

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON
FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
Centre for Biological Sciences

Utilising Uracil DNA Glycosylase to detect the presence of 5-Methylcytosine

by Scott Terrence Kimber, BSc (Hons)

Thesis for the degree of Doctor of Philosophy

September 2014

UNIVERSITY OF SOUTHAMPTON
ABSTRACT
FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
CENTRE FOR BIOLOGICAL SCIENCES

UTILISING URACIL DNA GLYCOSYLASE TO DETECT THE PRESENCE OF 5-METHYLCYTOSINE

by Scott Terrence Kimber

DNA is regularly subjected to endogenous and exogenous reagents that cause mutations that can be detrimental to a cell if they are not repaired. One class of enzymes responsible for DNA repair is the family of DNA glycosylases and their role is to remove damaged bases. Uracil DNA Glycosylase (UDG) is a member of this family and is highly specific, removing only uracil, an RNA base, from DNA. Uracil arises in DNA through misincorporation of deoxyuridine monophosphate (dUMP) creating an A.U base pair, or through deamination of cytosine resulting in a G.U base pair. Though UDG acts on A.U pairs, this is not its primarily role as A.U pairings are not mutagenic. However the G.U mispair is highly mutagenic and leads to a G.C to A.T transition on subsequent rounds of replication. UDG only reacts with uracil and has no activity at thymine since the 5-methyl group on the base is excluded from the active site. This thesis examines mutants of UDG that can cleave cytosine but not 5-methylcytosine. Methylation of cytosine at CpG sites leads to gene silencing and is an important epigenetic signal. Knowing the methylation state of cytosines will therefore be important for understanding gene control and may be beneficial for treating many diseases. The most common method for detecting cytosine methylation uses a bisulphite reaction followed by normal DNA sequencing methods. This process has several drawbacks and the aim of this work is to create an enzyme that is capable of distinguishing between 5-methylcytosine and cytosine. It has been reported that mutation of a critical asparagine in UDG to an aspartate allows the enzyme to accommodate cytosine into its active site; generating a cytosine DNA glycosylase (CDG). Using the natural ability of UDG to distinguish between uracil and thymine due to the presence of the 5-methyl group, we hypothesised that the mutant enzyme should be able to discriminate between 5-methylcytosine and cytosine, which differ by the presence or absence of a methyl group in the same position. *E. coli* and human CDGs were prepared and their ability to remove cytosine or 5-methylcytosine examined when placed in different sequence contexts. *hCDG* was generated through complete gene synthesis of *hUDG* followed by the N204D mutation. The corresponding mutation in *E.coli* (N123D) generates a highly cytotoxic enzyme that cannot even be cloned in pUC19. As L191 aids base flipping, mutation to alanine (L191A) renders the enzyme inactive; activity can then be rescued using a bulky synthetic nucleoside that occupies the base pair and forces the target base into an extrahelical conformation. The L191A mutation was followed by N123D to generate an expressible and functional *eCDG*, denoted *eCYDG*. We demonstrate that these mutants have cytosine glycosylase activity when the cytosine is mispaired or unpaired, but not when paired with guanine, and show no activity against 5-methylcytosine in any context. The activity of these CDGs varies with the stability of the base pair, with the fastest cleavage rates being obtained with the least stable base pairs, and also varies with the local sequence context. As CDGs are able to discriminate between cytosine and 5-methylcytosine we began development of a real-time PCR assay for detection of 5-methylcytosine. This employed a hexaethylene glycol (HEG) linker opposite the target cytosine, as this produces one of the fastest cleavage rates and cannot be read by a DNA polymerase.

Publications

Kimber ST, Brown T, Fox KR (2014) A Mutant of Uracil DNA Glycosylase That Distinguishes between Cytosine and 5-Methylcytosine. *PLoS ONE* 9(4): e95394. doi:10.1371/journal.pone.0095394

Acknowledgements

I would firstly like to thank Keith Fox for his supervision throughout my research and Dave Rusling for his help, advice and tuition in the lab. Thanks to Tom Brown, Stuart Findlow and Mark Coldwell for their guidance and expertise, and thank you to the past and present lab members who have helped out along the way. Thanks to the many people who have made this PhD an enjoyable experience, you know who you are, and finally I would like to thank my family for their love and support over these many years.

Table of Contents

Chapter 1: General Introduction.....	1
1.1 DNA and Epigenetics	1
1.2 DNA Glycosylases	4
1.2.1 Repair of Damaged DNA	5
1.2.2 How Uracil Arises in DNA.....	6
1.3 Uracil DNA Glycosylases	7
1.3.1 Family 1 UDGs.....	7
1.3.1.1 Recognition of Uracil by UDG	8
1.3.1.2 Structure of UDG	11
1.3.1.3 DNA-Enzyme Binding.....	14
1.3.1.3.1 Serine Interactions	14
1.3.1.3.2 Effect of Flanking Regions.....	15
1.3.1.3.3 Rate Limiting Factor of Excision	16
1.3.1.4 The Role of Leucine.....	16
1.3.1.4.1 Introduction of Pyrene into the Substrate.....	16
1.3.1.5 Active Site Composition and Mechanism of UDG.....	18
1.3.1.6 Events after Excision	23
1.3.1.7 The Role of Phenylalanine 77 and Tyrosine 66 in Thymine Exclusion	25
1.3.1.8 Tyrosine 66 Mutations	26
1.3.1.9 Properties of Cytosine DNA Glycosylases	27
1.4 5-Methylcytosine Detection	35
1.4.1 Bisulphite Sequencing	35
1.4.2 Bisulphite and UDG.....	36
1.4.3 Methylation Sensitive Enzyme Restriction.....	37
1.4.4 Nanopores	37
1.4.5 Triplexes	37

1.4.6 Small Molecules	38
1.4.7 Other Methods	38
1.5 Purpose of this Research	39
1.5.1 Selection of Enzyme.....	39
Chapter 2: Materials and Methods.....	41
2.1 Materials.....	41
2.1.1 Oligonucleotides.....	41
2.1.2 Enzymes	41
2.1.3 Chemicals	41
2.1.4 Plasmids Generated in this Work	42
2.2 Methods for Studying Base Excision.....	43
2.2.1 Cleavage Studies and Rate of Reaction Determination.....	43
2.3 Protocols.....	44
2.3.1 Agarose Gel Electrophoresis	44
2.3.2 Site-Directed Mutagenesis	44
2.3.3 Preparation of Competent Cells	47
2.3.4 Transformation	47
2.3.5 Plasmid Purification	48
2.3.6 DNA sequencing	49
2.3.7 Gene Amplification	49
2.3.8 Gene Synthesis	50
2.3.9 PCR clean-up.....	52
2.3.10 Cloning	52
2.3.11 Colony PCR.....	52
2.3.12 Taq ^o I Digestion	53
2.3.13 Purification of Enzymes	53
2.3.13.1 Cell Harvesting	53

2.3.13.2	Sonication and Protein Preparation	54
2.3.13.3	Protein Purification	54
2.3.13.4	SDS-PAGE.....	54
2.3.14	5´ Radiolabelling of oligonucleotides	55
2.3.5	Excision Assays	57
2.3.15.1	Activity Assays	57
2.3.15.2	Quantitative Assays.....	57
2.3.15.3	Denaturing Polyacrylamide Gel Electrophoresis	57
2.3.15.4	Cleavage Quantification.....	58
2.3.16	Detection of 5-Methylcytosine	58
2.3.16.1	DNA Probing	58
2.3.16.2	Incubation with <i>e</i> CYDG	59
2.3.16.3	Real-time PCR	59
Chapter 3:	Production and Properties of <i>e</i> CYDG	61
3.1	Introduction	61
3.2	Experimental Design	62
3.2.1	Mutagenesis of UDG	62
3.2.2	Excision Assays	63
3.2.2.1	Initial Assays	63
3.2.2.2	<i>e</i> CYDG Rate of Reaction Determination.....	63
3.3	Results	64
3.3.1	N123D mutagenesis of pET <i>e</i> UDG and pUC19 <i>e</i> UDG	64
3.3.2	Determining the Effects of Mutagenesis	67
3.3.3	Cloning of <i>e</i> CDG into pUC18 and pUC19	68
3.3.4	Generating <i>e</i> CYDG.....	69
3.3.5	Cloning <i>e</i> CYDG into pET28a and pUC19	69
3.3.6	Enzyme Expression.....	70

3.3.7 Excision Activity of <i>eUDG</i>	73
3.3.9 Excision Activity of <i>eUYDG</i>	76
3.3.10 Excision Activity of <i>eCYDG</i>	77
3.3.10.1 Activity Determination	79
3.3.11 Rate of Reaction Determination	81
3.3.11.1 Excision of uracil	84
3.3.11.2 Excision of ssC substrates	85
3.3.11.3 Investigating the Effect of an Unpaired Cytosine.....	86
3.3.11.4 Investigating the Effect of the Flanking Base Pair on <i>eCYDG</i> Cleavage ..	88
3.3.11.5 The effect of pH on A.C excision	94
3.3.12 Mutating F77 and Y66	94
3.4 Discussion	95
Chapter 4: Exploring the inability to clone <i>eCDG</i>	101
4.1 Introduction.....	101
4.2 Experimental Design.....	102
4.2.1 Reverse Mutagenesis	102
4.2.2 Using different Cell Types to Generate a Stable <i>eCDG</i> Clone	102
4.2.2.1 Recombination Deficient Cells	102
4.2.2.2 Glucose Supplementation	102
4.2.3 Altering the Region Upstream of the <i>eUDG</i> Start Codon.....	102
4.2.4 Co-transformation of Ugi	103
4.2.5 Generating an <i>eUDGUgi</i> Fusion Protein.....	103
4.2.6 Further Mutagenesis of <i>eUDGUgi</i>	104
4.3 Results.....	104
4.3.1 Restoration of <i>eCDG</i> Activity	104
4.3.2 Effects of Different Cell Types	105
4.3.2.1 Using Sure Cells	105

4.3.2.2 Effects of Glucose	105
4.3.3 Alternative Initiation	105
4.3.4 Addition of Stop Codons	105
4.3.5 Introduction of Ugi	105
4.3.6 Cloning and Mutagenesis of UDG-Ugi Constructs	106
4.3.6.1 Generating a pET _e UDGUgi Fusion Construct	106
4.3.6.2 Generation of pET _h UDGUgi <i>via</i> complete gene synthesis	110
4.3.7 Expression of pET _e UDGUgiST	111
4.4 Discussion	113
Chapter 5: Excision Properties of <i>h</i> CDGs	115
5.1 Introduction	115
5.2 Experimental Design	115
5.2.1 Complete Gene Synthesis of <i>h</i> UDG	115
5.2.2 Mutagenesis of <i>h</i> UDG	116
5.2.3 Excision Properties	116
5.3 Results	116
5.3.1 Complete Gene Synthesis of <i>h</i> UDG Δ 81	116
5.3.2 <i>h</i> CDG Mutagenesis	117
5.3.3 Expression of <i>h</i> CDG Δ 81	118
5.3.4 Excision Activity Determination	119
5.3.5 Rate of Reaction values for <i>h</i> CDG with different DNA substrates	121
5.3.5.1 Excision of ssC substrates with <i>h</i> CDG	124
5.3.5.2 Examining the Effect of Flanking Regions on Excision	126
5.3.5.3 The Effect of pH on A.C excision	129
5.3.6 Rate of Reaction Determination of <i>h</i> CYDG	130
5.4 Discussion	134
Chapter 6: Developing an Assay for the Detection of 5-methylcytosine	137

6.1 Introduction.....	137
6.2 Experimental Design.....	138
6.3 Results.....	140
6.3.1 Designing the Assay.....	140
6.3.2 <i>e</i> CYDG:PCR Assay.....	142
6.3.3 <i>e</i> CYDG Reaction Analysis.....	143
6.4 Discussion.....	148
Chapter 7: General Discussion.....	151
References.....	155
Appendix I - List of General Oligonucleotides used.....	165
Appendix II - List of Oligonucleotides used for the Generation of <i>h</i> UDG Δ 81 and <i>h</i> UDG Δ 81Ugi.....	167
Appendix IIIa - List of Oligonucleotides used for the PCR assay.....	169
Appendix IIIb - List of Primers/Protection Oligonucleotides used for the <i>e</i> CYDG:rtPCR Assay.....	169
Appendix IV – Cleavage Assay Interpretation.....	170
Appendix V – DNA Sequences of ORFs of <i>e</i> UDG and <i>h</i> UDG Δ 81.....	172
Appendix VI - Sequencing Chromatograms.....	173

Abbreviations

A	adenine
aa	amino acid
ALKA	3-methyladenine DNA glycosylase II
AP	abasic site/apurinic site
APE1	apurinic endonuclease 1
BER	base excision repair
bp	base pair
C	cytosine
ddNTPs	dideoxynucleoside triphosphates
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTPs	deoxynucleoside triphosphates
ds	double-stranded
CDG	cytosine DNA glycosylase
CYDG	cytosine/pyrene DNA glycosylase
EDTA	ethylenediaminetetraacetic acid
<i>e</i> MUG	<i>E. coli</i> mismatch specific uracil DNA glycosylase
G	guanine
HEG	hexaethylene glycol
I	inosine
IPTG	isopropylthio- β - <i>D</i> -galactoside
kb	kilobase(s)
kDa	kilodalton(s)
MCS	multiple cloning site
MTP	mitochondrial targeting peptide
MUG	mismatch specific uracil DNA glycosylase
MUTY	MUTYH glycosylase
NER	nucleotide excision repair
NLS	nuclear localisation signal
OGG1	8-oxoguanine glycosylase
PAGE	polyacrylamide gel electrophoresis
Pn	pyrene
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SLS	sample loading solution
SMUG1	single-strand-selective mono-functional uracil DNA glycosylase
ss	single-stranded
T	thymine
TBE	tris, borate, EDTA; 1 x buffer
TDG	thymine DNA glycosylase
TET	ten-eleven translocation
TEMED	tetramethylethylenediamine
Tris	tris-hydroxymethyl-aminomethane
U	uracil
UDG	uracil DNA glycosylase
UV	ultraviolet
UYDG	uracil/pyrene DNA glycosylase

X-Gal
Z
εC

5-bromo-4-chloro-3-indoyl-β-*D*-galactosidase
anthraquinone pyrrolidine
3,N4-ethenocytosine

List of Figures

Figure 1.1 Gene regulation by cytosine methylation.	2
Figure 1.2 DNA bases.	4
Figure 1.3 The transition mutation from G.C to A.T, caused by cytosine deamination.	7
Figure 1.4 Cartoon alignment for comparison of human and <i>E. coli</i> UDGs.	8
Figure 1.5 A G.U wobble base pair.	9
Figure 1.6 Space filled model of showing the positive face of <i>h</i> UDG (arrow) next to its active site.	10
Figure 1.7 Structure of <i>h</i> UDG bound to dsDNA (PDB 1SSP) (Parikh <i>et al.</i> , 1998).	13
Figure 1.8 The “Push and Plug” roles of leucine.	17
Figure 1.9 Partial ClustalW sequence alignment for the three commonly studied Family 1 UDGs.	18
Figure 1.10 The critical residues lining the active site of <i>E. coli</i> uracil-DNA glycosylase.	19
Figure 1.11 N-Glycosidic bond cleavage mechanism.	22
Figure 1.12 The mechanism of the BER pathway (Adapted from Lindahl, 1993).	24
Figure 1.13 Thymine exclusion by UDG.	25
Figure 1.14 Alteration of the hydrogen bond pattern for N123D.	27
Figure 1.15 Pyrene rescue of CDG.	29
Figure 1.16 The ability of MUG to hydrogen bond with the opposing widowed base, mimicking that of the flipped out uracil base. DNA highlighted in red and the water molecule in blue.	30
Figure 1.17 Dissociative mechanism of N-Glycosidic bond cleavage by Family 4 UDGs.	33
Figure 1.18 The bisulphite reaction mechanism	35
Figure 2.1 The structure of anthraquinone pyrrolidine phosphoramidite.	41
Figure 2.2 Plasmid maps.	44
Figure 2.3 QuikChange site-directed mutagenesis	46
Figure 2.4 Oligonucleotides used in the amplification of genes investigated.	50
Figure 2.5 Protocol for <i>h</i> UDG Δ 81 gene synthesis.	51
Figure 3.1 Recognition of cytosine and uracil by the 180° rotation of Asp123’s side chain.	61
Figure 3.2 Mutation primers annealed to template plasmid.	62
Figure 3.3 Sequencing chromatogram showing multiple primer sequence repeats.	65

Figure 3.4 Theoretical plasmid digest with Taq ^I of pUC19 <i>e</i> UDG and pUC19 <i>e</i> CDG that would produce two and three restriction fragments respectively.	65
Figure 3.5 Gel electrophoresis of clones digested with Taq ^I	66
Figure 3.6 Sequencing analysis of pUC19 <i>e</i> CDG1 highlighting the N123D mutation (A) and extra cytosine base (B).	67
Figure 3.7 Universal primer sequences.	67
Figure 3.8 Gel electrophoresis of pUC clones digested with Taq ^I after <i>e</i> CDG cloning. ..	69
Figure 3.9 Gel electrophoresis of <i>e</i> CYDG clones digested with Taq ^I	70
Figure 3.10 SDS-PAGE purification of <i>e</i> UDG.	71
Figure 3.11 SDS-PAGE purification of <i>e</i> CYDG.	72
Figure 3.12 SDS-PAGE purification of <i>e</i> UDG (left) and <i>e</i> UYDG (right).	72
Figure 3.13 <i>e</i> CYDG Ni-NTA loading and elution profile.	73
Figure 3.14 Cartoon representation of the cleavage assay.	74
Figure 3.15 <i>e</i> UDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates.	75
Figure 3.16 <i>e</i> UYDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates.	76
Figure 3.17 <i>e</i> CYDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates.	77
Figure 3.18 Effect of pH (A) and temperature (B) on the activity of <i>e</i> CYDG against uracil, thymine, cytosine and 5-methylcytosine substrates.	78
Figure 3.19 <i>e</i> CYDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates.	80
Figure 3.20 <i>e</i> CYDG excision of G.C ^M C.	81
Figure 3.21 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates A.C, AP.C and Z.C.	82
Figure 3.22 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates G.U and A.U.	84
Figure 3.23 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates ssC(polyA) and ssC(GAT).	85
Figure 3.24 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates gap.C, long gap.C and HEG.C.	87
Figure 3.25 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrate I.C.	88
Figure 3.26 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates A.C(G), A.C(AG) and A.C(GA).	90
Figure 3.27 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates A.U(G) and HEG.C(G).	91

Figure 3.28 Kinetics of <i>e</i> CYDG cleavage of the 31mer substrates A(T).C(G), long HEG.C(G), APHEG.C(G), and I.C(G).	93
Figure 3.29 Kinetics of <i>e</i> CYDG cleavage the 31mer substrate A.C pH 7.4.....	94
Figure 4.1 Sequences produced to vary the distance between the LacZ and <i>e</i> UDG start codons.	103
Figure 4.2 The DNA sequences of the region between the start codons of LacZ (5´ blue) and <i>e</i> UDG (3´ blue).....	103
Figure 4.3 Oligonucleotide primers designed for mutation of the upstream thrombin cleavage site.	104
Figure 4.4 Cartoon of the pET <i>e</i> UDGUgi construct.	107
Figure 4.5 Colony PCR to determine positive pET <i>e</i> UDGUgi clones.....	108
Figure 4.6 Sequencing analysis of pET <i>e</i> UDGUgiST.	109
Figure 4.7 DNA and protein sequence alignments of bacteriophage PBS1 and <i>E. coli</i> to show optimisation for expression in <i>E. coli</i>	110
Figure 4.8 Complete gene synthesis of <i>h</i> UDGΔ81Ugi.	111
Figure 4.9 SDS-PAGE purification of <i>e</i> UDGUgiST.	112
Figure 5.1 Complete gene synthesis of <i>h</i> UDG.....	117
Figure 5.2 Generation of <i>h</i> CDGΔ81.	118
Figure 5.3 SDS-PAGE showing the purification of <i>h</i> CDG.	119
Figure 5.4 <i>h</i> CDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates.	120
Figure 5.5 <i>h</i> CDG excision of G.C ^M C.	121
Figure 5.6 Kinetics of <i>h</i> CDG cleavage of the 31mer substrates A.C, AP.C and Z.C.....	122
Figure 5.7 Kinetics of <i>h</i> CDG cleavage of the 31mer substrates G.U, A.U and HEG.C....	124
Figure 5.8 Kinetics of <i>h</i> CDG cleavage of the 31mer substrates ssC(polyA) and ssC(GAT).	125
Figure 5.9 Kinetics of <i>h</i> CDG cleavage of the 31mer substrates G.C(AT), A.C(G) and I.C.	127
Figure 5.10 Kinetics of <i>h</i> CDG cleavage of the 31mer substrate A(T).C(G).	128
Figure 5.11 Kinetics of <i>h</i> CDG cleavage of the 31mer substrates gap.C and long gap.C.	129
Figure 5.12 Kinetics of <i>h</i> CDG cleavage of the 31mer substrate A.C pH 7.4.	130
Figure 5.13 Kinetics of <i>h</i> CYDG cleavage of the 31mer substrate A.C.	131
Figure 5.14 Kinetics of <i>h</i> CYDG cleavage of the 31mer substrates AP.C, Z.C and HEG.C.	132

Figure 5.15 Kinetics of <i>h</i> CYDG cleavage of the 31mer substrates ssC(GAT), G.C and A.C(G).	133
Figure 5.16 Kinetics of <i>h</i> CYDG cleavage of the 31mer substrate G.U.	134
Figure 6.1 Proposed method for detecting 5-methylcytosine using <i>e</i> CYDG.	139
Figure 6.2 Oligonucleotide sequences of the cytosine target and cleaved cytosine product mimics.....	140
Figure 6.3 Real-time PCR analysis of a 1 nM C, ^M C, and a cleaved C mimic, target oligonucleotide with a 10 nM HEG probe.....	141
Figure 6.4 Real-time PCR concentration gradient analysis of 10 nM to 0.1 fM C.	141
Figure 6.5 Real-time PCR concentration gradient analysis of 10 pM to 10 fM C.	142
Figure 6.6 Binding of protection oligonucleotides to the target DNA.	144
Figure 6.7 <i>e</i> CYDG cleavage of the C- and ^M C-containing substrates.....	145
Figure 6.8 <i>e</i> CYDG cleavage of C- and ^M C-containing substrates in the presence of protecting oligonucleotides.....	146
Figure 6.9 <i>e</i> CYDG cleavage of C- and ^M C-containing oligonucleotide templates with different backbone cleavage agents.	147

List of Tables

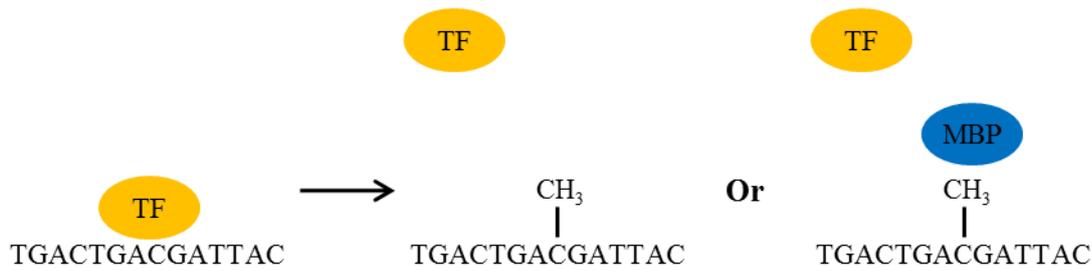
Table 1.1 Common DNA glycosylases.....	5
Table 2.1 Plasmid constructs used and generated in this thesis.....	43
Table 2.2 Sequences of oligonucleotide primers used in site-directed mutagenesis.	47
Table 2.3 Composition of agar plates.	48
Table 2.4 Oligonucleotides used in excision assays.	56
Table 2.5 Oligonucleotides used in the cytosine detection assay. H; HEG.	59
Table 3.1 Theoretical DNA sequences of the different enzymes at the mutagenic sites.	63
Table 3.2 <i>e</i> CYDG reaction rates.	83
Table 3.3 A.C oligonucleotides used to assess the effect of flanking regions on the rate of excision.	89
Table 3.4 Sequences of A(T).C(G) and long HEG.C.	92
Table 5.1 G.C oligonucleotides used in excision assays.....	120
Table 5.2 <i>h</i> CDG reaction rates.....	123
Table 5.3 <i>h</i> CYDG reaction rates.....	130
Table 6.1 Oligonucleotides used in the cytosine detection assay. H; HEG.	138
Table 6.2 Primers and protection oligonucleotides used in the cytosine detection assay..	143

Chapter 1: General Introduction

1.1 DNA and Epigenetics

DNA is the molecule that is responsible for the storage of genetic information. Gene expression is regulated by many cellular processes that usually involve protein interaction. This can cause up- or down-regulation of a specific gene leading to alterations in its expression. One way in which gene expression is regulated is through the methylation of cytosine bases within DNA, a form of epigenetics, which usually results in gene silencing. Errors in this have been implicated in diseases such as cancer (Jones and Baylin, 2002). Cytosine methylation most commonly occurs at CpG sites that are found in high density around promoter regions of genes, and are hotspots for mutations (Shen *et al.*, 1992) as 5-methylcytosine can be deaminated to create thymine, producing a G.T mismatch (Sartori *et al.*, 2002). Alterations to the methylation patterns of these regions can therefore affect gene expression and regulation. It has been found that methylation of cytosines inhibits gene expression through two main mechanisms. The first is whereby the methylation of cytosine itself prevents the binding of DNA binding proteins (Figure 1.1A), such as transcription factors (Watt and Molloy, 1988). This is because sequence recognition is disrupted by the methyl group protruding into the major groove. The second, and probably the most important, occurs through the binding of methyl-CpG-binding proteins (MBPs), to methylated cytosines. MBPs can either repress gene expression directly by occupying the binding domains of DNA binding proteins (Figure 1.1A), or recruit other repressor proteins to cause gene silencing through chromatin remodelling (Figure 1.1B) (Jones *et al.*, 1998). One such class of protein that is recruited by MBPs are histone deacetylases (HDACs). The recruitment causes the HDACs to come into close proximity to the N-terminal tail of histone 3 that is part of a nucleosome around which the DNA is wound. The HDACs are then able to cause deacetylation of lysine 4 and 9 (Figure 1.1B) and its subsequent methylation by a histone methyltransferase (HMT), also recruited by MBPs, which restores the positive charge on the lysine. This increases the affinity of DNA for the nucleosome due to charge interactions, causing stronger DNA binding and chromatin remodelling into a condensed and thus repressed form (Nan *et al.*, 1998).

A



B

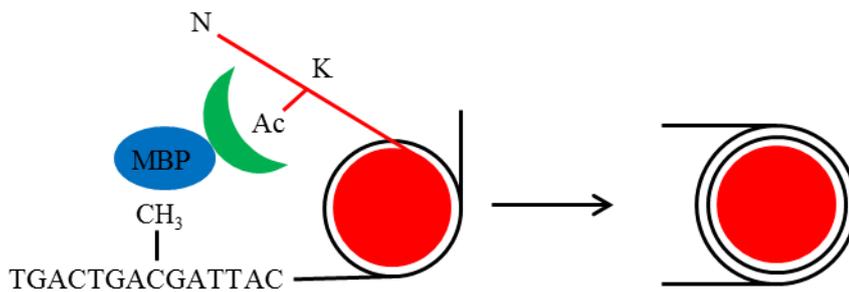


Figure 1.1 Gene regulation by cytosine methylation. A) Binding of methyl binding proteins (MBPs; blue) prevent the binding of transcription factors (TF; orange). B) MBPs recruit histone deacetylases (HDACS; green) that deacetylate lysines in the N-terminal tail of histones (red). This restores the lysines positive charge, increasing the affinity for DNA (black lines) creating a condensed chromatin form.

More recently a new “6th” base has been found that could also play a role in gene regulation, 5-hydroxymethylcytosine (Kriaucionis and Heintz, 2009, Tahiliani *et al.*, 2009). This base is generated through the addition of a hydroxyl group to 5-methylcytosine *via* Ten Eleven Translocation (TET) proteins. As no cytosine demethylation enzymes are known it has been proposed that this acts as an intermediate in the demethylation of 5-methylcytosine or possibly to act as another epigenetic marker (Tahiliani *et al.*, 2009). More recent studies seem to suggest that it is indeed an intermediate in demethylation (He *et al.*, 2011, Jin *et al.*, 2014) and further support comes from the role of thymine DNA glycosylases (Hashimoto *et al.*, 2012, Muller *et al.*, 2014). These are specific for uracil and thymine substrates when they are mispaired with guanine, but has no activity towards cytosine, even though 5-carboxylcytosine is a substrate for mammalian thymine DNA glycosylase (He *et al.*, 2011). They are therefore able to excise the deaminated products of cytosine, 5-methylcytosine and 5-hydroxymethylcytosine (*i.e.* uracil, thymine and 5-hydroxymethyluracil respectively) (Hashimoto *et al.*, 2012, Morera *et al.*, 2012). It is not surprising that this seems to be the route for demethylation as a DNA glycosylase capable of excising 5-methylcytosine and 5-hydroxymethylcytosine, might also be able to

accommodate cytosine, which would result in constant genome lesions, generating a high mutation rate.

The inclusion of 5-methylcytosine within the genome does not only affect gene expression regulation; it also has potent mutagenic properties, as deamination of this base generates thymine, which upon subsequent DNA replication results in a transition mutation (G.^MC to A.T). Since the resulting G.T mismatch is made from canonical bases it would then not be obvious which is the incorrect base that should be excised. The high rate of deamination of 5-methylcytosine, and its slow repair rate, accounts for the high mutation and relatively low abundance of CpG sites. It is for this reason that the dinucleotide CpG is underrepresented in all mammalian genomes (Lindahl, 1993, Mol *et al.*, 1995b). Deamination of unmethylated cytosine is also common at CpG sites generating UpG. This can also lead to a transition mutation, but uracil, unlike thymine, is not a standard DNA base and so can be recognised as a lesion to be repaired. The generation of uracil in DNA can arise from both enzymatic and spontaneous deamination (Mol *et al.*, 1995b).

Tools for the detection of 5-methylcytosine are important for assessing the epigenetic status of any CpG sites, and might be useful for identifying the cause of some diseases. The main method currently used in the detection of 5-methylcytosine is bisulphite sequencing (discussed in section 1.4). Although this is routinely used, it suffers from several drawbacks and requires DNA sequencing. We therefore plan to design an enzymatic approach for the detection of 5-methylcytosine that will prove to be accurate and reliable. The aim of this project is to generate an enzyme capable of discriminating between cytosine and 5-methylcytosine. This will be achieved by selective mutation of the enzyme Uracil DNA glycosylase (UDG; also known as UNG). In order to do this we must first understand how UDG discriminates between uracil and thymine, and apply it for the discrimination between cytosine and 5-methylcytosine (Figure 1.2).

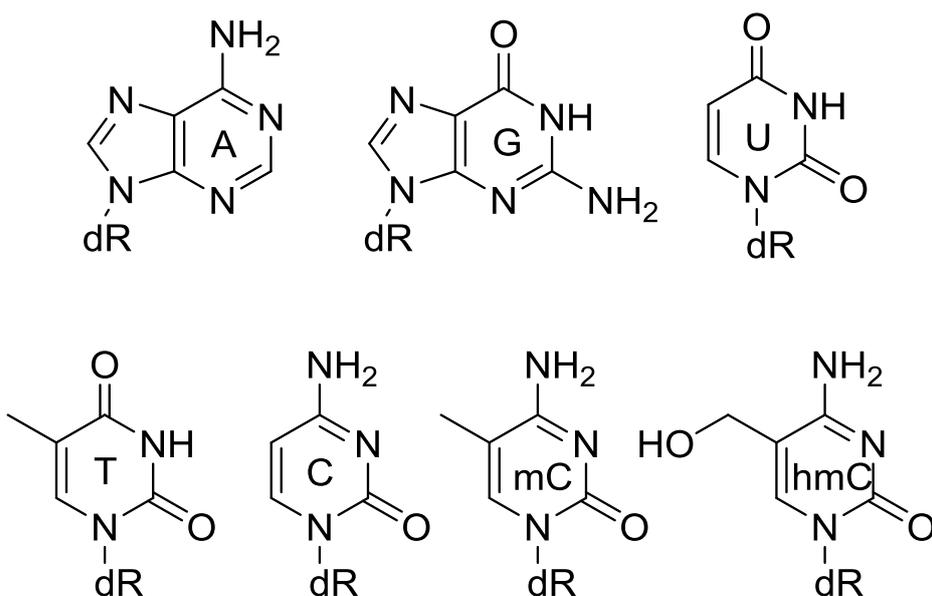


Figure 1.2 DNA bases. A: adenine, G: guanine, U: uracil, T: thymine, C: cytosine, ^mC:5-methylcytosine and hmC: hydroxymethylcytosine, dR: deoxyribose.

1.2 DNA Glycosylases

DNA is constantly under attack from endogenous and exogenous reagents, which cause damage through incorporation or formation of cytotoxic and mutagenic bases, that are either analogues or adducts of the four Watson-Crick bases (Lindahl, 1993). Surprisingly, endogenous agents cause the most damage, such as oxygen that creates 8-oxoguanine *via* hydroxyl radicals, and water that causes deamination of cytosine and 5-methylcytosine to uracil and thymine respectively (Lindahl, 1993, Sartori *et al.*, 2002). These reactions can be accelerated by higher temperatures (Lindahl and Nyberg, 1974) and by mutagens such as nitrous acid (Savva *et al.*, 1995). In eukaryotes, this can cause temporary cell-cycle arrest that then allows DNA repair to take place before replication begins (Krokan *et al.*, 1997). A list of common glycosylases and their substrates are displayed in Table 1.1.

Glycosylases that have broad substrate specificity are able to recognise the damaged base itself as it no longer forms hydrogen bonds to its complement or does not stack correctly within the duplex, enabling the base to be flipped out of dsDNA more easily into the enzyme's active site (Berdal *et al.*, 1998). Some of these glycosylases also admit normal, undamaged bases into their active site such as the *E. coli* enzyme 3-methyladenine DNA glycosylase II (ALKA) that admits adenine (Drohat *et al.*, 2002). Interestingly its counterpart 3-methyladenine DNA glycosylase I is specific for only 3-methyladenine (Berdal *et al.*, 1998).

The thermostability of DNA containing a damaged base may also affect the efficiency of the repair process, especially if this alters its ability to hydrogen bond with its complementary base or stack with the neighbouring bases. This is because many repair proteins work by flipping the damaged base out from the duplex into their active site; the greater the destabilisation of the lesion the less energy will be required for this step, thereby facilitating repair (Sagi *et al.*, 2000). Increased thermostability of DNA can be advantageous to thermophilic glycosylases that require high DNA melting temperatures to aid repair (Sagi *et al.*, 1999). The thermostability of damaged DNA can also be dependent on its flanking sequence (Sagi *et al.*, 2000), by influencing the base pair conformation and/or base stacking (Singer and Hang, 1997).

Glycosylase	Species	Substrates
UDG	<i>E. coli</i> , human	U
TDG	Human (MBD4), <i>E. coli</i> (MUG)	T.G, U.G
SMUG1	Human	U, hoU, hmU
ALKA	<i>E. coli</i> , human (MPG)	3-meA, hypoxanthine
OGG1	Human, <i>E. coli</i> (Fpg)	8-oxoG
MUTY	<i>E. coli</i> , human (NYH)	A.8-oxoG

Table 1.1 Common DNA glycosylases. Homologues for the different species are given in brackets. hoU: hydroxyuracil, hmU: hydroxymethyluracil, 3-meA: 3-methyladenine, 8-oxoG: 8-oxoguanine.

