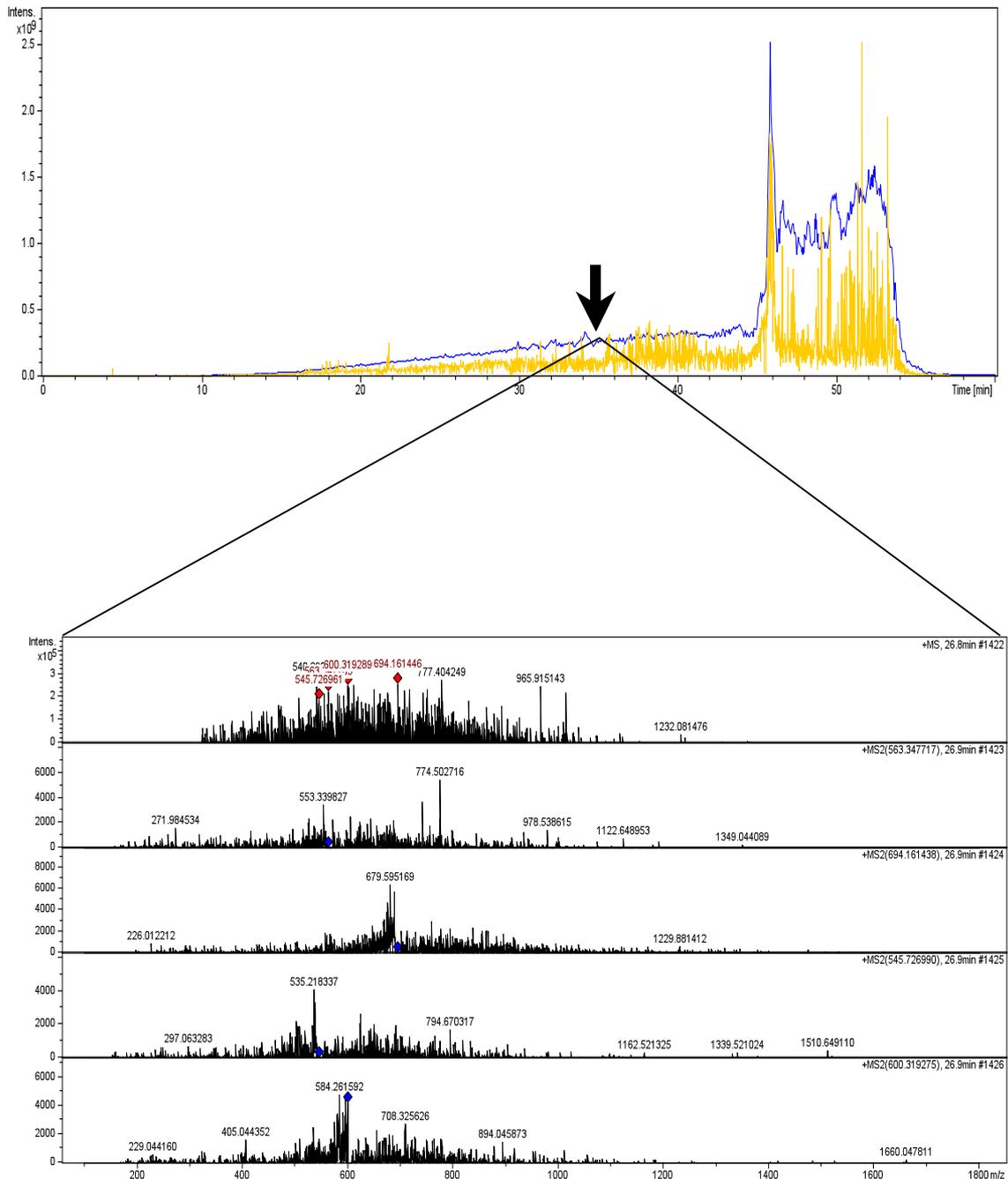


Figure 6.5: MS (blue) and MS/MS spectra (yellow) with an expansion of the collision-induced fragmentation pattern

A typical spectra is shown in figure 6.5. The top window contains the crude spectra, with the MS trace shown in blue and the MS/MS shown in yellow. Selection of a time point in the experiment allows the fragmentation patterns at that moment to be viewed, as shown in the bottom window. It is the information contained within these fragmentation patterns that are compared with theoretical patterns when *in silico* analysis and peptide matching takes place.

For the analysis of the affinity-based isolation of AtI1 and interacting proteins, a 'bottom-up' proteomics approach was used. After separation of the proteins in the sample by 1-D gel electrophoresis and subsequent digestion with trypsin, tandem mass spectrometry (MS/MS) experiments were used to analyse the peptides in the sample (section 6.3). The samples were initially fractionated by on-line reverse phase liquid chromatography (LC), after which electrospray ionisation coupled with a quadropole ion trap (ESI-TRAP) was used in order to ionise, separate and fragment the peptides and produce CID spectra for subsequent analysis and protein identification.

6.2.2 Interpretation of MS Data

The data generated by MS/MS experiments must be converted into peptide identifications in order to produce meaningful results. For MS/MS, the peak patterns in the CID spectrum corresponding to fragment ions provide information about the peptide sequences as well as the masses. The types of product ion produced by MS/MS depends on a number of factors including the internal energy, primary structure and charge state and are shown in figure 6.6.(130). Ions that retain the charge on the N-terminus are labelled a,b,c and

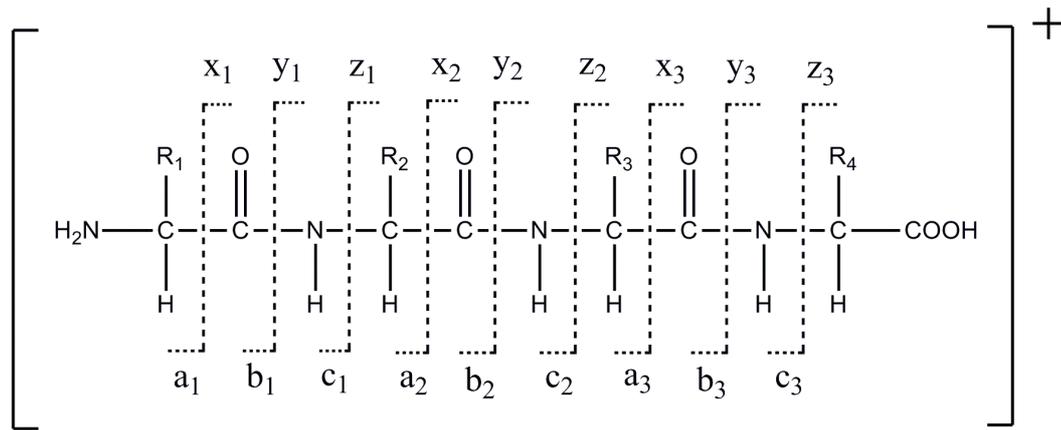


Figure 6.6: Peptide fragment ions generated by MS/MS experiments

those on the C-terminus x,y,z. Certain fragmentation techniques generate characteristic types of ion; in this instance (i.e. using ESI-TRAP), CID generates b and y ions which result from specific cleavage of the peptide bonds. Alternatively, the softer ionisation technique electron transfer dissociation (ETD) produces c and z ions.

Treatment of the spectral data with specific algorithms followed by comparison to databases which contain theoretical fragmentation patterns allow the identity of the peptides to be elucidated.(131,132) Furthermore, if a number of unique peptides corresponding to those from a known protein are found, the presence of that protein in the original sample is inferred with some certainty. Due to the nature of the matching process the peptides must correspond to those from known proteins, in the sense that their sequences are in a protein database (such as Swiss-Prot (<http://www.uniprot.org/uniprot/>) or NCBI Protein Databank (<http://www.ncbi.nlm.nih.gov/protein/>)). Due to the potential for false positives within this method (i.e. mis-identification of proteins based on erroneous analysis) a number of filtering criteria are used which gives the protein a database search identification score. This score,

which is based on the Mowse algorithm (126), is a measure of how reliable and significant the peptide product ion identifications are, based on various parameters such as product ion abundance, favored cleavage sites, product ion type, precursor ion charge state and polarity. It is usually accepted within the proteomics field that a match of two or more unique peptides is sufficient for accurate protein identification. However in the case where only one peptide is found then two good quality MS/MS spectra must be produced and the peptide must not be shared (i.e. not occur in any other known protein) (detailed in the Molecular and Cellular Proteomics journal at mcponline.org).

6.3 Affinity-based Isolation of AtI1 and Interacting Proteins

In an attempt to isolate AtI1 along with any associated proteins from a whole-cell extract of *S.pombe*, an immobilised ODN containing an O⁶-alkylguanine residue was used as bait in a classical pull-down assay (figure 6.7). It was reasoned that any interacting proteins would be more likely to recognise the AtI1-DNA complex, rather than AtI1 alone, because it is the damaged DNA that will be involved in the next stage of NER repair. The ability to affinity purify any proteins that interact with AtI1 assumes that a molecular 'handoff' between AtI1 and NER proteins takes place during the repair process, and that these protein-protein interactions are non-transient and of sufficient affinity to survive the conditions of the assay. The ODNs used were biotinylated at the 5'-terminus, in order that they could be immobilised on streptavidin beads which facilitated separation of the ODNs and any proteins isolated from the WCE after the incubation period used in the assay. Once the beads had been removed from the WCE they were washed to remove any

non-specifically bound proteins. The proteins of interest were then cleaved from the beads, separated by 1-D PAGE, digested with trypsin and analysed by ESI-TRAP mass spectrometry.

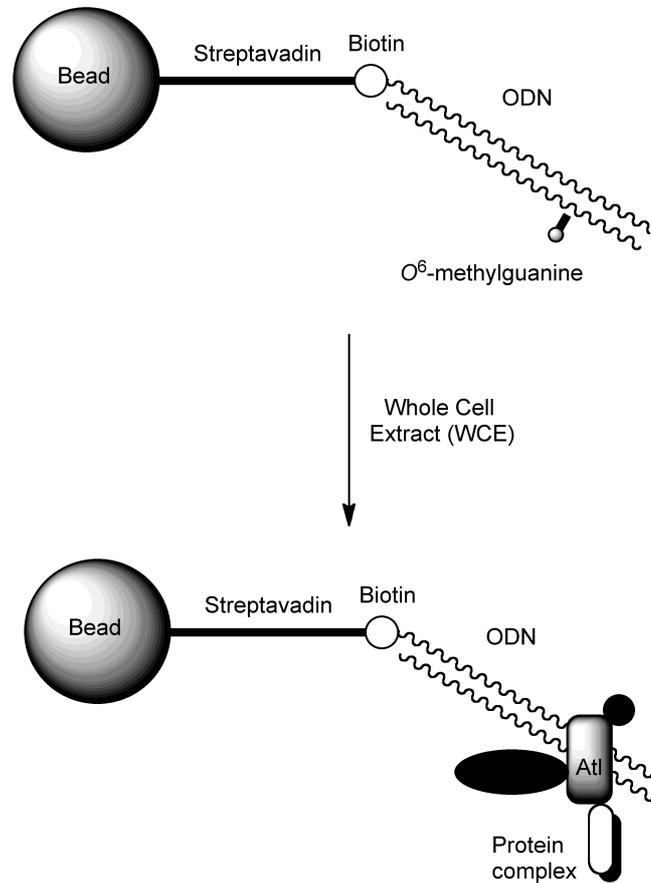


Figure 6.7: Pull-down assay using immobilised ODNs as bait

The ODNs used to isolate At1 in the assay contained an O⁶-alkylguanine residue in order that they would be specifically bound by the protein. It has been previously shown that At1 recognises both single-stranded and double-stranded DNA with similar affinity but it is likely that dsDNA would be preferable for use in these assays. The proteins that associate with the At1-DNA complex may not interact directly with At1 but with the DNA itself, and it is also possible that the repair proteins recognise

the undamaged strand opposite the lesion, which is the case for Rad4, the XPC orthologue in *S.cerevisiae*.(38) In addition, using dsDNA allowed placement of the O^6 -alkylguanine residue in one ODN, and the biotin label in the other, which was more convenient in terms of the preparation and synthesis of the ODNs. It has been shown that during repair of UV-induced DNA damage, the NER machinery of *S.cerevisiae* excises a patch of between 24 and 27 nucleotides in length.(96) Therefore, it was considered likely that an ODN longer than this would be more successful at pulling down any NER protein complexes.

Three strategies were chosen for making the modified ODN substrates used in these pull-down assays: standard DNA synthesis using the phosphoramidite method to produce 102-mer ODNs; ligation of overlapping single-stranded ODNs followed by primer extension using a Klenow fragment (a highly tolerant DNA polymerase) to form 219-mer double-stranded ODNs; and ligation of three short ODNs after annealing to a 102-mer complementary strand to produce 102-mer duplex ODN. The synthesis and modification of these ODN substrates has been described in detail in section 3.3.

The 219-mer ODNs shown in figure 6.8 were initially used in the pulldown assays. OW61 contains O^6 -methylguanine which is known to be bound with high affinity by AtI1 and therefore this ODN was expected to isolate AtI1 from the mixture of proteins in the WCE. OW60 was used as the control sequence as AtI1 should have very little affinity for this ODN. The ODNs OW60 and OW61 were already duplexes by the nature of their synthesis (i.e. using a Klenow fragment, see section 3.3) and therefore no annealing step was required. The binding of the ODNs to the beads was

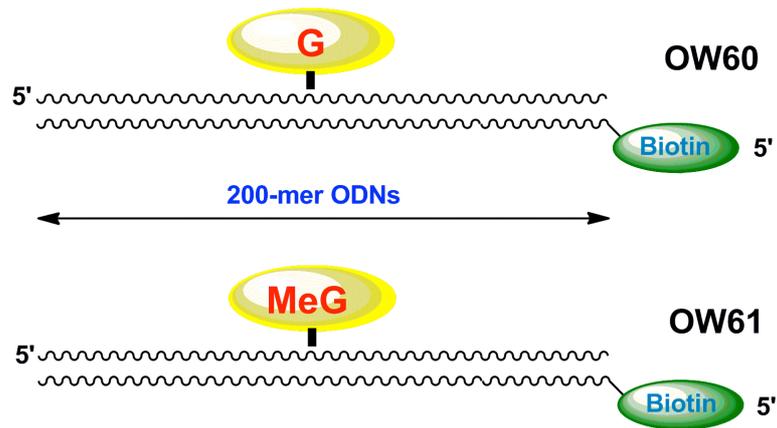


Figure 6.8: Double-stranded 200-mer ODNs used in the pull-down assays

performed in binding buffer (section 9.12), where 500pmols of each ODN was incubated with pre-equilibrated streptavidin beads for 2 hours at room temperature. The binding efficiency was quantified by measurement of relative DNA concentration in the supernatant before and after incubation using UV absorption at 260nm. In both cases 250-300 pmols of ODN was successfully bound to the beads which gave a binding efficiency of between 50-60%. This was considered a reasonable quantity of bound ODN and sufficient substrate to remove a suitable amount of AtI1 from the WCE for subsequent MS analysis. The beads were then washed with binding buffer and pre-equilibrated with extraction buffer (section 9.12).

The next stage was the preparation of *S.pombe* whole-cell extracts (WCE). Two separate cultures of *S.pombe* were cultivated: a wild-type strain (WT) and an AtI1 deletant strain (Δ AtI1). Both were treated identically during growth (in YES media, harvested at $OD_{595} \approx 0.6$), and were subsequently made into small cryogenically frozen pellets by dripping resuspended cell extract into liquid nitrogen. After storage at -80°C , these pellets were ground into a fine powder (using a pre-cooled grinding press) and this was

immediately used to make the WCEs by resuspension in extraction buffer. The proteins were extracted from the powder almost immediately to avoid potential degradation of proteins in the ground cells: if there is a need to keep the *S.pombe* cells for any length of time, they should be stored at -80°C as small cryogenically frozen pellets and only ground into a powder when required for experiments.

Approximately 8g of each powder was resuspended in cold extraction buffer by mixing for one hour at 4°C . The cells were centrifuged and the supernatant (i.e. the WCE) removed. The total protein concentration was determined by Bradford assay ($\sim 30\text{mg/mL}$) and hence the extracts were ready to be used in the pull-down assays. The pre-equilibrated beads were then added to the extracts in the following combinations:

- Wild-type extract (WT WCE) with OW61 (MeG) (**Pd1**)
- Wild-type extract (WT WCE) with OW60 (G) (**Pd2**)
- AtI1 deletant extract (ΔAtI1 WCE) with OW61 (MeG) (**Pd3**)
- AtI1 deletant extract (ΔAtI1 WCE) with OW60 (G) (**Pd4**)

These reactions (**Pd1-4**) were then incubated for one hour at room temperature, after which the beads were separated from the WCEs using a magnetic rack. The selective removal of AtI1 from the WT WCE by OW61 but not by OW60 was demonstrated and quantified by ELISA (approximately 80% of total AtI1 in the wild-type WCE was bound to OW61). The collected beads were washed three times with 100uL of extraction buffer, and then boiled in

1X Laemmli buffer to cleave all the proteins from the DNA. Finally, the samples were separated and analysed by 1-D SDS PAGE (figure 6.9).

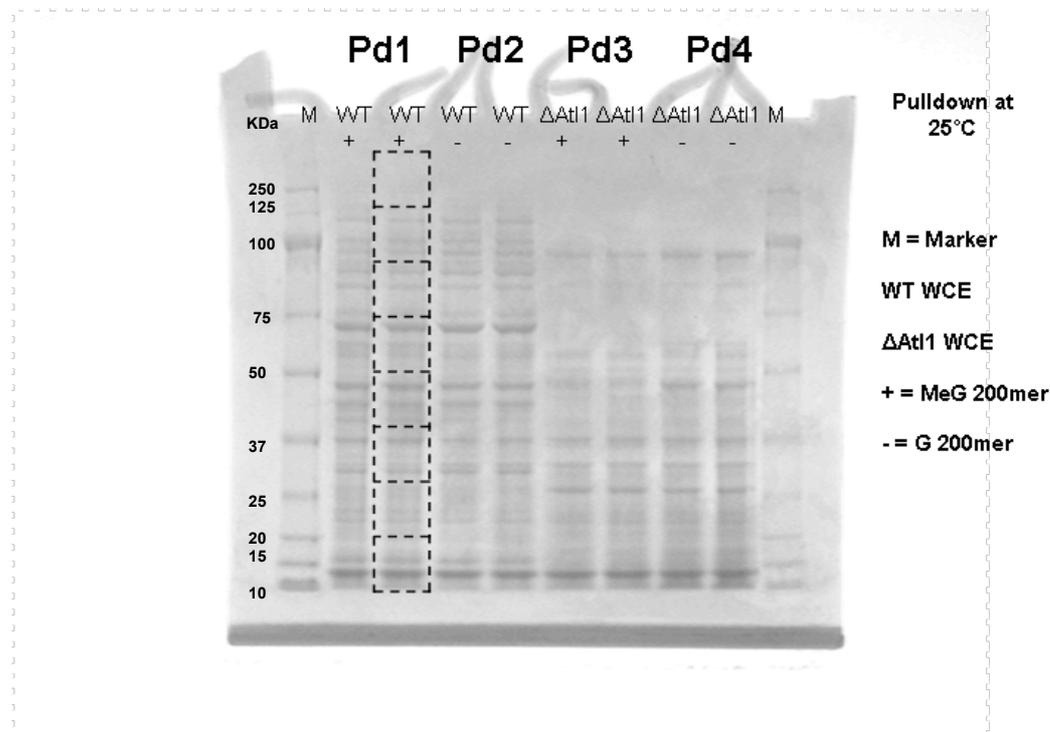


Figure 6.9: 1-D PAGE gel of the protein mixtures isolated during the pulldown assays. The dotted lines indicate the excision of the lanes into gel pieces for MS sample preparation

The PAGE analysis demonstrates the presence of many proteins in the samples, as perhaps would be expected when using WCEs. In order to prepare the samples for analysis by MS,(133) it was necessary to excise the bands from the gel. Two lanes were initially analysed: those containing the proteins from the assays **Pd1** (with OW61 (MeG)) and **Pd2** (with OW60 (G)) that had been incubated with wild-type whole-cell extract (WT WCE). To gain the maximum amount of information from the experiments, it was decided to attempt to identify all of the proteins in the samples and hence each lane was cut into eight separate pieces. After excision, the gel pieces were destained, digested (in-gel) overnight with trypsin at 37°C and the peptides extracted from the gel bands by treatment with a number of solutions (section 9.13.1).

Finally, this solution of peptides was dried under vacuum to give a pellet which was stored at -20°C until the mass spectrometer was available. Before loading, the peptides were resuspended in Switchos buffer (section 9.13.1) and sonicated. The samples were then placed on the mass spectrometer ready for loading to perform the MS/MS experiments.

The exact details of the experimental set-up are described in section 9.13.2. Firstly, the peptide solution enters a reverse-phase liquid chromatography column (C18) which fractionates the peptide mixture. The eluant is then ionised by electrospray (ESI) and enters a quadrupole ion trap (Q-TRAP) mass analyser where the peptide ions are initially separated and selected according to m/z , fragmented by collision with an inert gas (N_2) and

{
MATRIX
}
SCIENCE

MASCOT MS/MS Ions Search

Your name	Oliver Wilkinson	Email	o.j.wilkinson@shef.ac.uk
Search title	OW61 MeG WT WCE Gel Piece 1		
Database(s)	<div style="border: 1px solid gray; padding: 2px;"> Invertebrates_EST Human_EST Fungi_EST Environmental_EST SwissProt </div>	Enzyme	Trypsin
		Allow up to	1 missed cleavages
		Quantitation	None
Taxonomy	Schizosaccharomyces pombe (fission yeast)		
Fixed modifications	--- none selected ---		
	Display all modifications <input type="checkbox"/>		
Variable modifications	Oxidation (M)	<div style="border: 1px solid gray; padding: 2px;"> Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term) </div>	
Peptide tol. \pm	1.2 Da	# ^{13}C	0
Peptide charge	1+, 2+ and 3+	MS/MS tol. \pm	0.6 Da
Data file	Browse...		
Data format	Mascot generic	Precursor	m/z
Instrument	ESI-TRAP	Error tolerant	<input type="checkbox"/>
Decoy	<input type="checkbox"/>	Report top	AUTO hits
Start Search ...		Reset Form	

Figure 6.10: MASCOT MS/MS Ions search

separated again to generate a CID spectrum (section 6.2). This raw data needs to be processed before a database search for peptide matches can take place and therefore the spectrum was converted to a generic MS interrogation file format (.mgf). Essentially, a number of the most intense peaks are selected (5000 in the case of these experiments) and then this data is converted to a format compatible with the software search engine. The search engine uses an algorithm to match experimental peptide fragmentation patterns with theoretical ones calculated from primary sequence databases.

To search for peptide and protein matches from the data, the MASCOT algorithm (Matrix Science) was initially used. An MS/MS ions search was conducted using the parameters shown in figure 6.10. The software was set up to search all *S.pombe* protein sequences in the SwissProt databank for possible matches. The input parameters instructed the software that the proteins were digested with trypsin, the MS/MS experiment was run using ESI-TRAP (i.e. the peptide ion fragments will be of the type b and y), and that it should search for peptide fragment ions that carried +1, +2 and +3 charges. Methionine residues may have become oxidised during the process and this was accounted for in the variable modifications input. In addition, a mass tolerance of ± 0.6 Da was specified for the fragment ions, which was appropriate for the mass accuracy capability of the mass spectrometer.

For the first searches, gel pieces from the bottom of the lanes corresponding to **Pd1** and **Pd2** were selected. At11 would be expected to be in one of the lowest bands in the gel as it has a low molecular weight (12.6 kDa) and the primary success of the assays could be evaluated by whether At11

had been isolated and then identified by these methods. The results for gel piece 1 from **Pd1** are shown in figure 6.11.

Protein hits	Gene	Description	OS	GN	PE	SV
RIM1	SCHPO	Single-stranded DNA-binding protein rim1, mitochondrial	Schizosaccharomyces pombe	GN=rim1	PE=2	SV=2
MMF1	SCHPO	Protein mmf1, mitochondrial	Schizosaccharomyces pombe	GN=mmf1	PE=2	SV=1
G3P1	SCHPO	Glyceraldehyde-3-phosphate dehydrogenase 1	Schizosaccharomyces pombe	GN=tdh1	PE=1	SV=1
RS14	SCHPO	40S ribosomal protein S14	Schizosaccharomyces pombe	GN=rps14a	PE=1	SV=1
CYPH	SCHPO	Peptidyl-prolyl cis-trans isomerase	Schizosaccharomyces pombe	GN=ppi1	PE=2	SV=1
HSP16	SCHPO	Heat shock protein 16	Schizosaccharomyces pombe	GN=hsp16	PE=2	SV=1
RL22	SCHPO	60S ribosomal protein L22	Schizosaccharomyces pombe	GN=rpl22	PE=1	SV=3
ENO11	SCHPO	Enolase 1-1	Schizosaccharomyces pombe	GN=eno101	PE=1	SV=2
RFA1	SCHPO	Replication factor A protein 1	Schizosaccharomyces pombe	GN=ssb1	PE=1	SV=1
NHP6	SCHPO	Non-histone chromosomal protein 6	Schizosaccharomyces pombe	GN=nhp6	PE=2	SV=1
RS16	SCHPO	40S ribosomal protein S16	Schizosaccharomyces pombe	GN=rps16a	PE=2	SV=1
RS22	SCHPO	40S ribosomal protein S22	Schizosaccharomyces pombe	GN=rps22a	PE=2	SV=1
RLA1	SCHPO	60S acidic ribosomal protein P1-alpha 1	Schizosaccharomyces pombe	GN=rpa1	PE=2	SV=1
RL28E	SCHPO	Probable 60S ribosomal protein L28e	Schizosaccharomyces pombe	GN=rpl28e	PE=1	SV=1
RS17A	SCHPO	40S ribosomal protein S17-A	Schizosaccharomyces pombe	GN=rps17a	PE=2	SV=1
TRX1	SCHPO	Thioredoxin-1	Schizosaccharomyces pombe	GN=trx1	PE=2	SV=3
UBI10	SCHPO	Ubiquitin	Schizosaccharomyces pombe	GN=ubi1	PE=1	SV=1
ATL1	SCHPO	Alkyltransferase-like protein 1	Schizosaccharomyces pombe	GN=atl1	PE=1	SV=1
RS19A	SCHPO	40S ribosomal protein S19-A	Schizosaccharomyces pombe	GN=rps19a	PE=2	SV=1
RL30A	SCHPO	60S ribosomal protein L30-1	Schizosaccharomyces pombe	GN=rpl30a	PE=2	SV=2
RL27A	SCHPO	60S ribosomal protein L27-A	Schizosaccharomyces pombe	GN=rpl27a	PE=2	SV=2
RL35	SCHPO	60S ribosomal protein L35	Schizosaccharomyces pombe	GN=rpl35	PE=1	SV=1
RLA5	SCHPO	60S acidic ribosomal protein P1-alpha 5	Schizosaccharomyces pombe	GN=rpa5	PE=1	SV=1
RS18	SCHPO	40S ribosomal protein S18	Schizosaccharomyces pombe	GN=rps18a	PE=1	SV=1
RS23	SCHPO	40S ribosomal protein S23	Schizosaccharomyces pombe	GN=rps23a	PE=2	SV=2
PD2C	SCHPO	Probable pyruvate decarboxylase C1F8.07c	Schizosaccharomyces pombe	GN=SFAC1F8.07c	PE=1	SV=2
RS27	SCHPO	40S ribosomal protein S27	Schizosaccharomyces pombe	GN=rps27	PE=2	SV=1
YNGG	SCHPO	Uncharacterized protein C2D10.16	Schizosaccharomyces pombe	GN=SFBC2D10.16	PE=2	SV=2
RLP1	SCHPO	60S ribosomal protein L31	Schizosaccharomyces pombe	GN=rpl31	PE=2	SV=1
HSP9	SCHPO	Heat shock protein hsp9	Schizosaccharomyces pombe	GN=hsp9	PE=2	SV=1
PGK	SCHPO	Phosphoglycerate kinase	Schizosaccharomyces pombe	GN=pgk1	PE=1	SV=1
FKB9	SCHPO	FK506-binding protein	Schizosaccharomyces pombe	GN=SPBC839.17c	PE=2	SV=1
CHI0	SCHPO	10 kDa heat shock protein, mitochondrial	Schizosaccharomyces pombe	GN=hsp10	PE=1	SV=1
RS10B	SCHPO	40S ribosomal protein S10-B	Schizosaccharomyces pombe	GN=rps10b	PE=2	SV=1
RS12A	SCHPO	40S ribosomal protein S12-A	Schizosaccharomyces pombe	GN=rps12a	PE=2	SV=1
RL43B	SCHPO	60S ribosomal protein L43-B	Schizosaccharomyces pombe	GN=rpl43b	PE=2	SV=1

ATL1	SCHPO	Mass:	12662	Score:	117	Queries matched:	9	emPAI:	1.05
Alkyltransferase-like protein 1 OS=Schizosaccharomyces pombe GN=atl1 PE=1 SV=1									
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
389	467.2018	932.3890	932.3950	-0.0060	0	29	0.1	1	R.MDEFYTK.V 2784 390
523	498.2013	994.3880	994.5083	-0.1203	0	55	0.0002	1	K.VSTYGEIAR.Y 524 525 528
1449	645.8493	1289.6841	1289.6478	0.0363	0	35	0.024	1	K.LNLPEYMWKP.-
2290	753.2477	2256.7212	2257.0750	-0.3538	1	11	3.6	1	K.DRLEEKGVEIYQTSLGEYK.L

Figure 6.11: Protein hits from the MS/MS data for gel piece 1 (Pd1) (top) and peptide matches for Atl1 (bottom)

The results show all the proteins identified in the gel band. Clearly, Atl1 was identified in this lane, and whilst the other searches (for gel piece 2 from **Pd1** and gel pieces 1 and 2 from **Pd2**) resulted in a similar number of protein hits, Atl1 was not present in these samples. Thus it was confirmed that Atl1 was successfully isolated from the WCE in **Pd1** (using the ODN containing O⁶-MeG (OW61)), and not in **Pd2** (using the control ODN containing G (OW60)). Also shown in figure 6.11 (bottom window) are the peptide matches for Atl1. Four unique peptides were identified and their masses and

sequences are shown. The probability-based Mowse score was 117 for this hit which is reasonably high (section 6.2) and this means it can be concluded with some certainty that AtI1 was present in this sample (and not in the others).