1.2.1 Repair of Damaged DNA

DNA repair is therefore crucial for maintaining the integrity of the genome. The simplest form of repair is through direct dealkylation in a one-step mechanism; i.e. the removal of alkyl groups from O6-methylguanine (O6-meG) by Ada (Volkert, 1988). The other forms of repair are more complicated and require the excision of a nucleotide; nucleotide excision repair (NER), or just the damaged base itself; base excision repair (BER). NER is the most complex system that involves around 30 proteins (Krokan *et al.*, 1997) and is responsible for removing long stretches of DNA as an oligonucleotide and works by recognising the large distortions produced upon the DNA duplex, for example by XPA (Robins *et al.*, 1991). New DNA is then incorporated into the unpaired excised region using the complementary strand as a template to complete repair. The BER pathway, however, removes only the base, and not the nucleotide as in the NER pathway. DNA

glycosylases in the BER pathway remove damaged, cytotoxic or mutagenic bases through cleavage of the N-glycosidic bond between the target base and the deoxyribose sugar; thus releasing the base and leaving an abasic (AP) site (Pearl, 2000). The base is only excised when it is in an extrahelical conformation (Jiang and Stivers, 2002) in which it is able to access the active site of the glycosylase. Glycosylases do not cause major helical distortions to DNA, though some distortions are inevitable in order to aid formation of the extrahelical conformation (Pearl, 2000). DNA glycosylases stay tightly bound to their target, to protect the abasic site after base excision, until the next repair enzyme, an apurinic endonuclease, displaces the glycosylase in order for a polymerase to incorporate the correct base (Hoseki *et al.*, 2003). This is to prevent further mutations from occurring. A glycosylase that has been extensively studied is UDG, of which five families have currently been identified. This glycosylase excises uracil from DNA and is an enzyme involved in the BER pathway (Krokan *et al.*, 2001).

1.2.2 How Uracil Arises in DNA

Uracil can arise in DNA through misincorporation of dUMP, though this causes no adverse effects as it is replicated in exactly the same way as thymine. Though not mutagenic, the incorporation of uracil can affect the interaction of some DNA binding proteins (Handa *et al.*, 2002) thereby affecting regulatory DNA processes. Uracil also arises in DNA from deamination of cytosine, producing a guanine:uracil mismatch that results in a transition mutation (Figure 1.3). The role of UDG is therefore to excise uracil so that the correct base can be incorporated thereby preventing a transition mutation of G.C to A.T (Pearl, 2000). The importance of UDG is apparent when looking at a genome of 10^{10} base pairs (Pearl, 2000), in which up to 500 uracil bases per cell will be produced by deamination every day (Lindahl, 1993). Deamination is also approximately 4000 times faster in ssDNA, explaining why it is enhanced during transcription and replication (Mosbaugh, 1988).

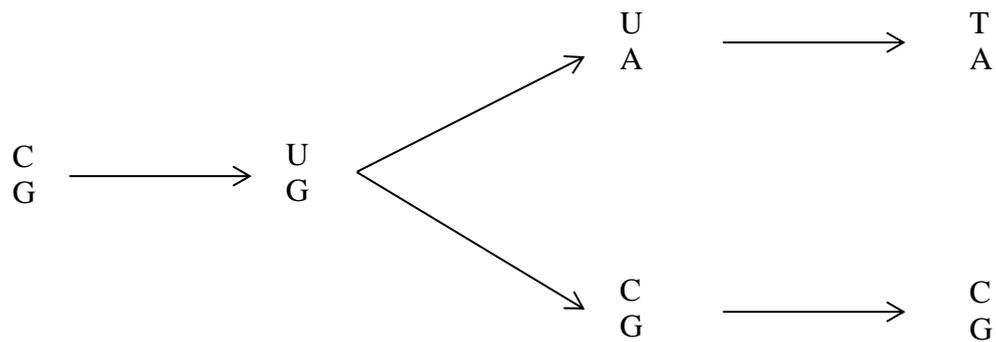


Figure 1.3 The transition mutation from G.C to A.T, caused by cytosine deamination.

UDG is responsible for protecting the genome and has a much higher enzymatic rate than other glycosylases (Boiteux *et al.*, 1990, Bjelland *et al.*, 1994, Neddermann and Jiricny, 1994, Roy *et al.*, 1994). This is probably due to its high selectivity for uracil and to the tight and specific active site that allows for quick catalysis. It may also be because most other glycosylases have activity towards two or more substrates and therefore have a lower specificity per substrate in comparison to uracil and UDG. This would result in weaker binding and a lower rate of catalysis and excision (Kavli *et al.*, 1996).

Uracil is a natural component of RNA, though this is not excised by UDG as the 2'OH group prevents admission into the active site (Mol *et al.*, 1995b, Slupphaug *et al.*, 1996), where there would be a steric clash with a phenylalanine (Savva *et al.*, 1995), in a similar fashion to the way that the methyl group that excludes thymine (Kavli *et al.*, 1996).

1.3 Uracil DNA Glycosylases

1.3.1 Family 1 UDGs

UDG found in humans, *E. coli* and *Herpes Simplex Virus 1* (HSV1) belongs to Family 1 of the UDG superfamily. Within humans there are two forms of UDG created from alternative splicing of the UDG gene, UNG1 and UNG2; UNG1 is targeted to the mitochondria and UNG2 the nucleus (Nilsen *et al.*, 1997, Haug *et al.*, 1998). This family of enzymes acts on uracil in both single stranded DNA (ssDNA) and double stranded DNA (dsDNA), but has no detectable activity towards dUMP, deoxyuridine or uridine (Lindahl, 1974). They are able to recognise uracil in an extrahelical conformation and cause excision of the base through cleavage of the N-glycosidic bond (Mol *et al.*, 1995b).

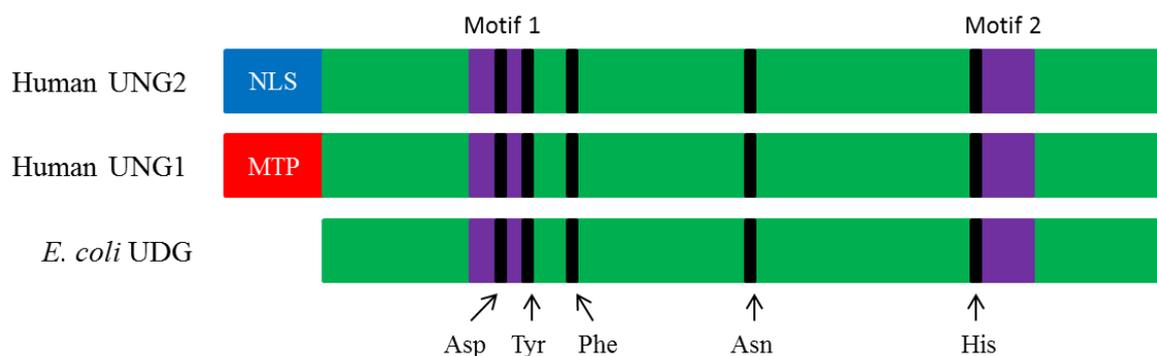


Figure 1.4 Cartoon alignment for comparison of human and *E. coli* UDGs. The human forms contain a nuclear localisation signal (NLS) or mitochondrial targeting peptide (MTP) at their N-terminals. Critical residues (black boxes) and motifs 1 and 2 (purple boxes) highlighted.

1.3.1.1 Recognition of Uracil by UDG

It is still unclear as to how UDG locates its substrate but the first stage is non-specific DNA binding (Dong *et al.*, 2000). So far three models have been suggested. The first is the inherent extrahelicity model that suggests that uracil has weak binding with its opposing base and spontaneously flips out from the helix, enabling enzyme binding. This seems unlikely as the mismatch pairing of G.U, or even A.U, is very similar to the correct base pairings with similar bond energies, suggesting that spontaneous base flipping is highly unlikely to occur (Pearl, 2000). The two-stage recognition is a hybrid model whereby UDG recognises uracil in DNA and then causes the base to flip into an extrahelical conformation (Pearl, 2000). This is similar to the mechanism of methyltransferases, which recognise a specific DNA sequence *via* major groove interactions of the fully stacked base-paired B-DNA conformation, resulting in the flipping out of the base so that it can undergo methylation (Klimasauskas *et al.*, 1994). UDG has an approximately 20,000 fold larger rate of activity than methyltransferases, suggesting that it plays an active role in flipping of the nucleotide (Slupphaug *et al.*, 1996). A G:U pair has a “wobble” structure (Figure 1.5) as the N7 and O6 of guanine and the O4 of uracil protrude into the major groove; suggesting a means for recognition (Pearl, 2000). The “wobble” created from abnormal base stacking may be more easily recognised from the minor groove that has less variation, as UDG has most contact with the minor groove. This is because most base modifications, *i.e.* methylation, protrude out of the major groove making detection of changes harder to recognise. Minor groove interactions by UDG may be an important factor in the recognition of a damaged and/or mismatch base (Slupphaug *et al.*, 1996). The A.U pair also has a distinct major groove conformation, which is different to that of G.U. This

makes the two stage recognition model also unlikely to be sufficient as the enzyme would have to have multiple recognition processes in order to recognise uracil opposite guanine, adenine or in ssDNA.

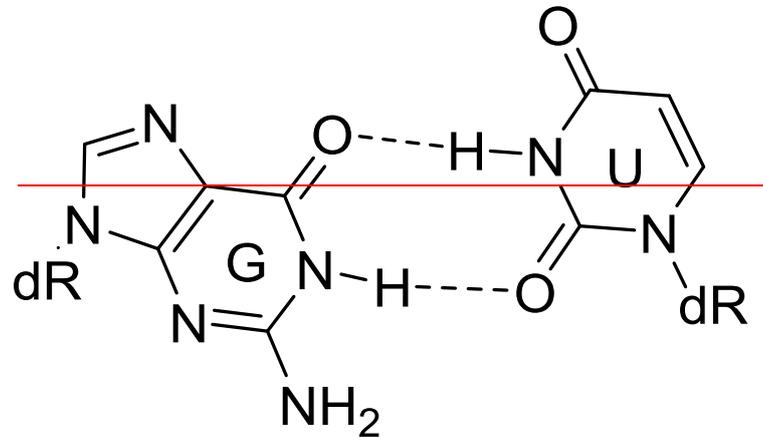


Figure 1.5 A G.U wobble base pair. Above the red line, indicating the axis of the pair, are the three hydrogen bond acceptors protruding into the major groove: guanine's N7 and O6, uracil's O4. dR: deoxyribose.

The base sampling model on the other hand suggests that UDG samples bases along the DNA and flips them out in order to check their interaction within the active site (Pearl, 2000). From this, the mechanism suggested for UDG is stochastic "hopping" (Slupphaug *et al.*, 1996, Jiang and Stivers, 2002, Rowland *et al.*, 2014) whereby it slides along the DNA *via* a charge interaction between the negative phosphate backbone (Friedman and Stivers, 2010, Zharkov *et al.*, 2010, Schonhoft *et al.*, 2013) of the DNA and the positive face (Figure 1.6) of UDG, where the active site is located.

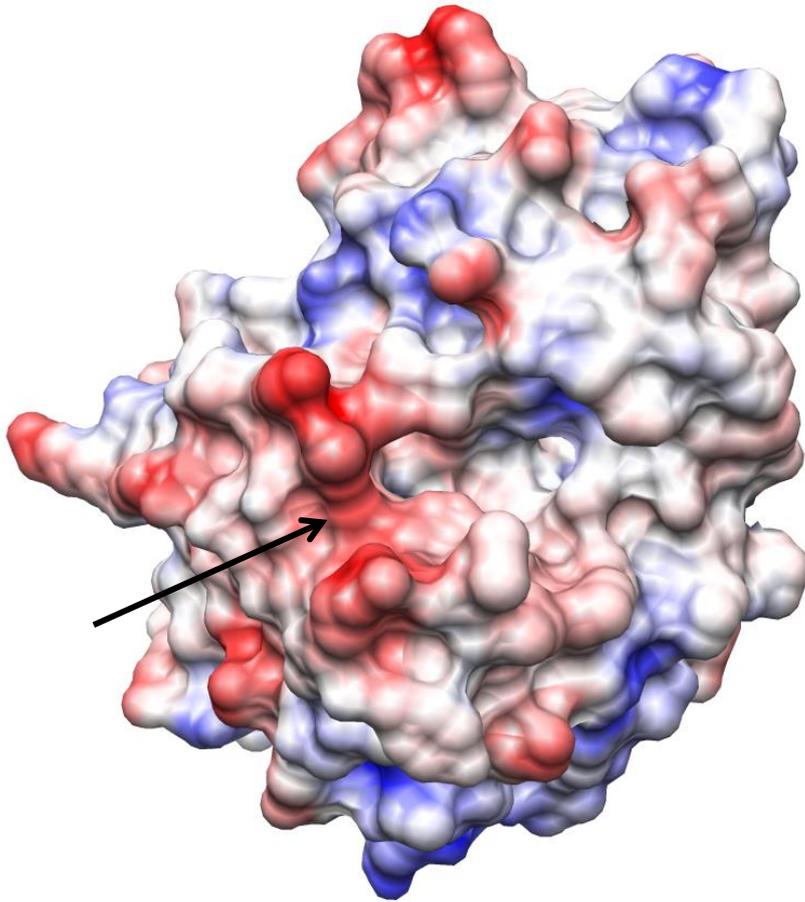


Figure 1.6 Space filled model of showing the positive face of *hUDG* (arrow) next to its active site. (PDB 1AKZ) (Parikh *et al.*, 1998). Positive charge: red, neutral; white and negative; blue. (Diagram created using Chimera V1.5.3.)

UDG scans for uracil bases in sections in a distributive manner (Purmal *et al.*, 1994), which is affected by the context of uracil distribution (Schonhoft and Stivers, 2013). UDG has a scan range of approximately 1.5 – 2 kb before dissociating (Higley and Lloyd, 1993). This means that UDG is not able to sample every base, as the time it spends in contact with the DNA is very short (Panayotou *et al.*, 1998, Jiang and Stivers, 2002). UDG therefore acts in a selective manner and would have to dissociate after excision in order to release the free uracil base (Slupphaug *et al.*, 1996). This mechanism is possible provided UDG does not completely detach from the DNA and maintains its active site orientation in order to excise uracil once it has been located. The scanning process is halted when UDG arrives at a mismatch, such as a G.U. The displacement caused by this wobble base pair causes a clash with a leucine residue (Leu275 in *hUDG*) on the surface of UDG. This in turn produces a rotation in the side chain of a tyrosine (Tyr275 in *hUDG*) causing UDG to halt and allow insertion of the leucine residue (Parikh *et al.*, 1998). The enzyme therefore acts

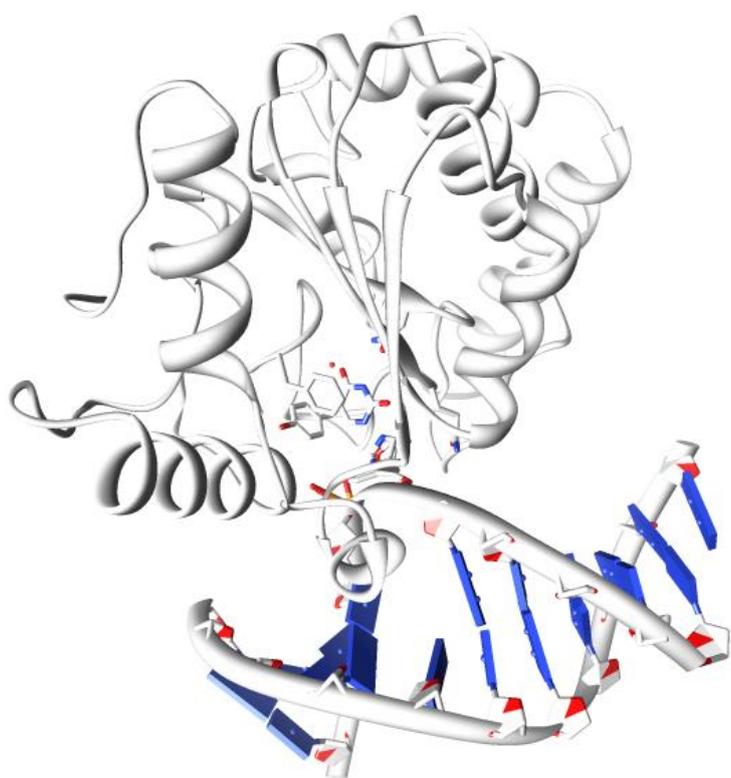
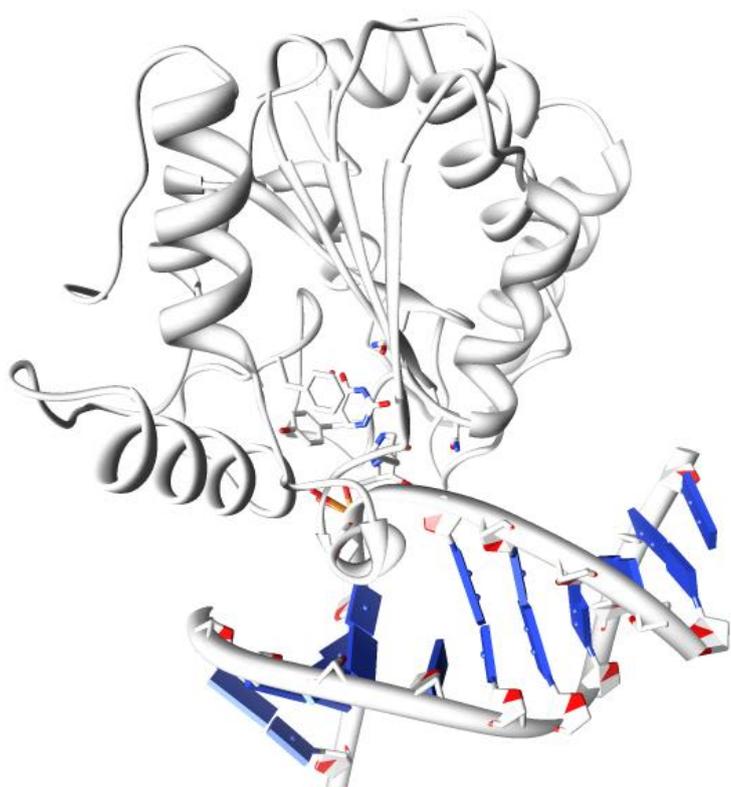
through what is known as a “Pinch, Push, Plug and Pull” mechanism (Jiang and Stivers, 2002). UDG has a preference for cleaving uracil in ssDNA as there is no opposing base and it can easily be disrupted and flipped into the active site. Base pairing stability determines the efficiency of substrate interaction in dsDNA (Krosky *et al.*, 2004, Krosky *et al.*, 2005); a uracil paired with guanine is less stable than uracil paired with an adenine. This is because an A.U pair is most similar to standard Watson-Crick base pair stacking. This makes disrupting the base pair interactions more difficult, hindering the flipping of the base into an extrahelical conformation. A guanine opposite uracil would be favoured by UDG due to its weaker base pair interactions within the DNA duplex. This has been supported by experimental data and provides the substrate preference for UDG as ssU > dsG.U >> dsA.U (Panayotou *et al.*, 1998, Pearl, 2000).

UDG therefore recognises DNA in three stages: (i) weak non-specific binding, (ii) destabilisation of uracil into an extrahelical conformation, and (iii) rapid docking of the uracil into the active site. This was first shown through the use of a deoxyuridine analogue, 2'-fluorouridine (2'-FU), positioned adjacent to 2-aminopurine (2-AP). A change in fluorescence is seen by 2-AP when 2'-FU substitute is flipped out into an extrahelical state (Stivers *et al.*, 1999), due to changes in local base stacking. This mechanism has since been shown using natural deoxyuridine (Wong *et al.*, 2002).

1.3.1.2 Structure of UDG

The crystal structure of UDG (Figure 1.4) has shown that it contains no disulphide bridges and has its C and N-termini on opposite sides of the enzyme. It has an α/β fold that creates a groove with an approximate diameter of 21 Å. This also happens to be the approximate width of a DNA duplex, and hence suitable as its binding site. The groove narrows to about 10 Å and suggests that the DNA duplex does not fit entirely and that a conformational change is required. However, the 10 Å end has three loops that are proline rich and provides the groove with a rigid structure, suggesting that a conformational change is unlikely (Mol *et al.*, 1995b). The active site appears to be flexible as there is an 11° difference in the plane of uracil between *e*UDG and *h*UDG (Werner *et al.*, 2000). Therefore this is a perfect site for DNA binding due to its shape, size and charge, which is created by the basic amino acids that line the groove and provide a positive electrostatic potential in which to complement the DNA and allow for binding of duplex DNA (Mol *et al.*, 1995b).

A



B

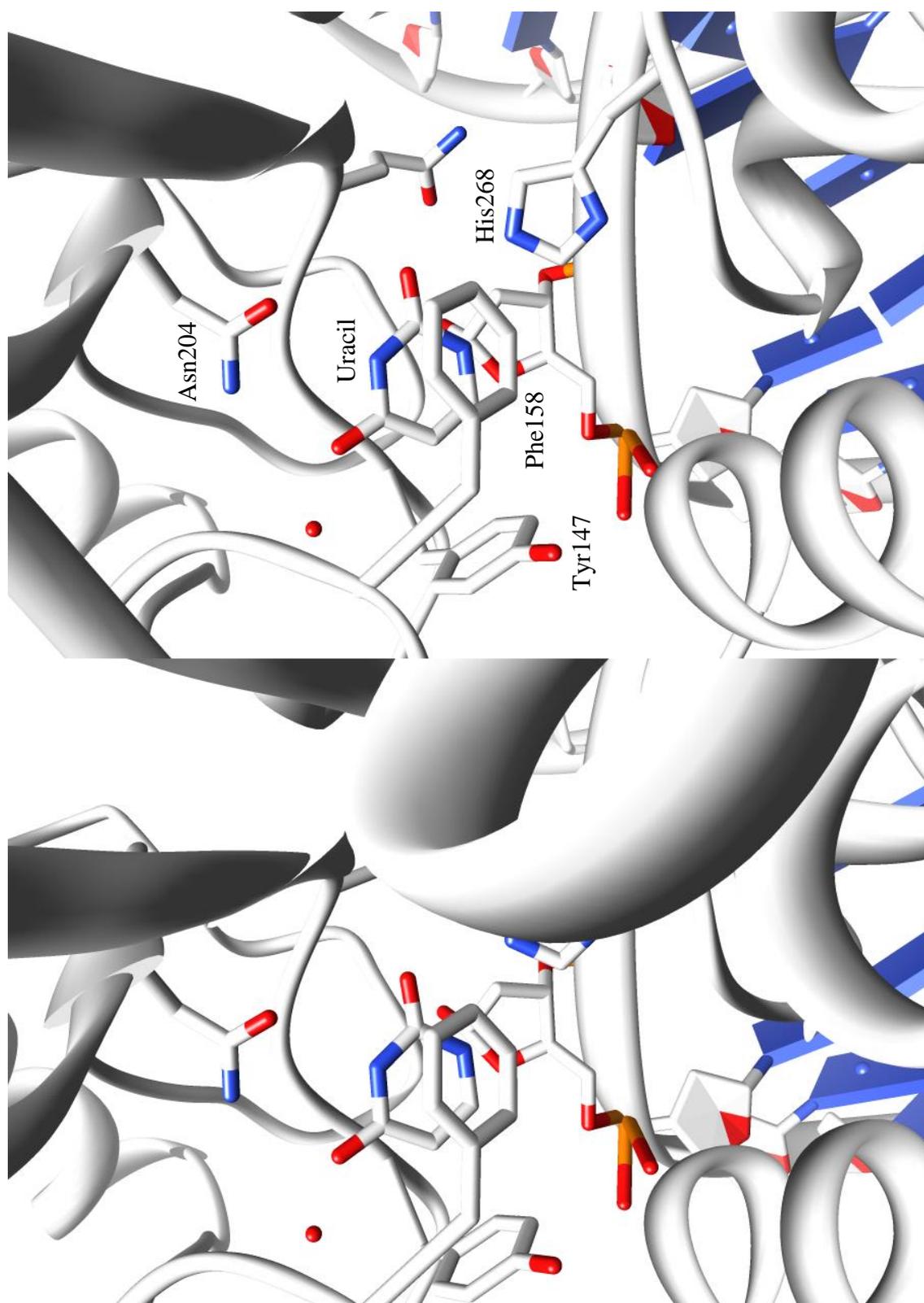


Figure 1.7 Structure of *h*UDG bound to dsDNA (PDB 1SSP) (Parikh *et al.*, 1998). A) Stereo view of UDG bound to uracil containing DNA. B) Stereo view of uracil bound in the active site of UDG with the critical residues highlighted. Oxygen atoms coloured red and nitrogen atoms coloured blue. *e*UDG counterparts for the critical residues are Tyr66, Phe77, Asn123 and His187. (Diagrams created using Chimera V1.5.3.)

1.3.1.3 DNA-Enzyme Binding

UDG only has a weak interaction with non-uracil containing DNA and upon addition of uracil into DNA the enzyme's affinity increases; favouring a model in which uracil recognition is due to recognition of the base itself (Panayotou *et al.*, 1998). This is shown through the first stage of DNA recognition whereby UDG recognises the DNA non-specifically, upstream of a mispaired uracil, causing distortion of the phosphodiester backbone and consequently helical strain (Parikh *et al.*, 1998). Binding of UDG to DNA results in DNA bending, which is important as the binding of undamaged DNA would result in steric clashes with the serine loops (described below), if it were not to be bent (Parikh *et al.*, 1998). This is because the 3' end of the DNA binds tightly to the enzyme and stays in a stacked conformation. This would result in the 5' end clashing with leucine 191 (which acts as a wedge), the nucleotide and the critical asparagine. As a result UDG is unable to excise uracil at the 3' end of a DNA fragment, but is able to excise uracil at the 5' end, provided it is phosphorylated (Krokan and Wittwer, 1981, Varshney and van de Sande, 1991). This mechanism is conserved throughout the UDG superfamily.

1.3.1.3.1 Serine Interactions

It has been demonstrated with *E. coli* UDG that once the uracil has been located, three important conserved serine residues, Ser88, 189 and 192, bind *via* their hydroxyl side chains to the +1, -1, -2 and -5 phosphates of the uracil base (reading 5' to 3') (Werner *et al.*, 2000, Handa *et al.*, 2002). The binding of Ser88 and Ser189 to the flanking 5' and 3' phosphates causes a decrease of 4 Å in the phosphate-phosphate distance, which is the approximate distance caused by the unwinding of a single nucleotide (Jiang and Stivers, 2002). The decrease in distance causes compression of the DNA around the area in which the uracil base is located. DNA torsion, caused by the serine interaction, is relieved through the bending of the DNA by approximately 45° towards the major groove (Parikh *et al.*, 1998, Jiang and Stivers, 2002), causing disruption of duplex stacking. This causes flipping of the mismatched uracil nucleotide into an extrahelical conformation (van Aalten *et al.*, 1999) that is then able to be admitted into the highly specific active site of UDG (Werner *et al.*, 2000), resulting in excision of the base. This binding around the deoxyuridine is referred to as the pinching mechanism (Parikh *et al.*, 1998, Stivers *et al.*, 1999, Jiang and Stivers, 2002). Substrate binding brings about a clamping motion of the serine loops which aid positioning of the substrate within the active site (Werner *et al.*,

2000). Mutation of Ser88 and Ser189 results in decreased binding, as also seen with the removal of the hydroxyl of Ser88, while mutation of both Ser88 and Ser189 reduces UDG's binding affinity significantly. The conformational change of UDG is the rate limiting step for these mutants showing that their role is in the formation of uracil's extrahelical conformation. These mutants have no effect on the enzyme's ability to flip the base into an extrahelical conformation or the ability of UDG to clamp around the base to form the final conformation before excision occurs. Thus, the serines play a crucial role in the early stages of UDG's excision mechanism.

It has also been suggested that the serines help to attain the final conformation after glycosidic bond cleavage, through cooperatively acting together (Jiang and Stivers, 2002). Their importance is also highlighted through a decrease in enzyme activity when they are mutated to alanine. Serine interactions with the 5' phosphodiester of uracil was shown to be most important in binding and catalysis while surprisingly having no significant effect on lowering the activation energy (Werner *et al.*, 2000). Though upstream of the target uracil base, the 5' phosphodiester of the -5 nucleotide contributes significantly to uracil excision (Handa *et al.*, 2002). The phosphodiester pinching caused by the serine residues contributes to the bending of the N-glycosidic bond along with aromatic stacking forces. This therefore weakens this bond and suggests that these residues may play an important role in lowering the activation energy for N-glycosidic bond cleavage, independent of DNA sequence effects (Werner *et al.*, 2000). These serines therefore play a role in the later stages of the cleavage mechanism, as well as initial binding of UDG.

1.3.1.3.2 Effect of Flanking Regions

As well as the flanking phosphates the flanking bases themselves also affect catalysis and excision. Uracil is best excised when the -2, -1, +1 and +2 bases are adenines. The efficiency is greatly reduced when these flanking bases are guanines (Slupphaug *et al.*, 1995). Complete disruption of the preferred sequence for optimal excision activity by UDG found that the -1 and -2 5' flanking bases have interactions that are most important for catalysis. However, the +1 base is not crucial for catalysis and only contributes to decreasing the activation energy (Jiang and Stivers, 2001). Therefore sequence specificity is important for excision and it has been shown that local and global interactions, such as DNA bound to histones, can greatly affect excision rates irrespective of sequence (Ye *et al.*, 2012).

1.3.1.3.3 Rate Limiting Factor of Excision

The flipping of the base is the rate limiting step as the rate of dissociation of UDG is similar for ssDNA and dsDNA, though ssDNA cleavage of uracil is three times faster than that of dsDNA. This is the same for different base pairings as the association of uracil in a G.U mismatch is approximately 15-fold faster than that of an A.U (Panayotou *et al.*, 1998). This is due to base pair stability whereby uracil is most similar to thymine and therefore has better binding with its natural partner, adenine. It is of note that UDG has a similar association rate (k_{on}) of recognition with both G.U and A.U (Parikh *et al.*, 1998), suggesting that the faster rate for G.U mismatches is not due to extrahelical recognition and that uracil spends more time in this conformation in a G.U mismatch than an A.U. This is also seen in enzymes from different species, which have different rates of excision for dsDNA but have similar rates for ssDNA (Eftedal *et al.*, 1993, Krokan *et al.*, 1997). The rate of uracil displacement from dsDNA can be affected by its surrounding sequence (Jiang and Stivers, 2002), most notably the adjacent 5' nucleotide (discussed previously) (Abu and Waters, 2003). The rate of excision can also be affected by the chain length suggesting that some interactions occur at a distance from the uracil base whereby longer 5' flanks of oligonucleotides facilitate uracil release (Handa *et al.*, 2002).

1.3.1.4 The Role of Leucine

Since flipping of the base by UDG is more frequent than spontaneous base flipping, it has been suggested that Leu191 (*e*UDG) protrudes into the minor groove and acts as a wedge to facilitate the extrahelical conformation. The leucine then acts as a plug to occupy the space left by the extrahelical base thereby increasing the time for the enzyme to act on the uracil, through maintaining its extrahelical conformation. In this way the leucine also serves to prevent reinsertion and the flipping back of uracil into the DNA duplex (Krokan *et al.*, 1997, Parikh *et al.*, 1998). Mutation of this leucine to alanine affects the enzyme's ability to bind and flip the base. This is because it loses the ability to "push" out the base and "plug" the space created and thus, this leucine is involved throughout UDG's excision mechanism (Jiang and Stivers, 2002).

1.3.1.4.1 Introduction of Pyrene into the Substrate

It was proposed that the introduction of a pyrene (Pn) residue (Figure 1.8D) opposite uracil (instead of a natural base) would force the uracil into an extrahelical formation, while not

compromising duplex stability. Due to the bulk of pyrene it is able occupy the space of the opposing base as well as its own (Jiang *et al.*, 2001). The pyrene thus acts like a wedge and is able to rescue the mutational effects of L191A/G (Figure 1.8).

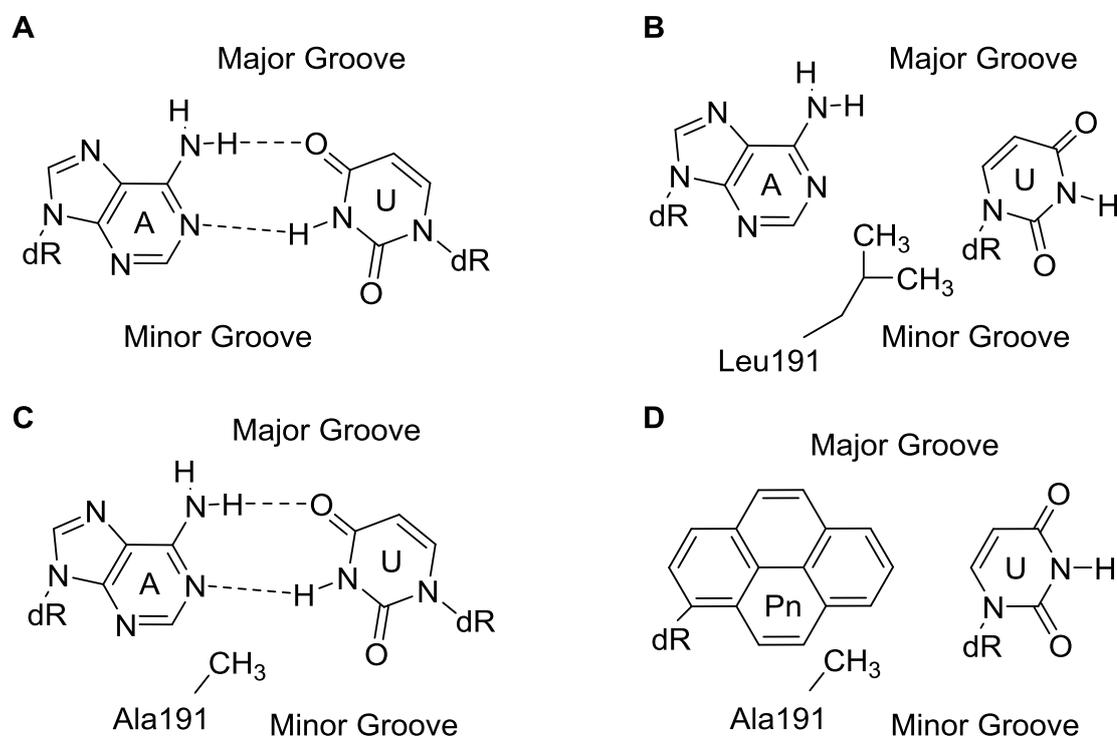


Figure 1.8 The “Push and Plug” roles of leucine. A) A.U base pair. B) Leucine acts as a wedge to force uracil into an extrahelical conformation, occupying the AP site. C) L191A mutation causes inhibition of UDG by preventing cytosine flipping. D) Incorporation of a pyrene (Pn) nucleotide to act as a wedge to force the uracil into an extrahelical conformation, rescuing the effects of the L191A mutation. dR: deoxyribose.

The L191G mutant shows a 60-fold decrease in activity, which is recovered by placing pyrene opposite uracil. Furthermore, UDG is rapidly able to attain the final state with uracil, as shown by a near 4-fold increase in catalytic rate of Pn.U compared to A.U. This is because the pyrene maintains the uracil in its extrahelical conformation thus increasing the stability (Jiang *et al.*, 2001, Jiang and Stivers, 2002). Fluorescence studies using 2'-FU have confirmed that pyrene acts as a surrogate for the leucine side chain (Jiang *et al.*, 2002b). This confirms that leucine has an important role in the pushing and plugging stages of the flipping process.