The other gel pieces were prepared and analysed in an identical manner. Although AtI1 was successfully and selectively isolated from **Pd1** and not **Pd2**, there were none of the NER recognition factors in either sample (such as Rhp7 and Rhp16 (UV-DBD), Rhp41 and Rhp42 (XPC), Rhp23 (hHR23b) or Rhp26 (CSB), see figure 1.13). Only one other protein of interest was found in the MASCOT search. This was Rad15 (the *S.pombe* homologue of human XPD), the NER helicase which forms part of the TFIIH core. Unfortunately this protein appeared as a significant hit in both lanes analysed (i.e. it was isolated from **Pd1** and **Pd2**) and so was likely to have been isolated by affinity for the ODN and not specifically by interaction with AtI1. This finding is consistent with results from co-immunoprecipitation experiments performed by Andrew Marriot in the Margison laboratory, who found that Rad15 was isolated in the assay but further investigation showed that this was due to interaction with DNA and not AtI1 (unpublished results).

The .mgf files were also analysed using an alternative search engine, Phenyx (Genebio). This also produced a significant hit for AtI1 in **Pd1** (with 15 unique peptides identified) which confirmed the previous result. However, in common with MASCOT, almost all of the other proteins found in this search were present in both samples and were abundant cellular proteins (such as ribosomal and mitochondrial proteins). The only protein of interest found in **Pc1** and not **Pc2** that was different to the MASCOT searches was rpc40, the

S.pombe homologue of RPAC1, which is part of the Pol core element of RNA polymerases I and III. However, only one unique matched peptide was found for this protein which questions the validity of the result (and in addition it did not appear in the MASCOT searches).

One of the problems with these experiments was that there were a large number of proteins (approximately 200) isolated by both pull-down assays. Most of these proteins were present in both lanes, leading to the conclusion that they were isolated non-specifically by the ODN or beads under the assay conditions. Some proteins only appeared in the **Pd1** or **Pd2** lane but had no significance as they were not related in any way to DNA repair. It was possible that our proteins of interest (i.e. NER recognition factors) were not being detected or identified due to the sheer number of proteins in the background that were masking low abundance proteins. This is because the MS/MS experiment has limitations in detecting low abundance peptides.

Hence, we decided to re-analyse some of the samples but using a slightly different MS/MS experiment, known as pseudo-selective reaction monitoring (pSRM), where certain peptides are targeted by the mass spectrometer for analysis. When looking for low abundance proteins, this technique can markedly enhance the sensitivity of the experiment and generate a low signal to noise ratio.⁽¹³⁴⁾ It essentially involves instructing the mass spectrometer to select for a number of predicted peptide ions from a protein of interest. The sequence of this protein must be known so that the expected masses and m/z values of the peptide ions that will be generated by the experiment can be calculated. The mass spectrometer then specifically isolates precursor ions of these m/z values for further fragmentation. This

```

      10          20          30          40          50          60
MNLTFKLNQQ QKFVISDVSA DTKISELKEK IQTQQNYEVE RQKLIYSGRI LADDKTVGEY

      70          80          90          100         110         120
NIKEQDFIVC MVSRPKTSTST TPKSAASPAP NPPASVPEKK VEAPSSTVAE STSTTQTVAA

      130         140         150         160         170         180
AAPSNDPTTA TSEAPIDANT LAVGAQRNVA VENMVEMGYE RSEVERAMRA AFNNPDRAVE

      190         200         210         220         230         240
YLLTGIPEDI LNRQREESAA ALAAQQQQSE ALAPTSTGQP ANLFEQAALS ENENQEQPSN

      250         260         270         280         290         300
TVGDDPLGFL RSIPQFQQLR QIVQQNPQML ETILQQIGQG DPALAQAITQ NPEAFLLQLLA

      310         320         330         340         350         360
EGAEGESALP SGGIQIQITQ EESESIDRLC QLGFDNRNIVI QAYLACDKNE ELAANYLFEH

GHESEDEP

```

Figure 6.12: Primary amino acid sequence of Rhp23 protein

Fragment sequence	Fragment Mass	m/z Value (+2 Ion)
IQTQQNYEVER	1407.5	704.75
SAASPAPNPPASVPEK	1519.67	760.835
EQDFIVCMVSRPK	1551.84	776.92
NVAVENMVEMGYER	1640.85	821.425
AVEYLLTGIPEDILNR	1816.08	909.04

Table 6.1: Peptide fragments from Rhp23 with masses and m/z values for +2 ions

drastically reduces the background by cutting out all ions of other m/z values.

This approach was used to search for the NER recognition factors Rhp7, Rhp23 and Rhp41. Rhp23 has a mass of 40,135Da and so would therefore be expected to be found in gel piece 4 (based on the gel marker, and the mass range (36,000-48,000) of the proteins identified in the previous MS/MS ions search for this gel piece). Input of the sequence of Rhp23 (figure

6.12) into a proteomics toolkit (expasy.org) allowed a list of the peptide fragments expected to be generated by ESI-TRAP to be calculated, and five peptide fragments were chosen with a RMM of between 1,400 and 2,000 Da, which are shown in table 6.1. These peptides were selected because their +2 ions will have m/z values in the range 700-1000 which aids their selection as this is where the mass spectrometer is most sensitive. These m/z values are then programmed into the mass spectrometer during the MS/MS experiment, so that only peptides ions with the m/z values in table 6.1 (or very close to them) will be selected for fragmentation. The search then attempts to match the experimental and theoretical fragmentation patterns as before.

Unfortunately none of the NER recognition factors were identified using this method. This may be because they are not present in the sample rather than a lack of detection due to low abundance. However, it may have been that the large number of non-specific interactions from abundant proteins was preventing At1 from being co-purified with any interacting partners. Thus, it was decided to attempt the pull-downs again with some modifications to the procedure.

It seemed likely that many of the abundant proteins were binding non-specifically to the DNA, and therefore shorter ODN substrates (100-mers) were used in order to provide less surface area for these interactions. In addition, ODNs containing O^6 -methylguanine, O^6 -benzylguanine and guanine were used to evaluate whether the At1-ODN complex containing O^6 -BnG would be more successful in isolating the proteins suspected of interacting with At1. The affinity of the interaction between At1 and ODNs containing O^6 -BnG is around ten-fold greater than for ODNs containing O^6 -MeG and this

could increase the efficiency of isolation of At11, or possibly affect recognition of the At11-DNA complex by NER factors.

The three ODN substrates used in the assays are shown in figure 6.13. The single-stranded 100-mer ODNs containing either an O^6 -methylguanine or guanine residue were synthesised commercially along with the 5'-biotinylated complementary strand. These were annealed at a concentration of 10 μ M (in 50mM NaCl solution) by heating at 98°C for 5 minutes before being allowed to cool slowly to room temperature. The ODN containing O^6 -benzylguanine was already double-stranded by the nature of its synthesis (i.e. by three way ligation after annealing, section 3.3.2).

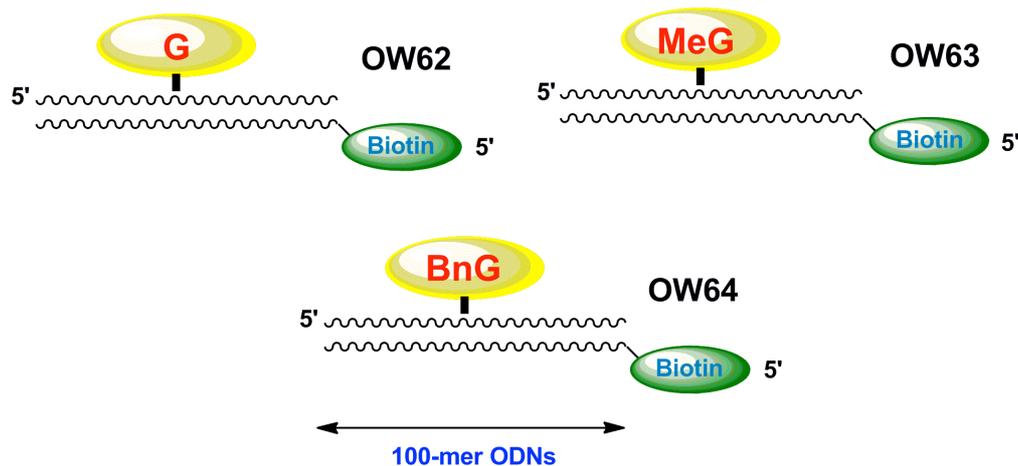


Figure 6.13: ODNs used in the second set of pull-down assays

The ODN duplexes were bound to the streptavidin beads in exactly the same manner as described previously. This time, however, the WCEs were prepared in a variety of different ways before the pulldown assays were carried out. Firstly, the WCEs were sonicated (6 x 30s pulses) to disrupt the nuclei of the cells and attempt to destroy the chromatin structure. This step was performed in an attempt to release any chromatin-associated proteins

(such as DNA repair proteins) and make them more freely available in the supernatant. It was hoped that this would facilitate the detection of NER factors that may have affinity for the At1-DNA complex. The sonicated WCEs were now split into three equal aliquots, which were prepared differently for the assays. The first was used in the assays without any further treatment. The second had ATP added to a final concentration of 1mM. The reasoning behind this was that formation of the desired complexes may be ATP-dependent and so by providing a high concentration in the extract, complex

Extract/ODN	G	MeG	BnG
WT + ATP	1	7	13
WT - ATP	2	8	14
WT + ATP + benzonase	3	9	15
ΔAt1 + ATP	4	10	16
ΔAt1 - ATP	5	11	17
ΔAt1 + ATP + benzonase	6	12	18

Table 6.2: Combinations of ODNs and treated WCEs used in the second set of pull-down assays

formation may be promoted. The third was treated with benzonase and 2.5mM MgCl₂. This nuclease was added in order to digest all the DNA and RNA in the WCE, and thus cause the release of any DNA-associated proteins of interest. After clearing the supernatant, EDTA was added to a final concentration of 10mM to inhibit any residual nuclease activity (by chelating any Mg²⁺ ions required for activity) and prevent digestion of the bait ODNs.

The pull-down assays were prepared by mixing the ODN substrates and WCEs in the combinations shown in table 6.2 and were then incubated at 25°C for 1 hour. The beads were then removed from the supernatant and washed twice with 750µL extraction buffer. This was a much larger volume than was used before, in an attempt to wash away more of the abundant and non-specifically bound proteins. The beads were then kept and stored at -20°. In order to avoid the time-consuming process of extensive sample preparation (1-D gel PAGE separation, band excision, digestion and peptide extraction), it was hoped that the proteins could be dissociated from the beads, digested directly, and then pSRM experiments performed to search for proteins of interest. Hence, the beads from selected samples (2, 8 and 14) were boiled for 5 mins in 50mM ammonium bicarbonate with 0.1% SDS solution to dissociate the proteins and then the supernatant was treated with trypsin (the standard buffer for trypsin digestion is 50mM ammonium bicarbonate and the manufacturer stated that this concentration of SDS would not affect protease activity). After digestion overnight at 37°C, the solution was dried down to a pellet, resuspended in Switchos buffer and analysed using the mass spectrometer as before.

Unfortunately this time no peptide identifications were made. It was difficult to tell at which stage the problem occurred: whether the proteins had not been cleaved from the beads efficiently, if there was a problem with the trypsin digestion, or if an unknown factor had been responsible. It was therefore decided to take samples 1, 7, 13 and 4, 10 and 16 (table 6.2) and prepare them for the MS/MS experiment according to the original method. The beads from these pulldown assays were boiled in 1x Laemmli buffer for 5 mins

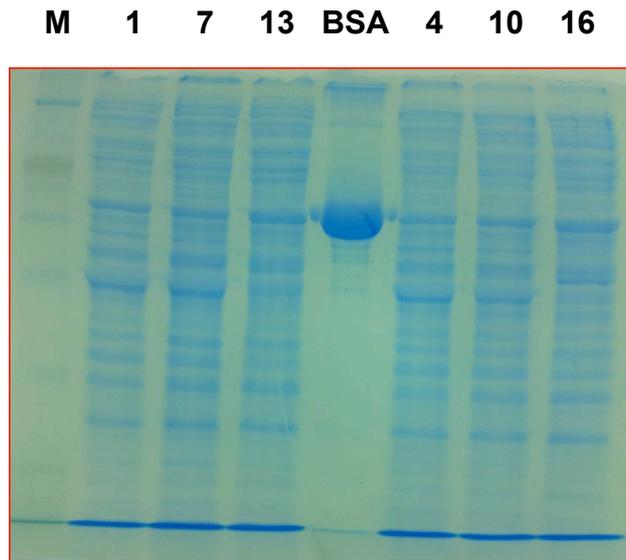


Figure 6.14: SDS-PAGE gel of protein mixtures from selected pulldown assays

and then separated and analysed by SDS-PAGE. The PAGE analysis of the samples is shown in figure 6.14. The gel appears to be similar to that from the first set of pulldown assays, possibly with slightly fewer proteins present. The lanes corresponding to experiments 1, 7 and 13 were each divided into eight pieces and then processed as before. However, this time a reduction and alkylation step were performed to irreversibly block any cysteine residues present in the proteins. This prevents disulphide bridges forming which can interfere with trypsin digestion. Hence, the gel pieces were treated with dithiothreitol (DTT) and then iodoacetamide before the trypsin digest. The samples were then analysed by MS/MS in the same way as before.

A number of proteins were identified by matched peptides, and there appeared to be a fewer number of proteins present than in the previous experiment. However, At1 was not identified in any of the samples (including those from the pull-down assays using ODNs containing O^6 -MeG and O^6 -BnG lesions). The reasons for this are difficult to explain: it may be that more

extensive washing with a larger volume of buffer resulted in AtI1 dissociating from the ODN during this step and not being present in the final sample, or that the treatment of the WCEs somehow led to the failure of the experiments. Unfortunately, due to time constraints these experiments were unable to be continued any further.

7.0 Attempted Synthesis of a Crosslinker

7.1 Introduction to Crosslinking of DNA to Proteins

Interactions between proteins and DNA are fundamental in biological systems and the understanding of them is important. There are various methods by which these interactions can be studied and one useful approach is that of cross-linking. Chemical cross-linking of protein to DNA involves forming a chemical bond between them, to generate a covalently bound DNA-protein complex. The formation of a covalent complex of this kind is especially useful when the interaction between the protein and DNA is transient or of not particularly high affinity, and allows the complex to be studied in more detail. A specific, cross-linked protein-DNA complex can be further characterised or manipulated, as it is likely to be considerably more stable than the native complex.

It is obviously desirable for the covalent complex to be as close in structure to the native complex as possible. If the disturbance of the true interaction can be kept to a minimum, then the information garnered from the studies will be more relevant. In this sense it is important that the cross-linking reaction is highly specific; that is, one covalent bond is made between the DNA and the protein, and that the interaction in the cross-linked complex reflects the actual interacting architecture.

The cross-linking reaction is generally achieved via nucleophilic or electrophilic substitution. As there are potentially many reactive sites in a protein or DNA molecule, it is desirable to create a unique site where the reaction will preferentially take place. In this way, a particular product can be

generated, one that will be useful for further investigations. A useful approach is to exploit the nucleophilicity of a specific cysteine residue in the protein of interest. If a site is created in the DNA substrate that will be exclusively attacked by the cysteine residue, it is likely that the desired product will be formed.

In 1996, Sigurdson and Eckstein reported the synthesis of a 2'-urea-modified oligoribonucleotide that is functionalised on an internal 2'-amino-2'-deoxyuridine nucleoside with a disulfide-containing reporter group (figure 7.1). This oligomer reacts with glutathione to form a single product.⁽¹³⁵⁾ Glutathione is a cysteine-containing tri-peptide (of sequence γ -Glu-Cys-Gly) and therefore this provides a demonstration of a site-specific chemical reaction involving a cysteine residue. In this case, the thiol group on cysteine is forming a disulphide bond with the modified decamer.

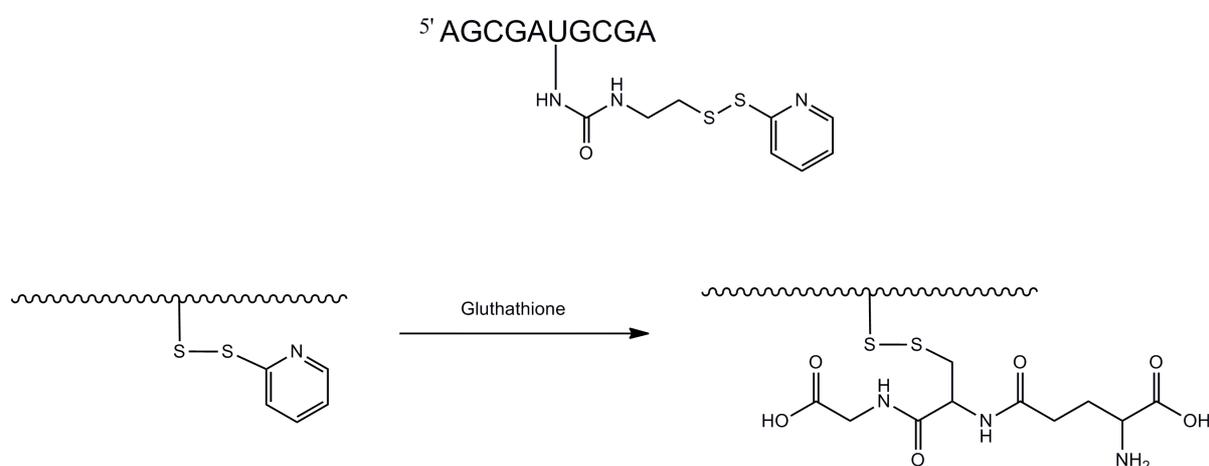


Figure 7.1: 2'-Urea-modified oligoribonucleotide (top) and schematic of its reaction with glutathione (bottom)

In addition, Xu has reported the preparation of an ODN containing 6-methylsulphoxypurine (figure 7.2). The modified pentamer was observed to

react with glutathione, displacing the 6-methylsulphoxy group on the purine and forming a cross-link via the sulphur atom of cysteine.(90) A 12-mer with the same modification also reacted in the same manner. It is an example of nucleophilic substitution by a peptide cysteine residue on a modified DNA base and could possibly be used to cross-link DNA with larger peptides and proteins.

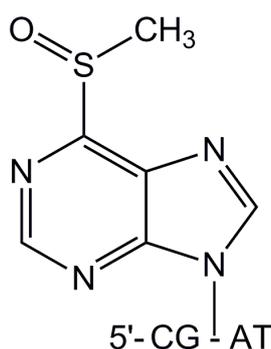


Figure 7.2: 6-methylsulphoxy-purine-containing oligodeoxyribonucleotide pentamer

MGMT, the human AGT protein homologue, removes O^6 -alkylguanine lesions from DNA by irreversible transfer of the alkyl group to a reactive cysteine residue in its active site.(48) The transfer is rapid, making observation of the DNA-protein complex extremely difficult under normal conditions. It was therefore useful to cross-link the MGMT with a DNA substrate to allow further biochemical and structural studies to take place. The first successful attempt at cross-linking to MGMT was published by Paalman, Noll and Clarke in 1997.(136) The covalent complex was formed using a 2'-deoxy-6-(cystamine)-2-aminopurine ($d^6\text{Cys}^2\text{-AP}$)-containing oligonucleotide (figure 7.3) and it was shown that the active site cysteine reacts to form a disulphide bond with the 6-cystamine group in the oligonucleotide to give the desired MGMT-DNA complex.

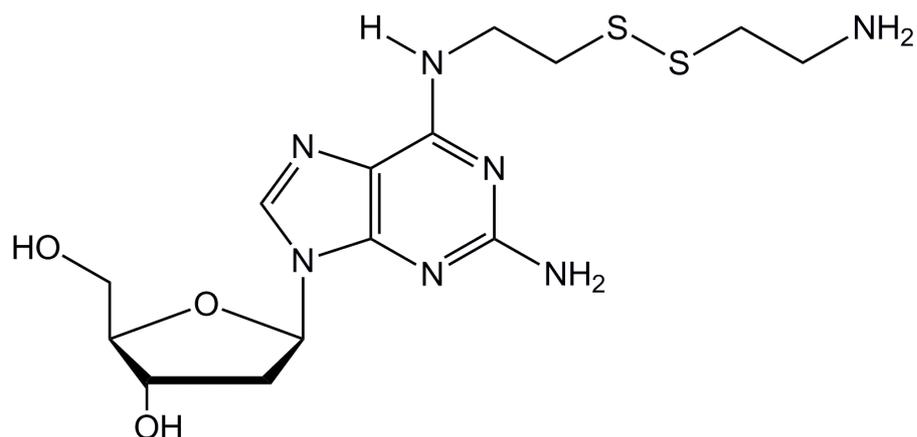


Figure 7.3: 2'-Deoxy-6-(cystamine)-2-aminopurine ($d^6\text{Cys}^2\text{-AP}$)

Unfortunately, this cross-linking reaction was inefficient and the product unsuitable for structural studies. Therefore, Noll and Clarke went on to covalently capture MGMT using an oligodeoxyribonucleotide containing N^1, O^6 -ethanoxanthosine (${}^e\text{X}$), an approach which resulted in cross-linking of MGMT to DNA in high yield.(137) The ${}^e\text{X}$ modified base is recognised for repair by MGMT, and subsequent nucleophilic attack by the active site cysteine on the O^6 -methylene carbon produces the covalent protein-DNA complex (figure 7.4). This cross-linked complex allowed the first crystal structure of MGMT bound to its DNA substrate to be elucidated and published which hugely increased understanding of the function and mode of action of the protein.(51)

It would be useful to form a complex of At11 (from *S.pombe*) covalently bound to an O^6 -alkylguanine-containing oligodeoxyribonucleotide (ODN). Although the interaction between At11 and a DNA substrate was sufficiently stable to generate a crystal structure, a covalent At11-DNA complex might be more appropriate for pull-down assays (sections 6.1 and 6.3) in order to find any proteins that interact with this complex. Such complexes might also be of

value to evaluate how such interactions lead to the repair of O^6 -alkylguanine lesions in *S.pombe*.

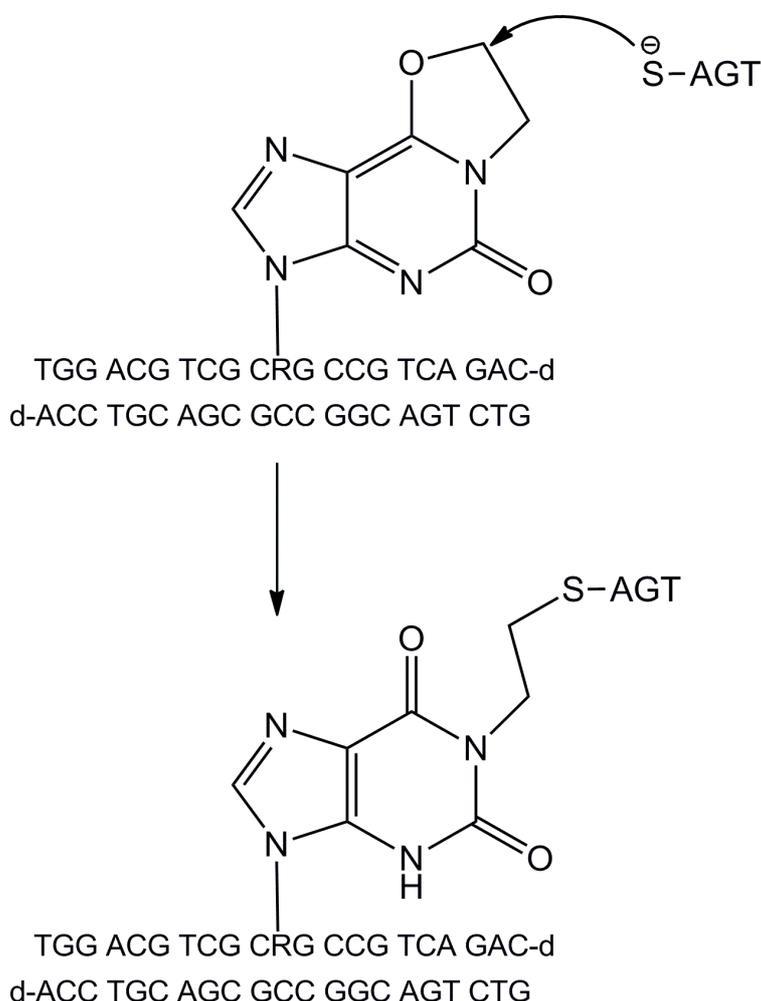


Figure 7.4: Crosslink formation between an oligodeoxyribonucleotide containing N^1,O^6 -ethanoxanthosine (eX) and MGMT

Therefore, for the purposes of biochemical investigation it was decided to create a covalently bonded At11-DNA complex. There were various possibilities as to how the desired chemical crosslink may be formed. The approaches described here consist of two main ideas for the development of a crosslinker: use of a disulfide linkage, and use of a maleimide moiety. The

hope with both approaches was that AtI1 would still recognise these modified bases as damaged DNA, bind the ODN accordingly, and finally react to generate a site-specific crosslink. This should result in an AtI1-DNA covalent complex that is close in nature to the usual bound state.

7.2 Disulfide Crosslinkers

Based on previous work by various researchers,(135,136) it was considered viable that a functionalised ODN containing an alkylguanine residue which contains a disulfide linkage would be recognised as a substrate by AtI1, and would also have the ability to form a disulfide linkage with a cysteine residue in the active site of the AtI1 W56C mutant. The first modified ODN (containing an *N*⁶-alkyl-2-aminopurine base) that was attempted to be synthesised was compound **1**, shown in figure 7.5.

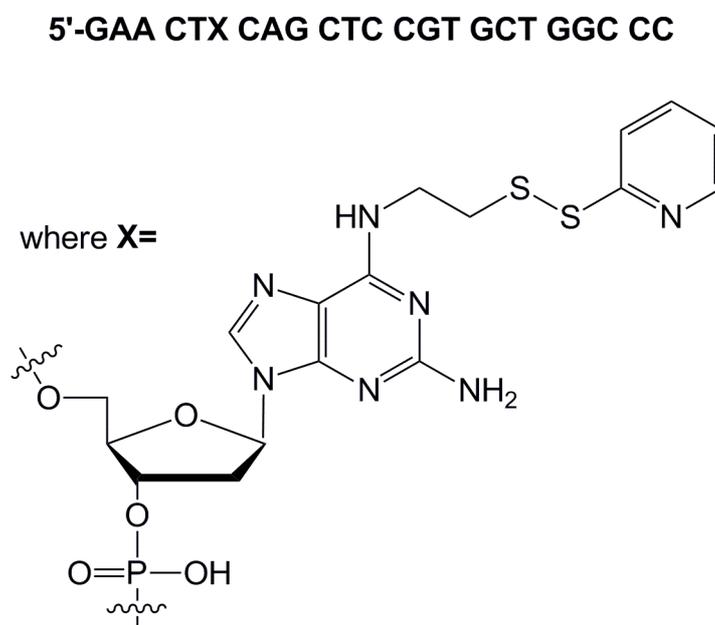


Figure 7.5: Structure of compound **1**

The synthesis of **1** was based on the post-synthetic modification chemistry developed and used extensively by the Williams group (20) (section 3.1.2 and figure 3.4). It was envisaged that S-(2-pyridylthio)cysteamine hydrochloride (PDA.HCl) (compound **2**) (figure 7.6) could be used as the nucleophile in this reaction to displace the sulfone group from the 2-amino-6-methylsulfonylpurine base in the ODN and give the desired product.

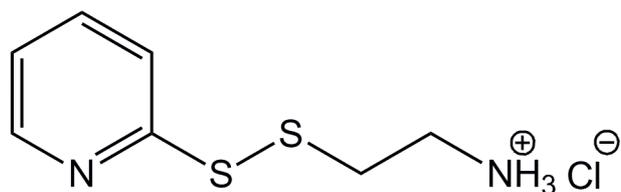


Figure 7.6: S-(2-Pyridylthio)cysteamine Hydrochloride (PDA.HCl) (**2**)

Thus **2** was prepared according to the literature procedure.(138) It has been documented that PDA (i.e. with a free amino group) is unstable (138) and so it was decided to use the hydrochloride salt (figure 7.6) in the reaction and deprotonate it *in situ* using DBU. Rather than a 2 day displacement at 37°C followed by a 3 day treatment with NH₄OH at room temperature (which are the standard conditions for an ODN displacement reaction), it was instead decided that the reaction would be incubated at 37°C for 5 days without any addition of aqueous ammonia. It was thought that these conditions would be sufficient to cause displacement, cleavage from the beads and deprotection of the nucleobases. In addition, due to the lack of solubility of PDA.HCl in acetonitrile, a total reaction volume of 600 µL was used which gave the final concentration of the amine as 0.83 M (usually 5M alcohol would be used in a displacement).

The HPLC traces of the crude reaction mixture showed a peak at around 26 min on a 0-40% gradient (from 100mM TEAB with 5% MeCN to pure MeCN), but this was subsequently shown to be starting material (PDA.HCl). Mass spectrometry indicated that there was no DNA in the sample. The spectrum showed only a mass ion at 222 (PDA.HCl), and no mass ions were present corresponding to an ODN of large molecular weight, either the sulfone containing starting material, or the expected product. It is probable that the basic conditions used did not cause cleavage of the DNA from the CPG beads upon which they were immobilised. There was also the issue with the stability of PDA, which may not make it a suitable reagent in the displacement reaction. The degradation of PDA, like that of other unsymmetrical disulfides (139,140) occurs through an irreversible disproportionation reaction where PDA is converted into its symmetrical counterparts. It was therefore decided to attempt the displacement reaction again, but this time using standard conditions and a symmetrical disulfide.

The symmetrical disulphide selected was cysteamine dihydrochloride (figure 7.7). It was expected that this diamine could react with the sulfone-containing ODN starting material through either terminal amine group to give rise to the same ODN product (compound **3**) shown in figure 7.8.

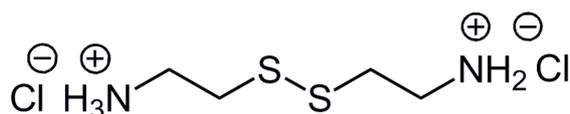


Figure 7.7: Cysteamine dihydrochloride

5'-GAA CTX CAG CTC CGT GCT GGC CC

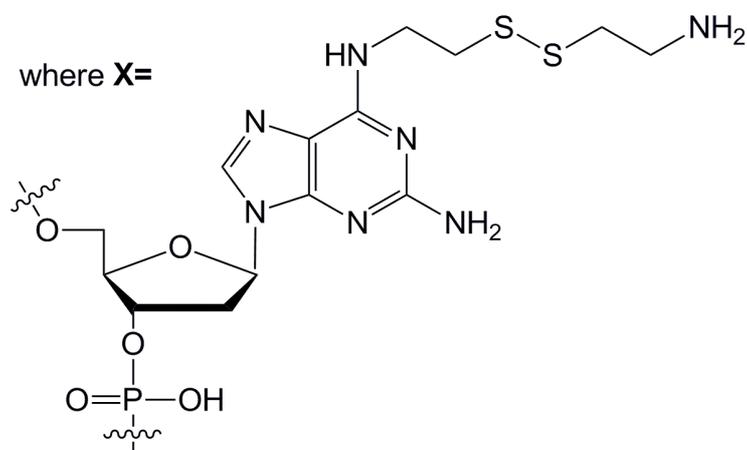


Figure 7.8: Structure of compound 3

This time, a 1.2 M solution of amine in acetonitrile was used along with 2 equivalents of DBU base (relative to the amine). After displacement and deprotection, the crude reaction mixture was purified by size-exclusion chromatography (NAP-5) to remove any amine present prior to being run on the HPLC. The product mixture, when analysed by RP-HPLC, gave a peak at 21 min retention time on a 0-40% gradient (from 100mM TEAB with 5% MeCN to pure MeCN), which is where a successfully modified ODN would be expected to run. This is consistent with a displacement reaction and subsequent cleavage from the CPG beads having had occurred. Unfortunately, mass spectrometry was unable to clarify which reaction had taken place (figure 7.9): there were many peaks at corresponding to high molecular weight products (i.e. ODNs) present in the spectrum, none of which corresponded to the mass of the desired product (7439 Da) or that of the ODN starting material (7382 Da).

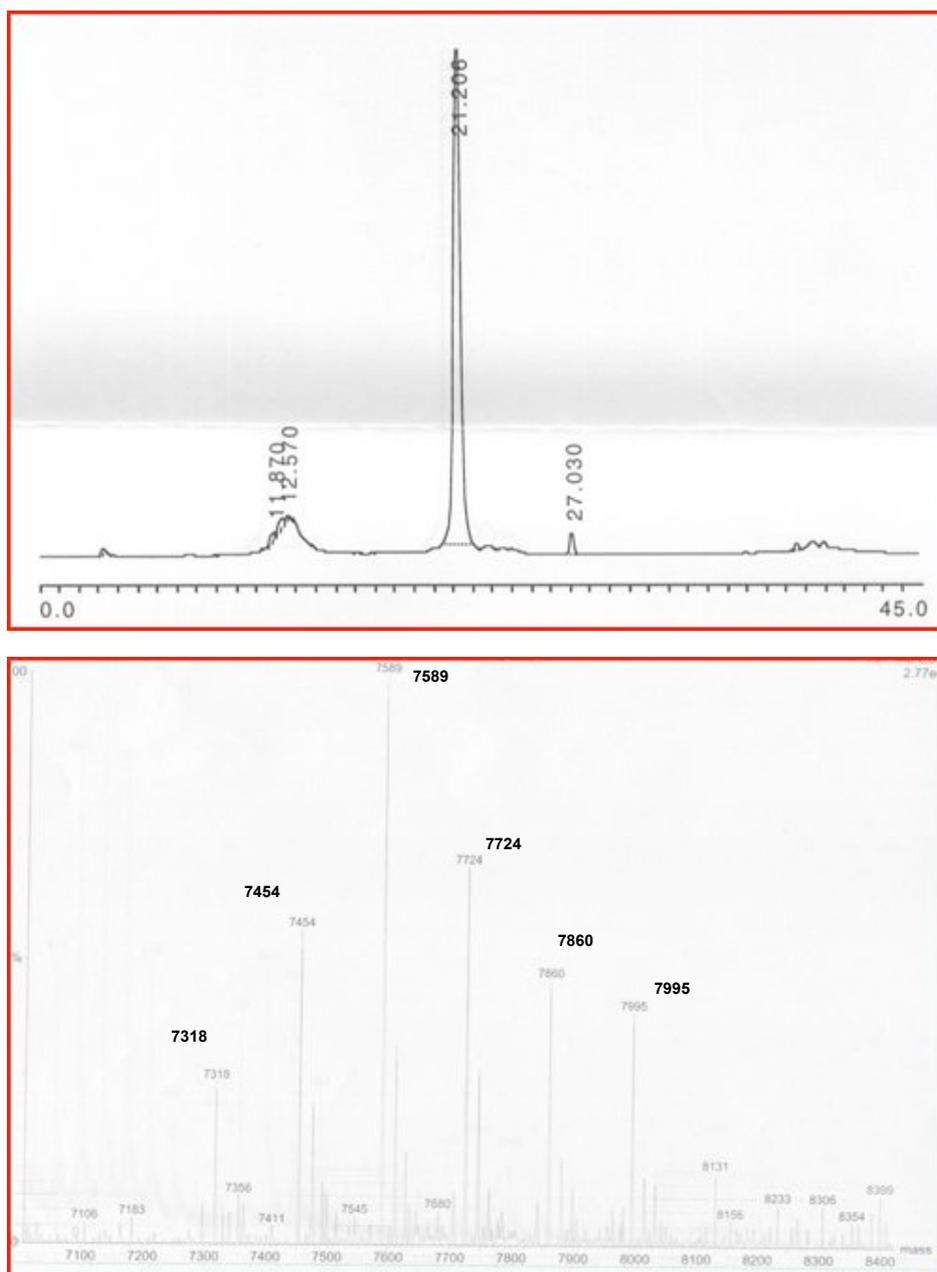


Figure 7.9: HPLC trace and MS spectrum of compound 3

To further investigate what may have happened, a small amount of the sample was digested with snake venom phosphodiesterase (SVPDE) and treated with alkaline phosphatase (AP). HPLC analysis of this sample was expected to produce four distinct peaks (corresponding to dG, dC, dA and dT) and a small extra peak for the alkylguanine base (figure 7.10) which was

indeed observed in the HPLC trace. However, analysis of the sample by mass spectrometry, whilst showing peaks for dG, dC, dT and the cleaved 5'-DMT group, failed to detect a peak of correct mass for the expected alkylguanine base (X in figure 7.8, mass (dX) = 404 Da). It was thus decided at this point to temporarily abandon the pursuit of a disulfide crosslinker in order to concentrate on preparing one containing a maleimide function.

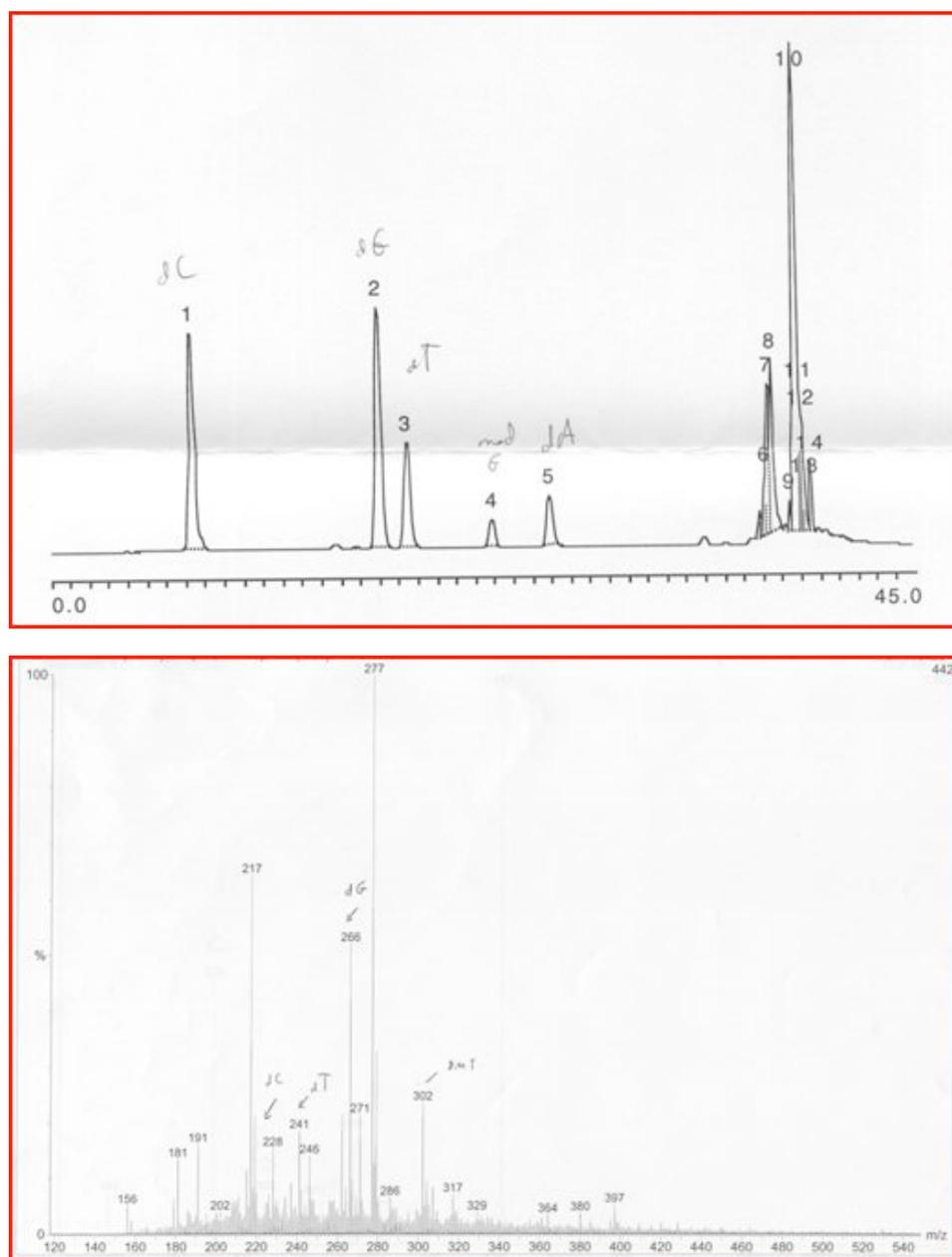


Figure 7.10: HPLC trace and MS spectrum of compound **3** after treatment with SVPDE and AP (which digests the ODN into its constituent nucleosides)

7.3 Maleimide Crosslinkers

The second approach was to incorporate a maleimide moiety at the 6-position on a purine base contained within the ODN, so it could react via a Michael addition with the cysteine residue in the active site of the AtI1 W56C mutant. This would be expected to generate the necessary crosslink between the protein and DNA. N^6 -alkyl-2-aminopurine or O^6 -alkylguanine was considered suitable for this purpose, with the alkyl group being *N*-ethylmaleimide (figure 7.11).

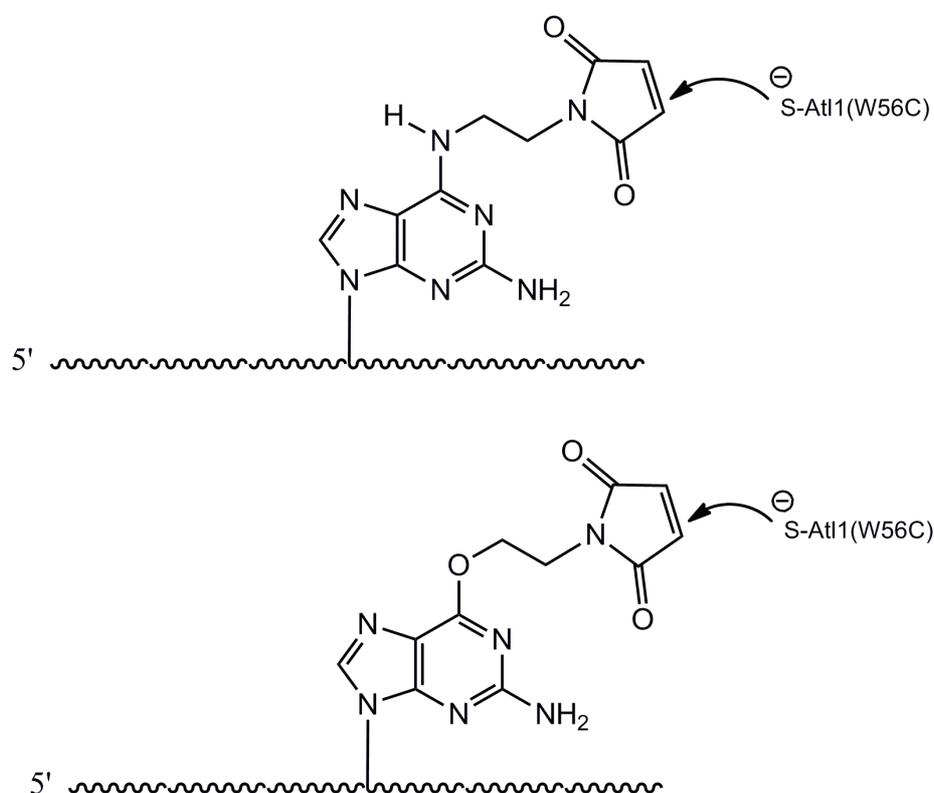


Figure 7.11: Proposed crosslink formation between maleimide-functionalised bases and W56C mutant AtI1 active site cysteine residue

The reaction scheme to produce the maleimide ethylamine (required for the ODN displacement reaction) is shown in figure 7.12. Briefly, it proceeds as follows: Boc-protection of the amino group on 2-aminoethanol

(55), condensation of the alcohol group with maleimide using the Mitsunobu reaction (56), and final deprotection with trifluoroacetic acid to give the product as its trifluoroacetate salt. The first step gave the protected amino alcohol (compound **4**) in 53% yield. The Mitsunobu reaction also proceeded as expected, and after subsequent deprotection with TFAA gave the product (compound **5**) as a yellow oil. The product was characterised by ^1H NMR and mass spectrometry. The mass ions of the cation (141 Da) and anion (113 Da) were present in the electrospray (+ve and -ve) mass spectra, confirming that the synthesis was successful. In addition, the ^1H NMR had all the signals expected for **5** (figure 7.13). However, there were additional peaks in the NMR spectrum that indicated the presence of impurities in the sample. After further ^1H NMR analysis it was confirmed that a significant amount of maleimide (seen at 6.9 ppm) was contaminating the product, and so it was decided to re-

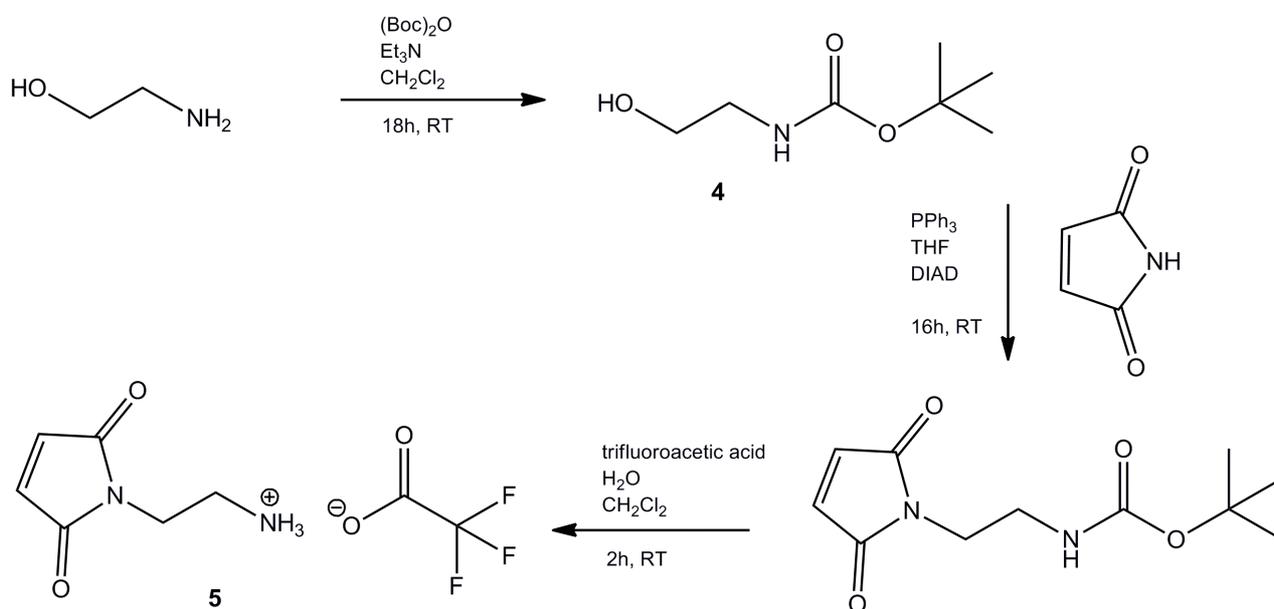


Figure 7.12: Reaction scheme for the synthesis of compound **5**

purify the mixture using ion-exchange chromatography (DOWEX-50 H⁺ form). The negatively-charged resin bound the desired product whilst the neutral maleimide flowed through the column. The immobilised product was then eluted with trifluoroacetic acid to give the same salt as previously. ¹H NMR confirmed the presence of **5** and the removal of the maleimide, but still had two additional peaks in the spectrum, corresponding to an ethyl group (-CH₂CH₂-). These additional peaks (a triplet and a multiplet at 3.65 and 2.98 ppm respectively) can be seen in the ¹H NMR spectrum taken prior to the ion-exchange purification (inset, figure 7.13). It is highly likely that this signal is due to the presence of the trifluoroacetate salt of 2-aminoethanol, which if still present in the sample would have co-eluted with **5** during the ion-exchange

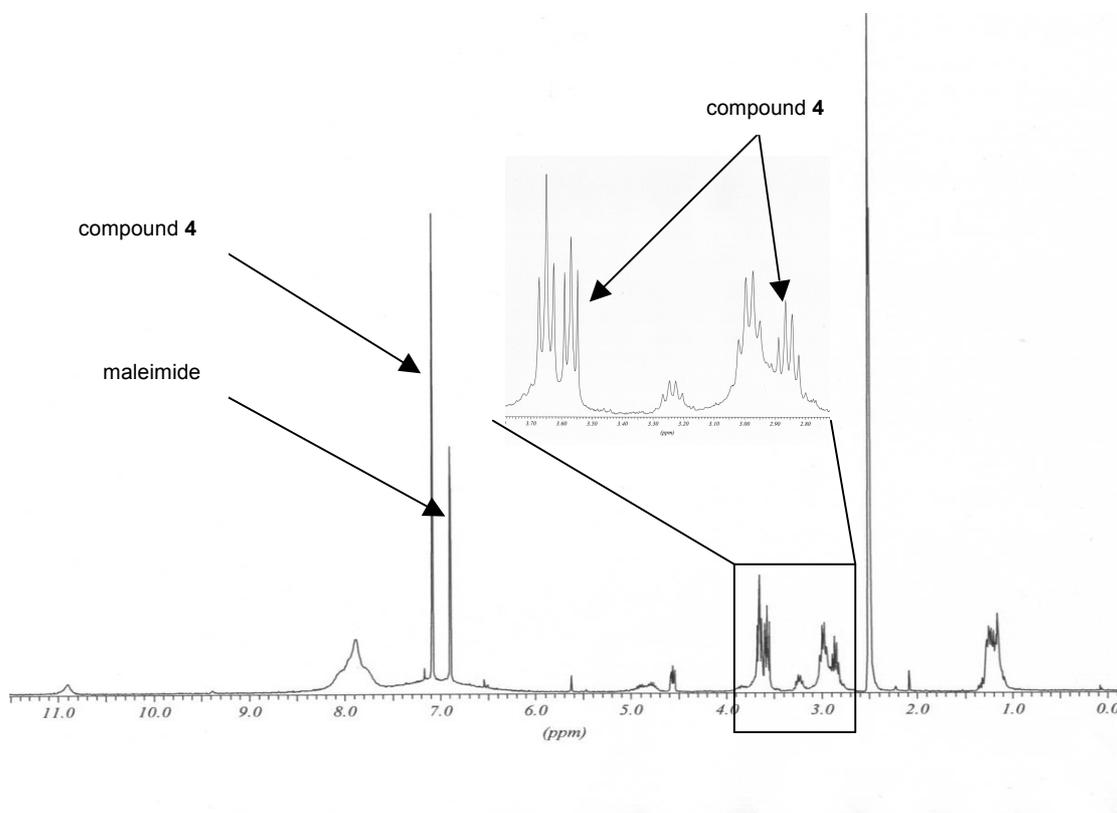


Figure 7.13: ¹H NMR spectrum (d⁶-DMSO) of compound **5**

purification. No further steps were taken at this stage to purify the product. Instead, attention was directed towards the synthesis of *N*-(hydroxyethyl)maleimide (compound **6**) (figure 7.14). The reaction was set up according to the literature procedure (57). After purification of the crude reaction mixture by flash column chromatography, the product was characterised by ^1H NMR and mass spectrometry. The mass ion $[\text{M}+\text{H}]^+$ at 142 Da indicated that **5** had been synthesised successfully. In addition, the ^1H NMR spectrum displayed the expected 2:1 ratio of CH_2 (3.75 ppm) to maleimide (6.75 ppm) proton signals and also showed that the product had a sufficient degree of purity.

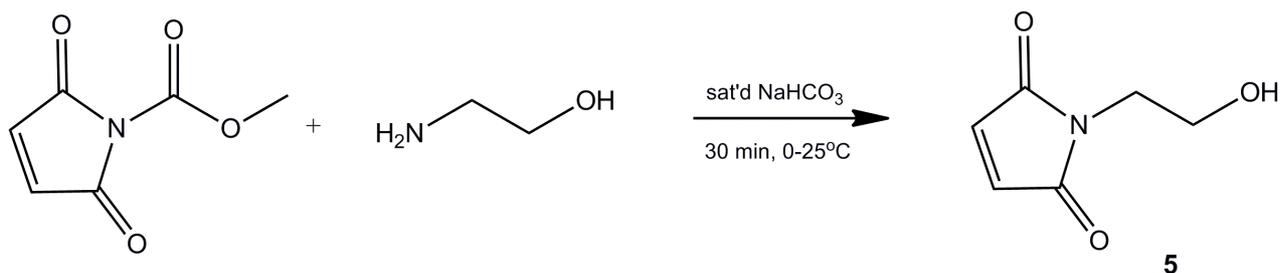


Figure 7.14: Synthesis of *N*-(hydroxyethyl)maleimide (**6**)

The next stage was to incorporate this maleimide alcohol into an ODN using the displacement chemistry mentioned previously. It was considered prudent at this stage to attempt the displacement reaction using an analogous nucleoside base (rather than the ODN itself), as this could be more easily identified by mass spectrometry and also characterised by ^1H NMR. Once this chemistry had been shown to be successful with the nucleoside, it could be performed on an ODN containing a S^6 -sulfonylpyrimidine residue with more confidence. The first step of the process was to convert the *N*⁹-benzyl- S^6 -

methylpurine base (**7**) by oxidation (40) into the corresponding S^6 -methylsulfonylpurine (**8**) which is much more reactive (figure 7.15).

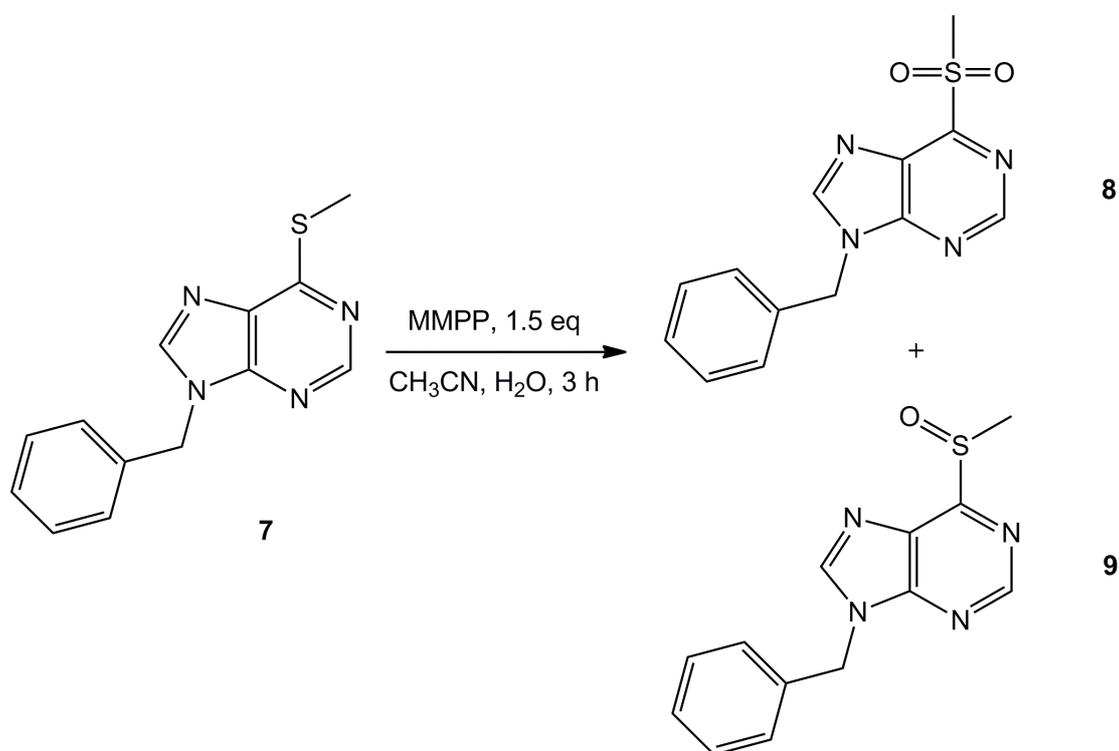


Figure 7.15: Oxidation of S^6 -methylpurine to prepare S^6 -sulfonylpurine (**8**) and S^6 -sulfoxypurine (**9**)

This reaction gave a mixture of both sulfone (**8**) and sulfoxide (**9**) which was confirmed by mass spectrometry (peaks at 273 and 289 for $[M+H]^+$) and by ^1H NMR. Attempts to fully oxidize compound **6** to the sulfone (**8**) by addition of more MMPP were unsuccessful. However in practical terms, since both sulfoxide and sulfone are electrophilic they should be sufficiently good leaving groups for the displacement reaction to work, and will theoretically give rise to the same product.