Pyrene is also able to rescue serine mutations involved with initial uracil recognition by preorganising the uracil into an extrahelical conformation. Though pyrene is able to rescue the effects of L191A by providing the driving force to deliver uracil in the active site, it is unable to rescue the effects of N123G or serine mutations. This is because it does not play a role in the stabilisation of uracil when bound in the active site (Jiang *et al.*, 2002b). This

is expected as pyrene is used purely as a substitute for leucine and forms no active site interactions (Mol *et al.*, 1995b, Parikh *et al.*, 1998). This therefore suggests that this asparagine binds late in the flipping process and only once the uracil is in the active site (Jiang *et al.*, 2002b).

1.3.1.5 Active Site Composition and Mechanism of UDG

Two short sequence motifs are conserved in the active site of Family 1 UDGs, even though some UDGs such as *h*UDG are larger enzymes. *h*UDG contains an extra 81 residues at the N-terminal that act as a signal sequence for post translational translocation to either the nucleus (UNG2) or mitochondria (UNG1) (Nilsen *et al.*, 1997, Krokan *et al.*, 2001). Motif 1 (also referred to as motif A) has the sequence GQDPY that contains the catalytic aspartate residue, Asp64, which activates a water molecule to form a nucleophile, and a tyrosine, Tyr66, that assists in active site specificity (Figure 1.9). Motif 2 (also referred to as motif B) has the sequence HPSPLSA and is responsible for complex stabilisation through minor groove interactions.



Figure 1.9 Partial ClustalW sequence alignment for the three commonly studied Family 1 UDGs. Active site residues are highlighted in red; motif 1 and 2 are boxed. * indicates positions which have a single, fully conserved residue; : indicates conservation between groups of strongly similar properties (scoring > 0.5 in the Gonnet PAM 250 matrix); . indicates conservation between groups of weakly similar properties (scoring =< 0.5 in the Gonnet PAM 250 matrix).

Motif 2 also contains a critical histidine, His187, responsible for stabilising the negative charge of uracil (Sartori *et al.*, 2002) in its anionic intermediate state (Werner and Stivers, 2000) upon cleavage of the glycosidic bond (Figure 1.10).

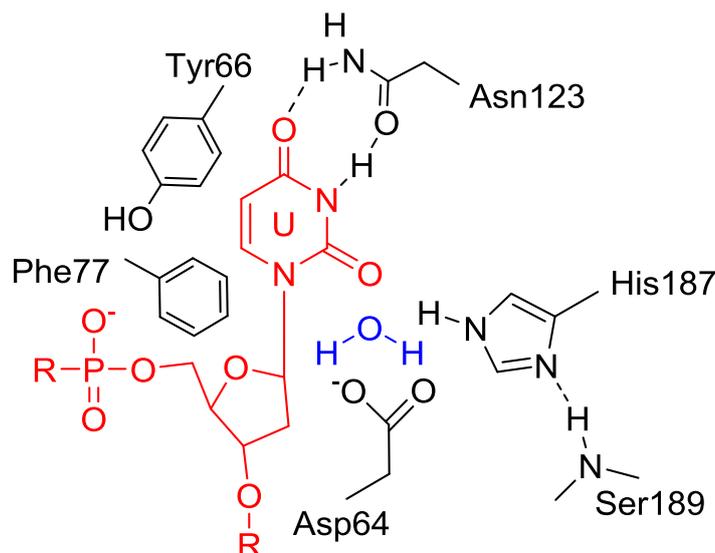


Figure 1.10 The critical residues lining the active site of *E. coli* uracil-DNA glycosylase. DNA highlighted in red and the water molecule in blue.

The energy gained upon UDG binding, through the serine and proline interactions with the DNA and uracil, starts an autocatalytic mechanism (Dinner *et al.*, 2001). Binding causes disruption of a hydrogen bond between Asp64 and His134 which allows the aspartate to rotate by approximately 120° (Parikh *et al.*, 1998, Xiao *et al.*, 1999). Asp64 is then correctly positioned to bind a water molecule between its side chain carboxyl and main chain carbonyl, (Dodson *et al.*, 1994, Savva *et al.*, 1995) which subsequently allows for the deprotonation of the water molecule to occur (Lindahl, 1974, Werner *et al.*, 2000) forming an OH^- nucleophile, which is then able to undergo nucleophilic attack on the C1' atom (Mol *et al.*, 1995b, Werner *et al.*, 2000, Jiang and Stivers, 2002) of the N-glycosidic bond. This is the bond that connects the base to the deoxyribose sugar of the DNA backbone, thus releasing the uracil base and creating an AP site (Pearl, 2000).

Cleavage of the glycosidic bond results in protonation of the O2 of uracil, thereby increasing the leaving ability of the base (Savva *et al.*, 1995) and creating the transition state (Mol *et al.*, 1995b, Drohat *et al.*, 1999b) that is stabilised by His187 (His268 in *hUDG*). His187 interacts with the 3' phosphate of uridine, enabling it to transfer a charged hydrogen bond to the O2 of uracil, after moving approximately 2 \AA to be in a position to interact with the O2, and thus aiding cleavage of the N-glycosidic bond through stabilisation of the transition state oxyanion (Slupphaug *et al.*, 1996, Krokan *et al.*, 1997, Drohat and Stivers, 2000, Werner and Stivers, 2000) (Figure 1.11). The movement of His187 by 2 \AA is restricted by the 2'OH and 3' *endo* sugar pucker of RNA uracil,

preventing excision (Slupphaug *et al.*, 1996). Asp64 has also been shown to bind to the extrahelical conformation of uracil suggesting, that along with His187, it plays a role in transition state stabilisation (Drohat *et al.*, 1999a, Jiang *et al.*, 2002a, Jiang and Stivers, 2002).

The flanking phosphates -2, -2, +1 and +2 of the uracil, that are involved in serine pinching, considerably lower the activation barrier for catalysis while also stabilising the transition state through interactions with the cationic sugar and repulsion of the oxyanion. It is of note that these phosphates also bind strongly to the ground state (Jiang *et al.*, 2003).

The N2 and O1 of Asn123's side chain binds to the O4 and N3 of the uracil base *via* donation and acceptance of a hydrogen bond respectively. This acts as the "pulling" mechanism to aid formation of the extrahelical state and the stability of the transition state. It is also possible for the backbone carbonyl of Asp64 and the backbone amides of Asp64 and Gln63 to form hydrogen bonds with the N3 and O2 of the uracil respectively to aid stability (Mol *et al.*, 1995b, Slupphaug *et al.*, 1996). The hydrogen bonds formed by Asn123 are critical as its mutation results in the enzyme being unable to actively flip the base due to the lack of binding stability (Jiang and Stivers, 2002).

His187 also has a large effect on lowering the activation barrier and is another critical residue for catalysis (Kavli *et al.*, 1996, Drohat *et al.*, 1999b, Xiao *et al.*, 1999), showing the importance of the hydrogen bonds formed by these residues (Xiao *et al.*, 1999, Jiang and Stivers, 2002). His187 and Asp64, along with Gln63, do not affect the stability of the active site-bound uracil, or the affinity of UDG for binding and flipping the uracil base. When mutated they show a decrease in activity, suggesting they have a role in glycosidic bond cleavage. Mutation of Asp64 to asparagine results in an increase in binding (Mol *et al.*, 1995b, Drohat *et al.*, 1999a, Jiang and Stivers, 2002) but this is most likely due to the more favourable interactions of the asparagine with the phosphodiester backbone (Jiang *et al.*, 2001). Other active site mutations have a large detrimental effect lowering the activity by 93-99.9% (Kavli *et al.*, 1996).

Once the uracil has been flipped from the helix and entered the active site, UDG undergoes a conformational change and envelops the base. (Parikh *et al.*, 1998, Stivers *et al.*, 1999, Werner *et al.*, 2000). This mechanism of closing around the uracil is probably a contributing factor in the specificity, forming the correct hydrogen bonds with Asn123 (Mol *et al.*, 1995b, Savva *et al.*, 1995), which keep the uracil in the correct orientation for

cleavage and are essential for leaving group activity (Drohatsky *et al.*, 1999a, Drohatsky *et al.*, 1999b). Only uracil-containing DNA produces a decrease in tryptophan fluorescence as this base is the only one that is able to induce the conformational change in UDG (Stivers *et al.*, 1999). Asn123, Leu191, Ser88 and Ser189 play important roles in the conformational change that occurs after the base has been flipped out, as seen with mutants that are defective in this mechanism (Mol *et al.*, 1995b, Savva *et al.*, 1995, Jiang *et al.*, 2002b).

Mutants of Asn123, Leu191, Ser88 and Ser189 acting against 2'-FU arrest with uracil in an extrahelical conformation without clamping and active site docking, and in turn prevent cleavage. A decrease in tryptophan fluorescence is not observed, but an increase in 2-AP (located adjacently 5' to the 2'-FU) fluorescence is still seen. This is because the enzyme is incapable of undergoing the conformational changes that are required to close around the base; a mechanism that alters the position and thus fluorescence of the tryptophan residue. It therefore appears that either the active site has become larger, allowing access by 2'-FU, or that the DNA is distorted, changing the base stacking and so producing an increase in 2-AP fluorescence (Jiang and Stivers, 2002). These observations show the importance of the critical asparagine in the later stages of UDG's mechanism.

The Asn123 in *e*UDG (and N204 in *h*UDG) could theoretically form hydrogen bonds with cytosine as well as uracil by rotating the amino acid side chain, since this does not interact with any other part of the protein. In order to maintain the correct orientation, the side chain amide of the asparagine makes a hydrogen bond with a water molecule, that in turn bonds with a further two water molecules. These molecules make further connections with the main chain carbonyls and peptide nitrogen's, and therefore fix its orientation. The opposite orientation would produce a repulsive force between the water and amide, and so would be energetically unfavourable. This water molecule also interacts with another group of water molecules that bind to the O4 of uracil. If this was a cytosine then a second repulsive interaction would occur between the protons. Fixing the orientation of asparagine by interaction with these water molecules is therefore critical for the enzyme's specificity (Pearl, 2000).

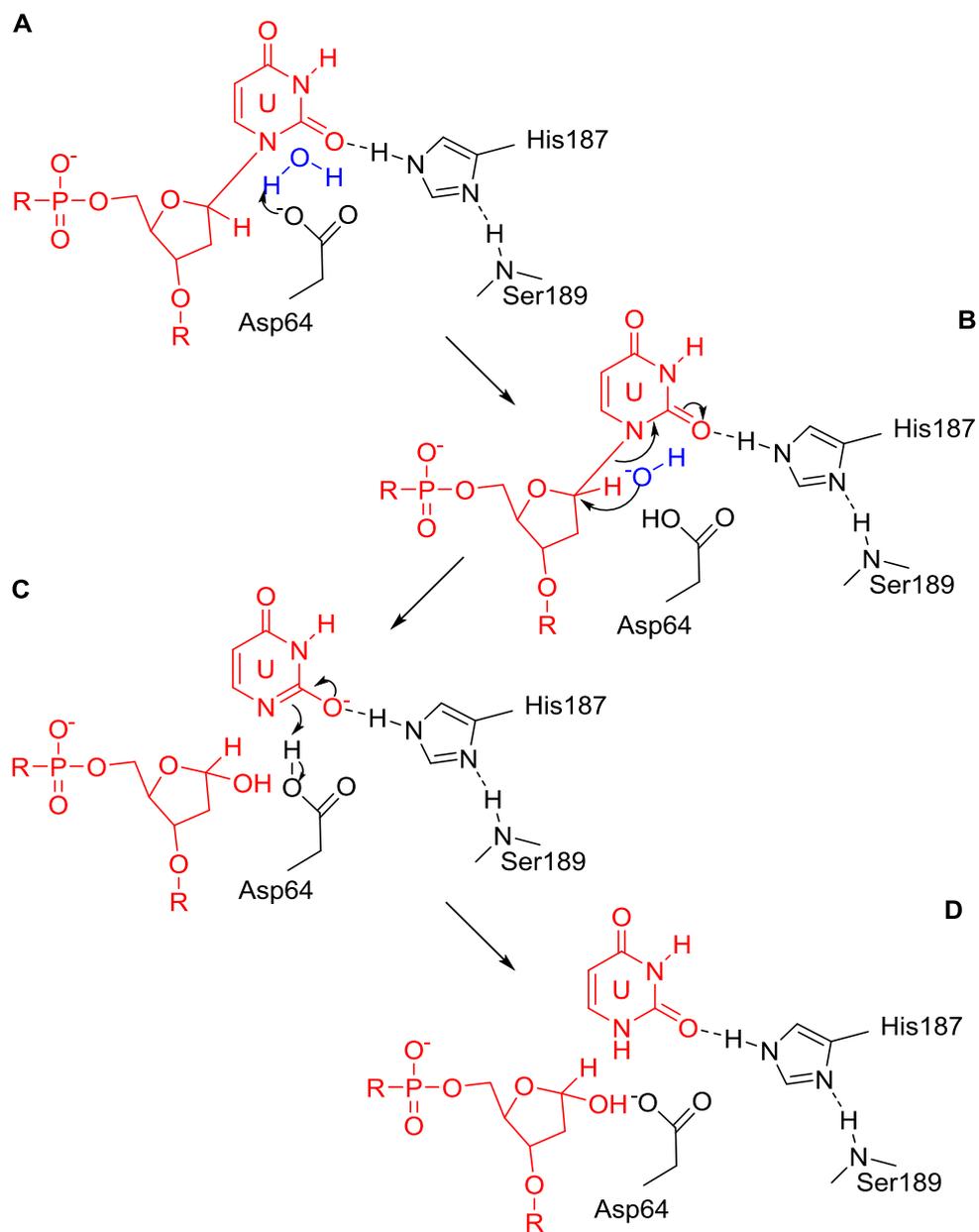


Figure 1.11 N-Glycosidic bond cleavage mechanism. A) Asp64 deprotonates a water molecule to create an OH^- nucleophile. B) The nucleophile attacks the C1' of the ribose ring. C) The N-Glycosidic bond is cleaved and the transition state of the free uracil is stabilised by His187. D) The uracil base in its native form due to rearrangement of electrons that also allows the regeneration of the catalytic Asp64. DNA highlighted in red and the water molecule/ OH^- nucleophile in blue.

1.3.1.6 Events after Excision

Dissociation of the released base is the major rate-limiting step of the reaction, and requires reversal of the conformational change that clamped the base (Jiang and Stivers, 2002). Dissociation of the base is also delayed if the next enzyme in the BER pathway is not available, as UDG stays bound to protect the AP site (Parikh *et al.*, 1998) in order to prevent misincorporation of another base. UDG also has nanomolar affinity for AP-DNA and is thus inhibited by its presence (Parikh *et al.*, 1998).

UDG can be inhibited by its product, free uracil, but this only occurs at millimolar concentrations (Slupphaug *et al.*, 1996). It is also inhibited by the *Bacillus subtilis* uracil DNA glycosylase inhibitor (Ugi) (Bennett and Mosbaugh, 1992), and other peptide inhibitors including ssDNA binding (SSB) proteins (Handa *et al.*, 2001) that are involved in DNA replication, repair and recombination. These peptide inhibitors mimic the interaction of UDG with DNA and in turn prevent binding of DNA or a polynucleotide (Mol *et al.*, 1995a, Savva and Pearl, 1995).

Once the free base has dissociated, the next enzyme in the BER pathway, apurinic endonuclease (APE), which has been suggested to directly interact with UDG (Pettersen *et al.*, 2007), removes the abasic sugar and the 3' phosphate. The correct nucleotide is then inserted by a DNA polymerase and finally the DNA backbone is sealed by DNA ligase (Krokan *et al.*, 1997, Pearl, 2000) (Figure 1.12).

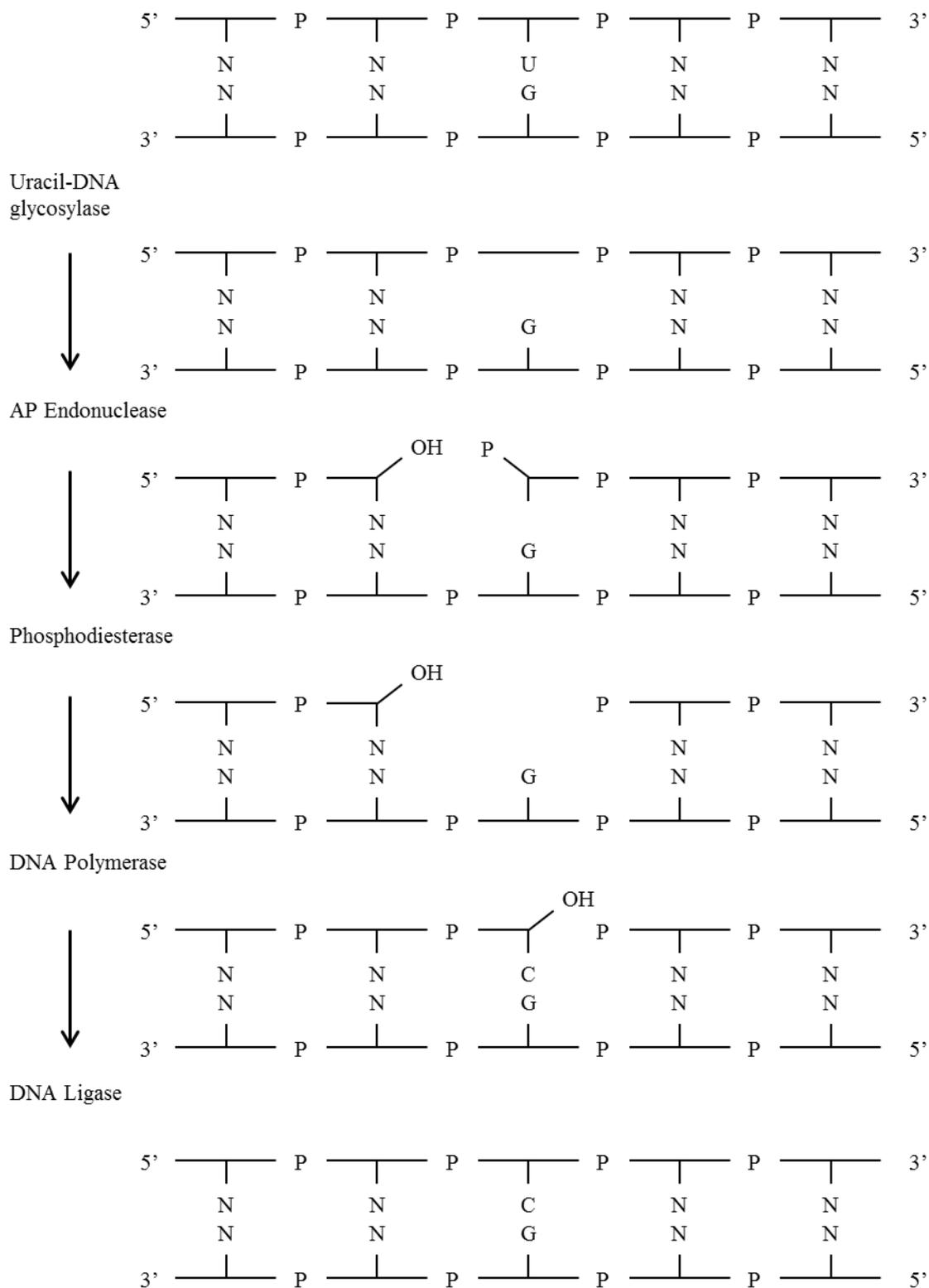


Figure 1.12 The mechanism of the BER pathway (Adapted from Lindahl, 1993). Uracil DNA glycosylase recognises the G.U mismatch in DNA and removes uracil. An AP Endonuclease breaks the phosphodiester backbone allowing the removal of the ribose and phosphate associated with the excised uracil base generating an abasic site. DNA polymerase then inserts cytosine (to correctly base pair with guanine) and DNA ligase seals the backbone completing repair.

1.3.1.7 The Role of Phenylalanine 77 and Tyrosine 66 in Thymine Exclusion

A conserved phenylalanine, Phe77, sits at the bottom of the active site (Mol *et al.*, 1995b) and stacks with uracil through van der Waals interactions (Savva *et al.*, 1995). This mimics the stacking of the pyrimidine ring and helps to stabilise the docked state (Mol *et al.*, 1995b, Krokan *et al.*, 1997). The phenylalanine's carbonyl oxygen forms a hydrogen bond with the OH of Tyr66, and both these residues are highly conserved in Family 1 UDGs. Phe77 and Tyr66 assist in shape complementarity, whilst Tyr66 also prevents thymine from entering the active site of UDG (Mol *et al.*, 1995b, Kavli *et al.*, 1996). Initial base sampling does not discriminate between uracil and thymine and it is only upon docking into the active site that discrimination occurs through Tyr66 (Parker *et al.*, 2007). This is because Tyr66 acts as a barrier preventing entry of thymine, since there would be a steric clash between the aromatic ring of tyrosine and the methyl group of thymine (Mol *et al.*, 1995b, Kavli *et al.*, 1996) (Figure 1.13).

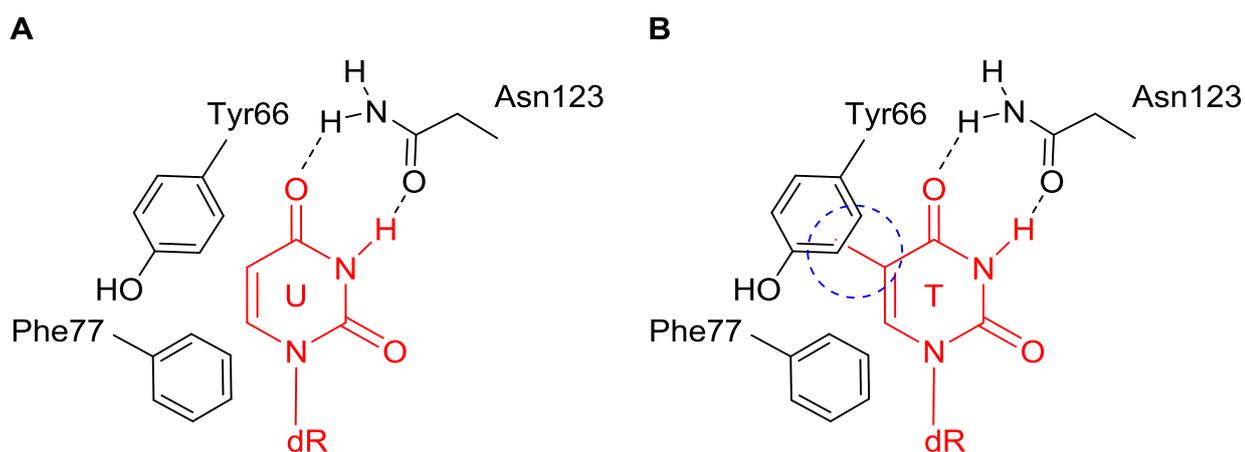


Figure 1.13 Thymine exclusion by UDG. A) Normal binding of uracil (red) within the active site. B) Exclusion mechanism of thymine (red) whereby if bound a steric clash (broken blue circle) occurs between thymine and Tyr66. dR: deoxyribose.

This is important in specificity as the 5-methyl group is the only difference between uracil and thymine. Tyr66 is kept in place *via* its H-bond with Phe77 and also through van der Waals interactions with Val160 and Pro167 (Mol *et al.*, 1995a). The tyrosine also excludes the larger purine bases, as their rings would also clash with the side chain of Tyr66. Although cytosine is also a pyrimidine it is excluded by repulsive interactions with the polar groups that line the active site, which would also interact unfavourably with purines. Thymine may also be excluded through favourable hydrophobic interactions with its ring and methyl group at the mouth of the active site. This may act as a “trap” that prevents

thymine from entering the active site. In order for thymine to bind it would have to displace the nucleophilic water molecule, which would result in the catalytic Asp64 rotating away from the substrate, and thus inhibiting the enzyme (Savva *et al.*, 1995).

1.3.1.8 Tyrosine 66 Mutations

The mutation of tyrosine at position 66 (*e*UDG) to a phenylalanine (Y66F) only causes a small reduction in UDG activity (Kavli *et al.*, 1996). These residues only differ by a hydroxyl group and both contain an aromatic ring that blocks the entry of thymine (Mol *et al.*, 1995a). However, mutation of Tyr66 to a smaller residue such as alanine (*e*UDG Y66A; *h*UDG Y147A) creates a larger space in the active site that is able to accommodate the 5-methyl group of thymine. This mutant is therefore able to excise thymine as well as uracil, generating a thymine DNA glycosylase (TDG) (Kavli *et al.*, 1996). Mutation of Y66 to cysteine or serine (Y66S/C), results in minimal uracil activity and a decrease in TDG activity in comparison to Y66A. This contradicts with the notion that the tyrosine's sole role is to provide a steric block towards thymine (Handa *et al.*, 2002). Though these TDGs have reduced activity, they still have a higher turnover rate than most other glycosylases (Kavli *et al.*, 1996, Handa *et al.*, 2002). This then suggests that Y66 plays a role in catalysis, possibly by stabilising the transition state through van der Waals interactions with Y66 and C5 of uracil (Handa *et al.*, 2002).

The weaker TDG activity of the Y66A/S/C mutants led to a 4-fold higher mutation frequency within *E. coli* in comparison to wild type UDG (Kavli *et al.*, 1996). Addition of wild type UDG decreased the mutation frequency in an UDG⁻ background (Kavli *et al.*, 1996), highlighting its importance in DNA repair. It is of note that these Y66 mutants are extremely cytotoxic and can only be produced *in vitro* (Handa *et al.*, 2002).

Interestingly, though still active towards uracil, mutation to tryptophan (Y66W) reduces its inhibition by either uracil or DNA containing an AP site (AP-DNA) (Acharya *et al.*, 2003). It has been proposed that a UDG deficient in its ability to bind AP-DNA could leave the site open, increasing mutation frequency (Bharti and Varshney, 2010). The tryptophan causes a widening of the uracil binding pocket and does not prevent access by uracil. However, free uracil does not inhibit this mutant in contrast to UDG or its Y66F/H/L mutants. It has been suggested that uracil may also bind to a second site resulting in non-competitive inhibition (Acharya *et al.*, 2003). As might be expected this mutant has no

TDG activity as entry of thymine into the active site would cause a steric clash between the 5-methyl group of thymine and the tryptophan's side chain (Bharti and Varshney, 2010).

These mutant UDGs with TDG activity are able to excise both uracil and thymine opposite either adenine or guanine, whereas natural TDGs, belonging to Family 2 (Waters and Swann, 1998), are only able to recognise the bases mismatched with guanine (Kavli *et al.*, 1996).

1.3.1.9 Properties of Cytosine DNA Glycosylases

UDG discriminates between uracil and cytosine, which only differ in hydrogen bonding at the C4 and N3 positions of the pyrimidine ring. A cytosine DNA glycosylase (CDG) can be created by mutating the critical asparagine to an aspartate (N123D for *e*UDG). This alters the hydrogen bond donor and acceptor pattern, allowing recognition of cytosine as well as uracil; shown by red hydrogen bonds in Figure 1.14. This occurs through hydrogen bonding between the carboxylate of aspartate and the 4-amino group of cytosine (Kavli *et al.*, 1996).

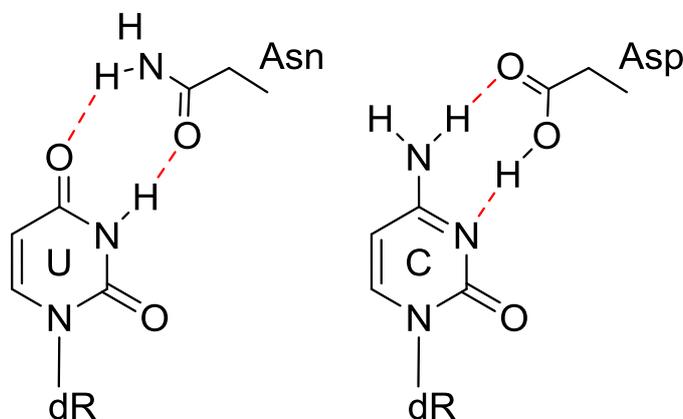


Figure 1.14 Alteration of the hydrogen bond pattern for N123D. dR: deoxyribose.

N123D is highly cytotoxic, as with Y66A (which has greater activity) as they excise the natural bases cytosine and thymine respectively. This will destroy the DNA and result in death of the host cell when expressing the proteins within a bacterial cell. When mutating *h*UDG, *h*CDG could only be produced in a *recA*⁺ strain of *E. coli* in which, in the absence of the IPTG inducer, the turnover rate of the enzyme was low enough for the host cell to repair the DNA damage (Kavli *et al.*, 1996). DNA degradation was still observed, and over time the DNA became too damaged and affected the survival of the host cell. The *E. coli* variant, *e*CDG appears to be even more cytotoxic in *E. coli* than the human counterpart as

no transformants were produced (Kwon *et al.*, 2003 and results presented in this thesis). *e*CDG can be produced through an *in vitro* transcription translation system (Handa *et al.*, 2002) and has only been produced in *E.coli* as a double mutant with L191A. As previously stated the leucine mutation prevents base flipping and excision, as the base is not in an extrahelical conformation and so cannot enter the active site of the enzyme. This renders the asparagine mutation inactive, though CDG activity is restored by placing the cytosine opposite a pyrene (Kwon *et al.*, 2003).

CDG still retains excision activity on uracil, which could be due to protonation of the aspartate's side chain, causing a change in the hydrogen bonding pattern, which is observed around a pKa of 6-7 (Kwon *et al.*, 2003). However, this residual uracil activity is most likely caused by the opposite rotamer of aspartate (Pearl, 2000). This could occur through the water molecules, which anchor the original asparagine in place, not acting in the same way towards the aspartate substitute. This allows the aspartate's side group to rotate and reverse its hydrogen bonding pattern, and thus be in the correct position to hydrogen bond to uracil.

It has been shown that the optimum pH for cytosine excision is approximately 6.2; this allows for protonation of Asp123 and for Asp64 to be deprotonated. This was due to a pronounced bell-shape pH-rate profile whereby it was suggested that the ascending limb represents the deprotonation of Asp64, and the descending limb the protonation of Asp123 (Kwon *et al.*, 2003). This neutralises the positive charge on the base in its anionic intermediate state during catalysis (Werner and Stivers, 2000). Protonation of Asp123 may be important in hydrogen bonding to the H4 and the donation of a proton to the N3 of cytosine, which aids release of the base from its transition state (Kwon *et al.*, 2003). This is consistent with protonation of the N3 that aids catalysis of the non-enzymatic hydrolysis of cytosine (Shapiro and Danzig, 1972).

Using the L191A and pyrene rescue (as mentioned in 1.3.1.4) it is possible to create a Cytosine/Pyrene DNA glycosylase (CYDG) (Kwon *et al.*, 2003). As with UDG, pyrene (Pn) is able to rescue the effects of the leucine mutation through mimicry (Figure 1.15). CYDG is able to excise cytosine from a Pn.C base pair but not a G.C pair. This is because the mutant enzyme is unable to flip the base without the bulky pyrene residue. The activity of CYDG is comparable to that of other DNA glycosylases, though it has 10^{12} fold lower catalytic power than UDG and 10^4 fold less than thymine DNA glycosylase (TDG) (Kwon

et al., 2003), a member of Family 2 UDGs (Waters and Swann, 1998, Abner *et al.*, 2001). CYDG has minimal G.C activity, 2600 fold less than Pn.C, as it is unable to perform the push and plug mechanism (Kwon *et al.*, 2003).

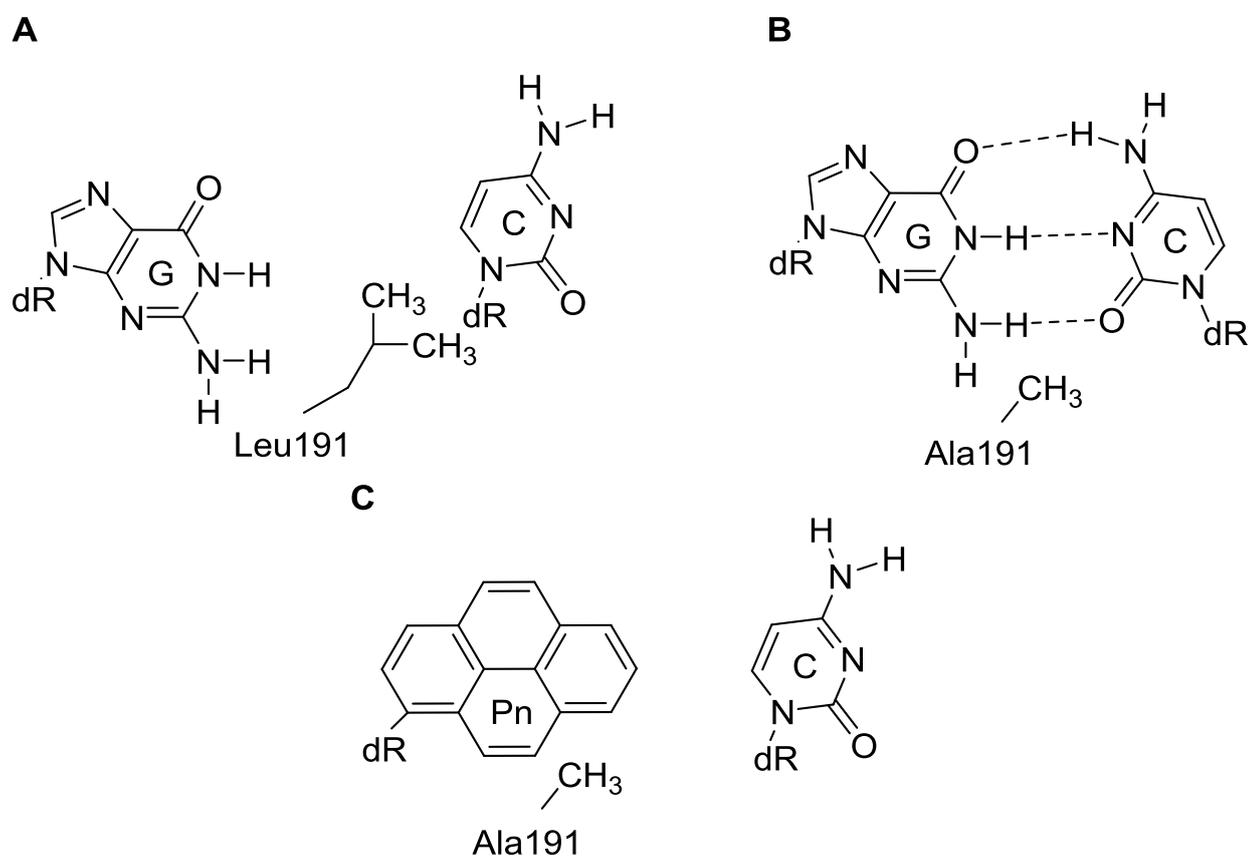


Figure 1.15 Pyrene rescue of CDG. A) Normal mechanism of CDG using Leu191 to force cytosine into an extrahelical conformation. B) L191A mutation causing inhibition of UDG through preventing cytosine flipping. C) Incorporation of a pyrene (Pn) nucleotide to act as a wedge to force the cytosine into an extrahelical conformation; rescuing the effects of the L191A mutation. dR: deoxyribose.

Y66A has also been created as a double mutant with L191A. In a similar fashion to that of CYDG, this creates a thymine/pyrene DNA glycosylase (TYDG) that is 100 fold less active than CYDG. TYDG has no activity towards A:T and is therefore specific for excision of thymine from Pn.T base pairs (Kwon *et al.*, 2003).