This purine analogue (and the corresponding residue in the ODN to be used) is deliberately lacking the 2-amino group that would be present in

guanine. This is in order to make the reaction simpler: in a displacement reaction this group must be subsequently deprotected using aqueous ammonia and there may be an issue with stability of the product under basic conditions. Replacing the amino group with a hydrogen atom eliminates the need for this step of the reaction. Instead, purified, deprotected ODNs containing 6-methylsulfonylpurine could be oxidized and then a displacement performed as described by Xu *et al.*(90).

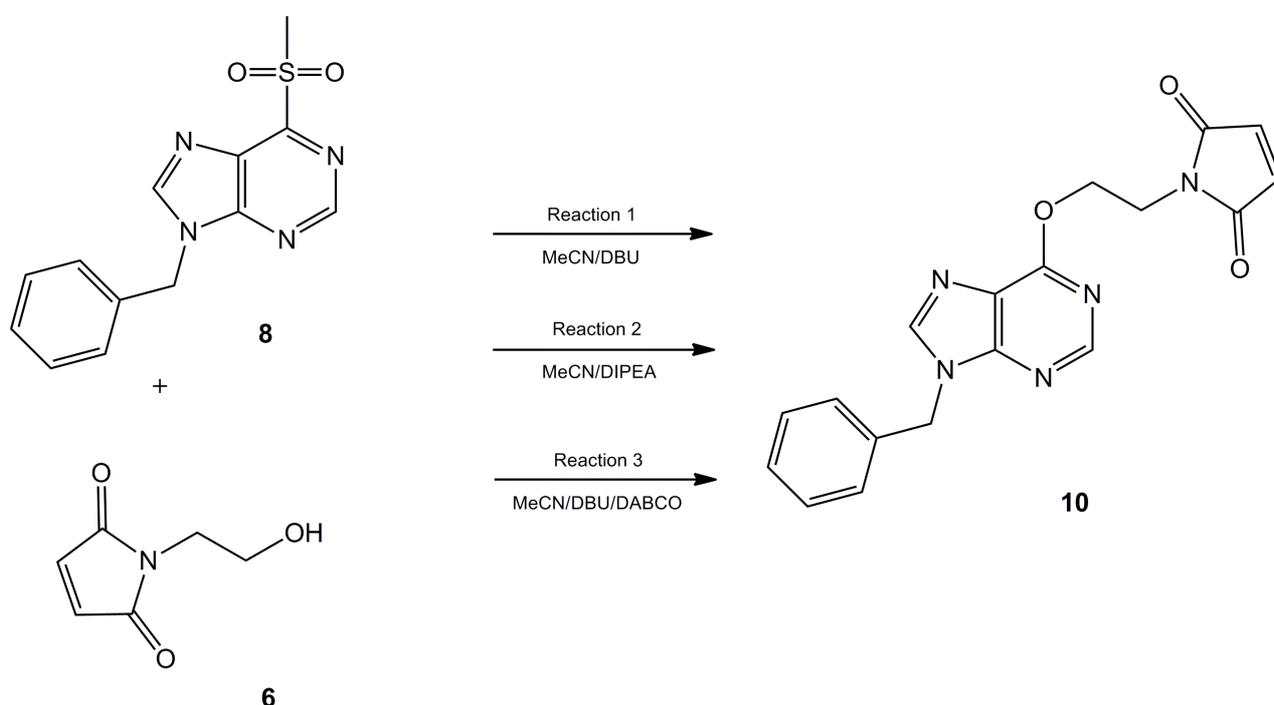


Figure 7.16: Three reactions carried out in the attempted synthesis of compound **9**

Three different reaction conditions were investigated for the synthesis of *N*⁹-benzyl-*O*⁶-(maleimidoethyl)hypoxanthine (compound **10**) (figure 7.16). All reactions were stirred overnight at room temperature. In reaction 1, 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) was used as the base. DBU was last to be added to the reaction mixture and immediately upon addition it was observed that the reaction turned a dark purple colour. This was unexpected

and would indicate that some change had instantaneously occurred to one or more of the reactants. A displacement reaction would not be expected to occur this fast. After the incubation period (overnight at RT) the crude reaction mixture was purified by column chromatography which resulted in separation of two compounds ($R_f = 0.52$ and 0.63). Unfortunately, the mass spectra (figure 7.17, $R_f = 0.63$) of these recovered products did not show a peak corresponding to the expected mass ion for compound **10** ($[M+H]^+ = 350$) and the ^1H NMR spectra did not show any evidence for the desired product.

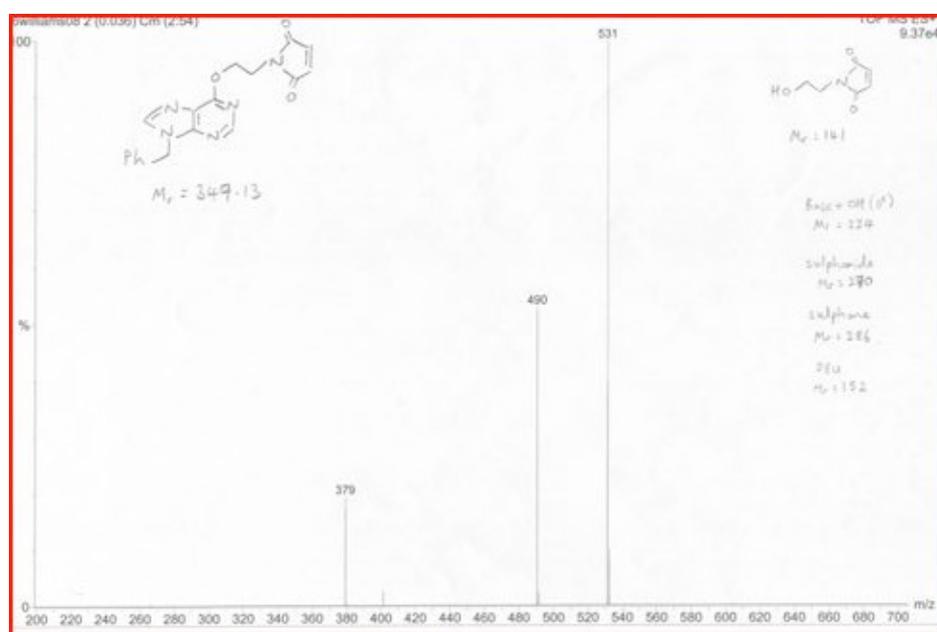


Figure 7.17: MS spectrum of compound **10** (reaction using DBU)

As a result of this, it was decided to repeat the reaction using a different base. If DBU was causing a problem in the reaction mixture, it was hoped that exclusion of it could resolve this difficulty. In reaction 2, *N,N*-diisopropylethylamine (DIPEA) was used with the same reaction conditions. Upon addition of DIPEA, there was no colour change of the reaction though

after 18 h the reaction had turned a red/pink colour. The crude product was analysed by ESI +ve TOF mass spectrum which contained many peaks, one of which was of the correct mass for the expected product ($[M+H]^+ = 350.26$) (figure 7.18). However, the ^1H NMR spectra was highly complex and inconclusive as to whether the reaction had worked. It was conceivable that all the major signals in the spectra were from unreacted starting materials rather than the product, especially if the reaction was inefficient and the product formed only in very small quantities.

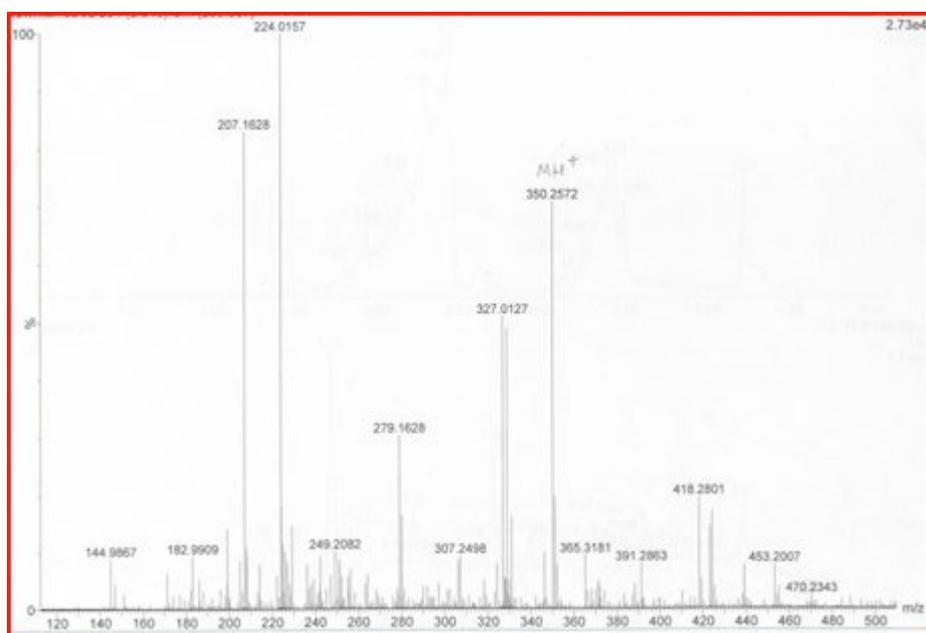


Figure 7.18: MS spectrum of compound **10** (reaction using DIPEA)

Thus, the reaction was attempted once more using a mixture of 1,4-diazabicyclo[2.2.2]octane (DABCO) and DBU. DABCO would be expected to initially displace the sulfonyl group (figure 3.19), before being itself displaced by the maleimide alcohol (**6**). As it is a better leaving group, it should make the

reaction more efficient and give the desired product (**9**). This time, the last component to be added to the reaction was **6**, and upon its addition the mixture turned the same dark purple colour. This indicates that there is some initial reaction between DBU and the maleimide alcohol that is possibly causing the problem and resulting in the failure of the displacement. However, TLC analysis indicated that the sulfonylpurine was being consumed as the reaction progressed and showed the evolution of a new product spot. After the overnight incubation, the crude product mixture was purified by column chromatography, which gave rise to two separate products with different R_f values. The mass spectrum of the faster-running product showed a peak at mass 350, though strangely there were many peaks of greater mass, each separated from each other by masses of 44 (figure 7.19) .

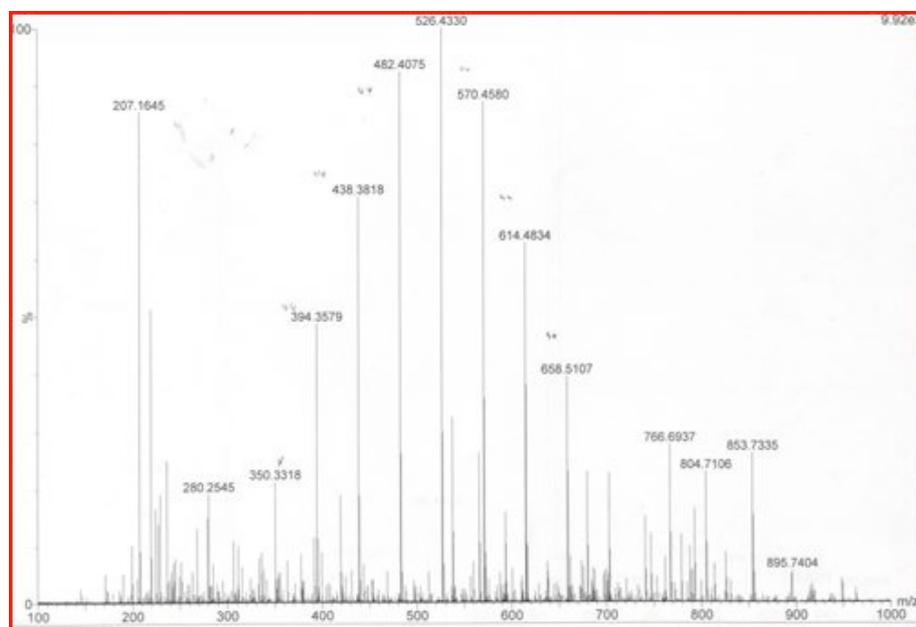


Figure 7.19: MS spectrum of compound **10** (reaction using DABCO and DBU)

However, the expected signals for **9** did not appear to be present in the ^1H NMR spectra. It appeared that all attempts to attach a maleimide-containing alkyl group to the 6-position of the purine base had thus far been unsuccessful and therefore an alternative strategy was pursued.

The final attempt that was made to introduce a maleimide group at the 6-position of a purine base utilised some of the chemistry that had previously been successful: the reaction to prepare compound **6**. It was hoped an ODN containing an aminoethyl group could be reacted with *N*-methoxymaleimide to produce the desired product (figure 7.20). An ODN containing an O^6 -aminoethylguanine residue had already been prepared to be used as a substrate in the AtI1 study, and although this ODN is SIMA-labelled the presence of the dye is unlikely to affect this reaction.

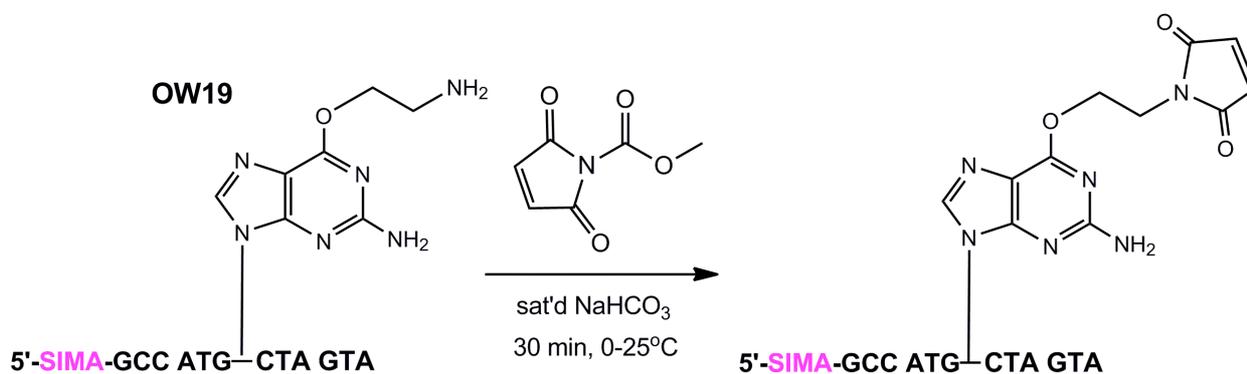


Figure 7.20: Synthesis of an ODN containing an O^6 -alkylmaleimide functional group

The reaction was performed in the same manner as the previous synthesis but with the following modifications. 15nmols of ODN OW19 was evaporated to dryness, resuspended in 200 μL of saturated sodium bicarbonate solution and cooled to 0°C. *N*-methoxymaleimide was added (2.5mg, ~1000-fold excess) and stirred at 0°C for 30 min, after which the

reaction was allowed to warm to r.t over a period of 10 min. The reaction mixture was purified by application to a size-exclusion column (NAP-5) and the eluate analysed by HPLC and MS.

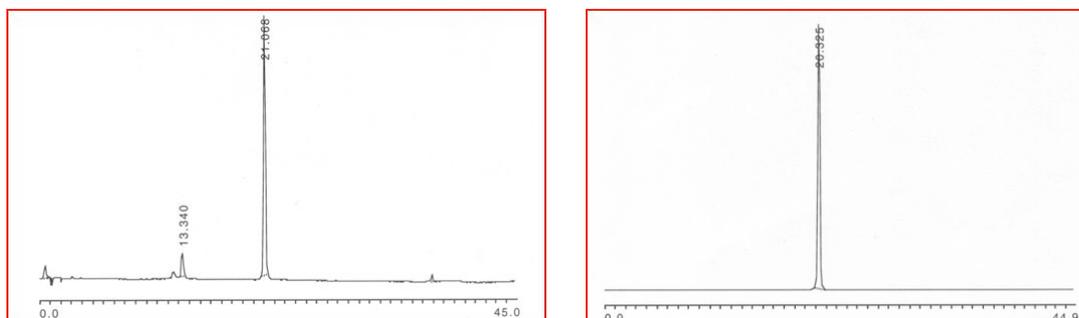


Figure 7.21: HPLC traces of the product of reaction (left) and a co-injection of the product with OW19 starting material (right)

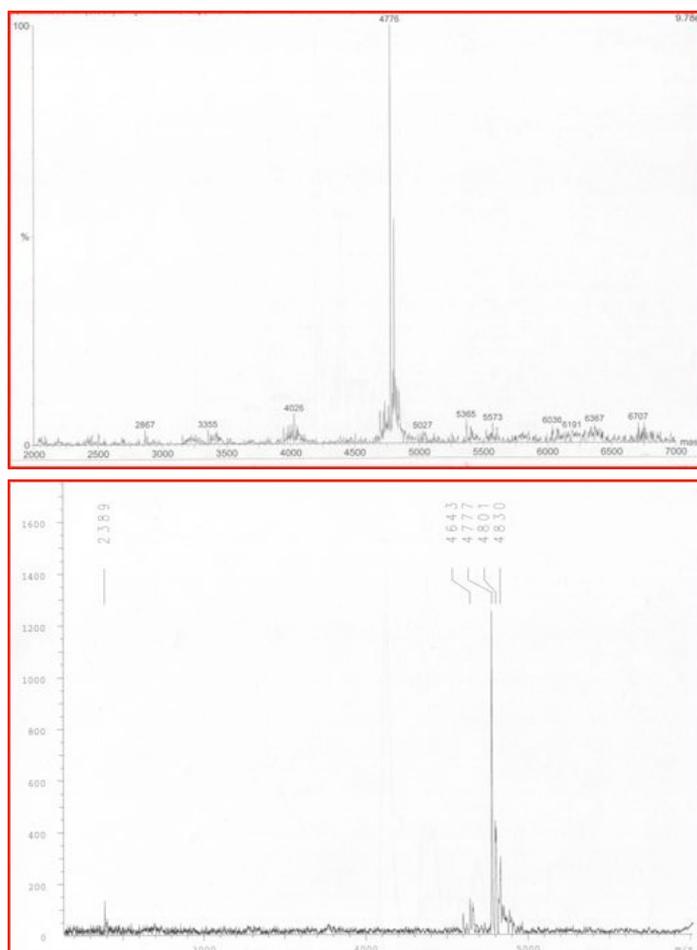


Figure 7.22: Mass spectra of the starting material (top) and product (bottom)

Although there was a peak corresponding to an ODN in the HPLC analysis the reaction appeared not to have been successful, demonstrated when co-injection of the product with OW19 (the ODN starting material) only showed one peak (figure 7.21). The product would be expected to have a different retention time to OW19 due to the presence of the maleimide group which will change the hydrophobicity of the ODN. The mass spectrum also confirmed that the reaction had not taken place and that only the starting material OW11 was present (calculated mass of starting material = 4776 Da, experimental mass of starting material = 4776 Da, calculated mass of product = 4856 Da, experimental mass of product = 4777 Da, figure 7.22). The reasons for the failure of this reaction are presently unknown and due to time limitations this part of the project was not continued.

7.4 Oxidation of ODN Containing S⁶-methylpurine

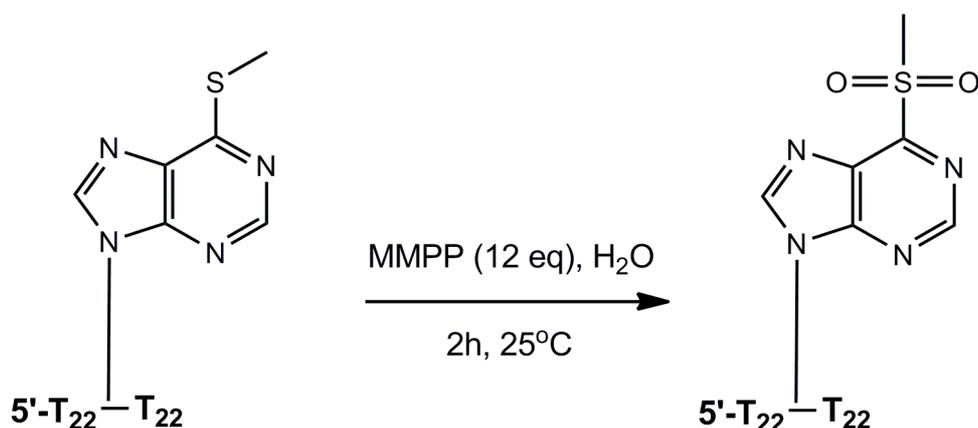


Figure 7.23: Oxidation of ODN OW1

The overall aim of this part of the project was to place a chemical group capable of reacting with cysteine into an ODN substrate. Since the oxidation

of the S⁶-methylpurine base to the more reactive sulphonyl form had been successful, a similar oxidation reaction was attempted with DNA (figure 7.22). The ODN sequence used contains exclusively thymine (apart from the modified base). This was because thymine is the only base that does not require base-catalysed deprotection (which is used during the second stage of the displacement process (section 3.2 and figure 3.4)) and therefore the treatment with ammonia could be omitted. This was considered an advantage as it was not certain that the displaced product (**9**) would be stable under basic conditions. For this reason, the 2-amino group usually present in guanine was also replaced with hydrogen on the modified purine base (as mentioned previously).

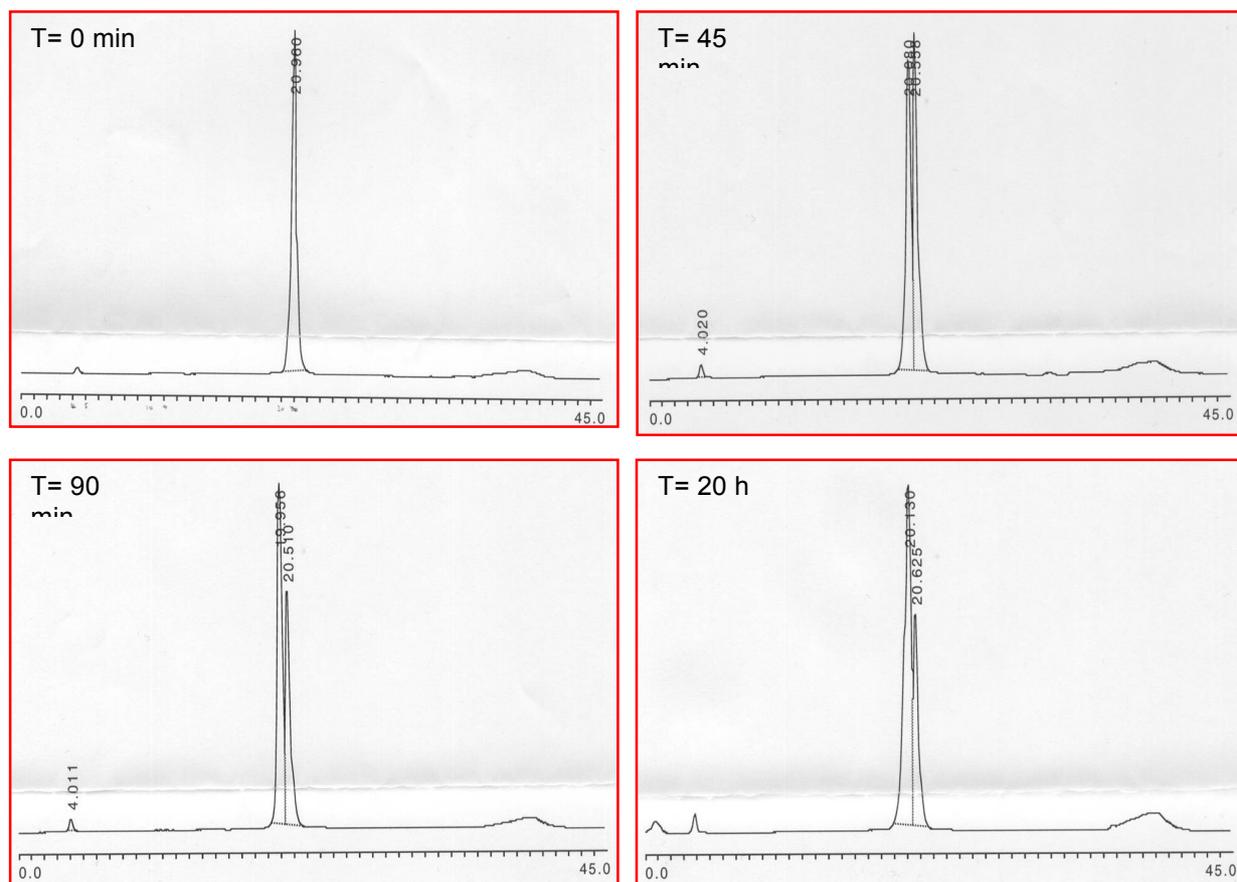


Figure 7.24: HPLC traces to show oxidation of ODN OW1

The reaction was monitored by RP-HPLC, the traces from which are shown in figure 7.23. These traces show that the oxidation reaction proceeds at a reasonably fast rate. Pure OW1 at T=0 (just before the addition of MMPP) shows a single peak at retention time 20.9 min. After 45 min, two new peaks have appeared, at 20.5 and 20 min retention time, presumably corresponding to the sulfoxide and sulfonyl products. At T=90, the peak at 20 min has grown larger and that at 20.5 min smaller, as sulfoxide is oxidised further to sulfonyl. At this stage, the sulfonyl peak has an area of 62% and the sulfoxide 37% (which gives an indication of the relative amounts of each product). After overnight incubation at room temperature, the area of the peaks is 69% and 31% respectively which shows that the reaction has only proceeded slightly further. It was found that even after incubation with additional MMPP, the ratio of these peaks (and hence the ratio of sulfonyl: sulfoxide) could not be made to exceed 69:31. The experiment was repeated a number of times with increasing amounts of MMPP and this was still found to be the case. At the end of the reaction, a co-injection of the reaction mixture with starting material (OW1) established for certain that these two peaks had evolved during the reaction and corresponded to products. However, although the correct mass was able to be established by MS analysis of OW1 (the starting material ODN), once the oxidation reaction had taken place the spectrum was very complex and it could not be established with any certainty whether the HPLC peaks observed were due to the desired products. The experiment was repeated a number of times with different quantities of MMPP but the results of the MS analysis was always highly inconclusive. It was therefore decided to abandon this approach.

8.0 Conclusions and Future Work

8.1 DNA Recognition by ATL proteins

The most extensive study to date of the recognition of DNA containing O^6 -alkylguanine and related purine lesions by ATL proteins has been carried out and is described in this report. Investigation of ATL proteins is valuable and relevant as current evidence suggests that they provide crosstalk between two previously unrelated DNA repair pathways (base repair and NER).(58) Therefore, increased knowledge of the function of ATL proteins may have profound implications in understanding the repair of O^6 -alkylguanine residues which are highly toxic, mutagenic and biologically relevant lesions.

Using evidence produced from fluorescence-based binding studies, along with analysis of new structures from Julie Tubbs and John Tainer, we have elucidated the mechanism by which AtI1, the *S.pombe* ATL homologue, is able to discriminate between guanine and O^6 -methylguanine (and other O^6 -alkylguanine) residues in its active site. Alkylation of guanine at the O^6 -position causes changes in the molecular electrostatic potentials of the atoms of the heterocyclic purine ring, and these differences are detected by an active-site arginine residue (Arg69), which forms positive interactions with the alkylated base and slightly repulsive ones with guanine. In addition, the hydrogen bonding interactions between the N2 and N3 atoms of the base and the AtI1 active site appear to be strengthened by the increased polarisation that accompanies the change in electrostatic potentials due to the presence of the O^6 -alkyl group and Arg69. Thus, Arg69 acts as a molecular probe of

electrostatic potential which acts in a novel and unprecedented mechanism to mediate discrimination between damaged and non-damaged guanine bases by AtI1.

In addition, AtI1 was shown to recognise with greater affinity bases that are large and hydrophobic. This is likely to be due to the presence of the active site tryptophan which can be seen interacting with bulky lesions in crystal structures. Further indirect evidence for this was provided by our results with the *T.thermophilus* ATL homologue TTHA1564 which has alanine replacing tryptophan in the putative active site and lacks the ability to discriminate between large and small O⁶-alkyl groups, whilst conserving the ability to recognise ODNs containing O⁶-alkylguanine residues with high affinity compared to the control. However, no structures exist of TTHA1564 and so it is hard to draw any direct conclusions about the mechanism by which this protein recognises damaged bases.

Finally, we conducted mutagenesis studies with AtI1 to provide further evidence of the role of Arg69 in base recognition. The ability of the AtI1 R69A mutant to recognise O⁶-alkylguanine residues was profoundly disturbed, and in addition this mutant showed very little ability to discriminate between guanine and O⁶-alkylguanines. However, both these abilities were partially restored in the R69F mutant. Interestingly, comparison of the sequence of TTHA1564 with other ATL proteins indicates that the residue corresponding to Arg69 in AtI1 would be Phe130, and in fact TTHA1564 shows a similar ability to discriminate between guanine and O⁶-methylguanine as the AtI1 R69F mutant protein. This is an interesting finding and warrants further investigation.

8.2 Affinity-based Isolation of Atl1 and Interacting Proteins

A series of pull-down assays were performed using modified ODNs containing O^6 -alkylguanine residues as bait, in an attempt to affinity purify and isolate Atl in complex with any interacting proteins from whole-cell extracts of *S.pombe*. These experiments had been previously attempted by the Margison laboratory using short ODNs (23-mers) but were unsuccessful, with no proteins at all being identified by the mass spectrometry-based analysis. Therefore, the experiments were modified and performed with longer ODNs (219-mers) containing O^6 -methylguanine and guanine. Atl1 was successfully isolated from the wild type *S.pombe* whole-cell extract using ODNs containing O^6 -methylguanine, and subsequently identified using MS/MS experiments and peptide matching. It was also demonstrated that Atl1 was not isolated from the whole-cell extract by the control ODN containing guanine by the same method. However, both general searches and targeted peptide searches (pSRM) failed to identify any interacting partners of Atl1 from the NER machinery or any other DNA repair system. Rad15 (the yeast homologue of human XPD, an NER helicase which is part of the TFIIH core) was identified but was present in both samples (i.e. the pull-downs with ODN containing O^6 -MeG and ODN containing G) and therefore would appear to have affinity for the DNA and not specifically for Atl1. The vast number of abundant proteins that were identified by MS methods appeared to be providing an unwanted background and so a second attempt at the pull-down assays was performed to rectify this. Shorter 100-mer ODNs were used in an attempt to reduce non-specific binding by abundant proteins to the DNA itself, and in addition ODNs containing O^6 -benzylguanine, O^6 -methylguanine and guanine were used, as

the interaction between AtI1 and ODN containing O^6 -BnG is of higher affinity than for O^6 -MeG. Furthermore, a more extensive washing stage was included to try and remove more of the unwanted proteins. However, the MS-based analysis of these samples was inconclusive (AtI1 could not be identified in any of the samples) and due to time constraints we could not optimise the pull-down assays or MS/MS experiments any further.

It would be worthwhile to continue with these experiments, as identification of the interacting partners would be expected to provide meaningful information about its cellular role. However, no-one has managed to glean any information thus far using these methods. Given more time, further analysis of the samples with more extensive targeted peptide searches (pSRM) may be of use. Unfortunately, it may be that the interactions between AtI1 and other proteins, if they exist at all (as it is quite possible that any recruited repair factors interact directly with the distorted DNA rather than AtI1 *per se*) are too weak or transient to be observed, certainly by *in vitro* methods such as were used here. It is still unclear whether our failure to produce results is because of a lack of detection of existing interactions, or whether they simply do not take place to be observed.

8.3 Attempted Synthesis of a Crosslinker

In the early stages of the project, attempts were made to synthesise a modified guanine or purine base that would be capable of reacting with a cysteine residue in the active site of the AtI1 W56C mutant protein to generate a covalent crosslink. Two ways to achieve this were envisaged: by incorporation of a disulfide-containing group, or a reactive maleimide moiety,

at the O⁶-position of the base. Unfortunately neither route was successful and so this part of the project was abandoned to concentrate on the promising results we were acquiring from the binding studies with ATL proteins. In hindsight, the value of a crosslinked AtI1-DNA complex is questionable in the context of the pull-down experiments: AtI1 was able to be isolated using an ODN containing O⁶-MeG by exploiting the tight binding affinity of the protein for the substrate, and so it is unlikely the use of a crosslinked AtI1-DNA complex would have improved these experiments. Also, some more recent results from the Margison laboratory (unpublished data) demonstrated that the AtI1 W56C mutant binds to lesion-containing ODNs with far less affinity than wild-type protein and therefore it may have been difficult to form the covalent bond, even if an ODN containing a suitably modified base had been successfully prepared.

8.4 Fluorescence-based Assays

8.4.1 MGMT Activity Assay

Radioactive assays have been previously used to measure the effect of various inhibitors on the alkyltransferase activity of the human AGT homologue MGMT.(68,141) It was our intention to develop an equivalent non-radioactive, fluorescence-based assay using molecular beacon ODNs. The assay involves using a hairpin ODN containing O⁶-MeG at a position in the sequence where it is blocking a PstI restriction site and thus prevents cleavage by the endonuclease. When MGMT demethylates the ODN, it unblocks the site and allows cleavage by PstI. The hairpin ODN has a dye-quencher pair at the termini of the ODN and thus the fluorescence will

increase as the dye becomes estranged from the quencher after cleavage. If the activity of MGMT used in the assay is consistent and known, then pre-incubation of MGMT with an inhibitor will decrease the activity of MGMT in the assay and therefore cause a more modest change in fluorescence after treatment with PstI. It was hoped that this change would be sensitive enough to accurately quantify the inhibitory effects on MGMT of various molecules and ODNs.

The preliminary findings were encouraging, as it was demonstrated that after treatment of ODN containing O^6 -MeG with MGMT and PstI the fluorescence significantly increased, whereas with PstI treatment only there was no significant change in fluorescence. However, due to time restrictions we were unable to optimise the assay. It is conceivable that with some more effort this assay can be made to work effectively, and it would certainly be worthwhile to do so.

8.4.2 FRET Analysis of ODN Hybridisation

In order to show that the 13-mer ODNs used in the ATL protein binding assays were duplexes at 1nM concentration, 50mM NaCl and 25°C, a series of FRET experiments were performed. Two complementary ODNs, one containing a 5'-FAM and the other a 5'-HEX label were annealed, and then cooled and diluted to 1nM under the same conditions as the binding assays. It was shown that the fluorescent signal of FAM was quenched and the signal of HEX enhanced when the duplex was annealed due to FRET occurring between the dye molecules in close proximity, and that this effect did not occur for non-complementary sequences containing the same dyes or for

complementary ODNs that had not been annealed together. Thus it was demonstrated that the ODNs had formed a duplex under these conditions.

However, to be absolutely certain of duplex formation and measure an accurate T_m , we attempted to perform a fluorescence-based DNA melt by monitoring the change in fluorescence of FAM which would be expected to increase as the strands dissociate during denaturation and the FRET-based quenching by the proximal HEX label ceases to occur. However, for some reason there appeared to be no change in signal whilst heating the annealed ODN duplex from 10°C to 70°C. The reasons for this are unclear given the previous results. The fluorimeter used for this experiment was not as sensitive as the Horiba Jobin-Yvon Fluoro-Max3 which may be an issue at such low dye concentrations. It may be that the DNA melting experiment would be more successful if a dye-quencher pair (instead of a FRET pair) was used as the change in fluorescence would be much greater. A fluorescence-based method of DNA melting to measure T_m has been successfully demonstrated by Morrison *et al.* where two complementary ODNs, one with a 5'-FAM label and the other containing a 3'-rhodamine quencher were used.(142)

8.5 Synthesis and Modification of ODNs

The work of Shibata, Millington and Williams was extended upon by using a synthetic ODN precursor containing 2-amino-6-methylsulfonyl-purine to prepare a wide variety of novel O^6 -alkylguanine and related purine bases in DNA. Whilst it was demonstrated that this post-synthesis modification chemistry was highly versatile when using primary alcohols and amines, it became somewhat more complicated when trying to use secondary and

tertiary alcohols, as well as extremely large and hindered ones. Whether the method can be adapted and improved to resolve these problems is open to question, though this may be a useful pursuit in the future. In addition, it was shown that O^6 -alkylpurine analogues could be prepared using a synthetic ODN precursor containing 6-chloropurine by displacement with alcohols.

It was also demonstrated that ODNs longer than those produced by standard phosphoramidite synthesis (i.e. 219-mers) were able to be prepared by annealing followed by primer extension with a Klenow fragment. In addition, double stranded DNA substrates for ATL proteins were successfully produced by annealing of short phosphorylated ODNs to a full-length complementary template strand, followed by treatment with DNA ligase to seal the nicks.

9.0 Experimental

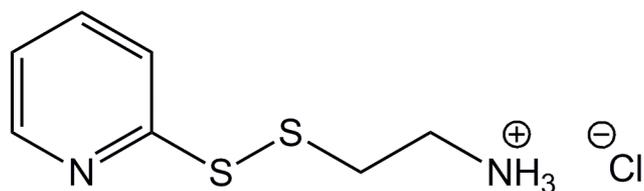
9.1 Chemical Synthesis

All reagents were obtained from commercial suppliers and used without further purification.

Column chromatography was performed using silica gel for flash column chromatography (particle size 30-70 μm). Thin layer chromatography (TLC) was performed on pre-coated Merck Keiselgel 60 F254 aluminium-backed plates.

NMR spectra were recorded on a Bruker AC250 spectrometer and chemical shifts are reported in δ values relative to tetramethylsilane as an external standard. ^1H NMR spectra were recorded at 250 MHz and ^{13}C NMR at 60 MHz.

All mass spectrometry was performed by the University of Sheffield Department of Chemistry Mass Spectroscopy Service.

S-(2-Pyridylthio)cysteamine Hydrochloride (2)

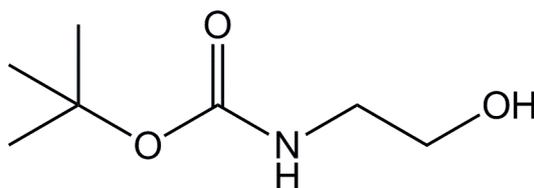
Prepared according to the literature procedure (143)

2-Pyridyl disulfide (4.0 g, 18.1 mmol) was dissolved in methanol (18 mL) and acetic acid (0.7 mL). A solution of cysteamine hydrochloride (1.03 g, 9.1 mmol) in methanol (9 mL) was then added dropwise over 30 min. The reaction mixture was stirred under argon at room temp. for 48 h, then evaporated to give a yellow oil. The oil was triturated with diethyl ether (50 mL) and redissolved in methanol (10 mL). The product was precipitated by the addition of diethyl ether (50 mL) and filtered to give a white crystalline solid (1.32 g, 65%)

TLC: R_f (NH₃OH: 95% ethanol 1:99 v/v) = 0.2

¹H NMR δ (CDCl₃): 2.95 (t, 2H), 3.2 (t, 2H), 7.15 (t, 1H), 7.55 (d, 1H), 7.65 (t, 1H), 8.3 (d, 1H) ppm

Mass Spec (ESI -ve): [M-H]⁻ = 222

1-Boc-aminoethan-2-ol (4)

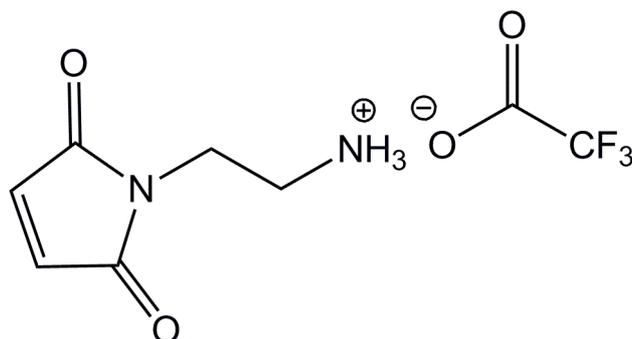
Prepared according to the literature procedure (144)

To a stirred solution of 2-aminoethanol (4 mL, 66.2 mmol) in dichloromethane (200 mL) was added triethylamine (9.28 mL, 66.2 mmol) and di-tert-butyl dicarbonate (14.4 g, 66.2 mmol). The reaction mixture was stirred at room temperature for 18 h, after which it was washed with saturated NaCl solution, and 1M aqueous HCl, then dried (MgSO_4) and evaporated to give a colourless oil (5.56g, 53%).

TLC: R_f (Petroleum ether: ethyl acetate 1:1 v/v) = 0.24

^1H NMR δ (CDCl_3): 1.45 (s, 9H), 3.25 (t, 2H), 3.7 (t, 2H), 5.1 (br s, 1H) ppm

Mass spec (ESI +ve): $[\text{M}+\text{Na}]^+ = 184$

***N*-(aminoethyl)maleimide hydrotrifluoroacetate (5)**

Prepared according to the literature procedure (145)

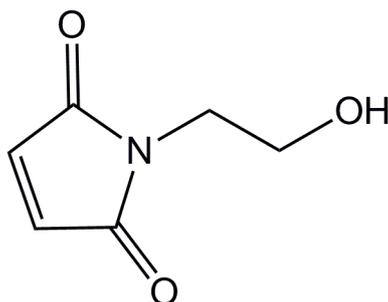
To a dry round-bottomed flask was added maleimide (3.31 g, 34 mmol), triphenylphosphine (8.77 g, 33.4 mmol) and dry THF (160 mL). *N*-(*tert*-butoxycarbonyl)ethanolamine (5 g, 31 mmol) and diisopropylazodicarboxylate (7.3 mL, 37.1 mmol) were then added, the mixture stirred under nitrogen at room temperature overnight, then evaporated. The crude product was filtered through a plug of silica gel using a 2:1 mixture of hexane: ethyl acetate as the eluent. The crude product was dissolved in 100 mL of a 60:35:5 mixture of dichloromethane: trifluoroacetic acid: water and stirred at room temperature for 2 h. The reaction mixture was then diluted with dichloromethane (50 mL) and water (50 mL), and the organic layer was extracted with water (3 x 25 mL). The combined aqueous layers were further extracted with dichloromethane (3 x 50 mL) and the aqueous layers concentrated on a rotary evaporator to give a yellow oil (6.46 g, 82%).

TLC: R_f (petroleum ether: ethyl acetate 1:1 v/v) = 0.2

^1H NMR δ (DMSO- d_6): 2.85 (m, 2H), 3.55 (t, 2H), 7.05 (s, 2H), 7.9 (br s, 3H)
ppm

^{13}C NMR δ (DMSO- d_6): 21.7, 34.9, 115.8, 135.0, 158.5, 171.8

Mass Spec (ESI +ve): $[\text{M}+\text{H}]^+ = 141$

***N*-(hydroxyethyl)maleimide (6)**

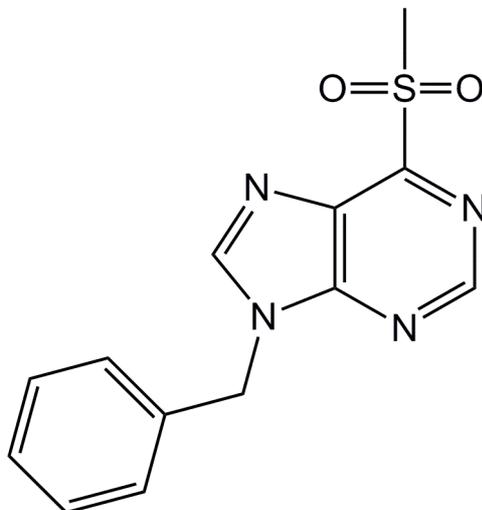
Prepared according to the literature procedure (146)

A stirred solution of 2-aminoethanol (0.6 mL, 10 mmol, 1 eq) in saturated aqueous sodium bicarbonate (50 mL) was cooled to 0°C in an ice bath. *N*-(methoxycarbonyl)maleimide (1.55g, 10 mmol) was then added portionwise with rapid stirring. The resulting solution was stirred at 0 °C for 30 min before the ice bath was removed and the solution allowed to warm to room temp. over 20 min. The aqueous layer was extracted with chloroform (3 x 25mL), the combined organic portions dried over MgSO₄ and evaporated to give a white solid with droplets of clear oil also present. This crude product was purified by flash column chromatography, eluting with dichloromethane-ethyl acetate (3:2 v/v) to give the product as a white solid (0.87g, 62% yield).

TLC: R_f (Dichloromethane: ethyl acetate 3:2 v/v) = 0.29

¹H NMR δ (CDCl₃): 3.42 (t, 2H), 3.48 (t, 2H), 4.8 (t, 1H), 7.02 (s, 2H)

Mass Spec: [M+H]⁺ = 142

***N*⁹-benzyl-(*S*⁶-methylsulfonyl)hypoxanthine (8)**

To a stirring solution of *N*⁹-benzyl-(*S*⁶-methyl)hypoxanthine (0.2 g, 0.8 mmol) in acetonitrile (4 mL) was added a solution of magnesium monoperoxyphthalate hexahydrate (MMPP, 0.99 g, 2 mmol) in water (4 mL). The reaction was stirred for 2 h, after which time more MMPP (0.16 g, 0.25 mmol) was added. After 1 h, the reaction was extracted into dichloromethane (100 mL) and washed with sat. aqueous sodium bicarbonate solution (2 x 20 mL). The organic layer was evaporated to give the product as a white solid (0.189 g, 86% yield)

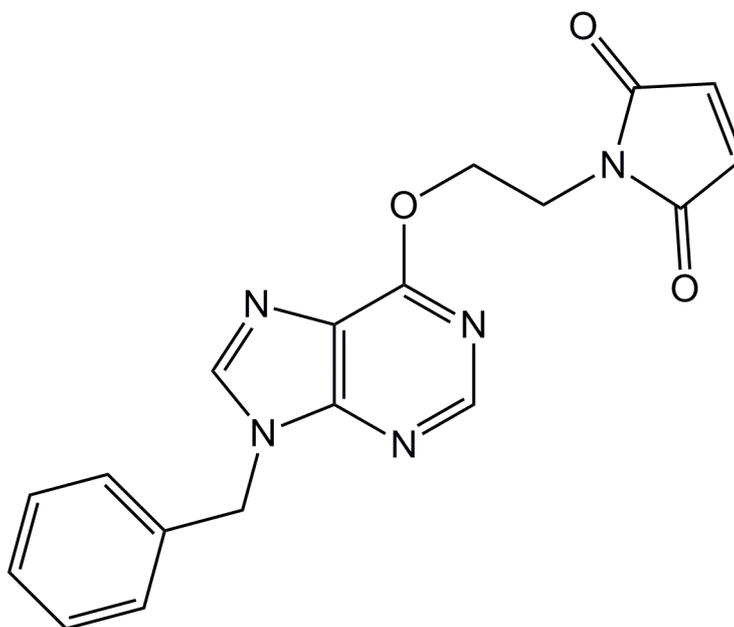
TLC: R_f (Dichloromethane: methanol 19:1 v/v) = 0.2

¹H NMR δ (CDCl₃): 3.5 (s, 3H), 5.52 (s, 2H), 7.35 (m, 5H), 8.38 (s, 1H), 9.1 (s, 1H)

Mass Spec: [M+H]⁺ = 289

This reaction gives a product mixture of sulfone (**8**) and sulfoxide (**9**). The sulfoxide has $[M+H] = 273$. Complete oxidation to completion to give exclusively sulfone product is very difficult (section 7.3).

The starting material *N*⁹-benzyl-(*S*⁶-methyl)hypoxanthine was synthesised by Hend Ismail, University of Sheffield.

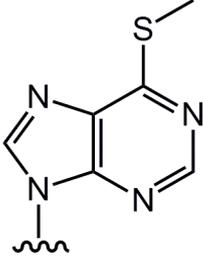
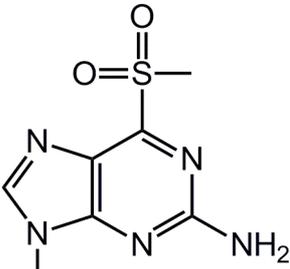
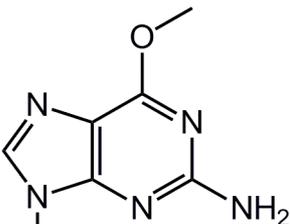
***N*⁹-benzyl-*O*⁶-(ethylmaleimide)hypoxanthine (10)**

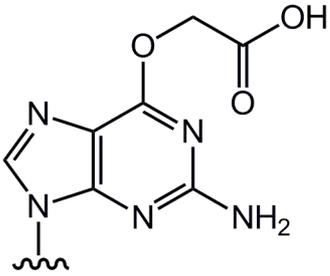
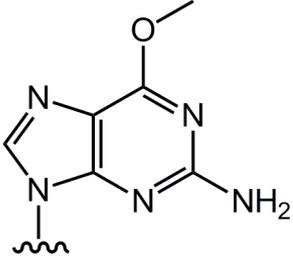
To a stirring solution of *N*⁹-benzyl-*O*⁶-(ethylmaleimide)hypoxanthine (compound **8**, 66mg, 0.23 mmol) in 0.5 mL acetonitrile was added 45 μ L 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU, 135 μ L, 0.69 mmol) and *N*-(hydroxyethyl)maleimide (compound **6**, 49mg, 0.35 mmol). The reaction was left stirring for 18 h after which it was extracted with 30 mL CH₂Cl₂ and washed with saturated NaHCO₃ solution (2 x 10 mL). The crude product was purified by column chromatography using CH₂Cl₂: MeOH (19:1) as the eluent.

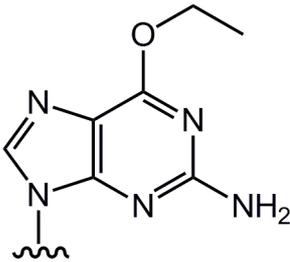
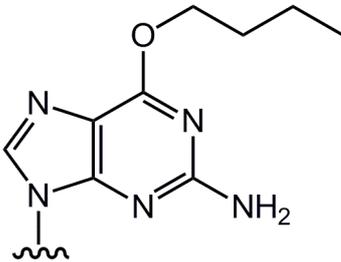
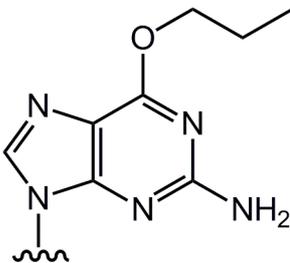
It is uncertain whether this synthesis was successful (section 7.3). Another attempt was made exactly as above but using *N,N*-Diisopropylethylamine (DIPEA) (120 μ L, 0.69mmol) instead of DBU. A final attempt was made exactly as above but using DBU (67 μ L, 0.35mmol) and 1,4-diazabicyclo[2.2.2]octane (DABCO) (52mg, 0.46mmol).

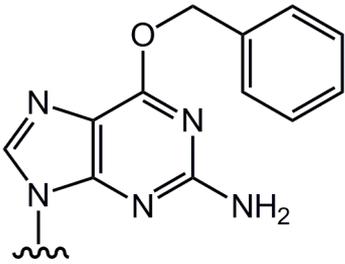
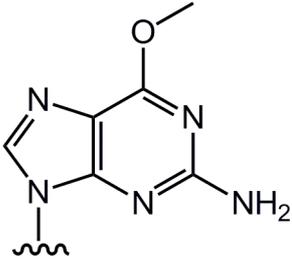
9.2 Synthesis of ODNs

Oligonucleotides (ODNs) were synthesised on an Applied Biosystems 394 automated synthesiser containing 4 column ports and 8 phosphoramidite ports. For the coupling steps, 0.1 M solutions of unmodified phosphoramidites and 0.15 M solutions of modified phosphoramidites in dry acetonitrile were used. The size of column used was always 1 μ mol. Standard deprotection phosphoramidites (benzoyl-dA, benzoyl-dC and isobutyryl-dG) and columns were purchased from Cambio, while the ancillary reagents and the mild/fast deprotection phosphoramidites/columns (phenoxyacetyl-dA and dG, acetyl-dC) were purchased from Glen Research. Fluorescently labelled phosphoramidites (FAM, HEX, Cy3 etc.) were also purchased from Glen Research. The SIMA-labelled ODNs including those containing 2-amino-6-methylsulfonyl-purine were made by ATB Biotech, and DNA Technology (Denmark). All ODNs which contained modified nucleotides were synthesised using mild/fast deprotection phosphoramidites for the natural bases of the sequence. The oligomers that were not 5'-labelled with a fluorophore were left with the 5'-dimethoxytrityl (DMT) group on for ease of purification unless otherwise stated. Standard cleavage of ODNs from the column was performed using concentrated aqueous ammonia (33%) on the DNA synthesiser. ODNs with standard deprotection groups were incubated overnight at 65°C, while ODNs prepared with fast/mild deprotection reagents were treated overnight at room temperature. Some modified ODNs were left on the column in order that they could be used for post-DNA synthesis chemistry. The procedures associated with these are detailed in section 9.3.

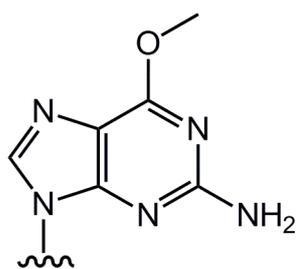
ODN name	Sequence
OW1	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTT X TTTTTTTTTTTTTTTTTTTTTTTTTTTTT
OW2	5'- <u>Biotin</u> TTTTTTTTTTTTTTTTTTTTTTTTTTTTT X TTTTTTTTTTTTTTTTTTTTTTTTTTTTT
	 <p>where X= </p>
OW10	5'-GCA TCA GCC ATG X CT AGT ACG G- <u>FAM</u> -3'
OW22	5'-GCC ATG X CTA GTA
	 <p>where X= </p>
OW6	5'-TAC TAG C CAT GGC
OW10MeG	5'-GCA TCA GCC ATG X CT AGT ACG G- <u>FAM</u> -3'
	 <p>where X= </p>

ODN name	Sequence
OW11–21 OW27 OW40 OW41 OW49 OW50 OW51	5'- SIMA -GCC ATG X CTA GTA
	where X = see figure 9.1 (pgs 253-254)
OW28	5'-TAC TAG T CAT GGC
OW29	5'-CGC G X ATT TGC G
OW30	5'-CGC X AAT TCG CG
	<div style="text-align: center;">  <p>where X=</p> </div>
OW31	5'- Cy3 -GGT CT X CAG AAA TTG ATT TCT GCA GAC C- BHQ2
	<div style="text-align: center;">  <p>where X=</p> </div>

ODN name	Sequence
OW39	5'- Cy3 -GGT CT G CAG AAA TTG ATT TCT GCA GAC C- BHQ2
OW33	5'-GAA CT X CAG CTC CGT GCT GGC CC
	 <p>where X=</p>
OW34	5'-GAA CT X CAG CTC CGT GCT GGC CC
	 <p>where X=</p>
OW43	5'-GAA CT X CAG CTC CGT GCT GGC CC
	 <p>where X=</p>
OW44	5'-GAA CT X CAG CTC CGT GCT GGC CC

ODN name	Sequence
OW36	5'- PO₄ -GAA CT X CAG CTC CGT GCT GGC CC
	 <p>where X=</p>
OW52	5'- HEX -GCC ATG GCT AGT A
OW53	5'- HEX - GCC ATG XCT AGT A
OW54	5'- HEX - GAA CT G CAG CTC CGT GCT GGC CC
OW55	5'- HEX - GAA CT X CAG CTC CGT GCT GGC CC
OW56	5'- FAM - GCC ATG XCT AGT A
OW57	5'- Cy3 -TAC TAG CCA TGG C
OW58	5'- HEX -TAC TAG CCA TGG C
	 <p>where X=</p>

ODN name	Sequence
OW70	5'- Biotin - CTT GAA TTC GGA AGC GTA ACT GGG AGT GAT TTC CCG GGG GCC AGC ACG GAG CTG CAG TTC CTC GTC TAC ATG CTT ATG CAG TCA TAC CTA ACT GGA TCC TGG
OW71	5'-CCA GGA TCC AGT TAG GTA TGA CTG CAT AAG CAT GTA GAC GAG
OW72	5'- PO₄ -C CGG GAA ATC ACT CCC AGT TAC GCT TCC GAA TTC AAG
OW73	5'- CCA GGA TCC AGT TAG GTA TGA CTG CAT AAG CAT GTA GAC GAG GAA CT G CAG CTC CGT GCT GGC CCC CGG GAA ATC ACT CCC AGT TAC GCT TCC GAA TTC AAG
OW74	5'- CCA GGA TCC AGT TAG GTA TGA CTG CAT AAG CAT GTA GAC GAG GAA CT X CAG CTC CGT GCT GGC CCC CGG GAA ATC ACT CCC AGT TAC GCT TCC GAA TTC AAG
OW75	5'-CTG GAT CCA TGG CAC ATC TCT AGT GTC GAC ATA CAC ATC GAC AAC CTG GGA GTG ACT CAA CAA GTG CAA TGG TGT TCC AGG TAC AAG CCA AGG CGC CTT CTC GAA TAG CAC TCA CT G CAG
OW76	5'-CTG GAT CCA TGG CAC ATC TCT AGT GTC GAC ATA CAC ATC GAC AAC CTG GGA GTG ACT CAA CAA GTG CAA TGG TGT TCC AGG TAC AAG CCA AGG CGC CTT CTC GAA TAG CAC TCA CT X CAG
OW77	5'-GG- Biotin dT - TCG GAT CCA CCG CGT CGA CAA TAG TTC GAC CAT AAG AAT ACA A GA TTG ATA GAC TCT CTA TCC ATC AGT CGA GAG ACC AAC CAA ATT CGG CGC CGG CTG CAG TGA GTG CTA TTC GAG
OW78	5'-GAA CT X CAG CTC CGT GCT GGC CC