1.3.2 Family 2 UDGs

Mismatch-specific uracil DNA glycosylase (MUG) and thymine DNA glycosylase (TDG) are similar enzymes that have the ability to excise thymine and are the two members of Family 2 UDGs. Both enzymes also have uracil excision activity, but only when it is mismatched opposite guanine (Neddermann and Jiricny, 1994); its activity is determined

by how easily the base pair can be disrupted. This is due to insertion of a leucine residue into the space left by the flipped out base (that is more extensive than Family 1) and intercalation of an arginine (Figure 1.16) into the distal strand that also provides selectivity towards G.U (Gallinari and Jiricny, 1996, Barrett *et al.*, 1998a, Barrett *et al.*, 1998b, Pearl, 2000). These enzymes show no ssDNA activity and are not greatly inhibited by Ugi (Gallinari and Jiricny, 1996, Lutsenko and Bhagwat, 1999, Waters *et al.*, 1999). They possess high affinity for the AP site produced (O'Neill *et al.*, 2003) that is most likely due to it not retaining uracil in the active site post excision (Pearl, 2000). The strong affinity for AP sites also means that it is only able to process one thymine at a time (Waters and Swann, 1998), and it is unable to move on to the next reaction until it has been displaced by APE1 (Waters *et al.*, 1999, Fitzgerald and Drohat, 2008), which increases the turnover rate of TDG. However, *Deinococcus radiodurans* MUG is able to excise uracil from G.U, A.U base pairs and from ssDNA (Moe *et al.*, 2006).

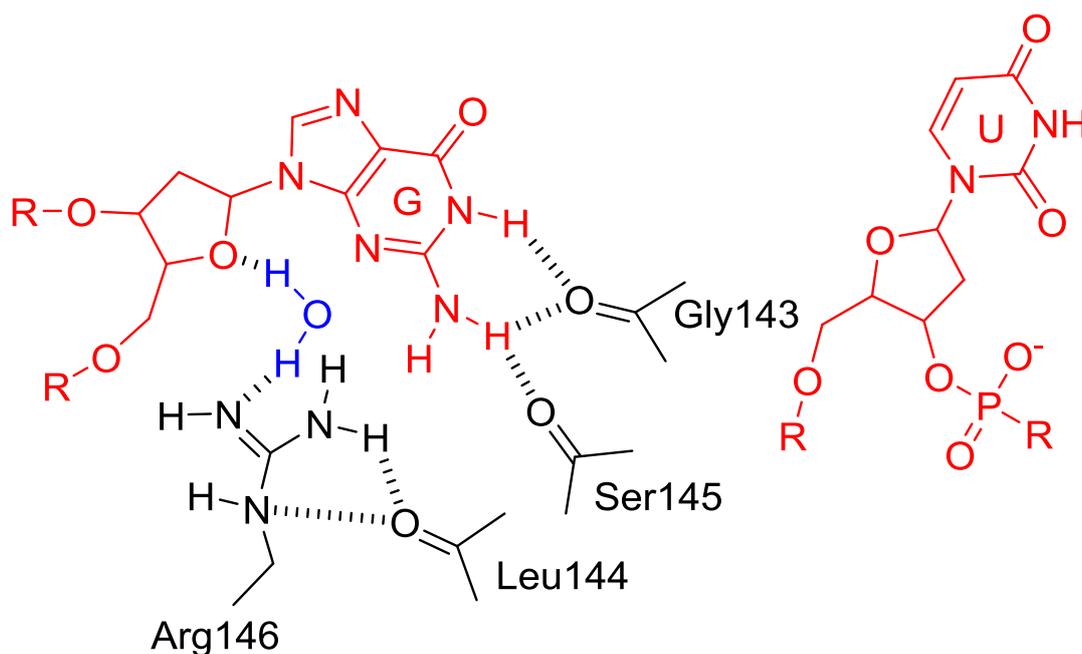


Figure 1.16 The ability of MUG to hydrogen bond with the opposing widowed guanine base, mimicking that of the flipped out uracil base. DNA highlighted in red and the water molecule in blue.

Unlike other UDGs, Family 2 enzymes function as dimers (Maiti *et al.*, 2008, Gripon *et al.*, 2011, Morgan *et al.*, 2011) and have slightly different residues within the conserved motifs of Family 1 UDGs resulting in lower activity and a different recognition and catalytic mechanism (Pearl, 2000). The lower activity is due to a lack of ability to activate a water molecule for nucleophilic attack of the N-glycosidic bond, and cleavage only occurs through a dissociative mechanism (Mol *et al.*, 1995b, Barrett *et al.*, 1998b, Barrett

et al., 1999, Bennett *et al.*, 2006). This explains why MUG has only 10% sequence homology to Family 1 UDGs but retains good structural homology (Barrett *et al.*, 1998b).

MUG is less able to excise thymine from G.T mismatches, than TDG. It has 120 fewer residues at its N-terminus (Gallinari and Jiricny, 1996), suggesting that thymine excision is not solely due to active site selectivity (Liu *et al.*, 2008). MUG also contains a serine (Ser23) that clashes with the 5-methyl group of thymine causing its weak thymine excision activity (Hardeland *et al.*, 2000, Moe *et al.*, 2006). TDG contains an alanine (Ala145 in *hTDG*), in the position of the critical tyrosine in Family 1 UDGs, and thymine can more easily be accommodated in the active site (Barrett *et al.*, 1999). It has been shown that the TDGs thymine activity is related to the balance of substrate interaction and the activation energy required for cleavage (Hardeland *et al.*, 2000). This also confirms the importance of tyrosine in Family 1 for thymine exclusion (Gallinari and Jiricny, 1996, Barrett *et al.*, 1998b). Both MUG and TDG are able to excise the cytosine adduct 3,N4-ethenocytosine (created through reaction with vinyl chloride (Hang *et al.*, 1998, Saparbaev and Laval, 1998), ethyl carbamate (Leithauser *et al.*, 1990), and also *via* lipid peroxidation (Borys-Brzywczy *et al.*, 2005)) with greater affinity than uracil and forms the same “wobble” pairing as G:T (Abu and Waters, 2003). Unlike G.T, 3,N4-ethenocytosine is displaced towards the major groove while the guanine remains stacked within the DNA duplex, though both are displaced vertical as well as horizontally (Cullinan *et al.*, 1997). It is able to excise 3,N4-ethenocytosine due to its larger active site (Hang *et al.*, 1998, Saparbaev and Laval, 1998, Lutsenko and Bhagwat, 1999, O'Neill *et al.*, 2003) producing greater destabilisation of the duplex requiring less energy to force it into an extrahelical conformation (Sagi *et al.*, 2000).

More recently *hTDG* (and MUG) have been shown to excise 7,8-dihydro-8-oxoadenine (8oxoA) when opposite any of the three standard bases (Talhaoui *et al.*, 2013), and like 8-hydroxyguanine (8oxoG), it is generated from reactive oxygen species; *e.g.* hydroxyl radicals (OH•) (Lindahl, 1993). *hTDG* excises 8oxoA with similar efficiency to thymine and 3,N4-ethenocytosine (Talhaoui *et al.*, 2013) further emphasising the role of these enzymes for repairing mutagenic base adducts.

As with Family 1 the rate of excision is dependent on the flanking sequences that alter the stability of the base pair (Hang *et al.*, 1998), most notably the nature of the 5' base (Abu and Waters, 2003). This is because the 5' base aids stabilisation of the transition state for

thymine but is able to stabilise both the transition state and enzyme-substrate complex for 3,N4-ethenocytosine.

In summary, Family 2 enzymes' main role appears to be the removal of cytosine adducts, as UDGs are far more active than MUG and are sufficient for excision of uracil for DNA repair. This is also supported by the observation that MUG does not affect the number of 5-methylcytosine to thymine or cytosine to uracil transition mutations (Lutsenko and Bhagwat, 1999). The primary role of TDG is most likely the removal of thymine from G.T mismatches that arise at CpG sites (Waters and Swann, 1998, Sagi *et al.*, 2000), as these are hotspots for mutations (Shen *et al.*, 1992) as 5-methylcytosine can be deaminated to create thymine, producing a G.T mismatch (Sartori *et al.*, 2002).

1.3.3 Family 3 UDGs

A third Family of UDGs was first identified in *Xenopus* (Haushalter *et al.*, 1999) and was first thought to be exclusive to eukaryotes but was later found in numerous bacteria (Pettersen *et al.*, 2007, Mi *et al.*, 2009). They are able to excise uracil from ssDNA with their affinity being only slightly lower affinity than that of Family 1 enzymes, with their affinity for G.U being considerably lower, resembling that of Family 2. It therefore appears to have hybrid properties of UDGs and TDG/MUGs (Pettersen *et al.*, 2007). Due to its greater preference for ssDNA in comparison to Family 1 UDGs, it was given the name Single-strand-selective mono-functional uracil DNA glycosylase (SMUG1) (Pearl, 2000), even though SMUG1 is up to 700 times more active against ds than ssDNA substrates (Wibley *et al.*, 2003) under physiological conditions (Doseth *et al.*, 2012). It is therefore essentially a double-strand-selective DNA glycosylase. It also shows activity towards xanthine (X) but has no activity towards hypoxanthine and oxanine; deaminated analogues of adenine and guanine respectively (Mi *et al.*, 2009).

As with Family 1 the asparagine (N58) is critical for catalysis and if mutated to an aspartate results in a loss of activity (Mi *et al.*, 2009), while it utilises another asparagine for the activation of the water molecule (Mol *et al.*, 1995b, Savva *et al.*, 1995). In contrast to the other families, SMUG1 has an arginine instead of leucine for plugging the space left by the flipped out base. This substitution does not affect the ability of the enzyme to occupy the space and an L191R mutation in UDGs only serve to increase the affinity for the abasic product (Slupphaug *et al.*, 1996).

1.3.4 Family 4 UDGs

Family 4 UDGs were initially discovered through homology screening of MUG in the thermophilic bacterium *Thermotoga maritima*. Studies on this class of enzyme have shown heat stability of up to 75°C and for up to 15 minutes at 85°C without the loss of activity, giving them the name “thermophilic UDGs”. This was designated as a new class of enzyme as no homologues have been found in eukaryotes, though others have been identified in other thermophilic bacteria (Sandigursky and Franklin, 1999, Sandigursky *et al.*, 2001). Despite low homology they are most similar to MUG (Sandigursky and Franklin, 1999), as they too do not contain asparagine or aspartate as the catalytic residue within the active site (Pearl, 2000), though they do contain a conserved asparagine which could support a water molecule for attack of the C1' of uracil (Kosaka *et al.*, 2007). Glycosidic bond cleavage occurs *via* a dissociative mechanism (Figure 1.17) (Hoseki *et al.*, 2003); DNA distortion causes strain on the glycosidic bond that weakens it and provides the required energy to aid cleavage (Parikh *et al.*, 2000). These thermophilic enzymes recognise uracil in the same manner as Family 1 and act on dsDNA regardless of the opposing base, and ssDNA through a positively charged active site (Hoseki *et al.*, 2003) while showing no activity towards thymine (Sandigursky *et al.*, 2001, Hoseki *et al.*, 2003).

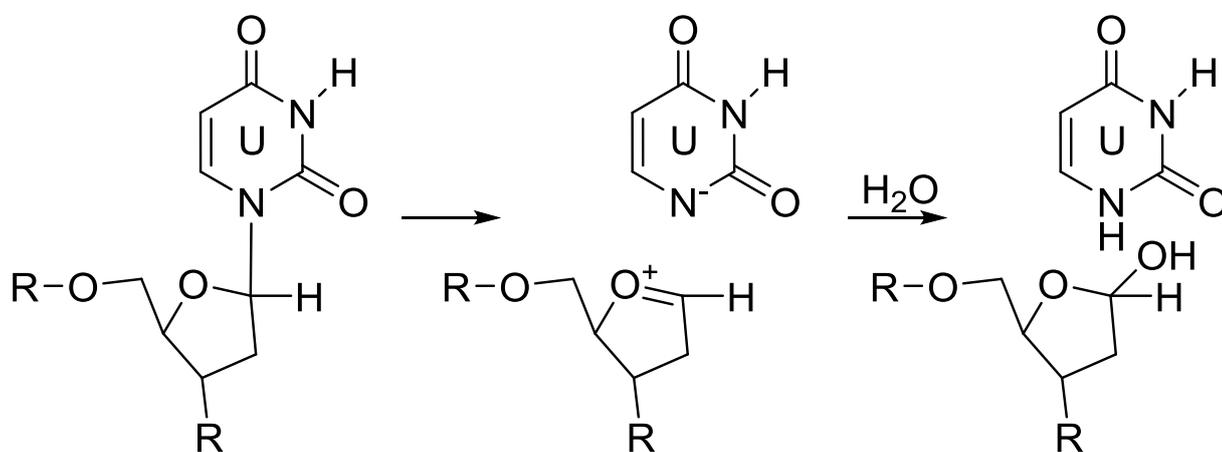


Figure 1.17 Dissociative mechanism of N-Glycosidic bond cleavage by Family 4 UDGs.

The majority of these thermophilic UDGs have a quartet of cysteine residues that form a loop structure and are thought to act as ligands for an iron-sulphur cluster (4Fe-4S). Since this is located away from the active site and the binding groove, it is presumed to have no effect on the binding or activity of the enzyme (Hoseki *et al.*, 2003). The metal ion coordination of the cluster, along with salt bridges, ion pairs and proline residues within

the loop region, may account for a balance between stability and flexibility required by a thermostable enzyme (Hoseki *et al.*, 2003, Engstrom *et al.*, 2012). It is also conceivable that the cluster may play a role in protein-protein interactions.

The first archaeal UDG to be discovered was from *Archaeoglobus fulgidus* and was designated a member of Family 4 due to its thermophilic properties. *A. fulgidus* UDG (*afUDG*) is most similar to *T. maritima* in sequence and enzymatic ability, and therefore suggests a similarity between BER in thermophilic bacteria and Archaea (Sandigursky and Franklin, 2000). The main similarity is that they have two conserved residues; a glutamate at positions 2 and 5 and a leucine at positions 4 and 7 for *afUDG* and *tmUDG* respectively. These could be important for their folding and thus their thermostability, though these residues are not conserved in any other thermophilic UDG's (Sandigursky *et al.*, 2001). As with other UDGs it is able to excise uracil when paired with either guanine or adenine and is able to act on ssDNA, whilst having no activity for excising thymine (Sandigursky and Franklin, 2000).

1.3.5 Family 5 UDGs

This family of enzymes is similar to Family 4, and also contain four cysteine residues that are highly conserved and are ligands for an iron-sulphur (4Fe-4S) cluster (Hoseki *et al.*, 2003). However they do not contain a polar residue in motif 1. The absence of a polar residue suggests that glycosidic bond cleavage occurs through a dissociative mechanism (Figure 1.17) that does not require water (Sartori *et al.*, 2002, Starkuviene and Fritz, 2002). Cleavage is therefore thought to occur *via* the energy that is gained from the unfavourable extrahelical conformation of the base, weakening the glycosidic bond (Parikh *et al.*, 2000), as well as by interactions between the negative phosphate backbone and the positive C1' atom (Dinner *et al.*, 2001). These enzymes have weak uracil excision activity that is most likely due to their broad substrate specificity (Kosaka *et al.*, 2007). The broad specificity may be due to their ability to distort the DNA by 60° (Kosaka *et al.*, 2007), which is greater than the 45° of UDGs and having a larger DNA binding region (Jiang and Stivers, 2002). As well as uracil they are able to excise thymine (Hoseki *et al.*, 2003), 5'-hydroxymethyluracil and 5'-fluorouracil (Starkuviene and Fritz, 2002), ϵ C and hypoxanthine (Sartori *et al.*, 2002).

1.4 5-Methylcytosine Detection

The detection of 5-methylcytosine within the genome is important in the understanding of gene regulation. Many techniques have been developed to detect 5-methylcytosine though they are limited by either sequence context or can generate false positives through incomplete chemical reactions. The most commonly used methods are hereby described.

1.4.1 Bisulphite Sequencing

The gold standard for 5-methylcytosine detection is bisulphite sequencing. The reaction of DNA with bisulphite, shown in Figure 1.19 (Baylin *et al.*, 2001, Arand *et al.*, 2012, Hackett *et al.*, 2012, Su *et al.*, 2013), causes the conversion of cytosine to uracil while bisulphite is unreactive towards 5-methylcytosine (Shapiro *et al.*, 1973).

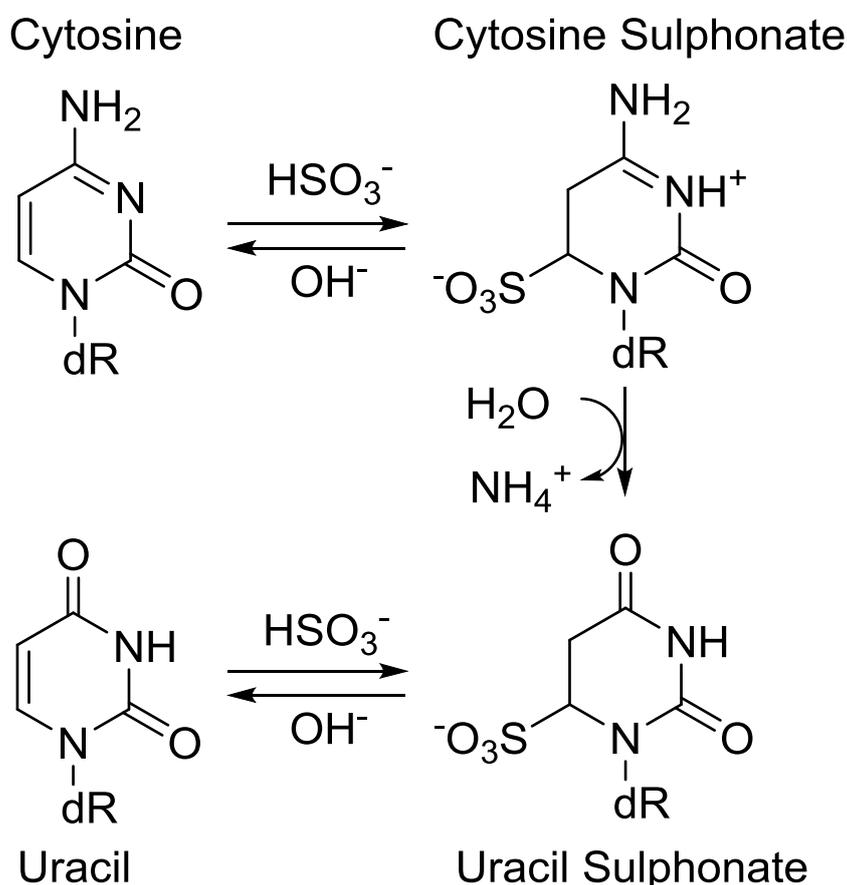


Figure 1.18 The bisulphite reaction mechanism (Adapted from Clark *et al.*, 1994). dR: deoxyribose.

The reaction products can then be sequenced and, since all the cytosines are converted to uracil, they appear as thymines in the sequence. In contrast any 5-methylcytosines are unreactive and still appear as cytosines in the sequencing reaction. The result, when run

alongside the sequence of an unreacted sample, detects the location of any methylated cytosines.

Though the bisulphite method is successful and is widely used, it has many limitations (Grunau *et al.*, 2001). The first is incomplete reaction of all the unmethylated cytosines, which then still appear in the cytosine sequencing lane, giving a false positive for cytosine methylation. The second is DNA degradation due to the bisulphite reaction itself resulting in a loss of the amount of DNA available for sequencing. The final problem is incomplete desulphonation during the bisulphite reaction, which can inhibit the sequencing polymerase (Clark *et al.*, 1994). To try and reduce the effects of these problems many different analysis methods have been developed; such as pyrosequencing (Colella *et al.*, 2003), methylation-sensitive single-strand conformation analysis (MS-SSCA) (Bianco *et al.*, 1999), high resolution melting analysis (HRM) (Wojdacz and Dobrovic, 2007) and microarray based methods (Adorjan *et al.*, 2002).

Variations of bisulphite sequencing have been developed to enable the detection of 5-hydroxymethylcytosine known as TAB-seq (Tet-assisted bisulphite sequencing) (Yu *et al.*, 2012) and OxBS-seq (Oxidative bisulphite sequencing) (Booth *et al.*, 2012). TAB-seq works through glycosylation of 5-hydroxymethylcytosine by β -glucosyltransferase (β GT) followed by reaction with TET that converts any 5-methylcytosines to 5-carbonylcytosines while being unreactive towards cytosine and the glucosylated 5-hydroxymethylcytosine. Upon bisulphite treatment cytosines and 5-carbonylcytosines are converted to thymine and appear so when sequenced and analysed while the glucosylated 5-hydroxymethylcytosine is unreacted and appears as cytosine; revealing sites of hydroxymethylation. OxBS-seq also detects the presence of 5-hydroxymethylcytosine through reaction with KRuO_4 followed by bisulphite treatment. 5-hydroxymethylcytosines are converted to 5-formylcytosines and like 5-methylcytosine are unaffected by reaction with bisulphite. The resultant products can then be compared to the same sample treated with bisulphite without having undergone a reaction with KRuO_4 . Additional cytosines shown upon sequencing analysis indicates sites of hydroxymethylation.

1.4.2 Bisulphite and UDG

Similar to bisulphite sequencing, UDG has been used in conjunction with the bisulphite reaction to detect 5-methylcytosine (Huang *et al.*, 2013). As cytosine is converted to uracil during bisulphite treatment the product can then be reacted with UDG under limiting

conditions, to remove the uracil and fragment the DNA. The resultant products can then be analysed by PAGE in comparison with a control sequence containing no 5-methylcytosines to determine which cytosines are methylated.

1.4.3 Methylation Sensitive Enzyme Restriction

Some restriction enzymes are methylation sensitive and will not digest DNA if methylated. The enzymes HpaII and MspI (Cedar *et al.*, 1979, McClelland *et al.*, 1994) are commonly used as they both recognise the sequence CCGG, with the former not being able to cleave if the central cytosine is methylated. Methylation can then be determined depending on which enzymes are able to cleave the DNA of interest, depending upon which cytosine and whether both strands of the DNA are methylated. Though this method detects methylation at CpG sites it has a major disadvantage in that it is limited to the restriction site sequence. Therefore the methylation state of cytosines at CpG sites that are not in a CCGG context or any other cytosine of interest not in a CpG context cannot be determined.

1.4.4 Nanopores

Nanopore sequencing (Kasianowicz *et al.*, 1996, Manrao *et al.*, 2011, Manrao *et al.*, 2012) has shown the ability to not only sequence DNA but also provide a way to detect methylation. Sequencing works by dividing a solution in two *via* a membrane containing a nanopore. A voltage is then applied across the membrane to generate a current that allows the DNA to enter and transcend the pore. The DNA causes a change in the potential difference, depending on its composition, allowing discrimination of bases. This technique can be further exploited to detect the presence of 5-methylcytosine (and 5-hydroxymethylcytosine) as the presence of 5-methylcytosine causes a local change in current compared to cytosine. The difference in current can then be used to detect the presence of 5-methylcytosine (Laszlo *et al.*, 2013).

1.4.5 Triplexes

Most recently it has been shown that triplexes can be used in the detection of 5-methylcytosine (and other cytosine modifications), most notably at CpA sites (Johannsen *et al.*, 2014). The technique works through determining the stability of the triplex by measuring its melting temperature. It was shown that the incorporation of 5-methylcytosine dramatically effects the binding of the third strand and thus reduces its stability, providing discrimination. Though this method is non-destructive towards DNA it is unable to

determine methylation at the more relevant CpG site, while also showing sequence dependency (it is restricted to sequences in which one strand is mainly purines); a common limitation with triplexes (Rusling *et al.*, 2005).

1.4.6 Small Molecules

Small molecules have also been used to detect the presence of 5-methylcytosine. The naphthyridine dimer (ND) analogue, NCD, binds to GG mismatches in the sequence context CGG/CGG in a 2:1 stoichiometry. The binding of NCD causes disruption and breakage of the flanking GC base pairs resulting in the cytosines being flipped out of the duplex. The exposed cytosines can then be reacted with hydroxylamine-sodium bisulphite and cleaved at these sites by piperidine. 5-methylcytosine does not react with hydroxylamine-sodium bisulphite but does with potassium permanganate and so cytosine and 5-methylcytosine can be distinguished in a CGG/CGG context (where C is cytosine or 5-methylcytosine) (Oka *et al.*, 2008, Oka *et al.*, 2009).

1.4.7 Other Methods

Thin layer chromatography (TLC) and mass spectrometry (MS) can be used to detect 5-methylcytosine, but these methods only allow for the detection of the presence of 5-methylcytosine, and not its specific location within DNA (Tahiliani *et al.*, 2009).

Immunoprecipitation (IP) with antibodies can also be used to detect the presence of 5-methylcytosine; once bound the DNA can be prepared for sequencing or microarray analysis (Weber *et al.*, 2005, Down *et al.*, 2008), providing an estimate to the degree of methylation within a region of interest. As the antibodies require more than one 5-methylcytosine for efficient binding, it requires 5-methylcytosines to be in close proximity to each other (Sano *et al.*, 1980, Sano *et al.*, 1988). Therefore this method does not allow for specific detection of 5-methylcytosine and only provides global and/or local methylation mapping.

1.5 Purpose of this Research

The purpose of this research was to find an alternative way of detecting the presence of 5-methylcytosine, by creating an enzyme that was able to discriminate between it and cytosine. The enzyme was created by mutating uracil DNA glycosylase (UDG) to generate a cytosine DNA glycosylase (CDG). Since UDG is able to exclude thymine from its active site as a result of steric clash with its 5-methyl group, we predicted that the mutant CDG should be able to exclude 5-methylcytosine by the same principle, since 5-methylcytosine is related to cytosine in the same way that thymine is related to uracil. Thus, CDG should be able to discriminate between cytosine and 5-methylcytosine.

This thesis therefore describes the production and characterisation of a series of mutants of *E. coli* and human UDG. The cleavage activity of these mutants is then examined using a range of radiolabelled synthetic oligonucleotides that contain different mismatches.

1.5.1 Selection of Enzyme

Initially all experiments were performed using the *E. coli* variant of UDG (*e*UDG), as the gene was easily obtained through PCR of the *E. coli* genome using appropriate UDG specific primers. This was subsequently cloned into a cloning/expression vector for mutagenesis purposes. In addition it was thought that any mutants produced, being derivatives of *E. coli* enzymes, would produce greater protein yields when grown and expressed in *E. coli*. We anticipated that the enzymes would be expressed more efficiently by having optimal codon usage selecting for the most abundant tRNAs for a specific residue. Due to problems of cytotoxicity we were unable to obtain *e*CDG, but worked with *e*CYDG instead (Kwon *et al.*, 2003). We then progressed onto using *h*UDG as it has been shown that this is more easily expressed in *E. coli* (Kavli *et al.*, 1996). As eukaryotes contain introns within their DNA a human variant of UDG cannot be obtained simply from PCR using genomic DNA as with *e*UDG. A clone of *h*UDG was therefore produced *via* complete gene synthesis with the N-terminal domain removed as this merely acts as a signal sequence for localisation to either the nucleus or mitochondria (Nilsen *et al.*, 1997). We were able to generate both *h*CDG and *h*CYDG.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Oligonucleotides

All oligonucleotides (Appendices I, II and III) were obtained from ATDBio Ltd (Southampton, UK). They were diluted in dH₂O and stored at -20°C until required. The anthraquinone pyrrolidine phosphoramidite (Figure 2.1) was obtained from Berry & Associates (Michigan, USA).

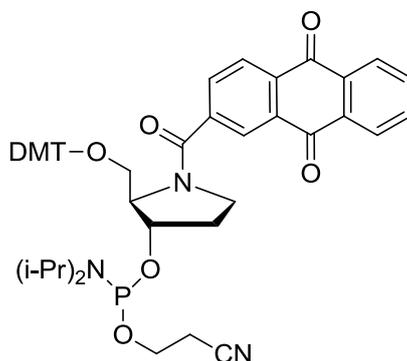


Figure 2.1 The structure of anthraquinone pyrrolidine phosphoramidite.

2.1.2 Enzymes

All restriction enzymes were purchased from either Promega (Southampton, UK) or New England Biolabs (NEB) (Herts, UK). *Pfu* DNA polymerase and thermosensitive alkaline phosphatase (TSAP) were also purchased from Promega. T4 polynucleotide kinase (PNK) and T4 DNA ligase were also purchased from NEB. All enzymes were stored at -20°C.

2.1.3 Chemicals

Radioactive [γ -³²P]ATP was purchased from Perkin Elmer (Cambridgeshire, UK) at a concentration of 3000 Ci/mmol and stored at 4°C. The vectors pUC18, pUC19 and pET28a were purchased from Sigma-Aldrich (Dorset, UK), and Merck Chemicals (Nottingham, UK) respectively. The QIAprep plasmid purification kit was purchased from Qiagen (Crawley, UK). UreaGel (20% acrylamide:bisacrylamide 19:1 containing 8 M urea) and Accugel (40% (w/v) acrylamide:bisacrylamide 19:1) were purchased from National Diagnostics (Hull, UK). The SYBR Green was purchased from Life Technologies (Paisley, UK). The GenomeLab™ DTCS Quick Start Kit, containing the master mix (composed of

the polymerase, buffer, dNTPs and fluorescent ddNTPs), and sample loading solution (SLS) were purchased from Beckman Coulter (High Wycombe, UK).

2.1.4 Plasmids Generated in this Work

Plasmid Name	Mutation	Notes
pUC19eUDG		Original construct obtained from Prof. K. R. Fox. Details of its generation are given in section 2.3.1.
pUC19eUYDG	L191A	
pUC19eCYDG	N123D, L191A	
pUC18eUDG		Subcloned from pUC19eUDG.
pUC18eCDG	N123D	
pUC18eUYDG	L191A	
pUC18eCYDG	N123D, L191A	
pETeUDG		Subcloned from pUC19eUDG into pET28a.
pETUgi		Ugi subcloned into pET28a from pRSETB; obtained from Dr. Renos Savva (Birkbeck, University of London, UK).
pETeUDGUgi		Fusion sequence of UDG and Ugi.
pETeUDGUgiS		As pETeUDGUgi, but with the internal UDG (between UDG and Ugi) stop codons removed.
pETeUDGUgiST		As pETeUDGUgiS, but with the 5' thrombin cleavage site mutated.
pETeUYDG	L191A	Inability to flip out target base due to leucine mutation.
pETeCYDG	N123D, L191A	As above, but with the N123D mutation to

		CDG.
pET <i>h</i> UDGΔ81		Created by total gene synthesis.
pET <i>h</i> CDGΔ81	N204D	
pET <i>h</i> CYDGΔ81	N204D, L272A	Used for comparison with <i>e</i> CYDG.
pUC19 <i>e</i> UDGBXE		Contains full length linker of BglII, XhoI and EcoRV restriction sites (Figure 4.1A).
pUC19 <i>e</i> UDGBX		As above, but with removal of EcoRV restriction site (Figure 4.1B).
pUC19 <i>e</i> UDGB		As above, but with removal of EcoRV and XhoI restriction sites (Figure 4.1C).
pUC19 <i>e</i> UDGIF		Contains a stop codon upstream and <u>In Frame</u> with UDG (Figure 4.2B).
pUC19 <i>e</i> UDGOF		Contains a stop codon upstream and <u>Out of Frame</u> with UDG (and in frame with the <i>LacZ</i> ; Figure 4.2C).

Table 2.1 Plasmid constructs used and generated in this thesis. For pUC plasmids the UDG gene was inserted between the HindIII and EcoRI sites, while for pET plasmids it was inserted between the EcoRI and NdeI sites. Details on individual plasmid constructs are presented in appropriate places in this thesis. The wild type sequences for *e*UDG and *h*UDGΔ81 can be seen in Appendix V.

2.2 Methods for Studying Base Excision

2.2.1 Cleavage Studies and Rate of Reaction Determination

The cleavage assay is a technique to study the ability of a DNA repair enzyme to act upon ds or ssDNA. In essence, the DNA is incubated with a repair enzyme to allow excision of a base or nucleotide. The DNA is then heated at 95°C, in the presence of 10% piperidine (Kwon *et al.*, 2003), to cause a break in the phosphodiester backbone, creating a shorter DNA fragment, showing that excision and therefore cleavage of the DNA had occurred. The products of the reaction are then run on a denaturing polyacrylamide gel to separate the cleaved and intact DNA strands. This assay can provide a quantitative estimate of the

enzyme's activity by comparing the intensities of the bands that corresponds to the uncleaved DNA and the cleaved products. The rate of reaction is determined from the assumption that all the substrate is bound to enzyme as the enzyme is in excess of substrate. It is therefore assumed that the initial velocity represents the rate of reaction as the amount substrate turned over per unit time.

2.3 Protocols

2.3.1 Agarose Gel Electrophoresis

Agarose gels of 0.7 – 1% (w/v) in 1x TBE buffer (unless stated otherwise) were prepared containing 0.1x GelRed, samples were ran in 1x TBE at 15 Vcm⁻¹ for 1 hour and imaged under UV light.

2.3.2 Site-Directed Mutagenesis

The wild type UDG plasmid was constructed by Professor K. R. Fox (University of Southampton) by PCR amplification of the *E. coli* UDG gene from genomic DNA and had been cloned into the multiple cloning site (MCS) of pUC19 between the HindIII and EcoRI sites using primers that included sites for NdeI and EcoRI. This was to allow the gene to be subcloned from pUC19 into pET28a as pUC19 doesn't contain an NdeI site within its MCS. The UDG gene was then removed and inserted into the expression vector pET28a using the restriction enzymes and sites of NdeI and EcoRI (Figure 2.2).

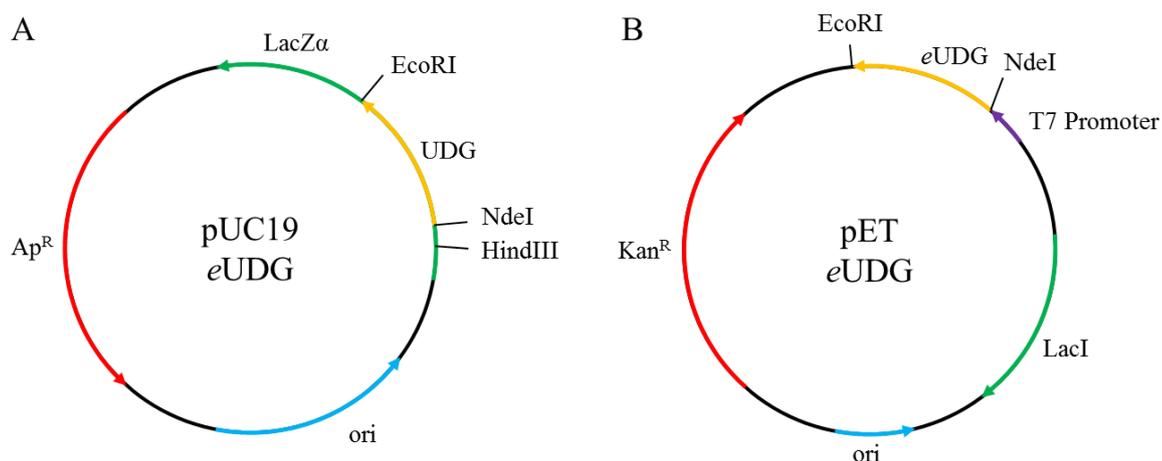


Figure 2.2 Plasmid maps. A) The structure of pUC19eUDG showing ampicillin resistance (red), origin of replication (blue), LacZα (green) and UDG (orange). B) The structure of pETeUDG showing kanamycin resistance (red), origin of replication (blue), LacI (green), T7 promoter (purple) and eUDG (orange). Arrows denote direction of transcription.

The different mutants of *eUDG* were generated by site-directed mutagenesis (SDM) using the QuikChange® protocol by Stratagene (Stratagene, 2004). The procedure, which is outlined in Figure 2.3, uses two complementary primers (Table 2.2) containing the bases to generate the mutation in the centre. These produce a mismatch when annealed with the target on the original plasmid. The mutagenic oligonucleotide primers must be of a suitable size so that they anneal to the target and tolerate the presence of the mismatch. The supercoiled plasmid is first denatured at 98°C. The mutagenic primers then bind during the annealing stage, when the temperature is lowered to approximately 55°C. The temperature is then increased to approximately 70°C and the primers are extended around the plasmid by the polymerase. This reaction proceeds through 12 - 18 cycles of denaturing, annealing and extension. The polymerase should have 3' exonuclease (proofreading) activity to greatly reduce the chance of secondary mutations occurring through misincorporation of an incorrect nucleotide. Unlike normal PCR, only linear amplification occurs as the products contain nicks (Figure 2.2) in the newly synthesised DNA, preventing the newly synthesised DNA from becoming a template in the next cycles of replication (usually 12 - 18). The parental/template plasmid is then removed by digestion with DpnI, as the enzyme acts only on methylated DNA. The newly synthesised, unmethylated annealed circular DNA can be transformed into competent *E. coli*.

The PCR reaction solution contained 500 nM of each primer, 500 nM dNTPs, 3% DMSO, 5 µl of 10x *Pfu* buffer, 1 µl of *Pfu* DNA polymerase (typically 2 - 3 u/µl) and approximately 100 ng of DNA template in 0.5 ml PRC tubes. The sample was made up to final volume of 50 µl with dH₂O. Samples were cycled using a TECHNE thermocycler using the following conditions:

Segment 1: Cycles 01: 98°C for 2 min.

Segment 2: Cycles 18: 98°C for 30 sec, 50°C for 30 sec and 72°C for 2 min per kb.

Segment 3: Cycles 01: 72°C for 10 min and hold at 4°C.

1 µl of DpnI (Promega: typically 10 u/µl) was added to each sample and mixed. The samples were left at 37°C for a minimum of 1 hour to digest the template DNA. Finally the plasmids were transformed into competent *E. coli* XL1-Blue cells.

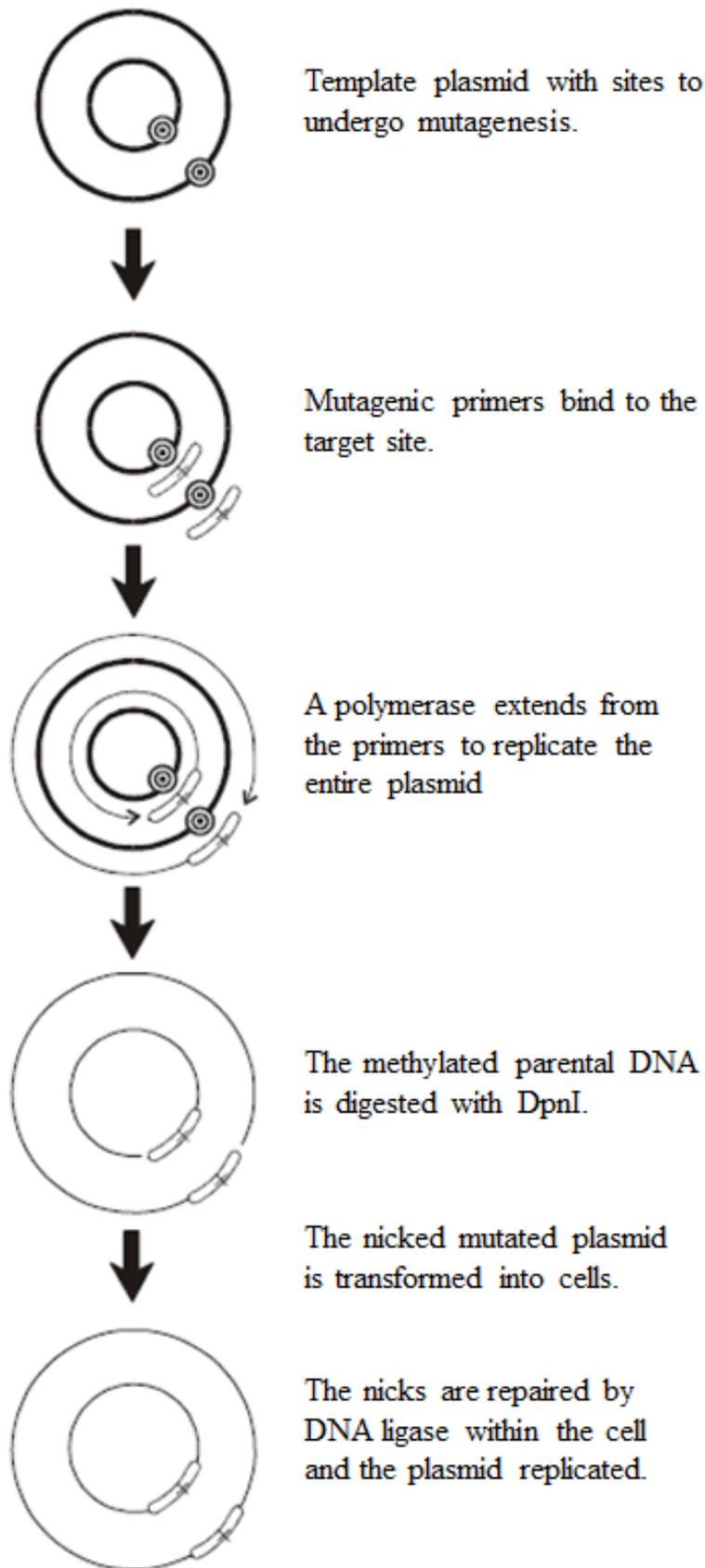


Figure 2.3 QuikChange site-directed mutagenesis (Stratagene, 2004).

Mutation	Primers
N123D	5' -GCGTCAGGGCGTTCTGCTACTC G ATACTGTGTTGACGGTACGCGC-3' 5' -GCGCGTACCGTCAACACAGTAT C GAGTAGCAGAACGCCCTGACGC-3'
L191A	5' -GCACCGCATCCGTCGCCG GCG TCGGCGCATCGTGGATTC-3' 5' -GAATCCACGATGCGCCG AGC CGGCGACGGATGCGGTGC-3'
A191L	5' -GCACCGCATCCGTCGCCG CTT TCGGCGCATCGTGGATTC-3' 5' GAATCCACGATGCGCCG AAG CGGCGACGGATGCGGTGC 3'
N204D	5' -GGTGTTCCTTCTC G ACGCTGTCCTCACGG-3' 5' -CCGTGAGGACAGCGT C GAGAAGGAGAACACC-3'
eUDGUgi Stop Codons	5' -TTACCGGCAGAGAGTGAG GGAGGA GAAATTCCTGGTGCCGCGC-3' 5' -GCGCGCACCCAGGAATTC TCCCTCC CTACTCTCTGCCGGTAA-3'
Thrombin	5' -AGCGGCCTGGTGCCG G GCGGCAGCCATATGG-3' 5' -CCATATGGCTGCCG C CGGCACCAGGCCGCT-3'

Table 2.2 Sequences of oligonucleotide primers used in site-directed mutagenesis. Base substitutions highlighted in red.

2.3.3 Preparation of Competent Cells

E. coli XL1-Blue cells were first grown on agar plates containing no antibiotic. A colony was selected and grown overnight at 37°C in 2YT media (16 g tryptone, 10 g yeast extract and 5 g NaCl per litre). 1 ml of these cells was transferred to 100 ml of 2YT media and grown to O.D. of 0.6 - 0.8, measured at 600 nm. The cells were transferred into 2 x 20 ml Sterilin tubes and centrifuged at 1950 x g for 10 minutes at 4°C. The supernatant was removed and the pellet resuspended in 20 ml of transformation buffer (50 mM CaCl₂, 10 mM Tris-HCl pH 7.4) and placed on ice for 30 minutes. The suspension was spun again, the supernatant was removed, and the pellet was resuspended in 5 ml of transformation buffer. Glycerol was added to the solution to give a final concentration of 15% (v/v) and made into 500 µl aliquots and stored at -80°C for future use when transforming plasmids.

2.3.4 Transformation

XL1-Blue competent cells (200 µl) were added to either 50 µl of the SDM reaction from the PCR mutagenesis, or 20 µl of the ligated construct. The solution was placed on ice for 30 minutes, heat shocked for 1 minute at 42°C and placed back on ice for a further 2 minutes. The bacteria were grown out with the addition of 500 µl of 2YT media to the solution and left at 37°C for 1 hour. The solution was plated onto blood agar (40 g blood

agar per litre) plates containing the relevant antibiotic (Table 2.3) and grown overnight at 37°C. Plates were stored at 4°C until required.

Procedure	Plate Composition
pET28a Mutagenesis/Cloning	30 µg/ml kanamycin
pUC18/19 Mutagenesis	100 µg/ml carbenicillin
pUC18 to pUC19 Cloning (Blue/White Selection)	100 µg/ml carbenicillin, 0.5 mM IPTG, 0.02% (w/v) X-Gal
Protein Purification in pET28a vector	30 µg/ml kanamycin, 50 µg/ml chloramphenicol

Table 2.3 Composition of agar plates.

2.3.5 Plasmid Purification

Single colonies were picked from the agar plates and shaken overnight at 37°C in 5 ml of 2YT media containing the same concentration of antibiotic as used in the agar plate that the colony had been taken from. The culture was transferred into 1.5 ml Eppendorf tubes and centrifuged at 6100 x g for 4 minutes. The plasmids were purified using a Qiagen QIAprep kit following the manufacturer's protocol. The supernatant was removed from the Eppendorf tubes and the cells resuspended in a total of 250 µl buffer P1. The cells were lysed under alkali conditions by the addition of 250 µl buffer P2 and neutralised with 350 µl of buffer N3. The solution was centrifuged at 17,900 x g for 10 minutes and the supernatant transferred into a spin column provided. The column was centrifuged at 17,900 x g for 1 minute and the supernatant discarded. 0.5 ml of buffer PB was added to remove any remaining cellular debris and centrifuged again at 17,900 x g for 1 minute. The supernatant was again discarded and 0.75 ml of buffer PE was added to the column and centrifuged twice at 17,900 x g for 1 minute discarding the supernatant in between each centrifuge. This was to remove salt within the solution and any remaining contaminants. Finally the column was placed into an Eppendorf tube and 50 µl of buffer EB was added. The solution was centrifuged for a final time at 17,900 x g for 1 minute and the flow-through containing the purified plasmid was collected in an Eppendorf tube. The

concentration and quality of the plasmid was determined using a Nanodrop 2000c (Thermo Scientific) and stored at -20°C until required.

2.3.6 DNA sequencing

Sequencing was either performed by MWG Eurofins (Germany) or manually using a Beckman Coulter CEQ8000 genetic analysis system. Chromatograms can be viewed in Appendix VI.

The samples for analysis by the CEQ8000 were prepared by first undergoing PCR. Between 100 - 195 ng pUC, or 260 - 390 ng of pET plasmids was taken and made up to 11 µl with dH₂O in 0.5 ml PCR tubes. The samples were heated at 95°C for 1 minute to denature the supercoiled DNA. 1 µl of 5 µM of primer (5 pmoles primer in the final reaction) and 8 µl of master mix were added to make a final volume of 20 µl. Samples were then cycled using a TECHNE thermocycler using the following conditions:

Segment 1: Cycles 30: 96°C for 20 sec, 50°C for 20 sec and 60°C for 4 min.

Segment 2: Cycles 01: Hold at 4°C.

The samples were cleaned-up by transferring into 1.5 ml Eppendorf tubes. 5 µl of stop solution (20 µl 3 M EDTA, 20 µl 100 mM NaOAc, 10 µl Glycogen and 60 µl 95% ethanol) was added to each sample and mixed thoroughly. The samples were centrifuged at 4°C for 5 minutes at 13000 x g. The supernatant was removed and 200 µl of 70% ethanol added followed by centrifugation at 4°C for 5 minutes at 13000 x g. This step was then repeated and the samples vacuum dried using a SpeedVac to remove any residual ethanol. Each pellet was re-suspended in 40 µl of SLS and transferred to a 96 well plate with a drop of mineral oil put on top of each sample. The samples were loaded into the CEQ8000 and sequenced.

2.3.7 Gene Amplification

Gene amplification was performed to obtain large amounts of the gene of interest for cloning purposes, rather than direct sub cloning of one plasmid to another. Gene amplification was also used after N123D mutagenesis to generate a product that could be sent for sequencing. This was to confirm whether site-directed mutagenesis was working when determining the cytotoxicity of *e*CDG. The *e*UDG mutant variants were amplified from pUC18 using flanking primers that also included the restriction sites required for cloning (Figure 2.4A). Ugi, obtained with thanks from Dr. Renos Savva (Birkbeck,

University of London, UK), was amplified using flanking primers designed to incorporate additional unpaired bases at the 5' terminus that included restriction sites required for cloning or the addition of a thrombin cleavage site (Figure 2.4B).

The PCR reaction solution was prepared in the same manner as for the site-directed mutagenesis protocol but with different cycling conditions:

Segment 1: Cycles 01: 98°C for 2 min.

Segment 2: Cycles 30: 98°C for 30 sec, 50°C for 30 sec and 72°C for 2 min per kb.

Segment 3: Cycles 01: 72°C for 10 min and hold at 4°C.

The amplified genes were stored at -20°C until required.

A

5' -TGTA AACGACGGCCAGT-3'

5' -TCACACAGGAAACAGCTATGAC-3'

B

5' -TCAGCTGAATTCCTGGTGCCGCGCGGCAGCATGACAAATTTATCTGACATCATTG-3'

5' -TAGTACAAGCTTATAACATTTTAATT-3'

Figure 2.4 Oligonucleotides used in the amplification of genes investigated. A) Oligonucleotides used for the amplification of *eUDG* mutant variants contained within pUC18. B) Oligonucleotides used for the amplification of Ugi cloned in the vector pRSETB. Additional unpaired overhang bases containing a thrombin cleavage and EcoRI site (red) and additional bases to allow for restriction enzyme docking (blue).

2.3.8 Gene Synthesis

hUDGΔ81 was prepared by complete gene synthesis using 18 oligonucleotides (each approximately 60 bases long; see Appendix II for sequences). These were designed to overlap at each terminus by approximately 20 bp (Figure 2.5) with the terminal oligonucleotides also containing an NdeI (5') and EcoRI (3') site. The oligonucleotides were mixed in equimolar amounts to give a stock concentration of 1 μM that was diluted in the PCR reaction mix to a final concentration of 100 nM. The PCR reaction mix was prepared as follows: 100 nM oligonucleotides, an additional 300 nM of the terminal oligonucleotides, 500 nM dNTPs, 3% DMSO, 5 μl 10x *Pfu* buffer, 1 μl of *Pfu* DNA polymerase (typically 2 - 3 u/μl) and made up to 50 μl with dH₂O in 0.5 ml PRC tubes. Samples were cycled using a TECHNE thermocycler using the following conditions:

Segment 1: Cycles 01: 98°C for 2 min.

Segment 2: Cycles 55: 98°C for 30 sec, 50°C for 30sec and 72°C for 2 min.

Segment 3: Cycles 01: 72°C for 10 min and hold at 4°C.

This produced a mixture of products, which included the desired full length, and was further amplified by PCR with the terminal oligonucleotides alone (Figure 5.1; see section 4.2.1 and 4.3.1 for more detail). 300 nM terminal oligonucleotides, 500 nM dNTPs, 3% DMSO, 5 μ l 10x *Pfu* buffer, 1 μ l of *Pfu* DNA polymerase (typically 2 - 3 u/ μ l) was added to 3 μ l of the PCR reaction into a new 0.5 ml PCR tube and underwent a further PCR reaction using the following cycling conditions.

Segment 1: Cycles 01: 98°C for 2 min.

Segment 2: Cycles 23: 98°C for 30 sec, 50°C for 30 sec and 72°C for 2 min.

Segment 3: Cycles 01: 72°C for 10 min and hold at 4°C.

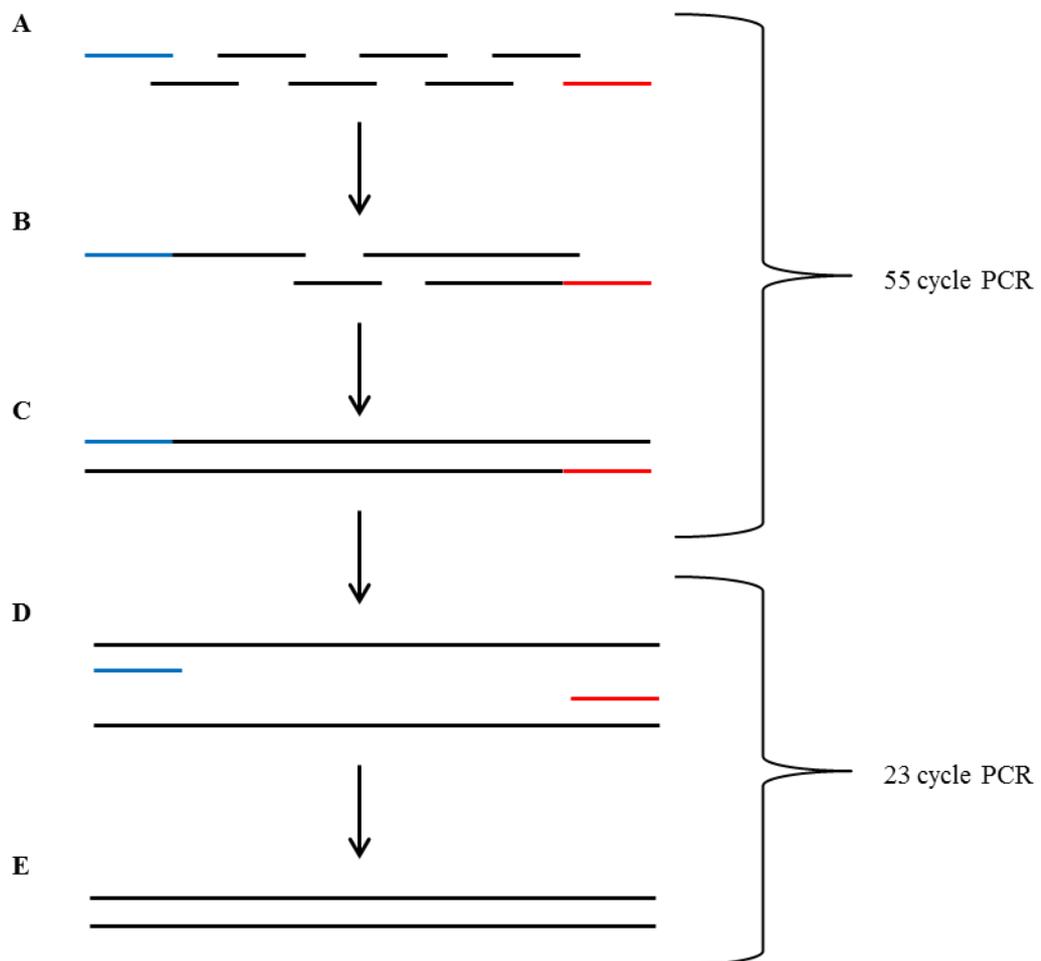


Figure 2.5 Protocol for *hUDGΔ81* gene synthesis. A) The 60mer overlapping oligonucleotides anneal and the unpaired regions are filled with DNA polymerase. B) Intermediate products. C) Full length gene product produced. D) Addition of the external oligonucleotides to amplify up the full length gene. E) Amplified full length gene available for cloning.

Each sample (2 μ l) was mixed with 3 μ l of ficoll loading solution (20% (w/v) ficoll, 10 mM EDTA and 0.1% (w/v) bromophenol blue) and loaded on a 1% agarose gel containing 0.1x GelRed, ran in 1x TBE buffer at 15 Vcm⁻¹ for 1 hour and imaged under UV light.

The resulting products underwent a PCR clean-up (QIAGEN) (see below) and stored at -20°C until required.

2.3.9 PCR clean-up

PCR reactions were cleaned-up using the Qiagen QIAquick PCR Purification Kit following the manufacturer's protocol. 5 volumes of Buffer PB to 1 volume of the PCR reaction was added and the solution mixed and transferred to spin column contained in a 2 ml collection tube. The sample was centrifuged for 1 min at 17,900 x g. The flow through was discarded and 750 µl Buffer PE was added to the spin column and the sample centrifuged as before for 1 min. Discard the flow through and centrifuge for a further 1 min. Transfer the spin column to a clean 1.5 ml Eppendorf tube and add 30 µl of Buffer EB and leave for 1 min followed by centrifugation as before for 1 min. The sample was then stored at -20°C until required.

2.3.10 Cloning

The *eUDG* mutant variants and the *hUDG*Δ81 constructs were cloned into their intended vector at an insert to vector ratio of 3 or 5:1. The insert and vector were double digested separately using 0.5 µl (Promega: typically 10 u/µl) of each enzyme; NdeI and EcoRI. Additionally 1 µl of 10x buffer D or B (Promega) respectively was added and made up to 10 µl with dH₂O and left at 37°C for 1 hour. An extra 0.5 µl of the appropriate buffer and 1 µl of TSAP (Promega: typically 2 u/µl) was added to the vector solution and left for 1 hour at 37°C to remove the 5' phosphates to prevent vector re-ligation. The insert was heated at 65°C and the vector at 75°C for 20 minutes to inactivate the enzymes. The samples were mixed followed by the addition of 2.5 µl of 10x ligase buffer (NEB) and 1 µl of ligase (NEB: typically 400 u/µl). The samples were then left at 37°C for a minimum of 1 hour to ligate before transformation.

2.3.11 Colony PCR

Blue/white selection of (pUC) colonies was used to determine potential positive clones. Where blue/white selection was not available (pET plasmids) a random selection of colonies were examined by colony PCR. Each colony was suspended in a PCR reaction mix containing 500 nM of each primer, 500 nM dNTPs, 3% DMSO, 10 µl of 5x GoTaq buffer, 0.2 µl of GoTaq DNA polymerase (typically 5 u/µl; Promega) and made up to final

volume of 50 μ l with dH₂O in 0.5 ml PCR tubes. Samples were cycled using a TECHNE thermocycler using the following conditions:

Segment 1: Cycles 01: 98°C for 2 min.

Segment 2: Cycles 30: 98°C for 30 sec, 50°C for 30 sec and 72°C for 1 min per kb.

Segment 3: Cycles 01: 72°C for 10 min and hold at 4°C.

The samples, after the addition of 3 μ l of ficoll, were loaded on a 1% agarose gel containing 0.1x GelRed, ran in 1x TBE buffer at 15 Vcm⁻¹ for 1 hour and imaged under UV light. An example can be seen in Figure 4.5.

2.3.12 Taq^qI Digestion

The N123D mutation generates an additional Taq^qI restriction site (TCGA) that would generate a different plasmid digestion pattern with Taq^qI; and which can be used as a diagnostic for successful mutagenesis. 1 μ l of 10x buffer B (Promega) and 0.5 μ l Taq^qI (NEB: typically 20 u/ μ l) was added to 1 μ l of DNA (from a typical plasmid preparation; approximately 100 ng) and made up to 10 μ l with dH₂O. The sample was left at 65°C for 1 hour for digestion to take place. 3 μ l of ficoll was added to the sample and loaded on a 1% agarose gel containing 0.1x GelRed, ran in 1x TBE buffer at 15 Vcm⁻¹ for 1 hour and imaged under UV light.

2.3.13 Purification of Enzymes

2.3.13.1 Cell Harvesting

The appropriate plasmid (100 ng) was transformed into 100 μ l of competent BL21(DE3)pLysS cells and plated onto blood agar (40 g blood agar per litre) plates as per the transformation protocol. A colony was selected and cultured overnight at 37°C in 5 ml 2YT media containing 30 μ g/ml kanamycin and 50 μ g/ml chloramphenicol. The overnight culture (5 ml) was transferred into 500 ml of 2YT media, containing the same concentrations of antibiotic, shaken at 37°C until an OD₆₀₀ of 0.5 - 0.8 was obtained. 1 ml of 100 mM IPTG was added, to induce the protein of interest, and shaken at 37°C for 2.5 hours. The solution was split into two 250 ml fractions and centrifuged at 5500 x g for 20 minutes. The supernatant was discarded and the pellets transferred into a Sterilin tube and buffered with 5 - 10 ml of 1x PBS [pH 7.4], which were then be stored at -20°C until required.

2.3.13.2 Sonication and Protein Preparation

The cell pellet was thawed on ice and 1x PBS [pH 7.4] was further added to 20 - 25 ml. The solution was sonicated in an ice bath for 10 minutes with pulse times of 30 seconds on and 40 seconds off. The solution was centrifuged at 31000 x g for 20 minutes and the supernatant collected and stored in a Sterilin tube, at 4°C if required, ready for purification.

2.3.13.3 Protein Purification

Purification was performed using an AKTA Prime (GE Healthcare). The sample was loaded in 1x PBS [pH 7.4], containing 20 mM imidazole and run through a His Trap FF Crude (GE Healthcare), to which the protein of interest binds, at a rate of 1 ml min⁻¹. The protein was eluted in 1x PBS [pH 7.4], containing 250 mM imidazole (see Figure 3.13) and the collected sample was subsequently concentrated using a 20 ml Vivaspin (Fisher) column by centrifuging at 1850 x g for 10 minutes. A volume of 5 ml 1x PBS [pH 7.4] was added to the column and centrifuged again as previously. A volume of 5 ml 1x PBS [pH 7.4] was again added and centrifuged at 1850 x g for 3 minutes cycles until approximately 0.5 ml remained. 2.5 ml of 1x PBS [pH 7.4] was added and all 2.5 ml removed into a Sterilin tube. A further 2 ml of 1x PBS [pH 7.4] was added, to obtain as much protein as possible, and the remaining 2.5 ml removed into the same Sterilin tube. The concentration was determined using a Nanodrop 2000c (Thermo Scientific) and glycerol was finally added to the solution to a concentration of 50% (v/v). The solution was divided into 500 µl aliquots that were stored at -20°C until required.

2.3.13.4 SDS-PAGE

The SDS-PAGE gel was comprised of two solutions denoted the “stacking” and “resolving” gels. The stacking gel contained 0.33 ml acrylamide (30% acrylamide: 0.8% bis-acrylamide (37.5:1)), 1.45 ml dH₂O, 0.62 ml gel buffer (3 M Tris, 0.3% (w/v) SDS pH 8.45) and 5 µl TEMED. The running gel contained 1.665 ml acrylamide (30% acrylamide: 0.8% bis-acrylamide (37.5:1)), 1.665 ml dH₂O, 1.665 ml gel buffer and 5 µl TEMED. 25 µl of 20% (w/v) ammonium persulphate was added to the running gel, which was applied first and allowed to set, and isopropanol was applied on top to prevent the formation of air bubbles. Once set, the isopropanol was removed and 12.5 µl ammonium persulphate was added to the stacking gel, which was applied on top of the running gel to create a complete gel. An example can be seen in Figure 3.10.

1 ml samples of pre and post protein induction cultures were taken and centrifuged at 17,900 x g for 3 minutes. 60 µl of 4x gel loading buffer (2.5 ml 1 M Tris pH 6.8, 4 ml 20% (w/v) SDS, 40 mg bromophenol blue, 2 ml glycerol, 1 ml dH₂O, 860 µl β-mercaptoethanol) was added and heated at 95°C for 15 minutes. 5 µl of 4x gel loading buffer was added to 15 µl samples of flow through, 250 mM imidazole protein elution and the protein solution after concentration. The samples were loaded and run in anode (200 mM Tris pH 8.9) and cathode (100 mM Tris, 100 mM Tricine, 0.1% SDS pH 8.2) buffers at 18 Vcm⁻¹ for 50 minutes. The gel was stained in coomassie blue (1 g Coomassie, 100 ml glacial acetic acid, 400 ml methanol, 500 ml dH₂O) on a rocking table for 30 minutes and was left in destain (400 ml methanol, 100 ml glacial acetic acid, 500 ml dH₂O) on a rocking table overnight.

2.3.14 5' Radiolabelling of oligonucleotides

Oligonucleotides (60 pmoles; Table 2.2) were radiolabelled at its 5'-end using 1 µl of 10 µCi/µl at 3000 Ci/mmol of [γ -³²P]ATP, 1 µl of PNK (NEB: typically 10 u/µl) and 2 µl of 10x PNK buffer and made up to 20 µl using dH₂O. The samples were left at 37°C for a minimum of 1 hour for the reaction to take place. The reaction was stopped using 20 µl stop solution (80% (v/v) formamide, 10 mM EDTA, 2 mM NaOH and 0.1% (w/v) bromophenol blue). The samples were run on a 12.5% denaturing polyacrylamide gel (20 cm by 0.3 mm) at 30 Vcm⁻¹ for approximately 1 hour. The gel was then exposed to X-ray film for 2 minutes and the band corresponding to the labelled fragment was excised. The gel slice was put into a P1000 tip containing glass wool (with the tip sealed with parafilm) and the DNA was eluted from the gel by the addition of 400 µl of Tris-EDTA 10:1 (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) and left overnight on a rocking table. The solution was removed from the tip using a P1000 pipette and pulsed centrifugation (\leq 6000 x g). The glass wool prevents the removal of the gel slice from the tip. The DNA was precipitated from the solution by adding 50 µl of 3 M NaOAc and 1ml of ethanol, leaving the sample on dry ice for a minimum of 1 hour. The solution was centrifuged at 17,900 x g for 10 minutes and the supernatant was removed; leaving a pellet of DNA that can be traced using a Geiger counter. The pellet was washed with 200 µl of 70% ethanol and centrifuged at 17,900 x g for 2 minutes. The solution was again removed to leave a pellet that was dried in a SpeedVac to remove any residual ethanol. The pellet was finally resuspended in Buffer 1 (10 mM Tris-HCl, 25 mM NaCl, 2.5 mM MgCl₂ [pH 8.0] (Drohat *et al.*, 1999a)) or Buffer 2 (10 mM NaMES, 25 mM NaCl, 2.5 mM MgCl₂ [pH 6.3] (Kwon *et al.*, 2003)),

for UDG and UYDG or CYDG respectively, to obtain approximately 10 cps/ μ l as estimated using a hand held Geiger counter.

Oligonucleotide	Sequence
U	5' -CCGAATCAGTGC GCA UAGTCGGTATTTAGCC-3'
T	5' -CCGAATCAGTGC GCA TAGTCGGTATTTAGCC-3'
C	5' -CCGAATCAGTGC GCA CAGTCGGTATTTAGCC-3'
^M C	5' -CCGAATCAGTGC GCA ^M CAGTCGGTATTTAGCC-3'
G	5' -GGCTAAATACCGACT G TGCGCACTGATTCGG-3'
A	5' -GGCTAAATACCGACT A TGCGCACTGATTCGG-3'
AP	5' -GGCTAAATACCGACT AP TGCGCACTGATTCGG-3'
Z	5' -GGCTAAATACCGACT Z TGCGCACTGATTCGG-3'
I	5' -GGCTAAATACCGACT I TGCGCACTGATTCGG-3'
ssC(GAT)	5' -GGATAAATAGGGAGT C TGAGAAGTGATTAGG-3'
ssC(polyA)	5' -AAAAAAAAAAAAAAAA C AAAAAAAAAAAAAAAA-3'
G2	5' -GCTAAATATATATAT G TTATATAATTATTCG-3'
C2	5' -CGAATAATTATATA A CATATATATATTTAGC-3'
^M C2	5' -CGAATAATTATATA A ^M CATATATATATTTAGC-3'
A2	5' -GGCTAAATACCGACC A CGCGCACTGATTCGG-3'
C(G)	5' -CCGAATCAGTGC GCG CGGTCGGTATTTAGCC-3'
gap	5' -GGCTAAATACCGACT-3' 5' -TGCGCACTGATTCGG-3'
long C	5' -CCGTACTGAATCAGTGC GCA CAGTCGGTATTTACGATAGCC-3'
long gap	5' -GGCTATCGTAAATACCGACT-3' 5' -TGCGCACTGATTCAGTACGG-3'
HEG	5' -GGCTAAATACCGACC HEG CGCGCACTGATTCGG-3'
APHEG	5' -GGCTAAATACCGACC APHEGAP GCGCACTGATTCGG-3'
U(G)	5' -CCGAATCAGTGC GCG UGGTCGGTATTTAGCC-3'
C(AG)	5' -CCGAATCAGTGC GCA CGGTCGGTATTTAGCC-3'
A(AG)	5' -GGCTAAATACCGACC A TGCGCACTGATTCGG-3'
C(GA)	5' -CCGAATCAGTGC GCG CAGTCGGTATTTAGCC-3'
A(GA)	5' -GGCTAAATACCGACT A CGCGCACTGATTCGG-3'

Table 2.4 Oligonucleotides used in excision assays. Central target base highlighted in bold. AP; abasic site, Z; anthraquinone pyrrolidine and gap; unpaired, HEG; hexethyleneglycol.

2.3.5 Excision Assays

Each radiolabelled oligonucleotide (U, T, C, ^MC) was split into equal aliquots and paired with 1 μ l (45 to 60 μ M; approximately 2-fold excess) of each of the complementary oligonucleotides (G, A, AP), determined by their central base (highlighted in bold in Table 2.4), to give a maximum of 12 oligonucleotide pairings. The samples were placed in a heating block at 95°C 10 minutes; the block was then removed and left to cool to ambient temperature overnight, allowing the strands to anneal and form a duplex. Occasionally the duplexes formed were further purified to remove the excess complementary oligonucleotide. This was performed by running the sample on a 16% native polyacrylamide gel (20 cm by 0.3 mm) at 15 Vcm⁻¹ for approximately 3 hours. The DNA was then purified as stated above.

2.3.15.1 Activity Assays

To determine whether an enzyme could excise a specific base from a mismatched duplex, enzyme activity excision assays were performed. Enzyme at various concentrations ranging from 3 - 3000 nM diluted in Buffer 2 (Buffer 1 was also used for *e*UDG and *e*UYDG) was added in equal volumes with 1 nM - 60 μ M radiolabelled duplex substrates on ice. The samples were incubated at 37°C for between 1 minute and 24 hours. The assays were stopped under alkali conditions with 3 μ l of stop solution (as previous) and heated at 95°C for 20 minutes. This causes cleavage of the ribose sugar and subsequent phosphodiester backbone cleavage.

2.3.15.2 Quantitative Assays

Assays were performed as described above in triplicate except that the reaction was stopped using 10% (v/v) piperidine before heating at 95°C for 20 minutes. The samples were lyophilised to dryness and resuspended in 10 μ l stop solution. Treatment with hot piperidine yielded single reaction products, while samples that were simply heated in stop solution yielded products with ragged ends and multiple bands on the denaturing polyacrylamide gels (see Figure 3.15).

2.3.15.3 Denaturing Polyacrylamide Gel Electrophoresis

The products of the cleavage reactions were run on a 12.5% denaturing polyacrylamide gels containing 8 M urea. The gels were run at 30 Vcm⁻¹ for approximately 1 hour in 1x

TBE and fixed in (v/v) 10% acetic acid. The gels were transferred onto Whatman 3MM paper, covered with Saran wrap and dried under vacuum at 86°C for 1 hour. The gels were exposed overnight to a phosphorimage screen and scanned using a Typhoon FLA 7000 phosphorimager (GE Healthcare).