**OW79**5'- Biotin-GGG CCA GCA CGG AGC TGC AGT TC**Table 9.1:** Code names and sequences of ODNs synthesised, modified and described in this report

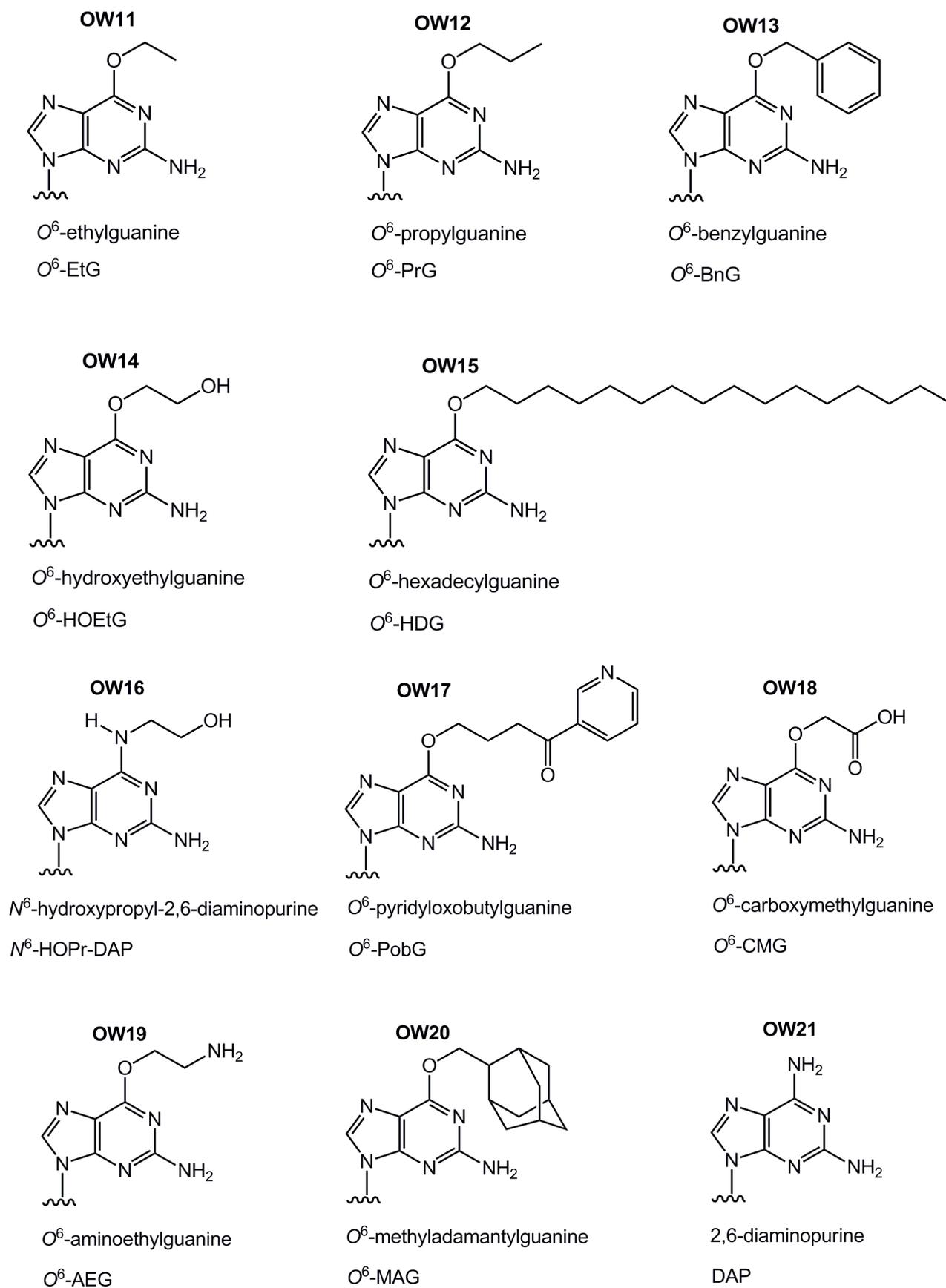


Figure 9.1 (continues): Modified bases incorporated into ODNs (see table 6.1)

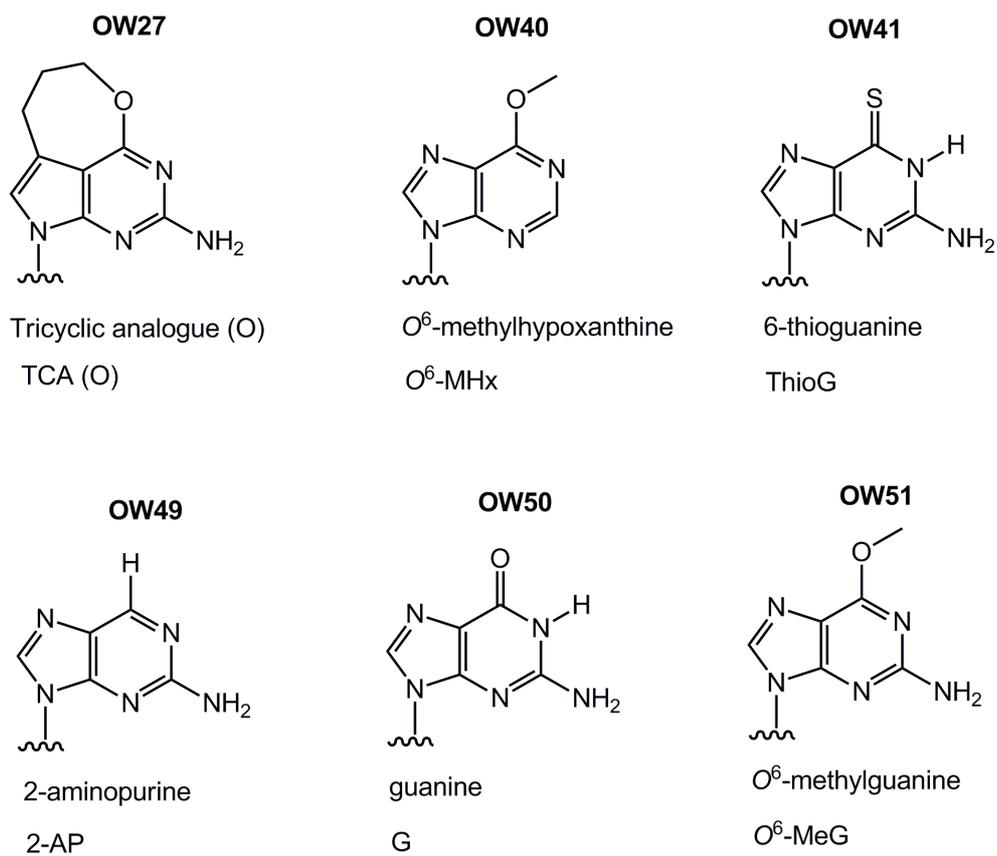


Figure 6.1: Modified bases incorporated into ODNs (see table 6.1)

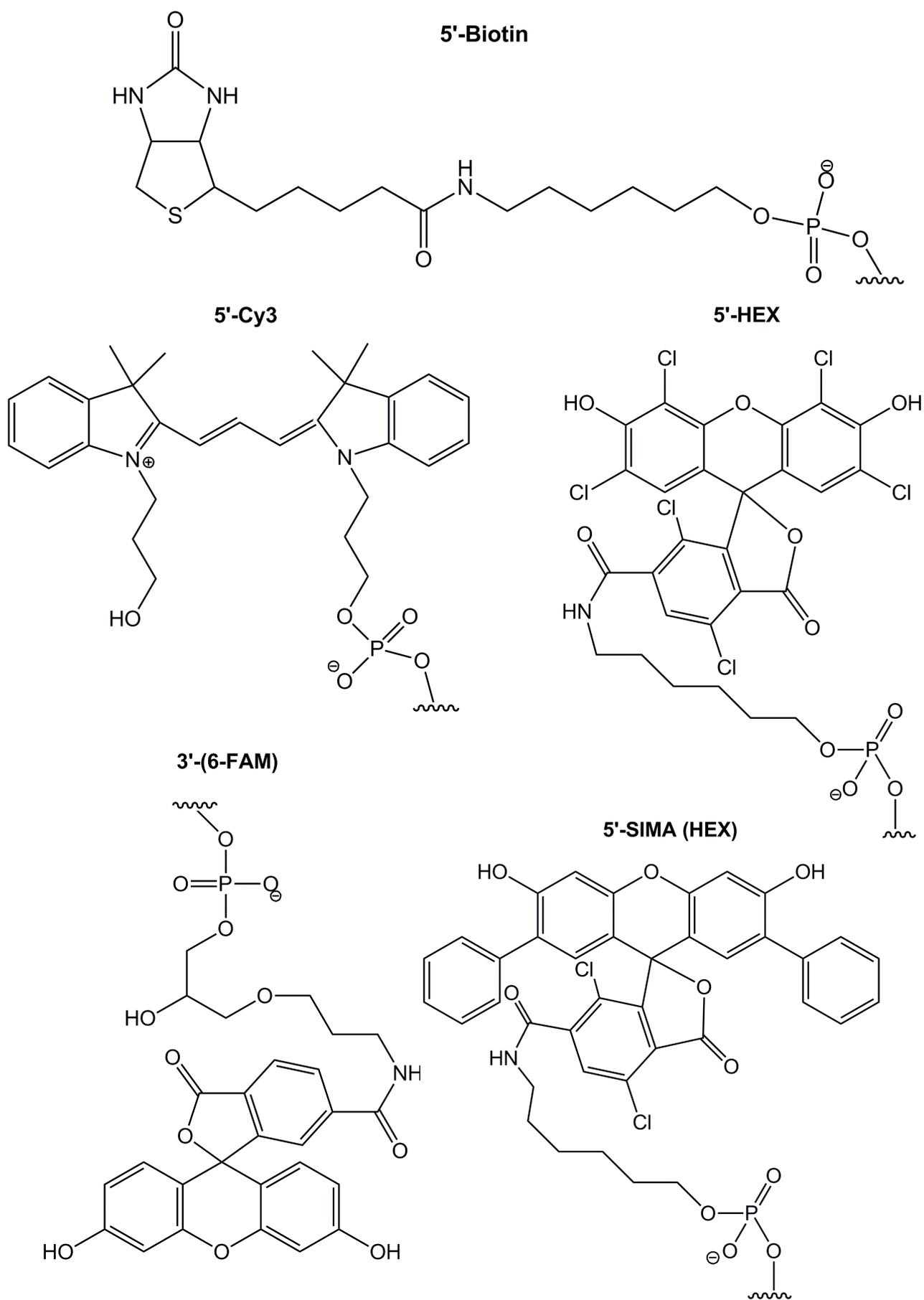


Figure 6.2: Reporter groups used to labelled ODNs

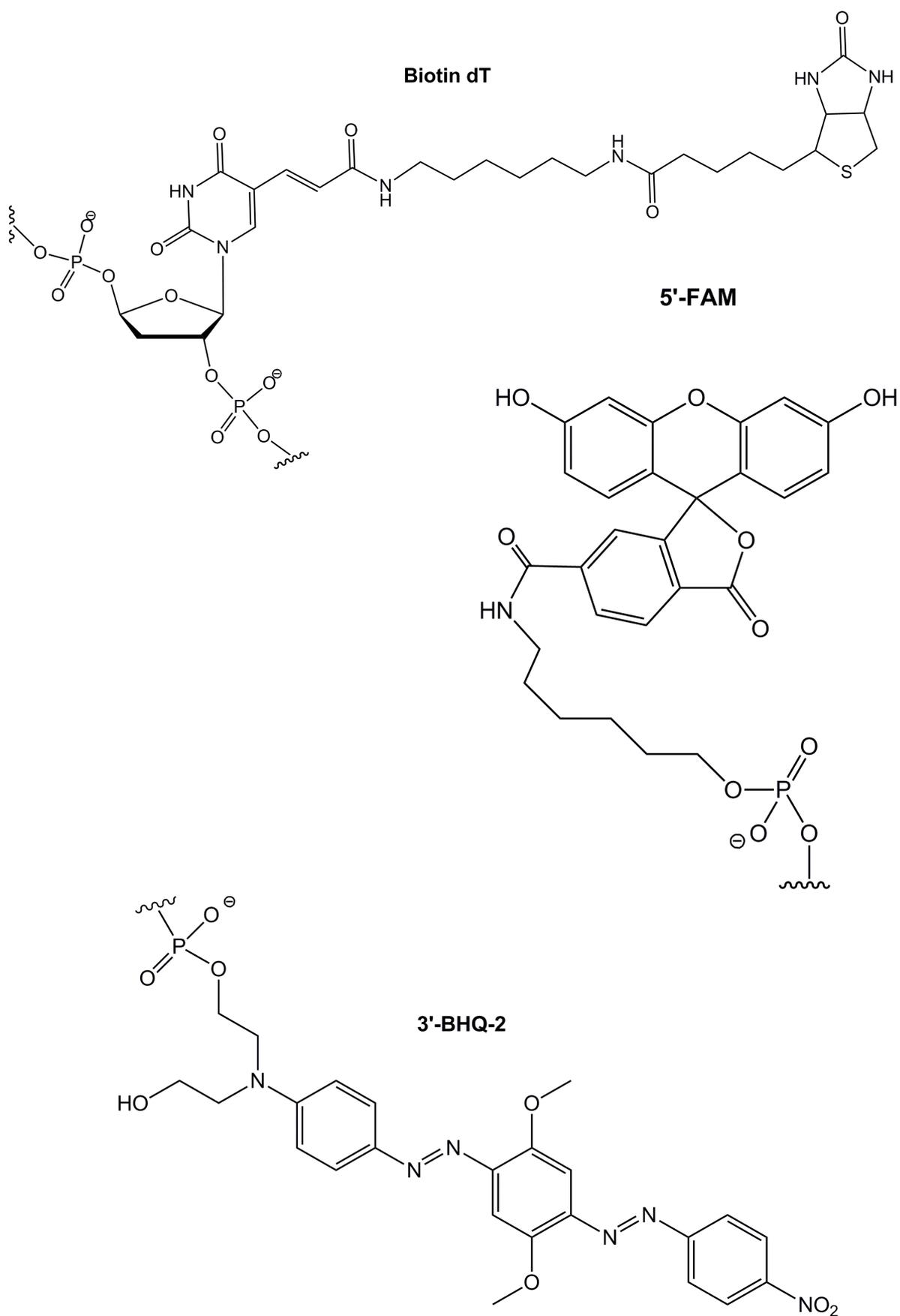
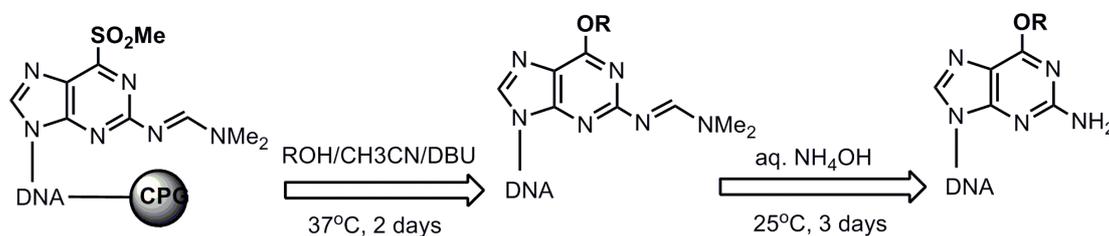


Figure 9.2 (continued): Reporter groups used to labelled ODNs

9.3 Post-DNA Synthesis Modification Chemistry – Chemical synthesis of ODNs containing O⁶-alkylguanine and related modified bases.

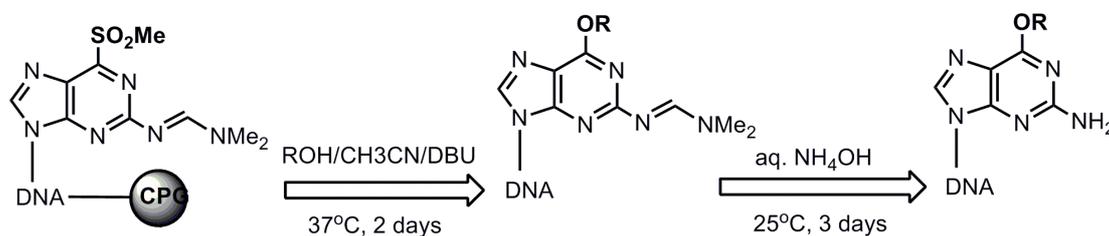
Post-DNA synthesis displacement chemistry using ODNs containing the reactive base 2-amino-6-methylsulfonyl-purine was used to introduce a variety of alkyl adducts onto the O⁶-position of guanine. Five procedures are described in this report, the first two of which are general protocols that were used depending on whether the alcohol was in solid or liquid form. The third procedure is that for the specific reaction to produce O⁶-carboxymethylated guanine-containing ODNs, which requires an additional step. The fourth procedure describes the displacement reaction with ammonia gas to prepare 2,6-diaminopurine and the fifth the displacement reaction of an ODN containing 6-chloropurine using methanol. The described reactions are suitable for use with both unlabelled and fluorescently labelled (SIMA-HEX) ODNs. When the displacement chemistry was undertaken with ODNs containing a fluorescent label these reactions were performed in the dark to preserve the integrity of the dye.

9.3.1 Synthesis of O⁶-alkylguanine containing DNA: displacement with liquid alcohols



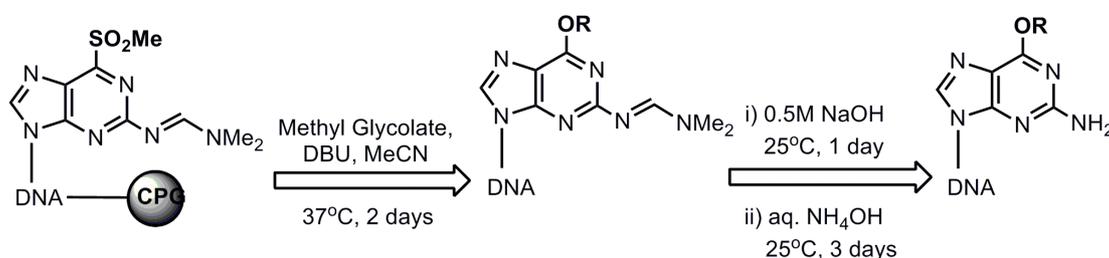
Prepared with modification to the literature procedure.(68) To a 1.5 ml eppendorf tube was generally added a third of a 1 μ mol DNA synthesis column's material (CPG). This was then treated with 200 μ L of dry MeCN/the appropriate alcohol/DBU in the ratio 9:9:2 (v/v/v) and in that order. If a whole column was used (1 μ mol) then a total volume of 500 μ L of the same solution was used. The eppendorf was then flushed with nitrogen and sealed and shaken for two days at 37°C. After this stage, 0.5 mL of concentrated aqueous ammonia (33%) was added and the reaction shaken at room temperature for a further three days. The mixture was then evaporated with a speed vac. concentrator. 150 μ L of water and 500 μ L of diethyl ether were added to the eppendorf followed by vigorous shaking. The organic layer was removed with a pipette and this step was repeated twice more. After the extraction, the beads were washed three times with 200 μ L of water and the aqueous layers combined. This crude product was then purified by RP-HPLC (see section 9.4).

9.3.2 Synthesis of O^6 -alkylguanine containing DNA: displacement with solid alcohols



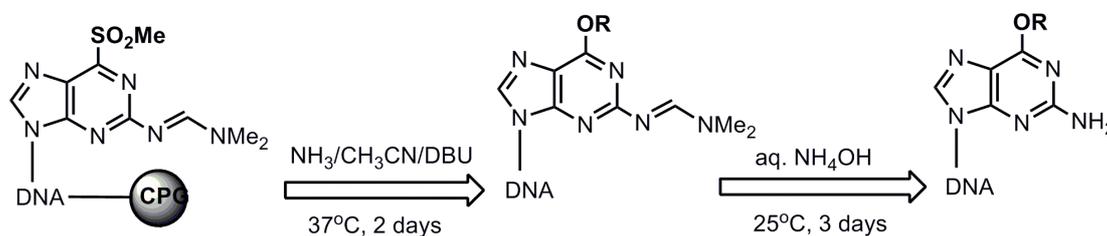
Prepared according to the same procedure used for liquid alcohols (section 9.3.1) but instead using $180\mu\text{L}$ of a 5M solution of the solid alcohol in dry MeCN, and $20\mu\text{L}$ DBU, in the displacement step.

9.3.3 Synthesis of O^6 -alkylguanine containing DNA: displacement with methyl glycolate



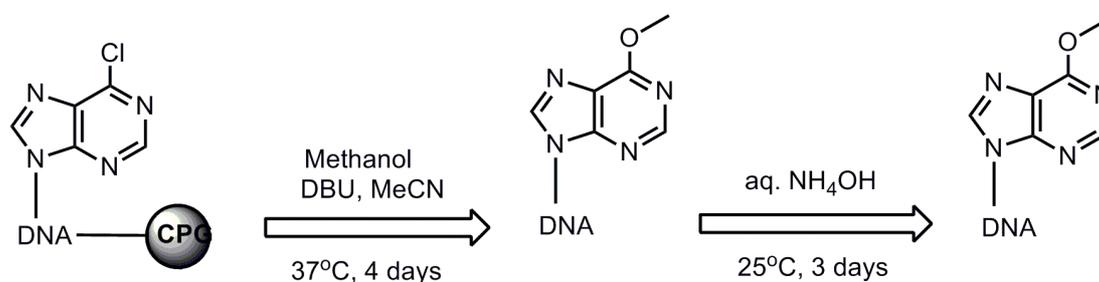
Prepared according to the same procedure used for liquid alcohols (section 9.3.1) but after incubation for 2 days at 37°C (i.e. the displacement step) was added $200\mu\text{L}$ of 0.5M NaOH and the mixture shaken for 1 day at room temp. After this stage was added 0.5mL conc. aq. ammonia solution. The procedure then continues as described in section 9.3.1.

9.3.4 Synthesis of 2,6-diaminopurine-containing DNA: displacement with ammonia gas



Prepared according to the same procedure used for liquid alcohols (section 9.3.1) but instead using 200 μ L of dry MeCN through which ammonia gas had been bubbled through for 1h in the displacement step.

9.3.5 Synthesis of *O*⁶-methylhypoxanthine containing DNA: displacement of 6-chloropurine with methanol



To a 1.5 ml eppendorf tube was added a half of a 1 μ mol DNA synthesis column's material (CPG). This was then treated with 180 μ L of a 5M solution of methanol in dry MeCN and 20 μ L DBU in that order. The eppendorf was then flushed with nitrogen and sealed and shaken for four days at 37°C. After this stage, 0.5 mL of concentrated aqueous ammonia (33%) was added and the reaction shaken at room temperature for a further three days. The mixture

was then evaporated with a speed vac. concentrator. 150 μL of water and 500 μL of diethyl ether were added to the eppendorf followed by vigorous shaking. The organic layer was removed with a pipette and this step was repeated twice more. After the extraction, the beads were washed three times with 200 μL of water and the aqueous layers combined. The sample was then desalted using a NAP-10 column. This crude product was then purified by RP-HPLC (see section 9.4).

(The chloropurine phosphoramidite is available is available from ChemGenes)

9.4 Purification of ODNs

All samples were purified by RP-HPLC using a Gilson Preparative HPLC system equipped with a 500 μL injection loop. The system was used at a flow rate of 1 mL min^{-1} with a reverse phase column ACE-5 C_{18} (250 x 4.6 mm) from Hichrom. The UV detector was set to a wavelength of 260 nm. The following buffers were used in the purification:

For unmodified DNA:

Buffer A = 100 mM triethylammonium acetate, 5% MeCN, pH 6.5

Buffer B = 100 mM triethylammonium acetate, 65% MeCN, pH 6.5

For modified DNA:

Buffer A = 100 mM triethylammonium bicarbonate, 5% MeCN, pH 7.5

Buffer B = MeCN

The ODNs were purified over a 30 min period with an increasing percentage of buffer B as an eluting solvent with buffer A. The gradient of elution used was optimised for each oligomer, with the retention time ideally being around 20 min for maximum separation from impurities. After 30 min, the column was cleaned by ramping to 100% buffer B over 5 min, back to 100% buffer A over 5 min, and finally re-equilibrating in buffer A for 5 min.

Usually around 200 μL of DNA solution was injected per cycle.

The specific RP-HPLC conditions and retention times are shown in table 9.3. ODNs that were made or modified and subsequently purified in Sheffield are shown (i.e. ODNs purchased from commercial suppliers are

generally not shown unless they were re-purified here). In the case where the ODN was purified with DMT-ON, the dimethoxytrityl group was removed by treatment with 1 mL of aq. AcOH (20%) for 1h at room temp. and then the sample was evaporated to dryness. The ODN was then redissolved in water and desalted using a NAP-10 gel filtration column. In the case of fluorescently labelled ODNs, the purifications were performed in the dark as far as was possible (e.g. using tin-foil to minimise light exposure etc.).

After RP-HPLC purification, the TEAB salts were removed under vacuum. The final concentration of the samples was determined by measuring the UV absorbance at 260 nm and using the Beer-Lambert equation:

$$A_{260} = \epsilon_{260} \cdot c \cdot l$$

where A_{260} = absorbance at 260 nm

ϵ_{260} = extinction coefficient at 260 nm ($M^{-1}cm^{-1}$)

c = concentration (M)

l = path length (cm)

The ϵ_{260} values used were as follows (units $\mu M^{-1} cm^{-1}$):

dA = 15400

dG = 11700

dC = 7300

dT = 8800

modified dG = 11700

ODN name	RP-HPLC conditions	Retention time (min)
OW1	0-50% B, 30 min	20.4 (DMT-ON)
OW2	0-50% B, 30 min	23.3 (DMT-ON)
OW6	0-50% B, 30 min	21.8 (DMT-ON)
OW10 MeG	0-20% B, 30 min	18.7
OW11	0-40% B, 30 min	21.5
OW12	0-40% B, 30 min	21.4
OW13	0-40% B, 30 min	19.7
OW14	0-40% B, 30 min	18.7
OW16	0-40% B, 30 min	20.3
OW17	0-40% B, 30 min	21.5
OW18	0-40% B, 30 min	20.7
OW19	0-40% B, 30 min	20.6
OW20	0-40% B, 30 min	20.2
OW21	0-40% B, 30 min	19.2
OW27	0-40% B, 30 min	20.5
OW28	0-50% B, 30 min	22.4 (DMT-ON)
OW29	0-20% B, 45 min	31.4
OW30	0-20% B, 45 min	31.3
OW33	0-15% B, 30 min	19.9
OW34	0-15% B, 30 min	21.3
OW36	0-15% B, 30 min	20.4
OW40	0-40% B, 30 min	21.6
OW41	0-40% B, 30 min	20.8

ODN name	RP-HPLC conditions	Retention time (min)
OW43	0-15% B, 30 min	20.3
OW44	0-15% B, 30 min	21.3

Table 9.2: RP-HPLC conditions and retention times for the ODNs that were purified and contained in this report

9.5 Characterisation of ODNs

Characterisation of the ODNs was performed by either MALDI-TOF or electrospray (ESI) mass spectrometry (MS). Only ODNs that were synthesised or modified in Sheffield are shown (i.e. ODNs purchased from commercial suppliers are not shown unless they were subsequently analysed by MS). This data is shown in table 9.3

ODN name	Calculated mass (Da)	Determined mass (Da)
OW1	13660	13666
OW2	14065	14072
OW6	3932	3934
OW11	4761	4761
OW12	4775	4775
OW13	4823	4823
OW14	4777	4777
OW16	4790	4790
OW17	4880	4880
OW18	4791	4791

ODN name	Calculated mass (Da)	Determined mass (Da)
OW19	4776	4776
OW20	4881	4881
OW21	4732	4732
OW27	4772	4771
OW28	3950	3950
OW29	3721	3721
OW30	3720	3720
OW31	9665	9666
OW33	7029	7029
OW34	7057	7057
OW36	7172	7172
OW40	4731	4732
OW41	4749	4749
OW43	7044	7044
OW44	7092	7092
OW51	4747	4747
OW50	4733	4733

Table 9.3: Calculated and determined masses for selected ODNs that were synthesised or modified and described in this report

9.6 Preparation of Double-stranded ODNs

9.6.1 Preparation of Long (219-mer) ODNs by Primer Extension

Single-stranded ODNs with short regions of overlapping complementary sequence (see 6.2) were annealed by mixing 25pmols of OW75 (containing G) or OW76 (containing O^6 -MeG) with 25pmoles OW77 (the biotinylated complement) in total volume of 50 μ L dH₂O containing 5 μ L polymerase I Klenow fragment buffer (10X buffer, Roche) and 1 μ L 10 mM dNTP mix and heating in a dry-block for 5 minutes at 95°C, then allowing to cool to room temperature. To fill in the 5'-overhanging regions left after annealing, 0.5 μ L (2.5 units) polymerase I Klenow fragment (Roche) was added and reaction mixture incubated for 15 minutes at 37°C. The double-stranded 219-mer ODN product was analysed by PAGE.

9.6.2 Preparation of 102-mer ODN by Ligation

5000pmols of single-stranded ODN OW70 in dH₂O was heated to 96°C and then cooled at RT for a few mins, after which 5000pmols each of OW71, OW72 and OW36 were added with 50 μ L 0.5M NaCl solution and dH₂O up to a total volume of 500 μ L (so final conc. of dsODN = 10 μ M). The annealing reaction was heated for 5 min at 96°C and then allowed to cool slowly to RT. To this was added 60 μ L DNA ligase 10X buffer and 40 μ L DNA Ligase (Roche, 5U/ μ L) and the ligation reaction was incubated for 16h at 16°C. The double-stranded 102-mer ODN product was analysed by PAGE.

9.6.3 Preparation of dsDNA Substrates for Fluorescence-based Assays

500pmols of the relevant single-stranded 13-mer ODN (e.g. OW51) was mixed with 600pmols of its ODN complement (e.g. OW6). To this was added 10 μ L 5X buffer (250mM Tris-HCl pH7.5, 250mM NaCl, 5mM EDTA)

and dH₂O up to a total volume of 50μL (so final conc. of dsODN = 1μM). This annealing reaction was heated to 80°C and cooled slowly to RT in the heating block for 1h, after which the reaction was mixed and centrifuged for 30s to give the dsODN in buffered solution.

9.6.3 Preparation of dsDNA Substrates for EMSA

5000pmols of ssODN was mixed with 5000pmols of its ODN complement in 50mM NaCl in a final volume of 50μL ([ODN]=10μM). The reactions were heated to 95°C for 5min and allowed to cool slowly to RT

9.7 Other ODN Reactions

9.7.1 ODN Digestion into Nucleosides

2500pmol of ODN was dried to a solid by vacuum concentration and resuspended in 50μL 10mM TrisHCl pH7.5 solution to which was added 1μL snake-venom phosphodiesterase (SVPDE, 10 Units) and the reaction incubated for 1h at 37°C. After this time, 2μL alkaline phosphatase (AP, 10 Units) was added and incubated for 16h at 25°C. The digested ODN was analysed by RP-HPLC and MALDI-TOF MS.

9.7.2 Oxidation of ODN containing S⁶-methylpurine

To a solution of ODN (OW1, 2500 pmol) in dH₂O was added 12 equivalents of MMPP (30,000pmol). The reaction was incubated at 25°C and the progress monitored by RP-HPLC (see section 7.3)

9.7.3 Reaction of ODN containing O⁶-aminoethylguanine with *N*-methoxycarbonylmaimide

15nmols ODN OW19 (containing O⁶-aminoethylguanine) was evaporated to dryness and resuspended in 200μL of saturated sodium

bicarbonate solution. This was cooled to 0°C and 2.5mg *N*-methoxycarbonylmaleimide was added (1000-fold excess). The reaction was stirred for 30min at 0°C and then allowed to warm to 25°C over 10min. The product mixture was purified by SEC (NAP-5 column, GE Healthcare) and analysed by RP-HPLC.

9.8 Polyacrylamide Gel Electrophoresis (PAGE)

9.8.1 SDS-PAGE for Protein Analysis

Protein samples were prepared for SDS-PAGE by addition of an equal volume of 2X Laemmli buffer (200 mM Tris-HCl, pH 6.8, 8% (w/v) SDS, 0.4% bromophenol blue (w/v) 40% glycerol (v/v), 400 mM DTT) and heating for 5 min at 100°C. Samples were then loaded onto a two-step SDS-PAGE gel consisting of a 5% stacking gel (total volume 10mL: 6.88mL dH₂O, 1.25mL 1 M Tris, pH6.8, 1.66mL 30% Protogel (37.5:1 acrylamide: bisacrylamide, National Diagnostics), 100µL 10% SDS, 100µL 10% APS, 10µL TEMED) over a 10-15% resolving gel (for 12%, (total volume 10 mL: 4.5mL dH₂O, 1.25mL 1 M Tris, pH8.8, 4mL 30% Protogel (37.5:1 acrylamide: bisacrylamide, National Diagnostics), 100µL 10% SDS, 100µL 10% APS, 10µL TEMED). Electrophoresis was at 200 V for approximately 45 minutes in running buffer (25 mM Tris base, 192 mM glycine, 0.1% SDS). A broad range molecular weight marker (NEB) was used to determine protein size. For Coomassie staining, the resulting gel was immersed in 0.05% (w/v) Coomassie Brilliant Blue R250, 30% (v/v) methanol, 10% (v/v) glacial acetic acid for 1h at 25°C and subsequently de-stained in (30% (v/v) methanol, 10% (v/v) glacial acetic acid).

9.8.2 Non-denaturing PAGE for ODN Analysis

ODN samples were prepared for PAGE by addition of 1/6 volume of 6X DNA non-denaturing loading buffer (10% glycerol, 0.0025% (w/v) xylene cyanate, 0.0025% (w/v) bromothenol blue). Samples were then loaded onto a PAGE gel ((for a 15% gel, total volume 10mL: 4.75mL dH₂O, 1.25mL 10X TAE, 3.75mL 40% Protogel (19:1 acrylamide: bisacrylamide, National Diagnostics), 0.5mL ethylene glycol, 75 μ L 10% APS, 7.5 μ L TEMED). Electrophoresis was at 200 V for approximately 45 minutes in running buffer (1X TAE, 0.05% (v/v) ethylene glycol). The gels were stained with Gel Red (3X solution) at 25°C for 1h and visualised using a UV transilluminator (λ =300nm).

9.8.3 Electrophoretic Mobility Shift Assays (EMSA)

ODNs (50pmol) were mixed with At11 protein (100pmol) in a total volume of 10 μ L 1X EMSA buffer (50mM TrisHCl pH8.3, 50mM NaCl, 1mM EDTA, 3mM DTT, 0.1 mg/mL BSA, 5% glycerol) and then incubated for 15min at 25°C (final [ODN]= 5 μ M, [At11] 10 μ M). The samples were then mixed with 2 μ L 6X DNA non-denaturing loading buffer and analysed by PAGE as described in section 9.8.2.

9.9 Protein Expression and Purification

9.9.1 Expression of Recombinant MBP-At11 Proteins in *E.coli*

At11 and At11 mutants (R69F, R69A and W56C) were expressed as maltose binding protein (MBP) fusion proteins from a pMAL-2c expression vector construct as described by Pearson *et al.*, (57) with minor modifications as follows. A single bacterial (DH5alpha) colony was inoculated into 100ml of

rich medium with glucose (w/v: 1% tryptone, 0.5% yeast extract, 0.5% NaCl, 0.2% glucose, supplemented with 100 µg/ml ampicillin (Sigma)) and incubated at 37°C overnight. Twenty milliliters of this culture were then used to inoculate 4 litres of the same medium. The culture was grown to $OD_{600} = 0.6$ then isopropylthiogalactoside (IPTG) from a 1M stock solution was added to a final concentration of 0.4 mM to induce protein expression. Cells were incubated for 4 hours at 37°C, and then harvested by centrifugation at 2500 x g for 10min.

9.9.2 Purification of AtI1 (Wild-type and Mutants)

Cell pellets were washed with 20mM Tris-HCl (pH8.3), and re-suspended in 20ml of binding buffer (BB: 20mM Tris-HCl (pH8.3), 200mM NaCl, 1mM EDTA). Extracts were prepared by sonication (four 20s pulses) with cooling on ice for 1 min between pulses. The extracts were then centrifuged at 20000 x g for 20min and the supernatants pooled. The protein concentration of the extract was determined by Bradford assay.

Amylose resin (NEB, binding capacity 3mg/ml) was pre-equilibrated with BB. The bacterial extract was diluted to 2.5mg/ml in a total volume of 50ml of BB, applied, washed with BB and eluted in 1ml fractions using BB containing 10mM maltose. The protein concentration of the eluted fractions was then determined by Bradford assay.

MBP-AtI1 fusion protein (40 mg) was cleaved with 0.1%w/v of Factor X_a (1mg/ml, NEB); at room temperature for 2 hours. The efficiency of the reaction was assessed by resolving the cleavage products on a 15% SDS-polyacrylamide gel. The cleavage reaction was then applied to a Superdex 200 column (HiLoad 16/60, Superdex 200 pg; GE Healthcare) that was pre-

equilibrated with 50 mM Tris-HCl (pH 8.3), 100 mM NaCl. The column was eluted at a flow rate of 0.8ml/ min and 1.6ml fractions were collected. Protein elution was monitored by absorption at 215 nm using a flow cell. Pooled AtI1 containing fractions were further purified through amylose columns to remove remaining uncleaved MBP-fusion protein.

Protein concentrations were determined by measuring the absorbance at 280 nm and then using the Beer-Lambert equation:

$$A_{280} = \epsilon_{280} \cdot c \cdot l$$

where A_{280} = absorbance at 280 nm

ϵ_{280} = extinction coefficient at 280 nm ($M^{-1}cm^{-1}$)

c = concentration (M)

l = path length (cm)

AtI1 was stored at $-20^{\circ}C$ in the SEC elution buffer (50 mM Tris-HCl (pH 8.3), 100 mM NaCl) and defrosted immediately prior to use in any assay.

(The mutant proteins were kindly purified by Gail McGown and Mary Thorncroft)

9.9.1 Expression of Recombinant MBP-TTHA1564 Protein in *E.coli*

TTHA1564 was expressed as maltose binding protein (MBP) fusion protein from a pMAL-2c expression vector construct as described by Morita *et al.*,(63) with minor modifications as follows. A single bacterial (Rosetta DE3) colony was inoculated into 10ml of rich medium with glucose (w/v: 1% tryptone, 0.5% yeast extract, 0.5% NaCl, 0.2% glucose, supplemented with 100 μ g/ml ampicillin (Sigma)) and incubated at $37^{\circ}C$ for eight hours. Twenty

milliliters of this culture were then used to inoculate 4 litres of the same medium. The culture was grown overnight (16h) with incubation at 37°C, and then harvested by centrifugation at 2500xg for 10min.

9.9.2 Purification of MBP-TTHA1564

Cell pellets were washed with and re-suspended in 20 ml of buffer I (50mM Tris-HCl (pH 8.0), 1mM EDTA, 1mM 2-mercaptoethanol) and sonicated on ice. The cell lysate was cleared by centrifugation at 30,000g for 20 min and the supernatant was then applied to an Amylose Resin column (15 ml) that had been pre-equilibrated in buffer I. After washing with 10 column volumes of buffer I, MBP-At11 was eluted in 1mL fractions using buffer I containing 10mM maltose. Fractions containing the MBP-TTHA1564 were pooled and applied to a TOYOPEARL-SuperQ column (15 ml) pre-equilibrated with buffer I. The proteins were eluted with a linear gradient of NaCl from 0 to 1.0M in a total volume of 200 ml of buffer I. Solid ammonium sulphate was added to the fractions containing MBP-TTHA1564 protein to a final concentration of 1.0 M. The protein solution was then applied to a TOYOPEARL-Ether 650M column (10 ml) pre-equilibrated with buffer I containing 1.0M ammonium sulphate. Proteins were eluted with a linear gradient of ammonium sulphate from 1.0 to 0M in a total volume of 150 ml of buffer I. Fractions containing the MBP-TTHA1564 were collected and pooled. The collected solution containing MBP-TTHA1564 was dialysed against buffer II [20mM Tris-HCl (pH 8.0), 50mM NaCl, 1mM EDTA, 1mM dithiothreitol (DTT), 5% (v/v) glycerol] at 4°C.

9.10 Site-directed Mutagenesis of AtI1

To introduce the Arg to Phe and Arg to Ala point mutations into the *atI1* gene of pMAL-2c-atI1 (57) vector, the Phusion site-directed mutagenesis kit was used (Finnzymes, NEB). Two 5'-phosphorylated primers atI1-R69F-M (CGTGGCACCATTCTCTAAATTCGATATCTCTGCTGGTG) and atI1-R69A-M (CGTGGCACCATTCTCTAAAGCTGATATCTCTGCTGGTG) were designed so that they would anneal “back to back” on the vector. The atI1-R69F-M primer contained AGA to TTC codon change (marked in bold) to create the *atI1-R69F* mutation, and the atI1-R69A-M primer contained AGA to GCT codon change (marked in bold) to create the *atI1-R69A* mutation. The 5' phosphorylation of the primers allowed blunt end ligation of the linear PCR product to form a circular plasmid. 2 units of Phusion Hot start DNA polymerase was used for PCR amplifications in a total volume of 50 μ L ddH₂O containing 10 μ L 5x Phusion HF buffer, 1 μ L 10 mM dNTP mix (Invitrogen), 50 pmol of each primer, 100 pg of template DNA (pMAL-2c-atI1). Amplification conditions were as follows: 1 cycle for 30 seconds at 98°C, 25 cycles of 10 seconds at 98°C, 30 seconds at 65°C and 3 minutes at 72°C and finally 1 cycle for 10 minutes at 72°C. Ligation of the PCR products was carried out for a minimum of 5 minutes at room temperature in a total volume of 10 μ L containing 5 μ L Quick T4 DNA Ligase buffer, 5 μ L of PCR product and 0.5 μ L Quick T4 DNA Ligase (NEB). The generated pMAL-2c-atI1-R69F and pMAL-2c-atI1-R69A constructs were expanded and purified. The correctness of the pMAL-2c-atI1-Y25F construct was verified by sequencing with commercial “MaIE forward sequencing primer (NEB).

(The majority of this work was performed by Vitaly Latypov)

9.11 Fluorescence-based Assays

9.11.1 Fluorescence-based ATL-DNA Binding Assays

Fluorescence emission intensity and fluorescence anisotropy measurements were made at 25°C on a Horiba Jobin Yvon FluoroMax-3 fluorimeter. For the SIMA and HEX dyes, the excitation wavelength was 530 nm (excitation slit width 5 nm), and the emission was detected at 560 nm (emission slit width 5nm). A 1 mL fluorescence cuvette with excitation and emission path lengths each of 10 mm, was used in the assays. A 1nM solution of ODN in titration buffer (50mM Tris–HCl (pH 7.5), 50mM NaCl, 1mM EDTA) was allowed 10 min to equilibrate with occasional mixing before small volumes of protein solution were added using a glass syringe (Hamilton) for added accuracy. The contents of the cuvette were mixed thoroughly with a pipette and left for 1 min before each reading was taken. For fluorescence emission intensity, 1 accumulation scan with an integration time of 1 sec was used for each measurement. For anisotropy, the time-trace function was used (total time = 60 sec with an integration time of 0.5 sec) to generate 5 data points which were averaged.

The ODNs used in the experiments were labelled with either 5'-HEX or SIMA (HEX) (see table 9.1). This allowed ODN concentrations to be used in the range of 2-20 nM for anisotropy and 1-5 nM for intensity. The concentration used for each experiment is given in the title of the relevant plot.

Anisotropy is defined by:

$$\text{anisotropy} = (I_{VV} - I_{VH}) / (I_{VV} + 2I_{VH})$$

where I_{VV} and I_{VH} are the intensities of the vertical and horizontal components of the emitted light using vertical polarised excitation.

For these direct titrations the binding isotherms were generated by plotting protein concentration against anisotropy or fluorescent intensity and were fitted by nonlinear least-squares regression using KaleidaGraph to the standard equation describing the equilibrium



D = ODN,

E = enzyme

DE = ODN-enzyme complex

For anisotropy

$$\mathbf{A = A_{min} + [(D+E+K_D) - ((D+E+K_D)^2 - (4DE))^{0.5}] (A_{max} - A_{min}) / 2D}$$

A = the anisotropy measured at a certain concentration of enzyme (E)

D = the oligonucleotide concentration

A_{min} = the lowest measured anisotropy (i.e. when no protein is added)

A_{max} = the highest measured anisotropy (i.e. when the binding is saturated)

K_D = the dissociation constant

For fluorescent intensity:

$$\mathbf{I = I_{max} + [(D+E+K_D) - ((D+E+K_D)^2 - (4DE))^{0.5}] (I_{min} - I_{max}) / 2D}$$

A = the intensity measured at a certain concentration of enzyme (E)

D = the oligonucleotide concentration

A_{min} = the lowest measured intensity (i.e. when the binding is saturated)

A_{max} = the highest measured intensity (i.e. when no protein is titrated)

K_D = the dissociation constant

Equations from (87)

9.11.2 FRET Analysis of AtI1-DNA Binding

To a cuvette containing 1mL of a 50nM solution of ODN OW56 (FAM-labelled) in 1X Buffer I (50mM TrisHCl pH7.5, 50mM NaCl, 1mM EDTA) was added increasing amounts of either OW57 (Cy3-labelled) or OW58 (HEX-labelled) in the following increments: 12.5nM, 25nM, 50nM (final [ODN]). Once a 1:1 duplex was formed, AtI1 was added to a final concentration of 250nM, and then 1000nM. The FAM label was excited at 490nm and the fluorescent emission intensity measured between 510-650nm with excitation and emission slit widths of 5nm, and an integration time of 1s. The assays were carried out at 25°C.

9.11.3 FRET Analysis of ODN Hybridisation

For the 'hybridised' ODNs: ODNs (e.g. OW56 and OW58) were mixed together in a solution of 1X Buffer I (50mM TrisHCl pH7.5, 50mM NaCl, 1mM EDTA) and incubated for 5 mins at 80°C before being slowly allowed to cool to RT. The dsODN solution was diluted to 1nM and 1mL of this solution was placed in a cuvette. The FAM label was excited at 490nm and the fluorescent emission intensity measured between 510-650nm with excitation and emission slit widths of 5nm, and an integration time of 1s. The assays were carried out at 25°C

For the 'unhybridised' ODNs: ODNs (e.g. OW56 and OW58) were mixed together in a solution of 1X Buffer I (50mM TrisHCl pH7.5, 50mM NaCl, 1mM EDTA) and immediately diluted to 1nM. 1mL of this solution was placed in a cuvette. The FAM label was excited at 490nm and the fluorescent emission intensity measured between 510-650nm with excitation and emission slit widths of 5nm, and an integration time of 1s. The assays were carried out at 25°C

9.11.4 Fluorescence-based MGMT Activity Assay

ODN OW31 (containing O⁶-MeG) or OW39 (containing G) was diluted to a concentration of 1µM in 1X NEB Restriction Digest Buffer 3 (New England Biolabs) and folded by heating at 95°C for 1min and cooling rapidly on ice for 5min. The ODN solution was diluted to 2nM in 1X NEB Restriction Digest Buffer 3 (New England Biolabs) to which was added MGMT to a final concentration of 100µM and incubated for 30min at 25°C. 1µL of PstI (20 Units, NEB) was added and the reaction incubated for 30min at 25°C. The emission intensity of the Cy3 dye was measured as a function of time throughout the reactions (excitation λ = 547nm, emission λ = 563nm, slit width= 5nm, integration time= 1s).

9.12 Affinity-based Isolation of AtI1

9.12.1 *S.pombe* Growth

S.pombe cultures were grown at 30°C in liquid rich media (YES, which consists of 0.5% yeast extract, 3% glucose, 225 mg/ml adenine, uracil, histidine, leucine and lysine). The cells were harvested at an OD₆₀₀ of approximately 0.5-1, centrifuged and the pellet resuspended in stop buffer.

This cell suspension was dripped into liquid nitrogen to make small cryogenically frozen pellets, and these stored unsealed overnight at -80°C to allow the liquid N_2 to evaporate.

9.12.2 Preparation of Streptavidin Beads (ODN Binding)

An aliquot (500 μL) of streptavidin (SA) beads was washed three times with an equal volume (500 μL) of 2X binding buffer (10mM Tris HCl pH7.5; 1mM EDTA; 2M NaCl) and then resuspended in half the volume of the same buffer (250 μL). To this was added an equal volume (250 μL) of 2 μM dsODN in dH_2O and the binding reaction incubated for 2h at RT (i.e. final conc.= 1 μM of ODN, 1X buffer). The supernatant was removed and the binding efficiency measured by comparison of ODN concentration (A_{260}) before and after incubation with the beads. Finally the ODN coated SA beads were resuspended in 500 μL 1X extraction buffer (20mM Tris-HCl pH 7.2; 100mM NaCl; 1mM DTT, 0.1mM PMSF, + protease inhibitor cocktail (Roche)) ready for use in the pull-down assay.

9.12.3 Pulldown Assays (1)

Pellets of *S.pombe* cells were ground into a powder using liquid nitrogen grinding apparatus. This powder (~8g) was added to 20mL 1X extraction buffer (20mM Tris-HCl pH 7.2; 100mM NaCl; 1mM DTT, 0.1mM PMSF, + protease inhibitor cocktail (Roche)), resuspended by mixing at 4°C for 1h and then centrifuged (20,000g, 10min, 4°C). The cleared supernatant was removed, mixed and the total protein concentration measured by Bradford assay (approx. 30mg/mL).

This whole-cell extract was divided into the appropriate number of aliquots and the relevant SA beads pre-coated with ODN were added. The

pull-down assays were incubated for 1h at 25°C, after which the beads were removed from the supernatant and washed three times with 100µL 1X extraction buffer. The beads were then resuspended in 40 µL Laemelli buffer and boiled at 100°C for 5min in preparation for SDS-PAGE separation and analysis.

9.12.4 Pulldown Assays (2)

Pellets of *S.pombe* cells were ground into a powder using liquid nitrogen grinding apparatus. This powder (~8g) was added to 20mL 1X extraction buffer (20mM Tris-HCl pH 7.2; 100mM NaCl; 1mM DTT, 0.1mM PMSF, + protease inhibitor cocktail (Roche)), resuspended by mixing at 25°C for 10min, sonicated at 4°C with 6 x 30sec pulses and centrifuged (12,000g, 20min, 4°C). The cleared supernatants (~15mg/mL total protein concentration) divided into aliquots and treated in the follow manner:

-ATL = no further treatment

+ATL = adenine triphosphate added to 1mM final concentration

+ATL (benz) = to 6mL extract was added MgCl₂ to 2.5mM final concentration and 11,000 Units Benzonase. After incubation for 1h at 25°C, EDTA was added to a final concentration of 10mM, and adenine triphosphate added to 1mM final concentration.

These whole-cell extracts were divided into the appropriate number of aliquots and the relevant SA beads pre-coated with ODN added. The pull-downs were incubated for 1h at 25°C, after which the beads were removed from the supernatant and washed twice with 750µL 1X extraction buffer. The beads were then stored at -20°C.

9.12.5 AtI1 Quantification by ELISA

ODNs OW78 and OW79 were annealed together in 50mM NaCl solution ([ODN]=10 μ M) by incubation at 95°C for 5min and were allowed to cool slowly to RT. The dsODN solution was diluted to 10nM, added to each well of a 96 well streptavidin-coated microtitre plate (Thermo) and incubated for 2h at RT. Plates were washed 4 times with 400 μ L PBS and blocked using 300 μ L of PBS/3% BSA for 2h at room temperature on a rotating platform. 100 μ L of a 1/500 dilution of the primary antibody (anti-AtI1) was added to the wells and incubated for 1h at room temperature on a rotating platform. The plates were then washed 4 times with 400 μ L of PBS and 100 μ L of horse-radish peroxidase-linked secondary antibody (goat-anti-rabbit, 1/2000 dilution, Dako) was added. The plates were incubated at room temperature for 1h on a rocking platform then washed 4 times with PBS. Western Lightning reagent (Perkin Elmer) was added and chemiluminescence measured as relative luminescence units using a TECAN GENios plate-reader.

9.13 Tandem Mass Spectrometry (MS/MS) Experiments

9.13.1 Sample Preparation

The protein samples from the affinity chromatography assays were separated and analysed by 1-D SDS PAGE (see section 9.7.1). The relevant bands were excised from the gel and destained by incubating at 37°C for 30min with solution 1. This step was repeated until all the coomassie blue was visibly removed. The gel piece was dried for 10min in a spin-vacuum concentrator, after which solution 4 was added with trypsin (20 μ g/mL) so that the approximate ratio of total protein: trypsin was 20:1 and the reaction

incubated overnight (16h) at 37°C with mixing. The supernatant was collected and kept aside, and then solution **5** was added to the gel piece, vortexed for a few seconds and incubated for 15min at 37°C. This supernatant was removed and then solution **6** was added to the gel piece, vortexed for a few seconds and incubated for 15min at 37°C. This supernatant was removed and then solution **7** was added to the gel piece, vortexed for a few seconds and incubated for 30min at 37°C. All the supernatants were pooled together and then evaporated to dryness by vacuum concentration to give the peptides as a solid.

Solution **1**: 200mM ammonium bicarbonate in 40% acetonitrile

Solution **2**: 10mM DTT solution

Solution **3**: 5mM Iodoacetamide solution

Solution **4**: 40mM ammonium bicarbonate in 9% acetonitrile

Solution **5**: 100% MeCN

Solution **6**: 5% Formic acid

Solution **7**: 5% Formic acid in 50% acetonitrile

Switchos: 0.1% Formic acid in 3% acetonitrile

In the case where the alkylation and reduction steps were carried out, the gel piece was treated after de-staining but prior to trypsin digestion in the following manner: the gel piece was incubated with Solution **2** for 1h at 56°C, after which the solution was discarded. The gel piece was then incubated with Solution **3** for 30min at 25°C in the dark, after which the solution was discarded. The gel piece was then washed twice with Solution **2** for 15 mins at

25°C, washed with Solution 3 for 15 mins at 37°C and then dried in a vacuum concentrator.

9.13.2 LC MS/MS Experimental Set-up

An Ultimate 3000 capillary HPLC operating in reverse-phase (RP) separation (Dionex, Surrey UK) was interfaced in tandem to an HCT Ultra PTM Discovery Q-IT MS (Bruker Daltonics, Coventry UK). Vacuum concentrated prefractionated samples were resuspended into transport buffer Switchos (3% acetonitrile, 0.1% formic acid), injected and captured onto a 0.35 mm trap column (3 m C18, Dionex-LC Packings). The trapped samples were then eluted onto a 0.075 150 mm analytical column (3 mm C18, Dionex-LC Packings) using an automated binary gradient with a flow of 300 nL min from 97% buffer I (3% acetonitrile, 0.1% formic acid) to 45% buffer II (97% acetonitrile, 0.1% formic acid) over 40 min followed by 90% buffer II for 5 minutes. Samples eluted from the high resolution RP separation were transferred onto a Bruker low-flow electrospray needle operating at 3600 V (Bruker Daltonics, Coventry UK)

Operating scan range for IT-MS was set from 300–1800 m/z using a dynamicion charge threshold (ICC) of 200 000 or a time-dependent accumulation no more than 200 ms using 'standard enhanced' scans. Operation of the IT-MS/MS threshold was set at 60–2700 m/z at an ICC of 200 000 with fragmentation energy of 0.5 V Data was acquired using Bruker software suite Compass v3.6 with Hystar DCMS-link v2.0 on a dual processor online workstation.

For the experiments where the aim was to identify specific peptides, ion trap pSRM measurements were designed and qualified under Ultrascan mode

operated under selective ion monitoring (Selective ion monitoring zooming (3 amu) and pSRM-MS/MS mode for intact and full fragment mass scan. Ion charge control was operated at 150,000 ions at pSRM-MS/MS conditions with a maximum accumulation time of 200 ms and 3 microscan averages.

9.13.3 MS/MS Data Analysis

Data analysis LC-MS data was collected and curated using Bruker Data-Analysis v4.0 into generic MS interrogation file format (.mgf). Further analysis was performed on an in-house Phenyx Search server (Genebio, Geneva) running on an 8-node Linux cluster and the remote access Phenyx vital-it multi-cluster server, or alternatively using an in-house MASCOT algorithm (Matrix Science, Boston). Analysis was performed by a variable modification of oxidation of methionines on a collective Swiss-Prot (Apr. 2011) *S.pombe* database. Mass threshold for interrogation of Q-IT data was set at 1.2 Da MS and 0.6 Da MS/MS. Acceptance parameters for search data was set at az-score of 6.0, p-value of 10^{-4} and a minimum peptide length of six amino acids.