2.3.15.4 Cleavage Quantification

The intensity of the bands from each gel was estimated using ImageQuantTL software (GE Healthcare). The rate of reaction (initial velocity) was determined from plots of percentage cleaved against time, using SigmaPlot, by fitting each set of data to a single exponential rise to maximum and averaged (Appendix IV). The rate of cleavage of some substrates was very low and in these instances an estimate of the rate constant was obtained from the fraction cleaved at a given time, assuming a simple exponential process.

2.3.16 Detection of 5-Methylcytosine

2.3.16.1 DNA Probing

An 80 mer oligonucleotide (Table 2.5) was designed encompassing a cytosine or 5-methylcytosine in a CpG context in the centre with primer binding sites at either end. The sequences flanking the CpG site are the same as those used for the cleavage assays. This region was targeted by a probe oligonucleotide containing a hexaethylene glycol (HEG) linker opposite the target cytosine or 5-methylcytosine and a 3' propyl to prevent polymerase read through and extension (Table 2.5). HEG was also used as this was found to provide one of the highest cleavage rates for the enzymes investigated (see section 3.3.12). The target and probe oligonucleotide in excess were mixed, heated to 95°C and allowed to cool and run on a 16% native polyacrylamide gel at 15 Vcm⁻¹ for approximately 3 hours. This was so that we could be certain that we had a fully annealed duplex and any excess oligonucleotides would be removed. The DNA was located using UV shadowing onto a TLC plate, eluted and purified as per 5' labelling (section 2.3.12) giving a final concentration of $\leq 20 \mu\text{M}$, and stocks of a maximum of 500 and 5 nM were made for experimental use.

Oligonucleotide	Sequence
Target	5' -TAGCGTAGGGAGATATACCATGGGCAGCAGCCAGTGCGCAC / ^M CG GTCGGTATTCTGAGATCCGGCTGCTAACAAAGTTGATCAT-3'
Probe	5' -GGCTAAATACCGACTHTGCGCACTGATTCGG-propyl-3'
Forward Primer (pET28a For)	5' -GGAGATATACCATGGGCAGCAGC-3'
Reverse Primer (pET28a Rev)	5' -CTTTGTTAGCAGCCGGATCTCAG-3'

Table 2.5 Oligonucleotides used in the cytosine detection assay. H; HEG.

2.3.16.2 Incubation with *e*CYDG

The 500 and 5 nM target-probe (TP) stocks were incubated with excess *e*CYDG for 24 hours to allow cleavage if the target cytosine was unmethylated. The reaction product was then reacted with 10% (v/v) piperidine and heated at 95°C for 30 minutes to cleave the phosphodiester backbone, and lyophilised to dryness.

2.3.16.3 Real-time PCR

The lyophilised samples were resuspended in 13.6 µL of dH₂O followed by the addition of 1 µL of 10 µM primers, 1 µL of 10 mM dNTPs, 4 µL of 5x GoTaq buffer, 0.2 µL of GoTaq DNA polymerase (typically 5 u/µl; Promega) and 0.2 µL 10X SYBR Green. The reactions were transferred to glass capillaries for use in a Roche LightCycler 1.5 qPCR machine. The cycling conditions are as follows:

Segment 1: Cycles 01: 98°C for 2 min.

Segment 2: Cycles 30: 98°C for 10 sec, 55°C for 10 sec and 72°C for 10s.

Segment 3: Cycles 01: 72°C for 4 min and hold at 4°C.

Chapter 3: Production and Properties of *e*CYDG

3.1 Introduction

Generating a glycosylase that has the ability to excise a normal DNA base might be expected to generate a protein that is cytotoxic to the host organism. This has previously been noted for *e*CDG (Kavli *et al.*, 1996, Kwon *et al.*, 2003). human CDG (*h*CDG) has previously been prepared using *recA*⁺ *E. coli* cells (Kavli *et al.*, 1996), while it has only been possible to express *E. coli* CDG using *in vitro* transcription and translation (Handa *et al.*, 2002). An alternative solution to expressing *e*CDG is to mutate leucine 191 to an alanine (L191A), that renders the enzyme inactive (Jiang *et al.*, 2001), followed by the N123D mutation allowing expression and purification of the enzyme (Kwon *et al.*, 2003). As described in the introduction (section 1.3.1.4) L191A disables the enzyme's ability to flip the target base into an extrahelical conformation. The enzyme's activity can then be restored by incorporating a bulky synthetic nucleoside (*i.e.* pyrene) opposite the target cytosine base. Pyrene, being planar, can stack within the DNA duplex while its large size causes it to encroach upon the cytosine's space, forcing it into an extrahelical conformation for enzyme binding and base excision. This enzyme was previously denoted a cytosine/pyrene DNA glycosylase (*e*CYDG) (Kwon *et al.*, 2003).

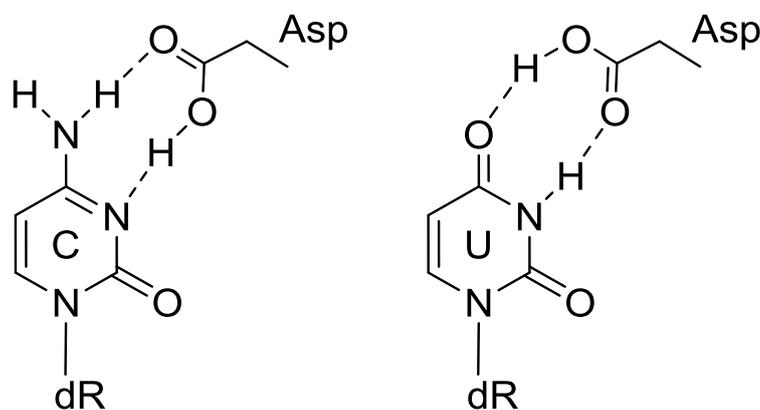


Figure 3.1 Recognition of cytosine and uracil by the 180° rotation of Asp123's side chain.

*e*CYDG excises cytosine at an optimal pH of 6.3 at which D123 and not D64 is in a protonated state (Handa *et al.*, 2002, Kwon *et al.*, 2003). This should also be applicable to any CDG in order to optimise its activity. In the wild type UDG asparagine 123 is held in a fixed conformation by a triad of water molecules that enables it to be specific for uracil recognition. Aspartate does not form these bonds with the water molecules and it is able to

freely rotate around its side chain (Figure 3.1), allowing it to retain activity against uracil as well as cytosine (Pearl, 2000).

The ability of *e*CYDG to excise cytosine from DNA led us to suggest that it should be able to discriminate between cytosine and 5-methylcytosine in the same way that UDG is able to distinguish between uracil and thymine, due to the steric hindrance between Tyr66 and the 5-methyl group.

3.2 Experimental Design

3.2.1 Mutagenesis of UDG

To generate a CDG and assess its properties, site-directed mutagenesis (SDM) was attempted on the wild type *e*UDG. Even though previous studies had failed to express *e*CDG in *E. coli*, as a result of its cytotoxicity, we hoped that this might be possible under the tight control of expression in the pET vector. The N123D mutation generates a third cleavage site for restriction enzyme Taq^qI (TCGA; mutated from TCAA in the wild type) (Figure 3.2A) compared to two found in pUC19*e*UDG. This difference was used to determine a potentially successful mutation, which was then confirmed by DNA sequencing. Mutagenesis for N123D was first performed on pUC19*e*UDG but this was unsuccessful due to cytotoxicity and was therefore performed on pUC18*e*UDG that produced a positive clone.

A

Forward Primer 5' -GCGTCAGGGCGTTCTGCTACTCGATACTGTGTTGACGGTACGCGC-3'
 Plasmid Template 3' -CGCAGTCCCGCAAGACGATGAGTATGACACAAACTGGCATGCGCG-5'

Plasmid Template 5' -GCGTCAGGGCGTTCTGCTACTCAATACTGTGTTGACGGTACGCGC-3'
 Reverse Primer 3' -CGCAGTCCCGCAAGACGATGAGCTATGACACAAACTGGCATGCGCG-5'

B

Forward Primer 5' -GCACCGCATCCGTCGCCGGCGTCGGCGCATCGTGGATTC-3'
 Plasmid Template 3' -CGTGGCGTAGGCAGCGGCGAAAGCCGCGTAGCACCTAAG-5'

Plasmid Template 5' -GCACCGCATCCGTCGCCGCTTTCGGCGCATCGTGGATTC-3'
 Reverse Primer 3' -CGTGGCGTAGGCAGCGGCCGCAGCCGCGTAGCACCTAAG-5'

Figure 3.2 Mutation primers annealed to template plasmid. The N123D (A) and L191A (B) base pair mismatches are highlighted in red with the Taq^qI restriction site underlined.

The L191A mutation (Figure 3.2) was performed on pUC18*e*CDG to generate *e*CYDG (to overcome cytotoxicity) that could then be cloned into the pET vector and expressed. The L191A mutation was also performed by itself to generate *e*UYDG. Table 3.1 shows the DNA sequences at the mutation sites for these enzymes. Once the gene was created it was excised from the pUC18 vector and cloned the pET28a expression vector using appropriate restriction enzymes. To confirm the cloning had been successful the plasmid was again mapped by Taq^I digestion and subsequent DNA sequencing.

Enzyme	Sequence
UDG	5' -TGCTACTCAATACTGTG--CGTCGCCGCTTTCGGCGCA-3'
UYDG	5' -TGCTACTCAATACTGTG--CGTCGCCGGCGTCGGCGCA-3'
CYDG	5' -TGCTACTCGATACTGTG--CGTCGCCGGCGTCGGCGCA-3'

Table 3.1 Theoretical DNA sequences of the different enzymes at the mutagenic sites. Sequence variations highlighted in red; -- denotes nucleotide sequence between the two mutation sites.

3.2.2 Excision Assays

3.2.2.1 Initial Assays

To assess the excision properties of *e*UDG, *e*UYDG and *e*CYDG, DNA cleavage assays were performed and analysed by denaturing PAGE. The oligonucleotide containing the target base was radiolabelled with ³²P at its 5' end and annealed to a complementary strand to form a duplex substrate for the enzyme. The substrates generated contained a target U, T, C or ^MC paired opposite a G, A, AP (abasic site). Varying concentrations of enzyme were incubated with the DNA for up to 30 minutes and the reaction quenched using a formamide stop solution (containing 10 mM NaOH) to cause cleavage of the phosphodiester backbone. The resulting products were run on denaturing PAGE, phosphorimaged and analysed.

3.2.2.2 *e*CYDG Rate of Reaction Determination

The DNA cleavage assays were performed in the same manner in triplicate, their catalytic rate calculated and averaged to determine the rate of reaction of the enzyme against a variety of different substrates. The assays were performed over a 24, four or one hour time course depending on the activity of the enzyme towards a particular substrate. As well as G, A and AP used to generate duplexes with the target base, Z (anthraquinone pyrrolidine),

HEG (hexaethylene glycol) and a gap (unpaired) duplex substrates were generated including the investigation of single stranded substrates. Anthraquinone pyrrolidine is another bulky synthetic nucleoside analogue that was used to mimic the reported effect of pyrene. After incubation with *e*CYDG the sample was placed on dry ice to stop the reaction followed by treatment with 10% (v/v) piperidine to cause specific cleavage at the 3' carbon-phosphate bond of the DNA to generate a single product band. The resulting products were run on denaturing PAGE, phosphorimaged and analysed.

3.3 Results

3.3.1 N123D mutagenesis of pET*e*UDG and pUC19*e*UDG

Initial mutagenesis experiments were performed on the pET28a cloning and expression vector containing the *e*UDG gene (pET*e*UDG). The aim was to mutate the wild type asparagine to aspartate (N123D) altering the hydrogen bonding pattern in the active site, allowing the enzyme to recognise cytosine as well as uracil. Even though this has previously been unsuccessful (Kavli *et al.*, 1996, Kwon *et al.*, 2003), it was attempted due to the tight control of expression given by the pET vector. This failed to produce any colonies upon transformation of products of the Quickchange mutagenesis reaction. It was therefore decided to perform mutagenesis on the sequence within the pUC19 cloning vector (pUC19*e*UDG) in conjunction with pET*e*UDG. This too proved difficult but eventually colonies were obtained upon transformation for both vectors. These were subjected to restriction mapping with Taq^qI digestion and the agarose gels are shown in Figure 3.5A; the presence of a different cleavage pattern to the wild type UDG for both vectors suggested that the mutagenesis had been successful. However, upon sequencing (MWG), although the clones were found to contain the mutated sequence, they contained multiple repeats of the primer sequence (Figure 3.3) or other mutations in addition to the desired A to G mutation. Further pUC19*e*CDG clones were obtained, which also gave the correct banding upon restriction mapping, as shown in Figure 3.5B. Theoretical plasmid digest maps are shown in Figure 3.4.

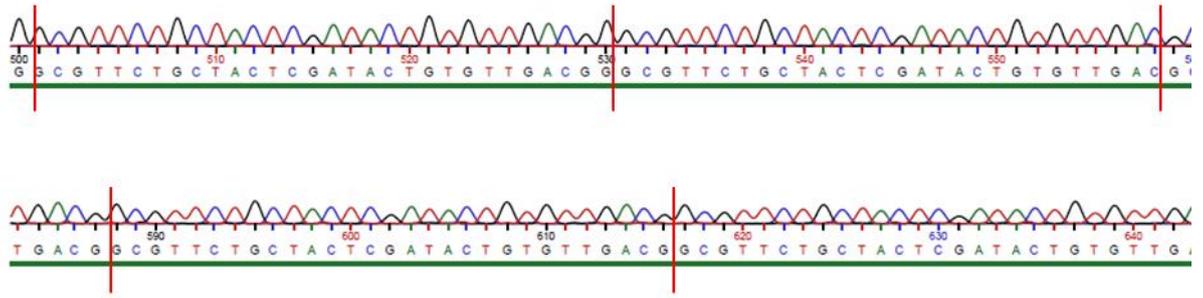


Figure 3.3 Sequencing chromatogram showing multiple primer sequence repeats.

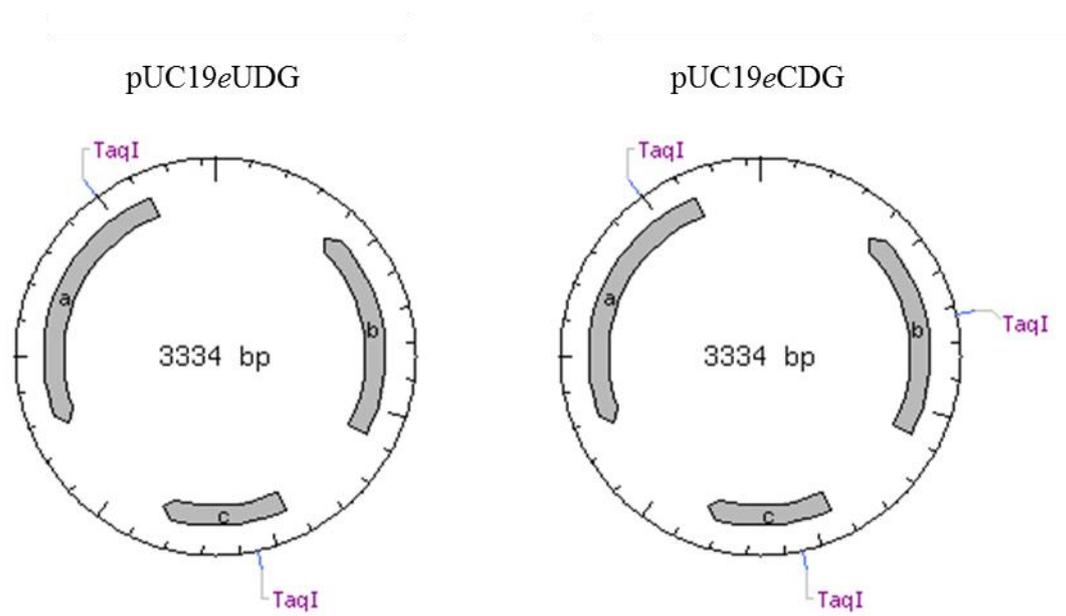


Figure 3.4 Theoretical plasmid digest with $TaqI$ of pUC19eUDG and pUC19eCDG that would produce two and three restriction fragments respectively.

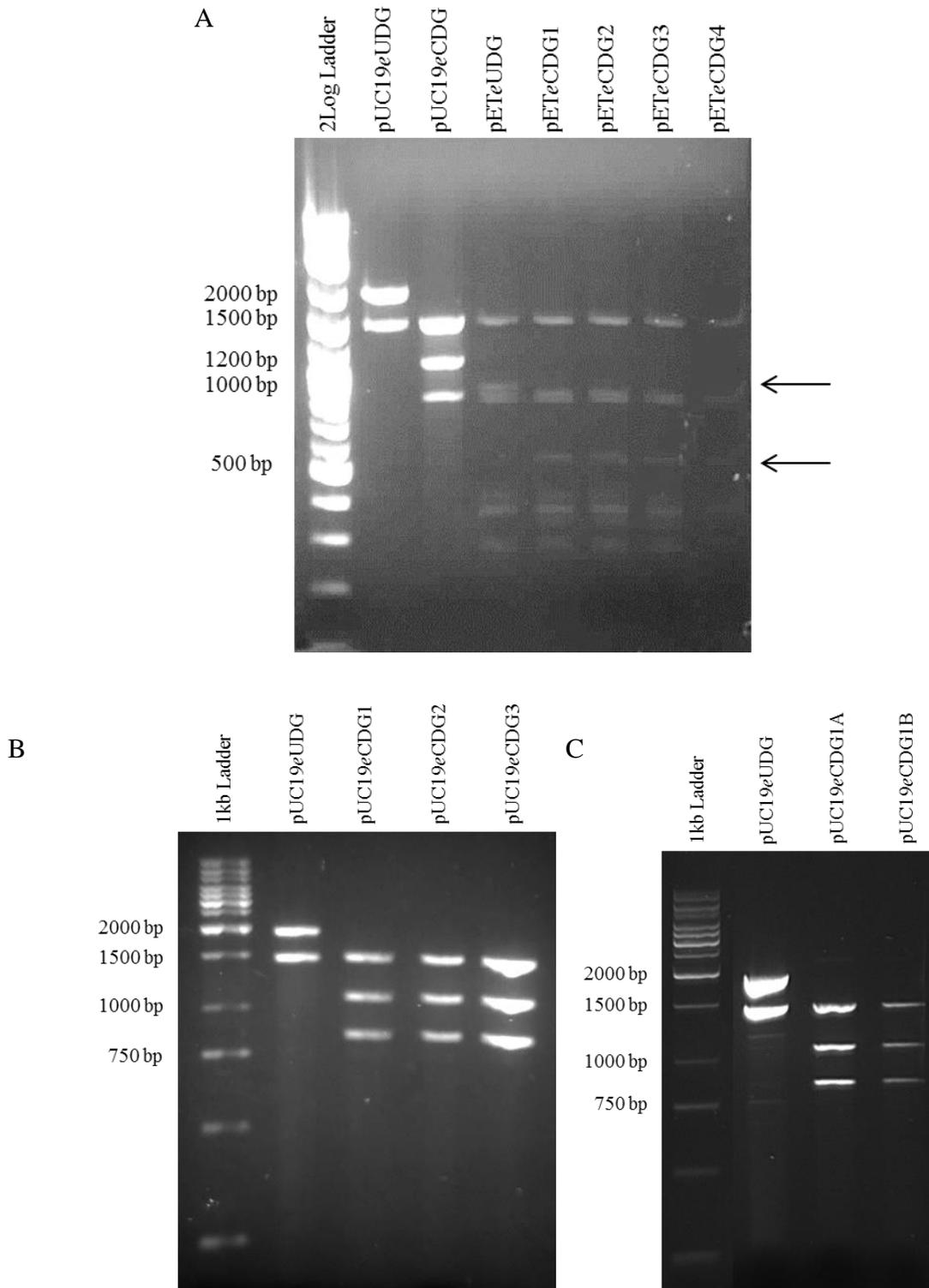


Figure 3.5 Gel electrophoresis of clones digested with $Taq^{\alpha}I$. A) pUC19eCDG and pETeCDG clones post mutagenesis. Arrows indicate loss of (top) and production (bottom) of a band of pETeCDG clones in comparison to pETeUDG. B) pUC19eCDG clones after first round of mutagenesis. C) pUC19eCDG clones after second round of mutagenesis to remove the extra cytosine residue. 0.7% agarose gel ran in 1 x TBE at 15 Vcm⁻¹ for one hour.

However, sequencing (MWG) showed that all three clones contained an additional secondary mutation (Figure 3.6; the addition of an extra cytosine within the region of the mutagenic primer) generating a frame shift mutation. Further clones, that also showed the correct restriction banding pattern, were all found to have either additions or deletions within the primer region. The clone pUC19*e*CDG1 (Figure 3.5B), containing the extra cytosine, was subjected to a second round of mutagenesis using the same primers in an attempt to remove the extra base. The colonies from this reaction also showed the correct pattern with Taq^αI (Figure 3.5C), but once again sequencing (MWG) revealed that they still contained the extra cytosine base. This strongly suggested that *e*CDG is highly toxic in *E. coli*, even when present out of frame in pUC19; this is further investigated and discussed in Chapter 6.

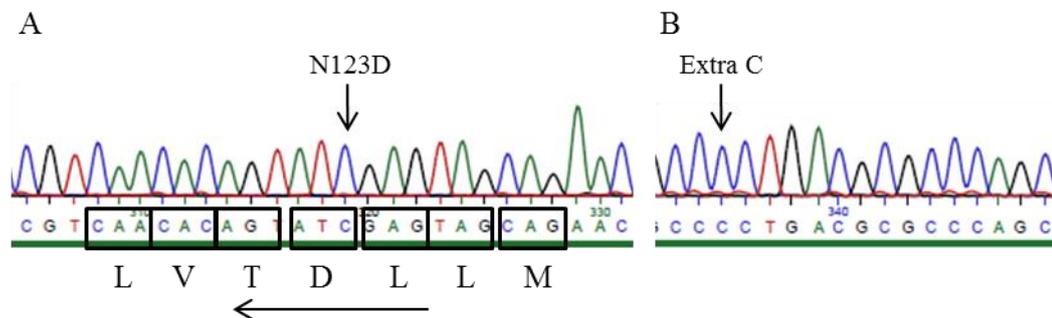


Figure 3.6 Sequencing analysis of pUC19*e*CDG1 highlighting the N123D mutation (A) and extra cytosine base (B). Amino acid sequence shown in A with direction of translation indicated by the arrow.

3.3.2 Determining the Effects of Mutagenesis

As a control to determine whether the mutagenesis was occurring correctly, pUC19*e*UDG was again subjected to N123D mutagenesis, followed by a second round of PCR using pUC18/19 universal primers (Figure 3.7) to amplify the potential *e*CDG gene.

Forward: 5' -GTAAAACGACGGCCAGT-3'
 Reverse: 5' -TCACACAGGAAACAGCTATGAC-3'

Figure 3.7 Universal primer sequences.

The PCR product was sequenced and showed that the N123D mutation had been successful and contained no secondary mutations, unlike that seen in Figure 3.6. It therefore appear that the failure to generate a clone of *e*CDG did not arise from problems in the site directed mutagenesis itself, but from the subsequent cloning and transformation, which only selected for colonies that contained a mutated (inactive) version of the gene. This is consistent with the suggestion that *e*CDG is extremely cytotoxic, as shown by

Kwon *et al.*, 2003 and Handa *et al.*, 2002. However, it was not expected that positive clones could not even be obtained when using the pUC19 vector, especially since the sequence is out of frame with the *lacZ* gene.

3.3.3 Cloning of *eCDG* into pUC18 and pUC19

The suggestion that *eCDG* might be very cytotoxic led us to attempt to clone the amplified *eCDG* gene into pUC18, as the multiple cloning site is in the opposite orientation. The gene would therefore be cloned in the opposite orientation relative to the *lacZ* and would produce a nonsense protein if transcribed and translated from the *lacZ* promoter. As a control we attempted to clone the amplified *eCDG* gene into pUC19. Colonies were obtained from both reactions and were subjected to Taq^qI digestion (Figure 3.8). The pUC19*eCDG* clone produced an incorrect banding pattern, whereas the pUC18*eCDG* clone produced the correct fragment sizes of approximately 1444, 1107 and 785 bp. This was subsequently sequenced (MWG) and confirmed that the clone was correct, containing the *eCDG* gene with no secondary mutations. These experiments suggest that the *eCDG* gene is extremely cytotoxic, even when cloned in pUC19. This was especially surprising since the *eCDG* gene is out of frame with the *lacZ* gene; this effect is explored further in Chapter 6.

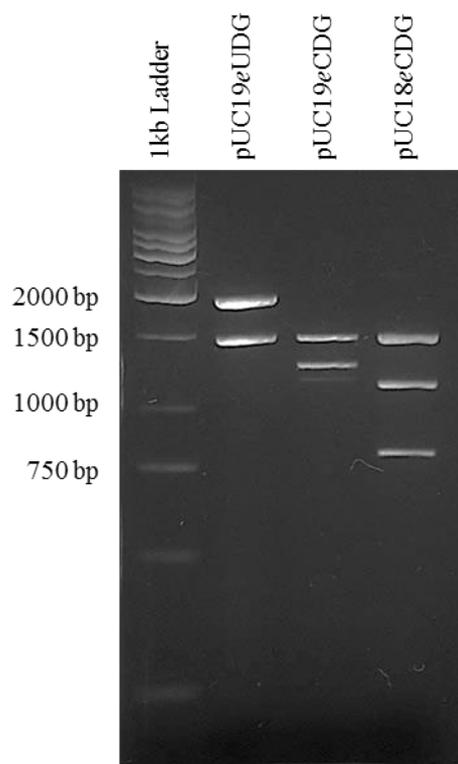


Figure 3.8 Gel electrophoresis of pUC clones digested with TaqI after *e*CDG cloning. 0.7% agarose gel ran in 1x TBE at 15 Vcm⁻¹ for one hour.

3.3.4 Generating *e*CYDG

In order to express an *E. coli* enzyme with CDG activity it was decided to further mutate pUC18*e*CDG to substitute the “pushing and plugging” leucine to an alanine (L191A). The primers used for this are shown in Figure 3.2B. This mutation creates an *e*CYDG that has no enzymatic activity towards cytosine unless rescued with a synthetic pyrene base opposing the cytosine, as shown previously by Kwon *et al.*, 2003. Transformation of pUC18*e*CDG after SDM to introduce the L191A mutation produced successful clones with both the N123D and L191A mutations, thus yielding pUC18*e*CYDG.

3.3.5 Cloning *e*CYDG into pET28a and pUC19

The *e*CYDG gene was then subcloned into pUC19 and pET28a, *via* the restriction enzymes HindIII/EcoRI and NdeI/EcoRI respectively. Clones were obtained upon transformation and subjected to Taq^qI digestion and sequencing where a positive digestion pattern was obtained for pET*e*CYDG3, shown in Figure 3.9A, which was then confirmed by sequencing, successfully generating pET*e*CYDG. Positive clones were also obtained for

pUC19 as shown in Figure 3.9A (pUC19eCYDG1 and 3), and again were confirmed through sequencing.

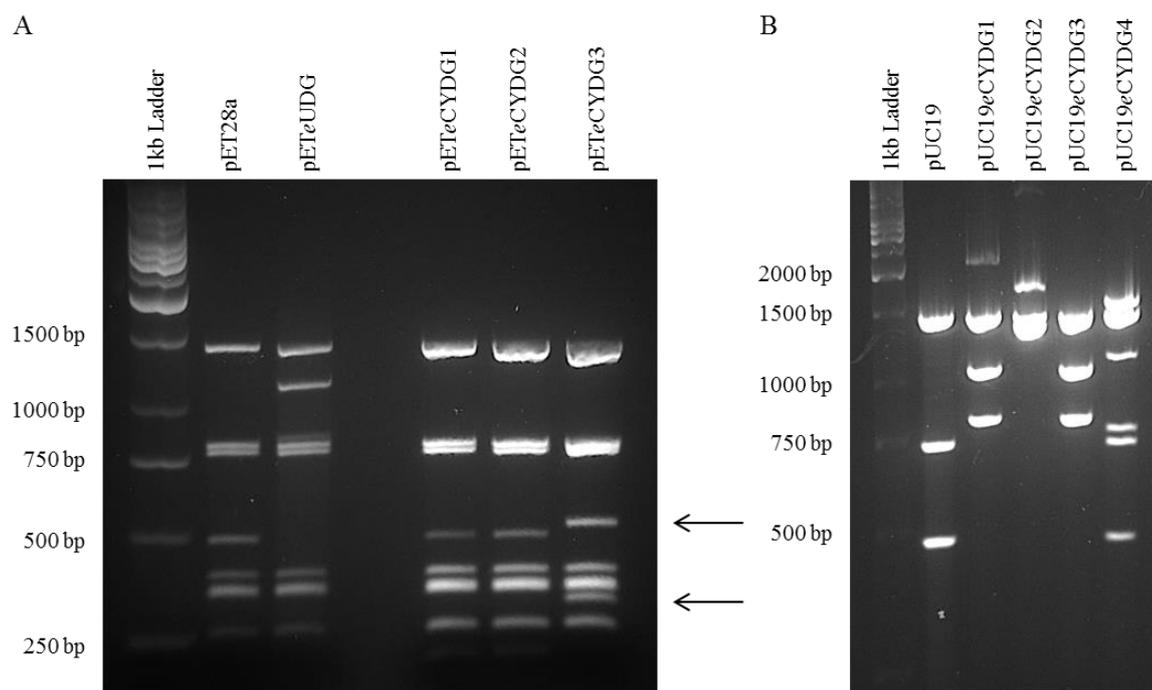


Figure 3.9 Gel electrophoresis of *e*CYDG clones digested with Taq^I. Digestion patterns of pETeCYDG (A) and pUC19eCYDG (B) clones. Arrows indicate the different banding pattern seen for the positive pETeCYDG3 clone from pET28a and the pETeUDG. 0.7% agarose gel ran in 1x TBE at 15 Vcm⁻¹ for one hour.

3.3.6 Enzyme Expression

The pET28a vector was used for expression of *e*UDG proteins as it contains a 6x His-Tag that can be used for purifying the protein of interest using a nickel chelating column. The purification of this His-tagged *e*UDG is shown in Figure 3.10. Expression of the protein was induced with IPTG, as seen by an increase in band intensity of approximately 27 kDa in the post-induction lane (Figure 3.10). A small amount of the enzyme was lost when the sample was pelleted. There was further loss when loading the sample onto the column as shown in the flow through lane. 20 mM imidazole was used to remove most of any protein contaminants remaining whilst UDG was finally eluted in 250 mM imidazole, clearly seen in the final lane as indicated by the arrow (Figure 3.10). The imidazole was removed from the 250 mM elution by buffer exchange using a PD-10 column, which also further purified and concentrated the enzyme giving a yield of 0.75 mg.

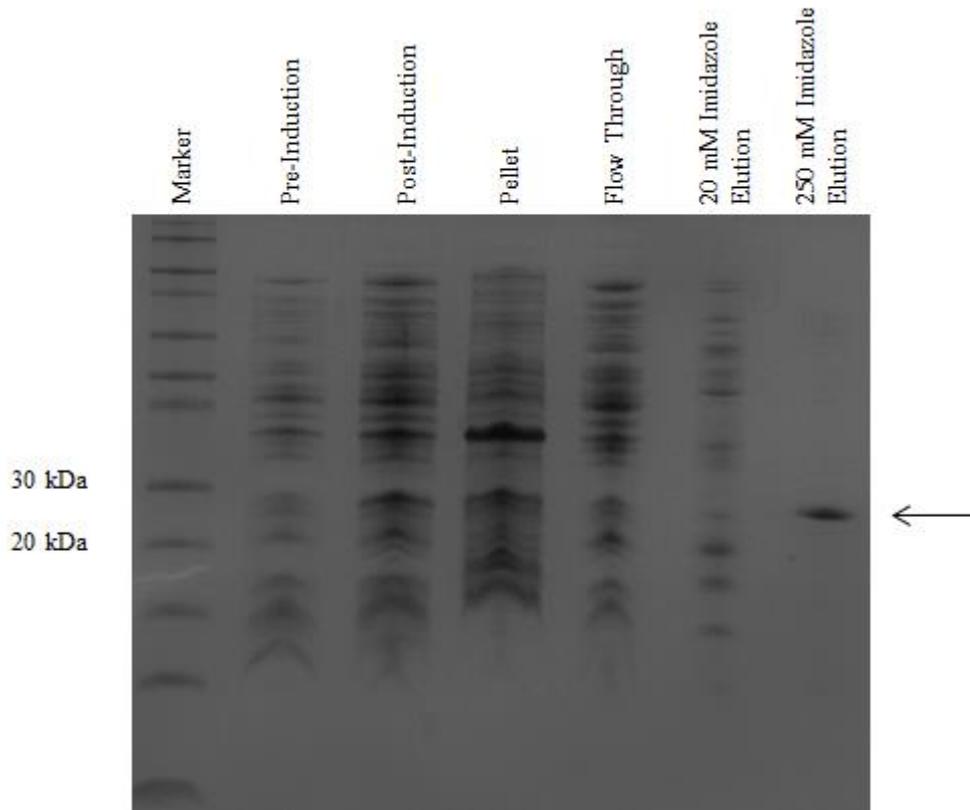


Figure 3.10 SDS-PAGE purification of *eUDG*. Samples were taken at different purification stages to assess enzyme expression and purity.

For purification of *eCYDG* it was decided to load the sample in 20 mM imidazole followed by enzyme elution in 250 mM imidazole. This was due to the tight enzyme binding to the column with minimal loss seen in the 20 mM imidazole elution when purifying *eUDG*. *eCYDG* was further purified and concentrated using a 10000 MW cut off Vivaspin centrifugal concentrator and 0.75 mg of *eCYDG* was obtained; the SDS-PAGE gel for this purification is shown in Figure 3.11. The same procedure was used for *eUYDG*, generated by performing the L191A mutation on pET*eUDG*, and further production of *eUDG* (Figure 3.12) which yielded 1.7 mg and 1.55 mg of enriched enzyme respectively. The purification trace for *eCYDG* is shown in Figure 3.13. The His-Tag was not removed during the purification process as it does not affect the activity of the enzymes as shown by our results and numerous previous studies (Liu and Liu, 2004).

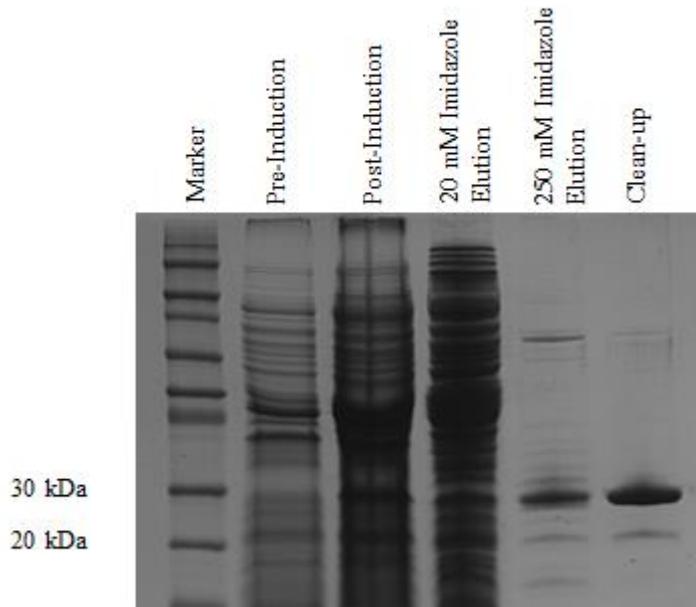


Figure 3.11 SDS-PAGE purification of *eCYDG*. Samples were taken at different purification stages to assess enzyme expression and purity.