11.0 References

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
2. Lohman, T.M. and Bjornson, K.P. (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annual review of biochemistry*, **65**, 169-214.
3. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. and Sugino, A. (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, **59**, 598-605.
4. Liu, Y., Kao, H.I. and Bambara, R.A. (2004) Flap endonuclease 1: a central component of DNA metabolism. *Annual review of biochemistry*, **73**, 589-615.
5. Crick, F. (1970) Central dogma of molecular biology. *Nature*, **227**, 561-563.
6. De Bont, R. and van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, **19**, 169-185.
7. Saul, R.L. and Ames, B.N. (1986) Background levels of DNA damage in the population. *Basic life sciences*, **38**, 529-535.
8. Blackburn, G.M. (2006) *Nucleic acids in chemistry and biology*. 3rd ed. / edited by G. Michael Blackburn ... [et al.]. ed. RSC Pub., Cambridge.
9. Kyrtopoulos, S.A., Anderson, L.M., Chhabra, S.K., Souliotis, V.L., Pletsa, V., Valavanis, C. and Georgiadis, P. (1997) DNA adducts and the mechanism of carcinogenesis and cytotoxicity of methylating agents of environmental and clinical significance. *Cancer detection and prevention*, **21**, 391-405.
10. Margison, G.P., Santibanez Koref, M.F. and Povey, A.C. (2002) Mechanisms of carcinogenicity/chemotherapy by O6-methylguanine. *Mutagenesis*, **17**, 483-487.
11. Friedberg, E.C. (2003) DNA damage and repair. *Nature*, **421**, 436-440.
12. Helleday, T., Lo, J., van Gent, D.C. and Engelward, B.P. (2007) DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA repair*, **6**, 923-935.
13. Rinne, M.L., He, Y., Pachkowski, B.F., Nakamura, J. and Kelley, M.R. (2005) N-methylpurine DNA glycosylase overexpression increases alkylation sensitivity by rapidly removing non-toxic 7-methylguanine adducts. *Nucleic acids research*, **33**, 2859-2867.
14. Margison, G.P., Povey, A.C., Kaina, B. and Santibanez Koref, M.F. (2003) Variability and regulation of O6-alkylguanine-DNA alkyltransferase. *Carcinogenesis*, **24**, 625-635.
15. Fang, Q., Kanugula, S., Tubbs, J.L., Tainer, J.A. and Pegg, A.E. Repair of O4-alkylthymine by O6-alkylguanine-DNA alkyltransferases. *The Journal of biological chemistry*, **285**, 8185-8195.
16. Hsieh, P. and Yamane, K. (2008) DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mechanisms of ageing and development*, **129**, 391-407.

17. Iyer, R.R., Pluciennik, A., Burdett, V. and Modrich, P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chemical reviews*, **106**, 302-323.
18. Branch, P., Aquilina, G., Bignami, M. and Karran, P. (1993) Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage. *Nature*, **362**, 652-654.
19. Karran, P. and Marinus, M.G. (1982) Mismatch correction at O6-methylguanine residues in *E. coli* DNA. *Nature*, **296**, 868-869.
20. Cejka, P., Mojas, N., Gillet, L., Schar, P. and Jiricny, J. (2005) Homologous recombination rescues mismatch-repair-dependent cytotoxicity of S(N)1-type methylating agents in *S. cerevisiae*. *Curr Biol*, **15**, 1395-1400.
21. Memisoglu, A. and Samson, L. (2000) Base excision repair in yeast and mammals. *Mutation research*, **451**, 39-51.
22. Sedgwick, B. (2004) Repairing DNA-methylation damage. *Nature reviews*, **5**, 148-157.
23. Rubinson, E.H., Gowda, A.S., Spratt, T.E., Gold, B. and Eichman, B.F. An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature*, **468**, 406-411.
24. Bruner, S.D., Norman, D.P. and Verdine, G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859-866.
25. Friedberg, E.C., Aguilera, A., Gellert, M., Hanawalt, P.C., Hays, J.B., Lehmann, A.R., Lindahl, T., Lowndes, N., Sarasin, A. and Wood, R.D. (2006) DNA repair: from molecular mechanism to human disease. *DNA repair*, **5**, 986-996.
26. Gillet, L.C. and Scharer, O.D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical reviews*, **106**, 253-276.
27. Reardon, J.T. and Sancar, A. (2003) Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease. *Genes & development*, **17**, 2539-2551.
28. Truglio, J.J., Croteau, D.L., Van Houten, B. and Kisker, C. (2006) Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical reviews*, **106**, 233-252.
29. Lin, J.J. and Sancar, A. (1992) Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. *The Journal of biological chemistry*, **267**, 17688-17692.
30. Verhoeven, E.E., van Kesteren, M., Moolenaar, G.F., Visse, R. and Goosen, N. (2000) Catalytic sites for 3' and 5' incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *The Journal of biological chemistry*, **275**, 5120-5123.
31. Orren, D.K., Selby, C.P., Hearst, J.E. and Sancar, A. (1992) Post-incision steps of nucleotide excision repair in *Escherichia coli*. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *The Journal of biological chemistry*, **267**, 780-788.
32. Selby, C.P. and Sancar, A. (1993) Molecular mechanism of transcription-repair coupling. *Science (New York, N. Y.)*, **260**, 53-58.

33. Hoeijmakers, J.H. (2009) DNA damage, aging, and cancer. *The New England journal of medicine*, **361**, 1475-1485.
34. Wittschieben, B.O., Iwai, S. and Wood, R.D. (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *The Journal of biological chemistry*, **280**, 39982-39989.
35. Riedl, T., Hanaoka, F. and Egly, J.M. (2003) The comings and goings of nucleotide excision repair factors on damaged DNA. *The EMBO journal*, **22**, 5293-5303.
36. Fitch, M.E., Nakajima, S., Yasui, A. and Ford, J.M. (2003) In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. *The Journal of biological chemistry*, **278**, 46906-46910.
37. Moser, J., Volker, M., Kool, H., Alekseev, S., Vrieling, H., Yasui, A., van Zeeland, A.A. and Mullenders, L.H. (2005) The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA repair*, **4**, 571-582.
38. Min, J.H. and Pavletich, N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570-575.
39. van Gool, A.J., Citterio, E., Rademakers, S., van Os, R., Vermeulen, W., Constantinou, A., Egly, J.M., Bootsma, D. and Hoeijmakers, J.H. (1997) The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *The EMBO journal*, **16**, 5955-5965.
40. van den Boom, V., Citterio, E., Hoogstraten, D., Zotter, A., Egly, J.M., van Cappellen, W.A., Hoeijmakers, J.H., Houtsmuller, A.B. and Vermeulen, W. (2004) DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *The Journal of cell biology*, **166**, 27-36.
41. Tantin, D. (1998) RNA polymerase II elongation complexes containing the Cockayne syndrome group B protein interact with a molecular complex containing the transcription factor IIH components xeroderma pigmentosum B and p62. *The Journal of biological chemistry*, **273**, 27794-27799.
42. Beaudenon, S.L., Huacani, M.R., Wang, G., McDonnell, D.P. and Huibregtse, J.M. (1999) Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, **19**, 6972-6979.
43. Sarker, A.H., Tsutakawa, S.E., Kostek, S., Ng, C., Shin, D.S., Peris, M., Campeau, E., Tainer, J.A., Nogales, E. and Cooper, P.K. (2005) Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Molecular cell*, **20**, 187-198.
44. Tremeau-Bravard, A., Riedl, T., Egly, J.M. and Dahmus, M.E. (2004) Fate of RNA polymerase II stalled at a cisplatin lesion. *The Journal of biological chemistry*, **279**, 7751-7759.
45. Mocquet, V., Laine, J.P., Riedl, T., Yajin, Z., Lee, M.Y. and Egly, J.M. (2008) Sequential recruitment of the repair factors during NER: the role

- of XPG in initiating the resynthesis step. *The EMBO journal*, **27**, 155-167.
46. Pegg, A.E., Wiest, L., Foote, R.S., Mitra, S. and Perry, W. (1983) Purification and properties of O6-methylguanine-DNA transmethylase from rat liver. *The Journal of biological chemistry*, **258**, 2327-2333.
 47. Daniels, D.S. and Tainer, J.A. (2000) Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O(6)-alkylguanine-DNA alkyltransferase. *Mutation research*, **460**, 151-163.
 48. Lindahl, T., Sedgwick, B., Sekiguchi, M. and Nakabeppu, Y. (1988) Regulation and expression of the adaptive response to alkylating agents. *Annual review of biochemistry*, **57**, 133-157.
 49. Margison, G.P., Cooper, D.P. and Potter, P.M. (1990) The E. coli ogt gene. *Mutation research*, **233**, 15-21.
 50. Sedgwick, B., Robins, P., Totty, N. and Lindahl, T. (1988) Functional domains and methyl acceptor sites of the Escherichia coli ada protein. *The Journal of biological chemistry*, **263**, 4430-4433.
 51. Daniels, D.S., Woo, T.T., Luu, K.X., Noll, D.M., Clarke, N.D., Pegg, A.E. and Tainer, J.A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature structural & molecular biology*, **11**, 714-720.
 52. Berg, J.M., Tymoczko, J.L., Stryer, L. and Stryer, L.B. (2007) *Biochemistry*. 6th ed. ed. W. H. Freeman ; Basingstoke : Palgrave [distributor], New York.
 53. Fang, Q., Kanugula, S. and Pegg, A.E. (2005) Function of domains of human O6-alkylguanine-DNA alkyltransferase. *Biochemistry*, **44**, 15396-15405.
 54. Pegg, A.E. (2000) Repair of O(6)-alkylguanine by alkyltransferases. *Mutation research*, **462**, 83-100.
 55. Tubbs, J.L. and Tainer, J.A. Alkyltransferase-like proteins: molecular switches between DNA repair pathways. *Cell Mol Life Sci*, **67**, 3749-3762.
 56. Pearson, S.J., Ferguson, J., Santibanez-Koref, M. and Margison, G.P. (2005) Inhibition of O6-methylguanine-DNA methyltransferase by an alkyltransferase-like protein from Escherichia coli. *Nucleic acids research*, **33**, 3837-3844.
 57. Pearson, S.J., Wharton, S., Watson, A.J., Begum, G., Butt, A., Glynn, N., Williams, D.M., Shibata, T., Santibanez-Koref, M.F. and Margison, G.P. (2006) A novel DNA damage recognition protein in Schizosaccharomyces pombe. *Nucleic acids research*, **34**, 2347-2354.
 58. Tubbs, J.L., Latypov, V., Kanugula, S., Butt, A., Melikishvili, M., Kraehenbuehl, R., Fleck, O., Marriott, A., Watson, A.J., Verbeek, B. *et al.* (2009) Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature*, **459**, 808-813.
 59. Wang, L., Spratt, T.E., Liu, X.K., Hecht, S.S., Pegg, A.E. and Peterson, L.A. (1997) Pyridyloxobutyl adduct O6-[4-oxo-4-(3-pyridyl)butyl]guanine is present in 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone-treated DNA and is a substrate for O6-alkylguanine-DNA alkyltransferase. *Chemical research in toxicology*, **10**, 562-567.
 60. Mijal, R.S., Thomson, N.M., Fleischer, N.L., Pauly, G.T., Moschel, R.C., Kanugula, S., Fang, Q., Pegg, A.E. and Peterson, L.A. (2004)

- The repair of the tobacco specific nitrosamine derived adduct O6-[4-Oxo-4-(3-pyridyl)butyl]guanine by O6-alkylguanine-DNA alkyltransferase variants. *Chemical research in toxicology*, **17**, 424-434.
61. Aramini, J.M., Tubbs, J.L., Kanugula, S., Rossi, P., Ertekin, A., Maglaqui, M., Hamilton, K., Ciccocanti, C.T., Jiang, M., Xiao, R. *et al.* Structural basis of O6-alkylguanine recognition by a bacterial alkyltransferase-like DNA repair protein. *The Journal of biological chemistry*, **285**, 13736-13741.
 62. Onodera, T., Morino, K., Tokishita, S., Morita, R., Masui, R., Kuramitsu, S. and Ohta, T. Role of alkyltransferase-like (ATL) protein in repair of methylated DNA lesions in *Thermus thermophilus*. *Mutagenesis*, **26**, 303-308.
 63. Morita, R., Nakagawa, N., Kuramitsu, S. and Masui, R. (2008) An O6-methylguanine-DNA methyltransferase-like protein from *Thermus thermophilus* interacts with a nucleotide excision repair protein. *Journal of biochemistry*, **144**, 267-277.
 64. Hu, J., Ma, A. and Dinner, A.R. (2008) A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 4615-4620.
 65. Duguid, E.M., Rice, P.A. and He, C. (2005) The structure of the human AGT protein bound to DNA and its implications for damage detection. *Journal of molecular biology*, **350**, 657-666.
 66. Morimoto, K., Dolan, M.E., Scicchitano, D. and Pegg, A.E. (1985) Repair of O6-propylguanine and O6-butylguanine in DNA by O6-alkylguanine-DNA alkyltransferases from rat liver and *E. coli*. *Carcinogenesis*, **6**, 1027-1031.
 67. Pletsas, D., Wheelhouse, R.T., Pletsa, V., Nicolaou, A., Jenkins, T.C., Bibby, M.C. and Kyrtopoulos, S.A. (2006) Polar, functionalized guanine-O6 derivatives resistant to repair by O6-alkylguanine-DNA alkyltransferase: implications for the design of DNA-modifying drugs. *European journal of medicinal chemistry*, **41**, 330-339.
 68. Shibata, T., Glynn, N., McMurry, T.B., McElhinney, R.S., Margison, G.P. and Williams, D.M. (2006) Novel synthesis of O6-alkylguanine containing oligodeoxyribonucleotides as substrates for the human DNA repair protein, O6-methylguanine DNA methyltransferase (MGMT). *Nucleic acids research*, **34**, 1884-1891.
 69. Shuker, D.E. and Margison, G.P. (1997) Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer research*, **57**, 366-369.
 70. O'Driscoll, M., Macpherson, P., Xu, Y.Z. and Karran, P. (1999) The cytotoxicity of DNA carboxymethylation and methylation by the model carboxymethylating agent azaserine in human cells. *Carcinogenesis*, **20**, 1855-1862.
 71. Watson, A.J., Millington, C., Marriott, A., Verbeek, B., Caetano, C., Butt, A., McGown, G., Thorncroft, M., Shibata, T., Santibanez-Koref, M.F. *et al.* (2009) The *S.pombe* protein, Atl1 recognises O⁶-alkylguanine residues in DNA that are not substrates for human O⁶-alkylguanine-DNA alkyltransferase *Manuscript in preparation*.

72. Fried, M.G., Kanugula, S., Bromberg, J.L. and Pegg, A.E. (1996) DNA binding mechanism of O6-alkylguanine-DNA alkyltransferase: stoichiometry and effects of DNA base composition and secondary structure on complex stability. *Biochemistry*, **35**, 15295-15301.
73. Chen, C.S., Korobkova, E., Chen, H., Zhu, J., Jian, X., Tao, S.C., He, C. and Zhu, H. (2008) A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. *Nature methods*, **5**, 69-74.
74. Mendonca, V.M., Kaiser-Rogers, K. and Matson, S.W. (1993) Double helicase II (uvrD)-helicase IV (helD) deletion mutants are defective in the recombination pathways of Escherichia coli. *Journal of bacteriology*, **175**, 4641-4651.
75. Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J. and Darst, S.A. (2006) Structural basis for bacterial transcription-coupled DNA repair. *Cell*, **124**, 507-520.
76. Morita, R., Hishinuma, H., Ohyama, H., Mega, R., Ohta, T., Nakagawa, N., Agari, Y., Fukui, K., Shinkai, A., Kuramitsu, S. *et al.* An alkyltransferase-like protein from Thermus thermophilus HB8 affects the regulation of gene expression in alkylation response. *Journal of biochemistry*, **150**, 327-339.
77. Lieber, M.R. (1997) The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays*, **19**, 233-240.
78. Beranek, D.T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutation research*, **231**, 11-30.
79. Rasmussen, L.J. and Samson, L. (1996) The Escherichia coli MutS DNA mismatch binding protein specifically binds O(6)-methylguanine DNA lesions. *Carcinogenesis*, **17**, 2085-2088.
80. Mazon, G., Philippin, G., Cadet, J., Gasparutto, D. and Fuchs, R.P. (2009) The alkyltransferase-like ybaZ gene product enhances nucleotide excision repair of O(6)-alkylguanine adducts in E. coli. *DNA repair*, **8**, 697-703.
81. Mazon, G., Philippin, G., Cadet, J., Gasparutto, D., Modesti, M. and Fuchs, R.P. Alkyltransferase-like protein (eATL) prevents mismatch repair-mediated toxicity induced by O6-alkylguanine adducts in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 18050-18055.
82. Margison, G.P., Butt, A., Pearson, S.J., Wharton, S., Watson, A.J., Marriott, A., Caetano, C.M., Hollins, J.J., Rukazenkova, N., Begum, G. *et al.* (2007) Alkyltransferase-like proteins. *DNA repair*, **6**, 1222-1228.
83. Tsuzuki, T., Sakumi, K., Shiraishi, A., Kawate, H., Igarashi, H., Iwakuma, T., Tominaga, Y., Zhang, S., Shimizu, S., Ishikawa, T. *et al.* (1996) Targeted disruption of the DNA repair methyltransferase gene renders mice hypersensitive to alkylating agent. *Carcinogenesis*, **17**, 1215-1220.
84. Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V. *et al.* (2007) Sea anemone genome reveals ancestral eumetazoan gene

- repertoire and genomic organization. *Science (New York, N.Y)*, **317**, 86-94.
85. Caruthers, M.H. (1985) Gene synthesis machines: DNA chemistry and its uses. *Science (New York, N.Y)*, **230**, 281-285.
86. Beaucage, S.L. and Iyer, R.P. (1993) The Functionalization of Oligonucleotides Via Phosphoramidite Derivatives. *Tetrahedron*, **49**, 1925-1963.
87. Reid, S.L., Parry, D., Liu, H.H. and Connolly, B.A. (2001) Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry*, **40**, 2484-2494.
88. Kierzek, E. and Kierzek, R. (2003) The synthesis of oligoribonucleotides containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines via post-synthetic modification of precursor oligomers. *Nucleic acids research*, **31**, 4461-4471.
89. Lakshman, M.K., Ngassa, F.N., Keeler, J.C., Dinh, Y.Q., Hilmer, J.H. and Russon, L.M. (2000) Facile synthesis of O6-alkyl-, O6-aryl-, and diaminopurine nucleosides from 2'-deoxyguanosine. *Organic letters*, **2**, 927-930.
90. Xu, Y.-Z. (1998) Reactive DNA: 6-methylsulphoxypurine used for site-specific and chemical crosslinking with cysteine and its peptide. *Tetrahedron*, **54**, 187-196.
91. Chuvilin, A.N., Serebryakova, M.V., Smirnov, I.P. and Pozmogova, G.E. (2009) Byproduct with Altered Fluorescent Properties Is Formed during Standard Deprotection Step of Hexachlorofluorescein Labeled Oligonucleotides. *Bioconjugate chemistry*.
92. Shibata, T., Glynn, N., McMurry, T.B.H., McElhinney, R.S., Margison, G.P. and Williams, D.M. (2006) Novel synthesis of O-6-alkylguanine containing oligodeoxyribonucleotides as substrates for the human DNA repair protein, O-6-methylguanine DNA methyltransferase (MGMT). *Nucleic Acids Research*, **34**, 1884-1891.
93. Kolman, A., Chovanec, M. and Osterman-Golkar, S. (2002) Genotoxic effects of ethylene oxide, propylene oxide and epichlorohydrin in humans: update review (1990-2001). *Mutation research*, **512**, 173-194.
94. Tornqvist, M., Gustafsson, B., Kautiainen, A., Harms-Ringdahl, M., Granath, F. and Ehrenberg, L. (1989) Unsaturated lipids and intestinal bacteria as sources of endogenous production of ethene and ethylene oxide. *Carcinogenesis*, **10**, 39-41.
95. Coulter, R., Blandino, M., Tomlinson, J.M., Pauly, G.T., Krajewska, M., Moschel, R.C., Peterson, L.A., Pegg, A.E. and Spratt, T.E. (2007) Differences in the rate of repair of O6-alkylguanines in different sequence contexts by O6-alkylguanine-DNA alkyltransferase. *Chemical research in toxicology*, **20**, 1966-1971.
96. Guzder, S.N., Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1995) Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIIH. *The Journal of biological chemistry*, **270**, 12973-12976.
97. Marras, S.A., Kramer, F.R. and Tyagi, S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic acids research*, **30**, e122.

98. Dickerson, R.E. and Drew, H.R. (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *Journal of molecular biology*, **149**, 761-786.
99. Drew, H.R. and Dickerson, R.E. (1981) Structure of a B-DNA dodecamer. III. Geometry of hydration. *Journal of molecular biology*, **151**, 535-556.
100. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 2179-2183.
101. LiCata, V.J. and Wowor, A.J. (2008) Applications of fluorescence anisotropy to the study of protein-DNA interactions. *Methods in cell biology*, **84**, 243-262.
102. Heyduk, T. and Lee, J.C. (1990) Application of fluorescence energy transfer and polarization to monitor Escherichia coli cAMP receptor protein and lac promoter interaction. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 1744-1748.
103. Powell, L.M., Connolly, B.A. and Dryden, D.T. (1998) The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *Journal of molecular biology*, **283**, 947-961.
104. Shuttleworth, G., Fogg, M.J., Kurpiewski, M.R., Jen-Jacobson, L. and Connolly, B.A. (2004) Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea. *Journal of molecular biology*, **337**, 621-634.
105. Tyagi, S. and Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature biotechnology*, **14**, 303-308.
106. Wang, K., Tang, Z., Yang, C.J., Kim, Y., Fang, X., Li, W., Wu, Y., Medley, C.D., Cao, Z., Li, J. *et al.* (2009) Molecular engineering of DNA: molecular beacons. *Angewandte Chemie (International ed)*, **48**, 856-870.
107. Li, J.J., Geyer, R. and Tan, W. (2000) Using molecular beacons as a sensitive fluorescence assay for enzymatic cleavage of single-stranded DNA. *Nucleic acids research*, **28**, E52.
108. Maksimenko, A., Ishchenko, A.A., Sanz, G., Laval, J., Elder, R.H. and Saparbaev, M.K. (2004) A molecular beacon assay for measuring base excision repair activities. *Biochemical and biophysical research communications*, **319**, 240-246.
109. Margison, G.P. and Santibanez-Koref, M.F. (2002) O6-alkylguanine-DNA alkyltransferase: role in carcinogenesis and chemotherapy. *Bioessays*, **24**, 255-266.
110. Stivers, J.T. (2008) Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry (Weinheim an der Bergstrasse, Germany)*, **14**, 786-793.
111. Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E. and Tainer, J.A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*, **384**, 87-92.
112. Lau, A.Y., Wyatt, M.D., Glassner, B.J., Samson, L.D. and Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG.

- Proceedings of the National Academy of Sciences of the United States of America*, **97**, 13573-13578.
113. Friedman, J.I. and Stivers, J.T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry*, **49**, 4957-4967.
 114. Meyer, A.S., McCain, M.D., Fang, Q., Pegg, A.E. and Spratt, T.E. (2003) O6-alkylguanine-DNA alkyltransferases repair O6-methylguanine in DNA with Michaelis-Menten-like kinetics. *Chemical research in toxicology*, **16**, 1405-1409.
 115. Duguid, E.M., Mishina, Y. and He, C. (2003) How do DNA repair proteins locate potential base lesions? a chemical crosslinking method to investigate O6-alkylguanine-DNA alkyltransferases. *Chemistry & biology*, **10**, 827-835.
 116. Blainey, P.C., van Oijen, A.M., Banerjee, A., Verdine, G.L. and Xie, X.S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5752-5757.
 117. Rasimas, J.J., Pegg, A.E. and Fried, M.G. (2003) DNA-binding mechanism of O6-alkylguanine-DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability. *The Journal of biological chemistry*, **278**, 7973-7980.
 118. Banerjee, A., Yang, W., Karplus, M. and Verdine, G.L. (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature*, **434**, 612-618.
 119. Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E. and Tainer, J.A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 869-878.
 120. Parikh, S.S., Walcher, G., Jones, G.D., Slupphaug, G., Krokan, H.E., Blackburn, G.M. and Tainer, J.A. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 5083-5088.
 121. Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357-369.
 122. Frosina, G. (2000) Overexpression of enzymes that repair endogenous damage to DNA. *European journal of biochemistry / FEBS*, **267**, 2135-2149.
 123. Glassner, B.J., Rasmussen, L.J., Najarian, M.T., Posnick, L.M. and Samson, L.D. (1998) Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 9997-10002.
 124. Bruns, C.M., Hubatsch, I., Ridderstrom, M., Mannervik, B. and Tainer, J.A. (1999) Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *Journal of molecular biology*, **288**, 427-439.
 125. Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198-207.

126. Pappin, D.J., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol*, **3**, 327-332.
127. Chernushevich, I.V., Loboda, A.V. and Thomson, B.A. (2001) An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom*, **36**, 849-865.
128. Edwards, N.J. Protein identification from tandem mass spectra by database searching. *Methods in molecular biology (Clifton, N.J)*, **694**, 119-138.
129. Chelius, D., Zhang, T., Wang, G. and Shen, R.F. (2003) Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry. *Analytical chemistry*, **75**, 6658-6665.
130. Roepstorff, P. and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry*, **11**, 601.
131. MacCoss, M.J., Wu, C.C. and Yates, J.R., 3rd. (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Analytical chemistry*, **74**, 5593-5599.
132. Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551-3567.
133. Aitken, A. and Learmonth, M. (2002) Protein identification by in-gel digestion and mass spectrometric analysis. *Molecular biotechnology*, **20**, 95-97.
134. Drabovich, A.P. and Diamandis, E.P. Combinatorial peptide libraries facilitate development of multiple reaction monitoring assays for low-abundance proteins. *Journal of proteome research*, **9**, 1236-1245.
135. Sigurdson, S.T. and Eckstein, F. (1996) Site specific labelling of sugar residues in oligoribonucleotides: reactions of aliphatic isocyanates with 2' amino groups. *Nucleic acids research*, **24**, 3129-3133.
136. Paalman, S.R., Noll, D.M. and Clarke, N.D. (1997) Formation of a covalent complex between methylguanine methyltransferase and DNA via disulfide bond formation between the active site cysteine and a thiol-containing analog of guanine. *Nucleic acids research*, **25**, 1795-1801.
137. Noll, D.M. and Clarke, N.D. (2001) Covalent capture of a human O(6)-alkylguanine alkyltransferase-DNA complex using N(1),O(6)-ethanoxanthosine, a mechanism-based crosslinker. *Nucleic acids research*, **29**, 4025-4034.
138. Zugates, G.T., Anderson, D.G., Little, S.R., Lawhorn, I.E. and Langer, R. (2006) Synthesis of poly(beta-amino ester)s with thiol-reactive side chains for DNA delivery. *Journal of the American Chemical Society*, **128**, 12726-12734.
139. Bellas, M., Tuleen, D.L. and Field, L. (1967) Organic disulfide and related substances. XXII. Substituted benzyl 2-(n-decylamino)ethyl disulfide hydrochlorides. A possible neighboring-group effect involving sulfur. *The Journal of Organic Chemistry*, **32**, 2591-2595.
140. Field, L., Parsons, T.F. and Pearson, D.E. (1966) Organic Disulfides and Related Substances. XVIII. Synthesis and Disproportionation of 2-

- (Aryldithio)ethylamine Hydrochlorides 1a,b. *The Journal of Organic Chemistry*, **31**, 3550-3555.
141. Watson, A.J. and Margison, G.P. (2000) O6-alkylguanine-DNA alkyltransferase assay. *Methods in molecular biology (Clifton, N.J)*, **152**, 49-61.
 142. Morrison, L.E. and Stols, L.M. (1993) Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution. *Biochemistry*, **32**, 3095-3104.
 143. Ebright, Y.W., Chen, Y., Pendergrast, P.S. and Ebright, R.H. (1992) Incorporation of an EDTA-metal complex at a rationally selected site within a protein: application to EDTA-iron DNA affinity cleaving with catabolite gene activator protein (CAP) and Cro. *Biochemistry*, **31**, 10664-10670.
 144. Callahan, J.F., Ashton-Shue, D., Bryan, H.G., Bryan, W.M., Heckman, G.D., Kinter, L.B., McDonald, J.E., Moore, M.L., Schmidt, D.B., Silvestri, J.S. *et al.* (1989) Structure-activity relationships of novel vasopressin antagonists containing C-terminal diaminoalkanes and (aminoalkyl)guanidines. *Journal of medicinal chemistry*, **32**, 391-396.
 145. Baskin, J.M., Prescher, J.A., Laughlin, S.T., Agard, N.J., Chang, P.V., Miller, I.A., Lo, A., Codelli, J.A. and Bertozzi, C.R. (2007) Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 16793-16797.
 146. Keller, K.A., Guo, J., Punna, S. and Finn, M.G. (2005) A thermally-cleavable linker for solid-phase synthesis. *Tetrahedron Letters*, **46**, 1181-1184.