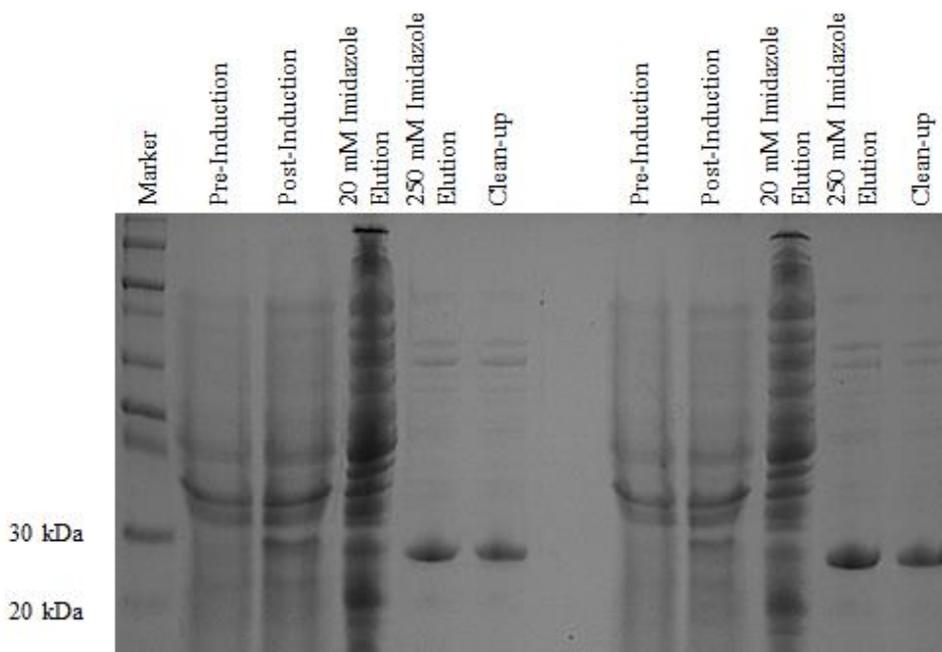


Figure 3.12 SDS-PAGE purification of *eUDG* (left) and *eUYDG* (right). Samples were taken at different purification stages to assess enzyme expression and purity.

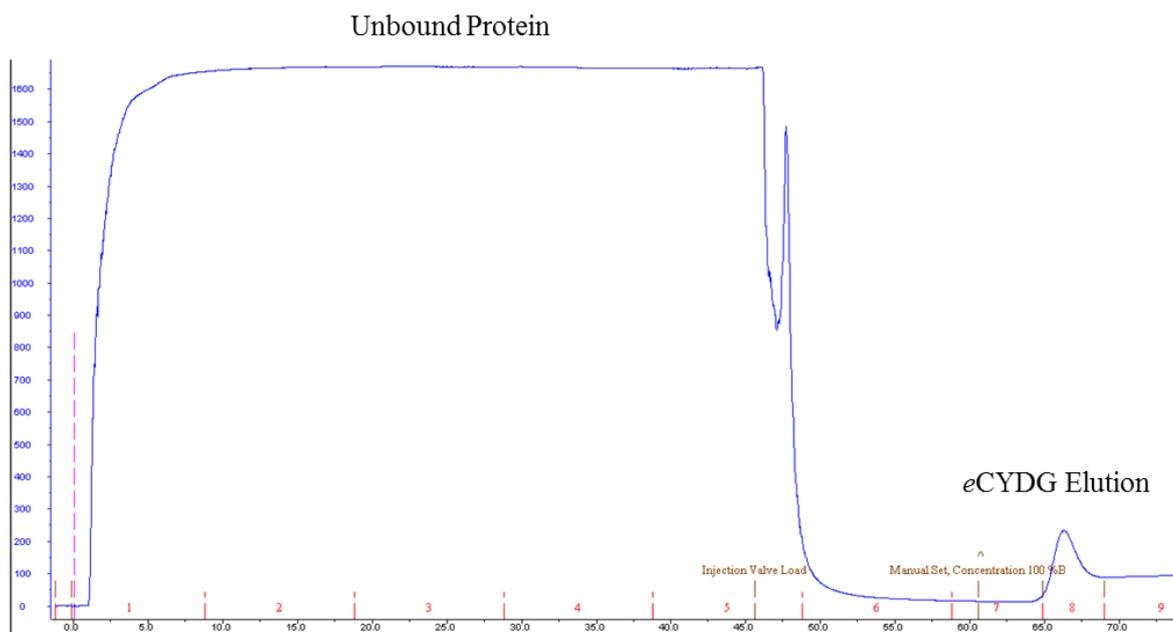


Figure 3.13 *eCYDG* Ni-NTA loading and elution profile. The blue trace is a measure of protein concentration absorbance measured at 280 nm. *eCYDG* is loaded onto the column in 20 mM imidazole allowing unbound proteins to pass through and elute. *eCYDG* was eluted using 250 mM imidazole. Fractions collected are shown in red along the x-axis.

3.3.7 Excision Activity of *eUDG*

Before generating a CDG that could be expressed and purified (*eCYDG*) we first examined the properties of *eUDG* in order to test the viability of our cleavage assay for determination of enzyme activity. A cartoon representation of this assay can be seen in Figure 3.14. Figure 3.15A shows the effect of *eUDG* on fragments in which uracil and thymine are paired with guanine, at enzyme concentrations of 300, 30 and 3 nM. The results show that *eUDG*, even at the lowest concentration (3 nM) and shortest reaction time (1 minute), produced assumed full cleavage of the uracil containing fragment, as shown by the appearance of the lower band(s) and disappearance of the full length substrate fragment (top band). This is in agreement with previous studies (Kwon *et al.*, 2003) where full cleavage occurred at a concentration of 10 nM. As expected, no cleavage was observed with a G.T mismatch, even at the highest enzyme concentration and longest incubation. *eUDG* at 3 nM was therefore used in further experiments.

Similar experiments were then performed using a wider range of substrates (Figure 3.15B) with different bases positioned opposite the target cytosine base including an abasic (AP) site. This substrate was also investigated with the addition of two intercalators, ethidium bromide and pyrene (in the form of 1-pyrenebutanol). This was to see whether they would

intercalate at the AP site and help to force the target base into the extrahelical conformation required for cleavage, mimicking the effect of the pyrene nucleoside (Kwon *et al.*, 2003). Intercalators were used to see whether the excision properties of *e*UDG would change when the target bases were T, C and ^MC (Figure 3.15). Unsurprisingly, activity against these bases was not observed, confirming *e*UDGs specificity for uracil. As expected full cleavage of U was observed with the addition of intercalators, which was as full cleavage of U occurred opposite an AP site.

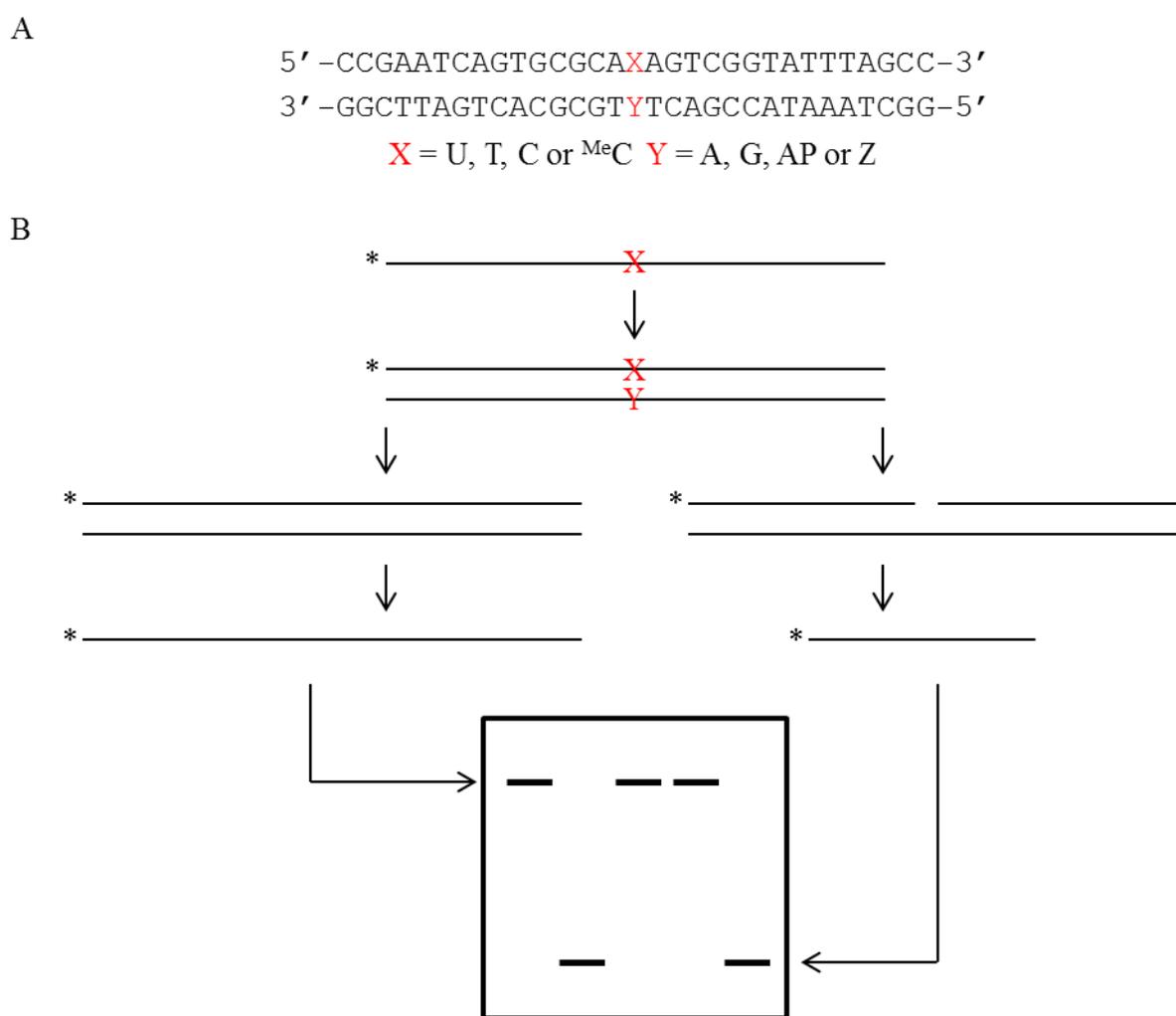


Figure 3.14 Cartoon representation of the cleavage assay. A) Duplex substrates generated from interchangeable oligonucleotides from Table 2.4. B) The oligonucleotide containing the target base (X) to be examined for excision is radiolabelled (*) at its 5' end and subsequently base paired with its complementary oligonucleotide, generating an X.Y base pair. After incubation with the enzyme, if the base is not cleaved (left hand side), the full length oligonucleotide under denaturing conditions appears at the top of the gel. If the base is cleaved (right hand side) then the smaller product oligonucleotide appears towards the bottom of the gel due to its greater mobility.

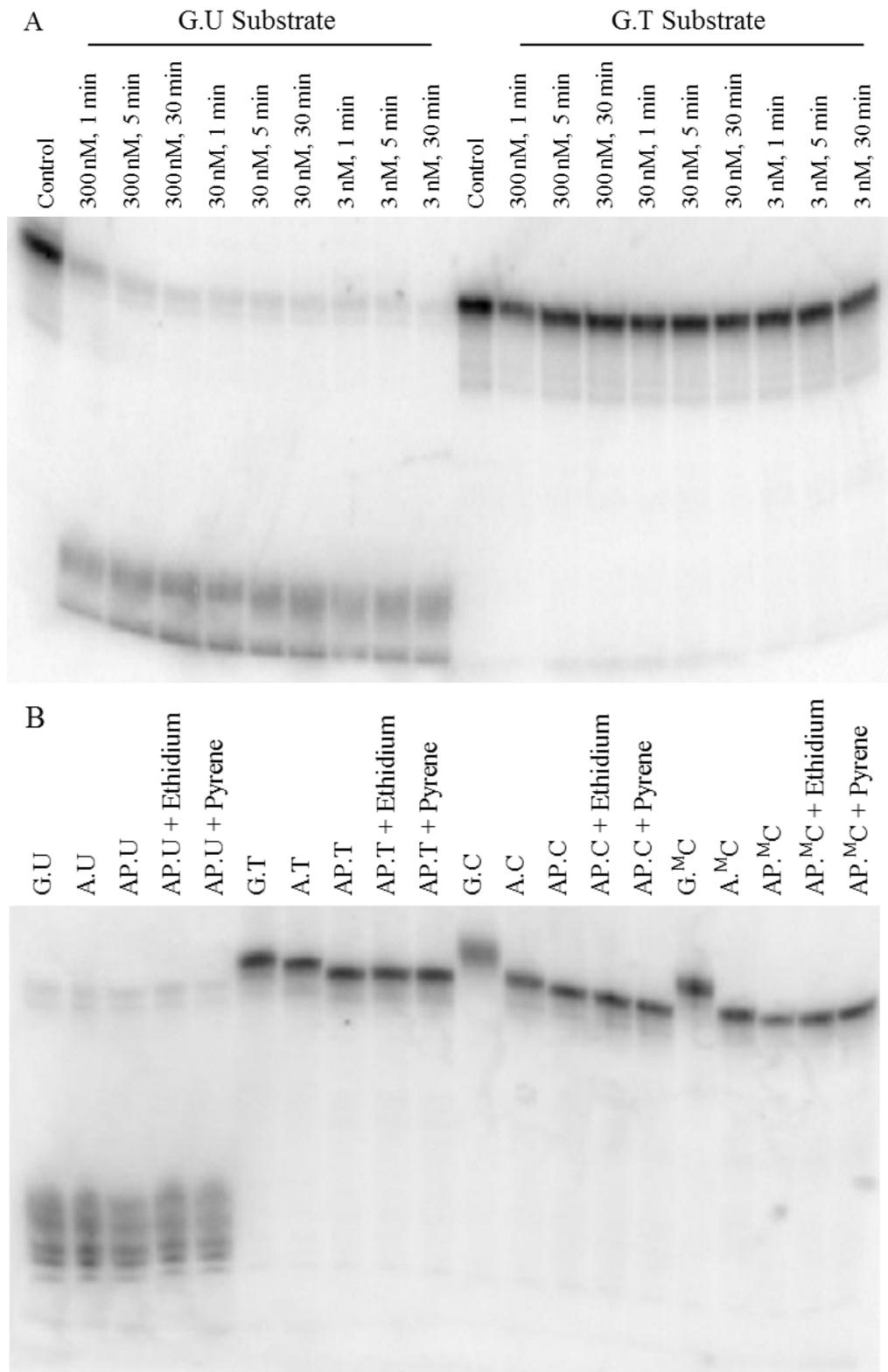


Figure 3.15 *e*UDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates. A) 31mer duplex (~50nM) substrates were incubated with 300, 30 and 3 nM *e*UDG for up to 30 minutes. B) 31mer substrates (~50nM), with the inclusion of 10 μ M free ethidium and 10 μ M pyrene with AP paired duplexes, were incubated with 3 nM *e*UDG for 30 minutes. All samples were quenched with 3 μ l stop solution and heated at 95°C for 30 minutes and resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging. The top and bottom bands correspond to the 31mer uncleaved or 15mer cleaved product. Control; unreacted G.U duplex, AP; abasic site.

3.3.9 Excision Activity of *e*UYDG

The activity of *e*UYDG against the same range of substrates as with *e*UDG, is shown in Figure 3.14. As with *e*UDG, *e*UYDG also showed complete cleavage of all uracil containing duplexes (G.U, A.U and AP.U). This is not surprising, as although the L191A mutation should severely reduce the enzymes' activity, previous studies (Jiang *et al.*, 2001, Kwon *et al.*, 2003) have shown that *e*UYDG still has the ability to excise U in an A.U and AP.U context, though at lower rates as the enzyme is deficient in base flipping. This is confirmed in the results shown in Figure 3.16 where *e*UYDG is able to excise all U containing substrates. As expected no activity was seen towards the T, C or ^MC, regardless of the addition of intercalators, which were included in an attempt to rescue the effect of the L191A mutation.

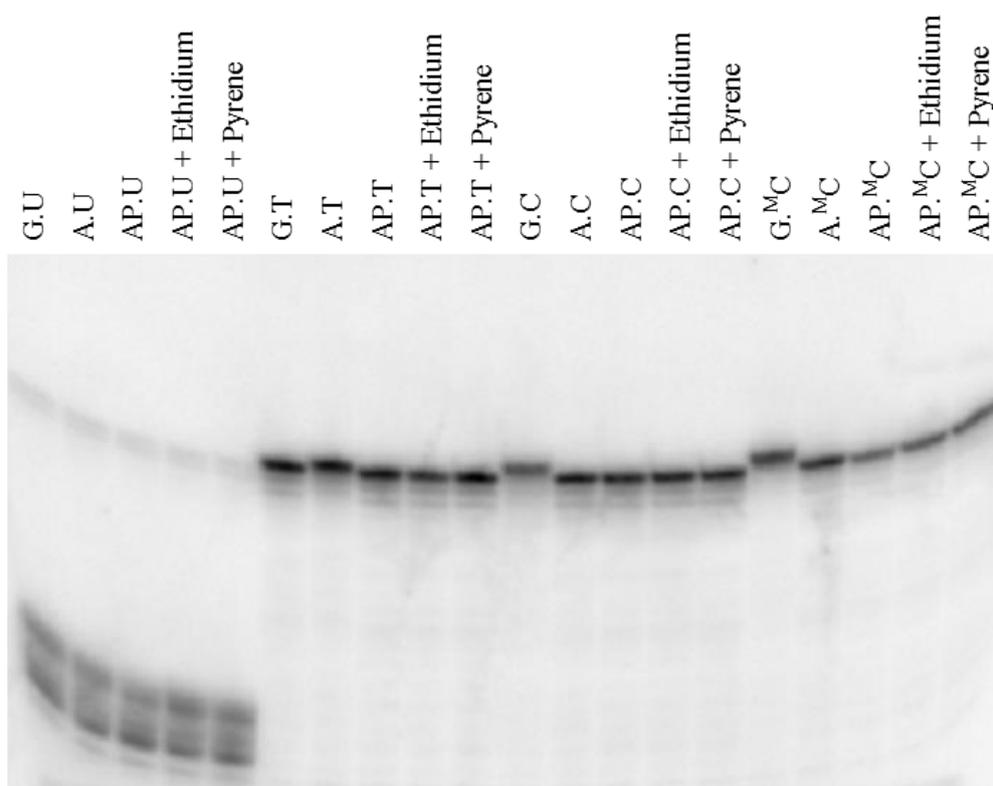


Figure 3.16 *e*UYDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates. 31mer substrates (~50nM), with the inclusion of 10 μ M free ethidium and 10 μ M pyrene with AP paired duplexes, were incubated with 150 nM *e*UYDG for 30 minutes. All samples were quenched with 3 μ l stop solution and heated at 95°C for 30 minutes and resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The top and bottom bands correspond to the 31mer uncleaved or 15mer cleaved product. AP; abasic site.

3.3.10 Excision Activity of *e*CYDG

Initial experiments with *e*CYDG were performed, as for *e*UDG/*e*UYDG, to gauge a rough estimate of *e*CYDG's activity towards U, T, C and ^MC containing substrates. We hypothesised that since *e*CYDG has a lower activity the addition of free ethidium and pyrene in conjunction with an AP site could have a greater effect. However, this not the case, as shown in shown in Figure 3.17 and no C excision activity was observed. These results showed that *e*CYDG retained residual U excision activity but has no activity against T, C or ^MC. It interesting to note that the fully based paired duplexes containing no mismatches (G.C, G.^MC, A.U and A.T) run at a slower mobility (Figure 3.17).

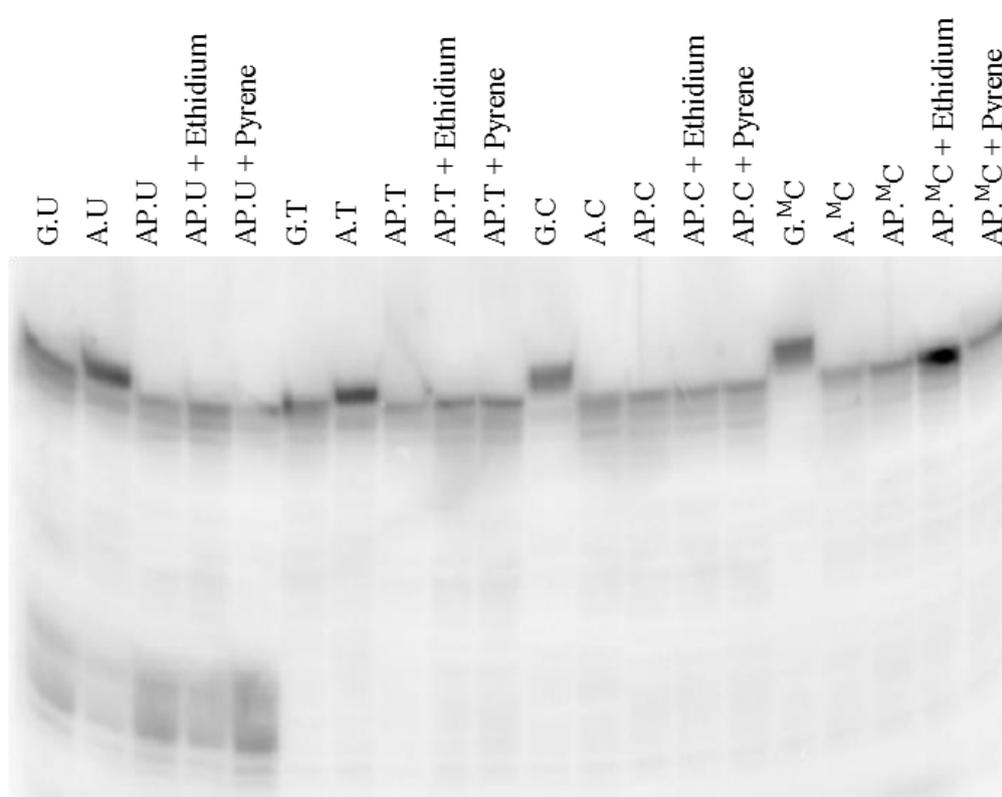


Figure 3.17 *e*CYDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates. 31mer duplex substrates (~50nM), with the inclusion of 10 μ M free ethidium and 10 μ M pyrene with AP paired duplexes, were incubated with 150 nM *e*CYDG for 30 minutes. All samples were quenched with 3 μ l stop solution and heated at 95°C for 30 minutes and resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The top and bottom bands correspond to the 31mer uncleaved or 15mer cleaved product. AP; abasic site.

Since these initial experiments were performed at pH 7.4, this could account for the absence of C excision activity, as *e*CYDG has been shown to excise C over a narrow pH range (Kwon *et al.*, 2003). We therefore investigated its activity at more acidic pHs (Figure 3.18A); the results again showed residual U excision activity but no C excision activity.

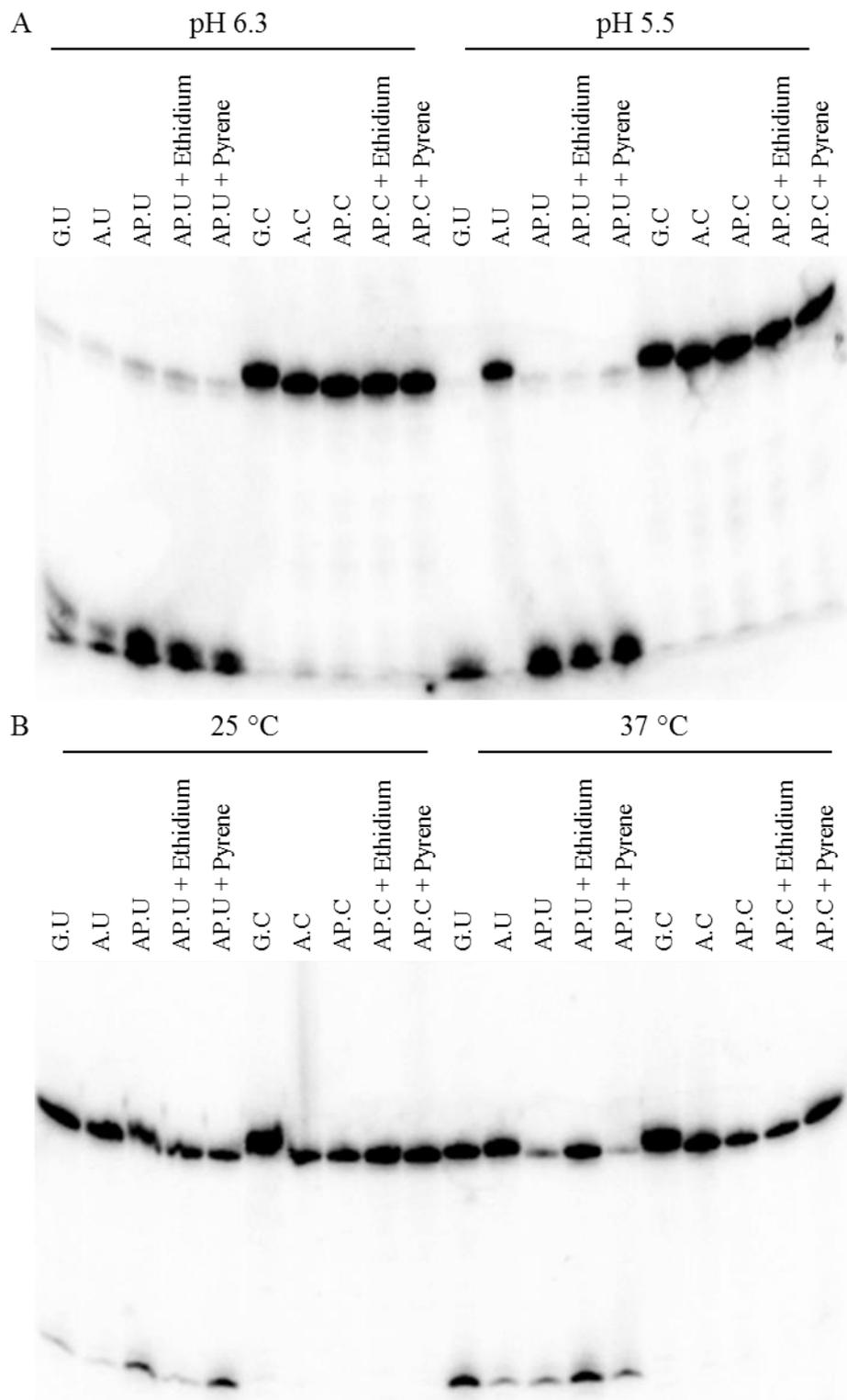


Figure 3.18 Effect of pH (A) and temperature (B) on the activity of *e*CYDG against uracil, thymine, cytosine and 5-methylcytosine substrates. 31mer duplex substrates (~50nM) with the inclusion of free ethidium and pyrene with AP paired duplexes were incubated with 150 nM *e*CYDG for 30 minutes. All samples were quenched with 3 μ l stop solution and heated at 95°C for 30 minutes and resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The top and bottom bands correspond to the 31mer uncleaved or 15mer cleaved product. AP; abasic site. B was performed at pH 6.3.

Since Kwon *et al.*, 2003 had determined the optimum reaction pH for *e*CYDG was 6.2, all further experiments were performed at pH 6.2 rather than 7.4. The apparent loss of excision for A.U at pH 5.5 (Figure 3.18B) is an experimental error as no other U containing substrate is affected.

Previous experiments with *e*CYDG (Kwon *et al.*, 2003) were performed at 25°C and we wondered whether the enzyme's activity was temperature dependent, as our experiments were conducted at 37°C; a more physiological temperature. We therefore investigated the cleavage 25°C and 37°C (Figure 3.18B). Although the rate at G.U appears to be greater at 37°C, no activity was observed against any of the C-containing duplexes.

These *e*CYDG experiments were performed comparatively with *e*UDG and *e*UYDG and it is unsurprising that it showed no C activity as it is at least 1000-fold less active (Handa *et al.*, 2002). Due to this and that *e*CYDG has been shown to have C excision activity (Kwon *et al.*, 2003), it was decided to incubate the enzyme with a range of substrates over a longer time course and at a higher enzyme concentration.

3.3.10.1 Activity Determination

The experiments in the previous section were performed with only 150 nM *e*CYDG, allowing 30 minutes for digestion. These experiments were repeated with longer digestion times (24 hours) and higher enzyme concentrations (1.25 µM) against a range of DNA duplexes and the results are shown in Figure 3.17. As expected all the U-containing substrates were fully cleaved by the enzyme (when this is paired with G, A or an abasic site), while none of the T-containing substrates were affected. However, most importantly, under these conditions *e*CYDG excised C when it is paired opposite A or an abasic site, though not when paired with G. No activity is evident against ^MC in any of these combinations. We also investigated placing a gap between two oligonucleotides opposite the target C, so that it was completely unpaired (gap.C); this again showed activity towards C, though interestingly only 50% cleavage was seen (discussed below). Again, no cleavage was seen with a similar ^MC-containing substrate. Cleavage was also observed when the C (but not ^MC) was placed opposite the non-nucleosidic linker hexaethylene glycol (see below).

The previous study with CYDG did not study base pair mismatches, but paired the target C with a pyrene nucleoside (Jiang *et al.*, 2001, Kwon *et al.*, 2003) in order to facilitate

expulsion of C into an extrahelical conformation. Since oligonucleotides containing this nucleotide were not available, we examined the reaction with an alternative commercially available nucleoside with a bulky intercalating group, anthraquinone pyrrolidine (Z; Figure 2.1). We predicted that it might act in the same way as pyrene, and force the opposing base out of the DNA duplex into an extrahelical conformation for cleavage by *e*CYDG. The results with this base pair combination (Z.C or Z.^MC) are also included in Figure 3.19 and again show cleavage with Z.C, but not Z.^MC.

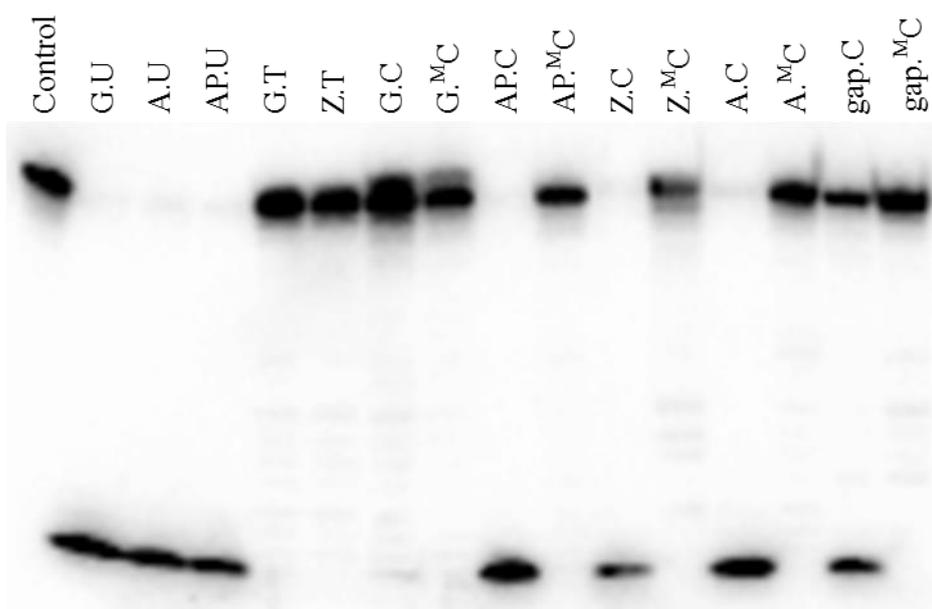


Figure 3.19 *e*CYDG excision of uracil, thymine, cytosine and 5-methylcytosine substrates. The ³²P labelled 31mer duplex substrates (~50 nM) were incubated with ~1.25 μM *e*CYDG for 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging. The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. Control refers to a G.U substrate unreacted with enzyme. AP; abasic site, Z; anthraquinone pyrrolidine, gap; unpaired C.

CYDG showed no activity towards either C or ^MC when paired with G. We were concerned that because this fragment contains several other G.C base pairs, *e*CYDG might have weak cleavage at these sites that were not detected. We therefore designed an alternative 31mer duplex that was AT rich and contained a single G.C/^MC base pair in the centre (sequences G2.C2 in Table 2.4). *e*CYDG incubation with this new G.C substrate is shown in Figure 3.20 and the results confirm that *e*CYDG does not cleave either substrate. The observation that a G.C base pair is not cleaved by *CYDG*, but that C in any other base (mis)pair combination is a substrate base pair, suggests that the rate of cleavage is profoundly affected by the stability of the base pair. Unstable base combinations, such as A.C, allow rotation of the base into the active site, which is now designed to accommodate

C. Presumably G.C is too stable to allow this rotation. However, it should be noted that eCYDG is able to excise U from an A.U base pair.

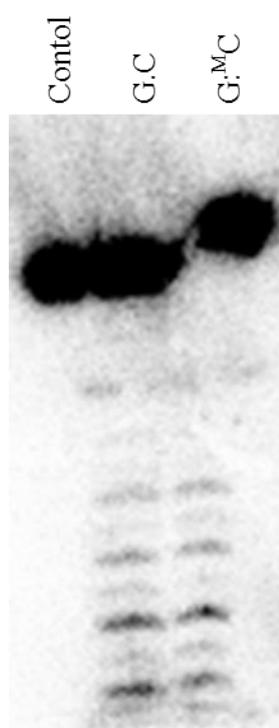


Figure 3.20 *e*CYDG excision of G.C^MC. The ³²P labelled 31mer duplex substrates (~50 nM) were incubated with ~1.25 μM *e*CYDG for 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The top band corresponds to the uncleaved 31mer product. Control refers to a G.C substrate unreacted with enzyme. The additional bands seen towards the bottom of the gel do not correspond to cleavage of other Cs and are degradation products, which most likely arise from treatment with piperidine.

3.3.11 Rate of Reaction Determination

In order to determine the best substrate for *e*CYDG for discriminating between C and ^MC we investigated the kinetics of the reaction when C is placed against a variety of opposing bases. As the enzyme is in excess of the DNA substrate it is assumed that all substrate is bound by enzyme and that the initial velocity represents the rate of reaction as substrate turned over per unit time. Representative cleavage profiles are shown for A.C, AP.C and Z.C in Figure 3.21A, C and E and the rate of reaction was derived from these and a further two repeats. The rates were then averaged and the error (standard deviation) calculated, as per Appendix IV, and are summarised in Table 3.2. *e*CYDG fully excised C from an A.C mismatch over a 24 hour period producing a single product at a rate of $0.006 \pm 0.001 \text{ min}^{-1}$.

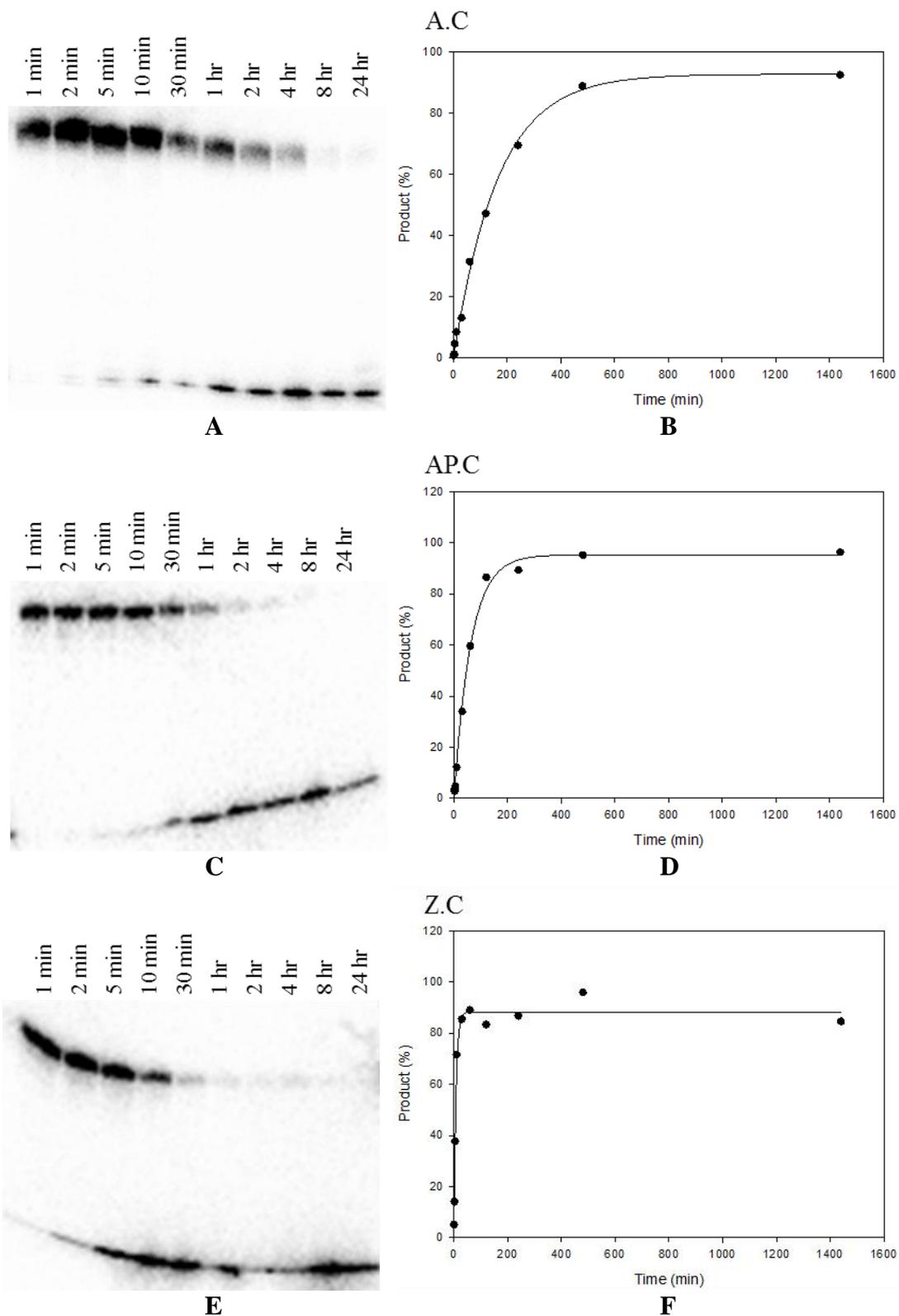


Figure 3.21 Kinetics of *e*CYDG cleavage of the 31mer substrates A.C, AP.C and Z.C. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *e*CYDG for up to 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, C and E). The top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B, D and F). AP: abasic site, Z: anthraquinone pyrrolidine.

As previously discussed, anthraquinone (Z) was incorporated opposite C to force the base into an extrahelical conformation. This Z.C pair produced the fastest cleavage rate of $0.1 \pm 0.02 \text{ min}^{-1}$; which is approximately 17-fold faster than A.C. This is consistent with previous studies (Kwon *et al.*, 2003), which used pyrene to force C into an extrahelical conformation. The rate of C cleavage from AP.C ($0.014 \pm 0.003 \text{ min}^{-1}$) is intermediate to A.C and Z.C. This is most likely because there is no hydrogen bonding with C, while an A.C mismatch has one H-bond, and this is not as destabilising as with Z. This gives an overall substrate preference of Z.C > AP.C > A.C.

<i>e</i> CYDG	Rate of Reaction (min^{-1})	Relative Activity (%)
G.U	0.36 ± 0.04	100
HEG.C	0.13 ± 0.01	36
Z.C	0.1 ± 0.02	29
A.U	0.02 ± 0.004	5.63
gap.C	0.017 ± 0.002	4.62
AP.C	0.014 ± 0.003	4.01
A.C pH 7.4	0.01 ± 0.002	2.71
HEG.C(G) ²	0.01 ± 0.001	2.3
long gap.C	0.007 ± 0.001	2.02
A.C	0.006 ± 0.001	1.72
A.C(AG)	0.005 ± 0.001	1.42
A(T).C(G) ²	0.004 ± 0.001	1.07
A.U(G)	0.0037	1.03
A.C(GA) ¹	0.0003	0.09
ssC(polyA) ¹	0.0003	0.07
A.C(G) ¹	0.0001	0.02
ssC(GAT) ³	0.0001	0.02
G.C(AT)	ND	
I.C	ND	
I.C(G)	ND	
Long HEG.C(G)1	ND	
APHEG.C(G)	ND	

Table 3.2 *e*CYDG reaction rates. Relative activity is in relation to the G.U substrate. The rate of reaction was determined as an average of three rates determined from cleavage profiles. Rate values were estimated from a single time point at 24 hrs¹, 5 mins² and 8 hrs³ assuming a simple exponential. Z; anthraquinone pyrrolidine, AP; abasic site, I; inosine, gap; unpaired C, long gap; unpaired C in a 41mer duplex, HEG; hexaethylene glycol, ND; not detectable.

3.3.11.1 Excision of uracil

Uracil excision from G.U ($0.36 \pm 0.04 \text{ min}^{-1}$) is approximately 60-fold faster than from A.C showing that *e*CYDG still has a large selectivity for U over C. Cleavage at A.U ($0.02 \pm 0.004 \text{ min}^{-1}$; again consistent with previous data (Kwon *et al.*, 2003)) is 20-fold lower than G.U and 5-fold slower than Z.C, further suggesting that base pair stability plays a major role in determining the excision rates.

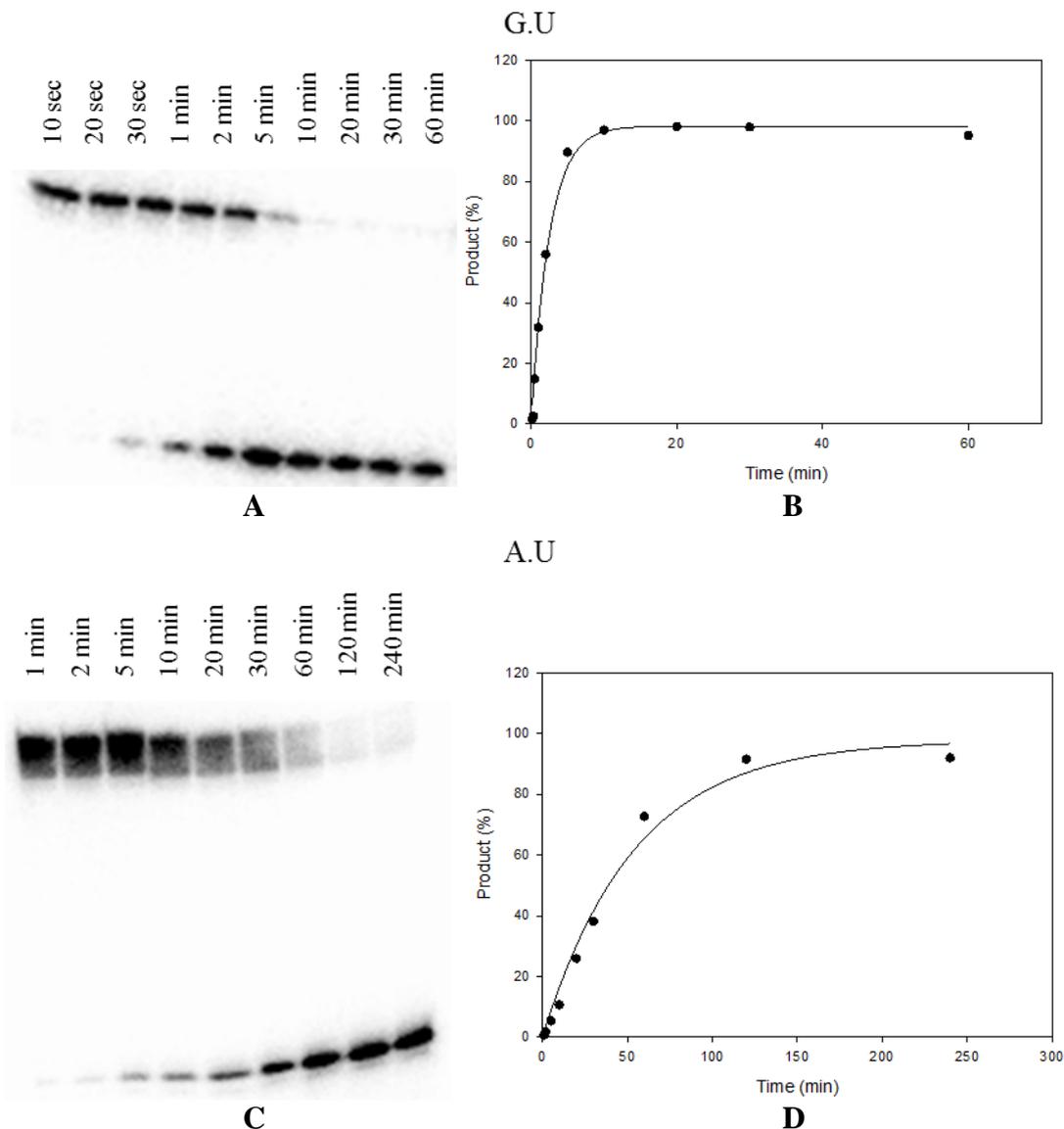


Figure 3.22 Kinetics of *e*CYDG cleavage of the 31mer substrates G.U and A.U. The ^{32}P labelled 31mer duplex substrates ($\sim 50 \text{ nM}$) were incubated with $\sim 1.25 \mu\text{M}$ *e*CYDG for up to 1 or 4 hours respectively followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A and C). The top and bottom bands correspond to a uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B and D).

3.3.11.2 Excision of ssC substrates

Since UDG is known to show good activity against single-stranded DNA substrates we investigate the activity of eCYDG against single stranded substrates. Since these substrates can only contain a single C we designed two different oligos; ssC(polyA) has the central C within a polyA tract while ssC(GAT) has the central C flanked by a sequence of G, A and T. Cleavage plots for these substrates are shown in Figure 3.23.

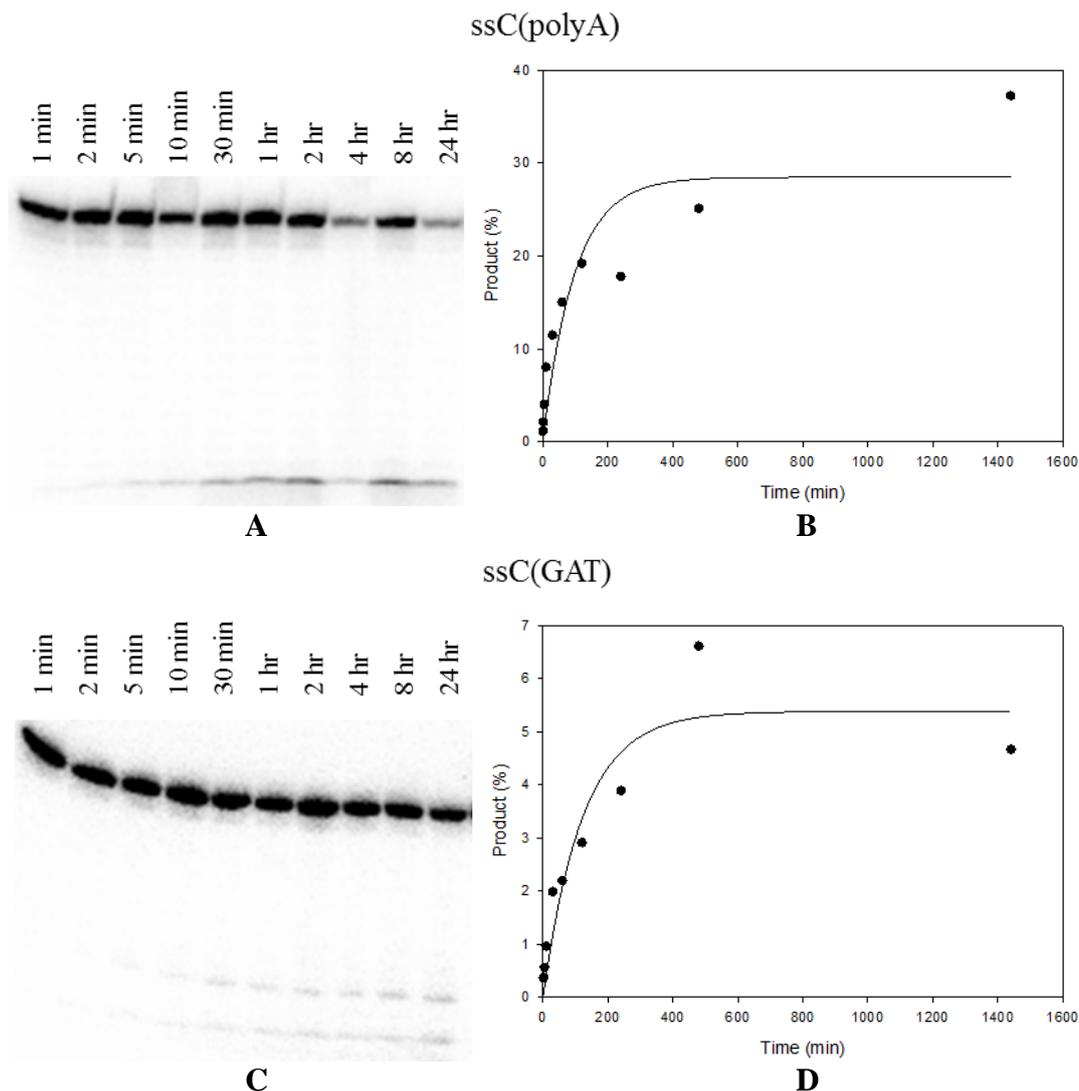


Figure 3.23 Kinetics of eCYDG cleavage of the 31mer substrates ssC(polyA) and ssC(GAT). The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM eCYDG for up to 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A and C). The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B and D).

Unlike UDG, *e*CYDG shows very weak activity towards single stranded substrates giving rates of 0.0003 min⁻¹ and 0.0001 min⁻¹ for ssC(polyA) and ssC(GAT) respectively. These rates were calculated from a single time point at 24 and 4 hours respectively assuming a simple exponential (Equation 1).

$$\text{Rate} = -\ln(1 - \text{Fraction cleaved}) / \text{Time}$$

Equation 1 Calculation of rate from a single time point.

These times were chosen as ssC(polyA) only reached a maximum substrate cleavage of approximately 31% after a long incubation time. Therefore its rate was calculated based on this value after 24 hours. The maximum substrate cleavage efficiency for ssC(GAT) varied greatly at the 8 and 24 hour time points, though was still very low, and the reaction rate was calculated after 4 hours at which only 4% of the substrate had been cleaved.

3.3.12.3 Investigating the Effect of an Unpaired Cytosine

Since the efficiency of CYDG cleavage at C seems to depend on the stability of the base pair we examined the enzyme's activity against double stranded substrates in which the C is unpaired. Figure 3.24A shows the reaction of *e*CYDG with a substrate containing a completely unpaired C (gap.C). This shows a rate of $0.017 \pm 0.002 \text{ min}^{-1}$, which is similar to that of AP.C. However, as noted above, only a maximum of approximately 50% of this substrate was excised. Since this substrate consists of two shorter duplexes, which will be less stable at the reaction temperature, we considered whether the reaction might be limited by the stability of the duplex substrate (melting temperatures (T_m s) of 39 and 45°C) and therefore designed a similar longer duplex containing an extra five base pairs either side of the central C (long gap.C), producing a 41mer duplex of greater stability (T_m s of 50 and 54°C) (Figure 3.24C). The reaction rate with this substrate decreased to $0.007 \pm 0.001 \text{ min}^{-1}$ (Figure 3.24D) and the amount of substrate cleaved increased to approximately 80%. We decided to investigate further the reaction at an unpaired C by incorporating hexaethylene glycol (HEG) opposite, acting as a "backbone linker" to hold the two oligonucleotides (Figure 3.24E). This produced a rate ($0.13 \pm 0.01 \text{ min}^{-1}$) similar to Z.C, which is one of the best substrates for this enzyme, again highlighting the dependency of the rate of excision on base pair stability.

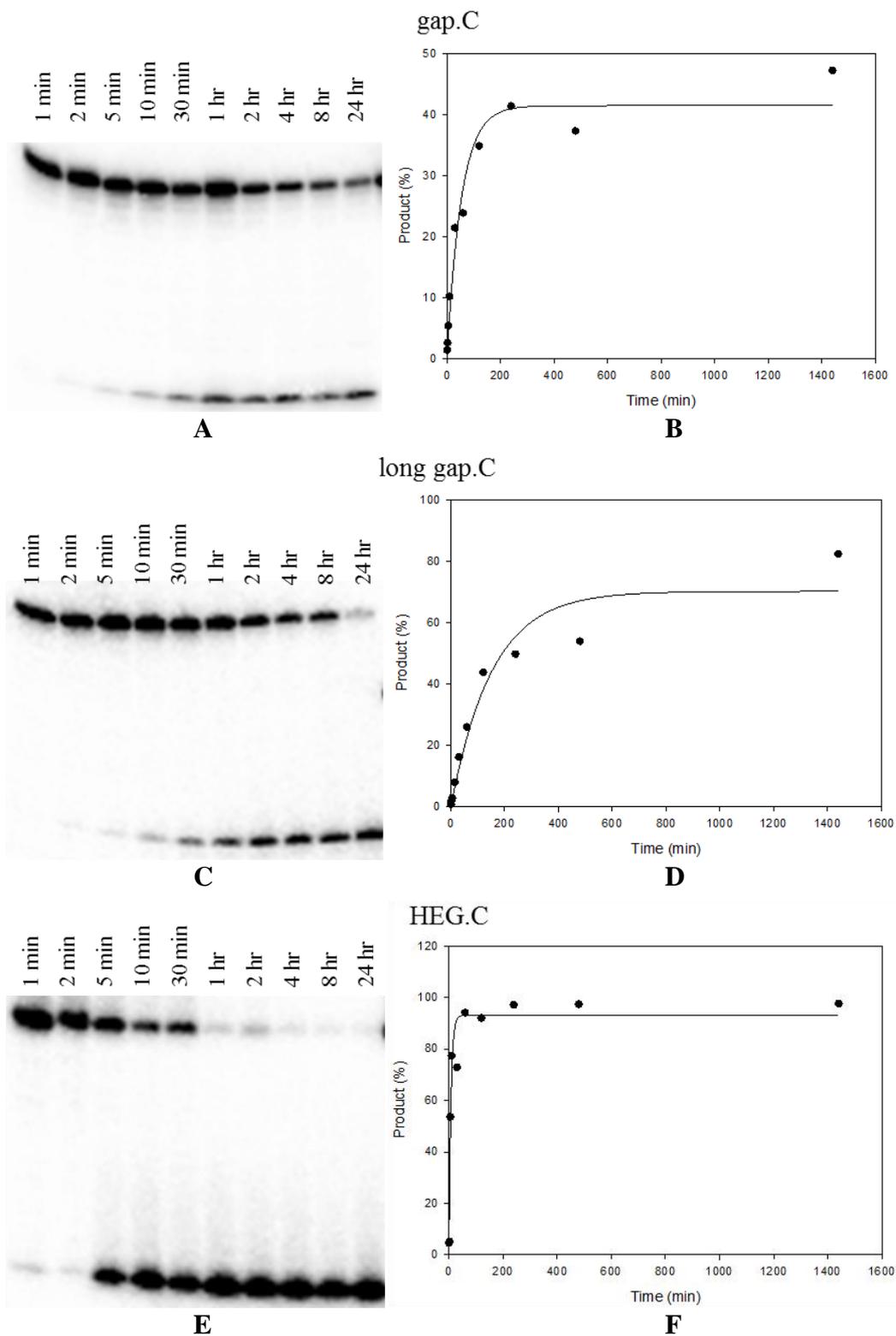


Figure 3.24 Kinetics of *e*CYDG cleavage of the 31mer substrates gap.C, long gap.C and HEG.C. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *e*CYDG for up to 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, C and E). The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves. gap: unpaired C, long gap: unpaired C in a 41mer duplex, HEG: hexaethylene glycol (B, D and F).

3.3.11.4 Investigating the Effect of the Flanking Base Pair on *e*CYDG Cleavage

As *e*CYDG showed no activity towards G.C, but maintained A.U activity, we investigated the interaction of C paired with inosine (I); a G.C-like base pair that is less stable as it contains only two H-bonds, similar to though more stable than A.U (Sponer *et al.*, 2004). We anticipated that this weaker base pairing would allow cleavage of C. The results are shown in Figure 3.25 and show that no significant activity can be detected. This suggests that other factors than simple base pair stability must have large contributions towards the rate of excision.

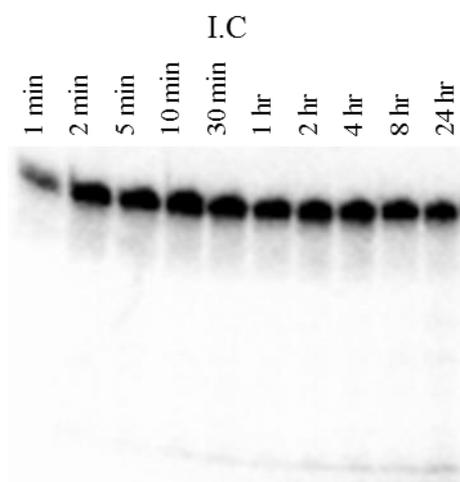


Figure 3.25 Kinetics of *e*CYDG cleavage of the 31mer substrate I.C. The ^{32}P labelled 31mer duplex substrate (~ 50 nM) was incubated with ~ 1.25 μM *e*CYDG for up to 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively.

The ability to excise C in any sequence context will be crucial for the development of an assay for $^{\text{M}}\text{C}$ detection (Chapter 6). The results above suggest that base pair stability has a strong effect on the cleavage efficiency of CYDG at C. Since this will be affected by the sequence context we investigated the excision at A.C when surrounded by different base pair combinations. For these experiments the duplex was mismatched so that the A.C mismatch was flanked by G.C base pairs (A.C(G); using oligos A2 and C3 in Table 2.4 and Table 3.3) instead of A.T. Figure 3.26 shows that there is little cleavage of this substrate and a rate of 0.0001 min^{-1} could only be calculated from the small amount of cleavage at 24 hours, assuming a simple exponential. This context reduced the rate of cleavage by approximately 100-fold in comparison to A.C in an AT context. The sequence was also altered so that the A.C mismatch was flanked by A.T and G.C base pairs on either side (A.C(AG) and A.C(GA); Table 3.3) so that the central sequence of the target C strand

reads 5'-ACG-3' or 5'-GCA-3' respectively. These results are shown in Figure 3.26 and reveal rates of A.C(GA) ($0.0003 \pm 0.0001 \text{ min}^{-1}$) and A.C(AG) ($0.005 \pm 0.001 \text{ min}^{-1}$), both of which are lower than A.C flanked by AT (substrate A.C) and had to be calculated from a single time point at 24 hours. This is consistent with previous results for UDG in which the flanking sequence AG is a better substrate than GA (Eftedal *et al.*, 1993).

Oligonucleotide	Sequence
A.C(G)	5' -CCGAATCAGTGC CGC CGGTCGGTATTTAGCC-3' 3' -GGCTTAGTCACGCGC A CCAGCCATAAATCGG-5'
A.C(AG)	5' -CCGAATCAGTGC CA CGGTCGGTATTTAGCC-3' 3' -GGCTTAGTCACGCGT A CCAGCCATAAATCGG-5'
A.C(GA)	5' -CCGAATCAGTGC CGC CAGTCGGTATTTAGCC-3' 3' -GGCTTAGTCACGCGC A TCAGCCATAAATCGG-5'

Table 3.3 A.C oligonucleotides used to assess the effect of flanking regions on the rate of excision. Central A.C base pair highlighted in bold.

These results suggest that the 5' flanking base pairs have a very important role in determining the cleavage rate. This could be as a result of changes in the local duplex stability and base stacking on the cytosine or it could be due to interactions with the enzyme itself and its ability to distort the duplex ready for base flipping (Parikh *et al.*, 1998). As placing a HEG opposite the C provided the greatest rate of excision, we investigated how this would be affected by flanking this HEG.C pair with G.C base pairs (HEGC.G). The results with this target are shown in Figure 3.27 and show an increase in amount excised (38%), but only a small effect on the rate of excision (0.01 min^{-1} ; calculated based on it reaching this level; of cleavage over 24 hours). To further explore the effect of flanking GC base pairs we examined the activity of CYDG at an A.U base pair with flanking GC base pairs (A.U(G); Figure 3.27). The rate decreased 5.5-fold to $0.004 \pm 0.0001 \text{ min}^{-1}$ compared to flanking AT base pairs, and only reached a maximum cleavage efficiency of 69%. This is also consistent with previous results (Eftedal *et al.*, 1993) and provides further evidence for the role of base stacking in determining the rate of excision.

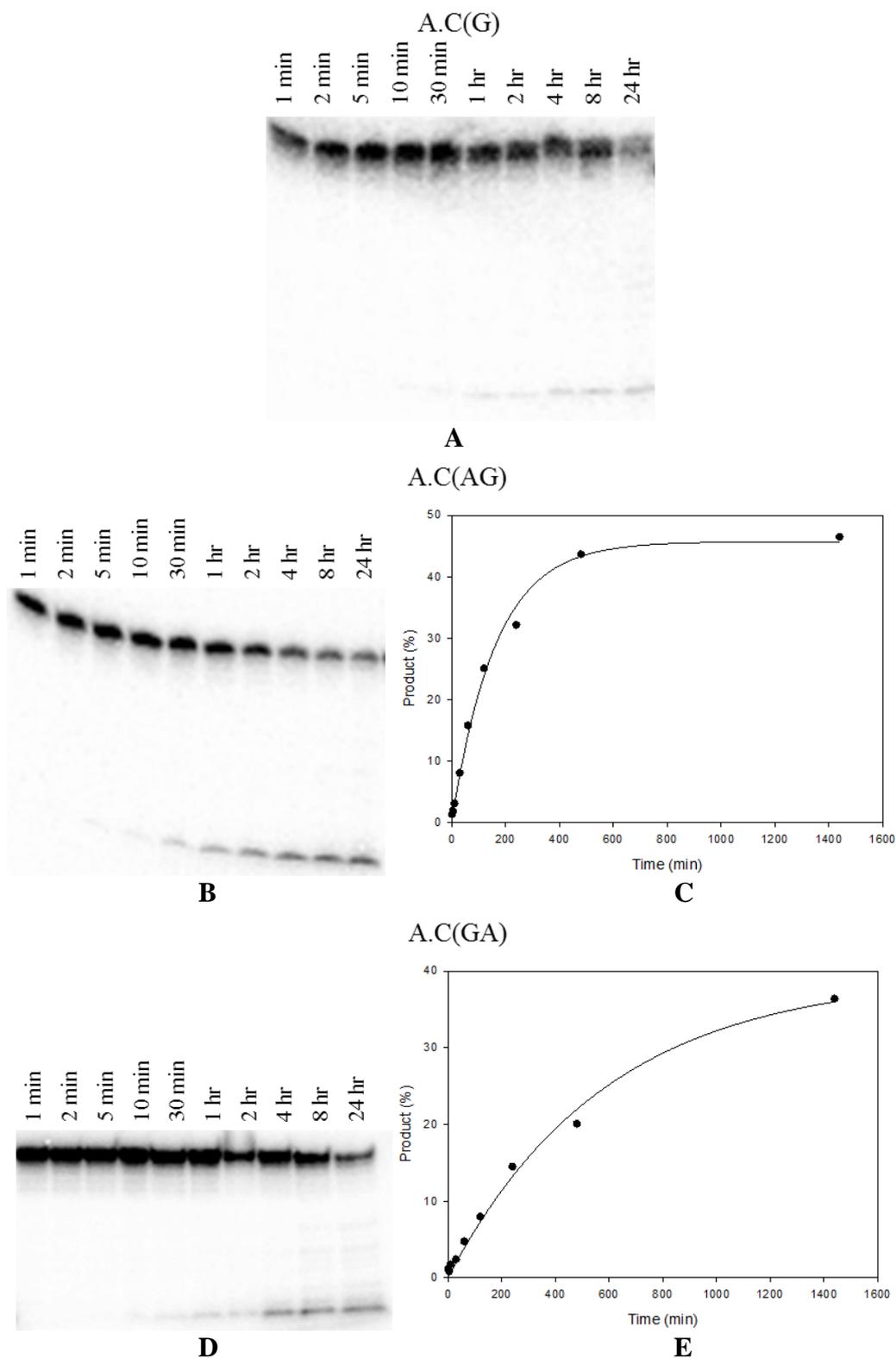


Figure 3.26 Kinetics of *e*CYDG cleavage of the 31mer substrates A.C(G), A.C(AG) and A.C(GA). The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *e*CYDG for up to 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (**A**, **B** and **D**). The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (**C** and **E**).

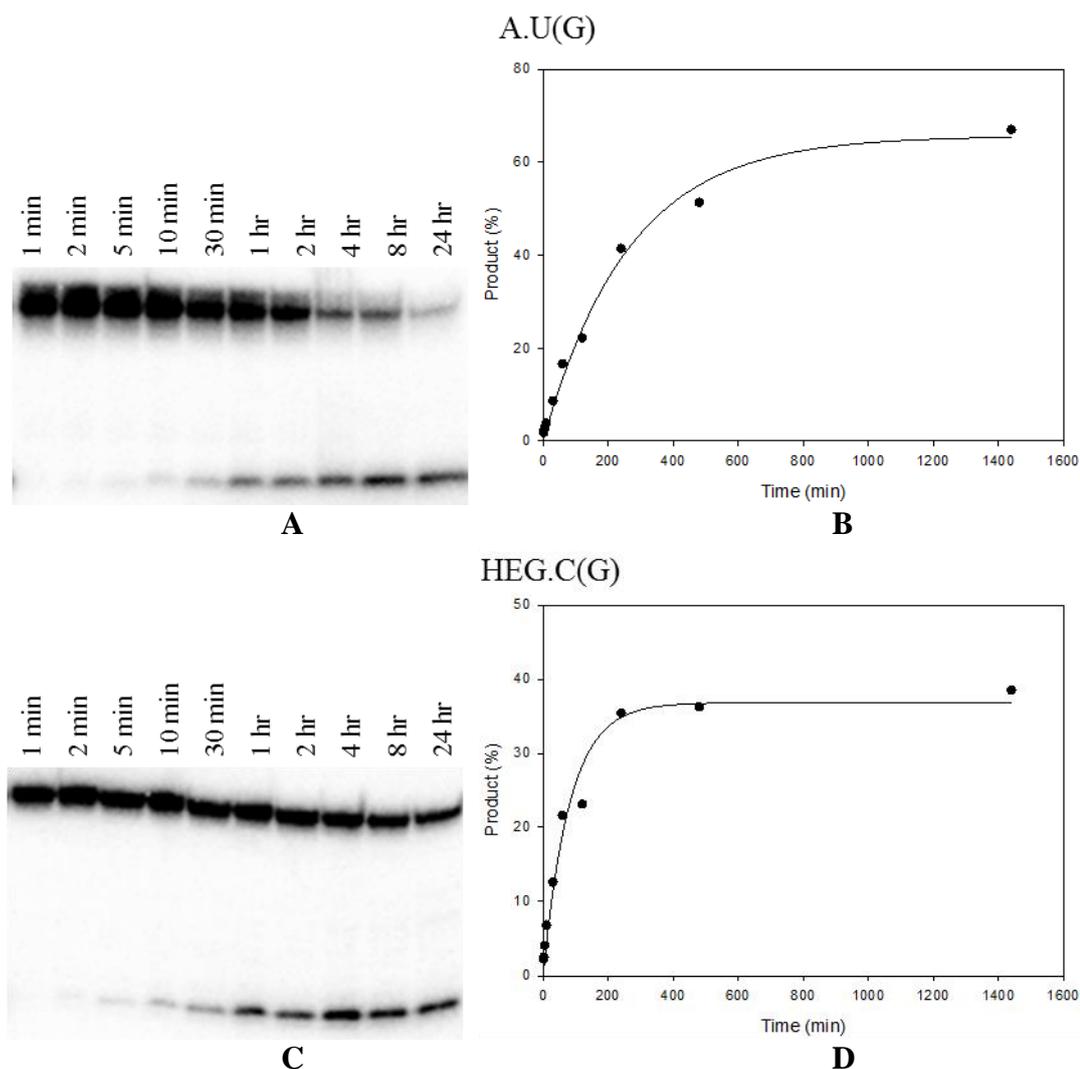


Figure 3.27 Kinetics of *e*CYDG cleavage of the 31mer substrates A.U(G) and HEG.C(G). The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *e*CYDG for up to 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (**A** and **C**). The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves. HEG: hexaethylene glycol (**B** and **D**).

As flanking G.C base pairs seem to have such a large effect on the rate of excision we investigated ways to reduce the stability of these flanking regions. Firstly we investigated using HEG further to oppose the central C as well as the flanking bases on each side, spanning three nucleotides (long HEG.C(G); Table 3.4). The HEG linker has the same number of bonds as three nucleotides, and so should be able to bridge this gap. However, no cleavage was detected (Figure 3.28C). Further studies examined the use of mismatches (G.T) flanking the central A.C mismatch (A(T).C(G); Table 3.4 and Figure 3.28A). This increased the amount of substrate cleaved from 12 to 31% over a 24 hour time period, though the estimated rate (0.0003 min^{-1}) is not significantly different to A.C(G) (0.0001

min⁻¹). It is worth noting that the loss of intensity seen for the substrate at 24 hours in Figure 3.28A is due to experimental error, *i.e.* loading error. The substrate is therefore not being degraded as a result of time, buffer conditions or contaminants as this is not seen in other gels (*i.e.* Figure 3.28C) and also from substrate unreacted with enzyme for 24 hours (controls in Figure 3.19 and Figure 3.20).

As G.T mismatches failed to increase the cleavage rate we investigated the combined effect of using HEG flanked by AP sites so that the flanking Gs of the target C were unpaired (APHEG.C(G)). This produced the same result as long HEG.C(G) and no activity was detected (Figure 3.28D). I.C(G) (an I.C central base pair flanked on either side by a G.C base pair) was also investigated, but as expected no activity was detected (Figure 3.28E).

Oligonucleotide	Sequence
A(T).C(G)	5' -CCGAATCAGTGCGCG C GGTCGGTATTTAGCC-3' 3' -GGCTTAGTCACGCGT A TCAGCCATAAATCGG-5'
long HEG.C	5' -CCGAATCAGTGCGCG C GGTCGGTATTTAGCC-3' 3' -GGCTTAGTCACGCG -H- CAGCCATAAATCGG-5'

Table 3.4 Sequences of A(T).C(G) and long HEG.C. Central base pairs highlighted in bold. -H-: hexaethylene glycol.

In summary *e*CYDG is able to excise C when paired with any base except guanine but showed no activity against ^MC in any sequence context. Although cytosine can be cleaved the rate of excision is context dependent and the more stable the base pair, the lower the rate of excision activity.

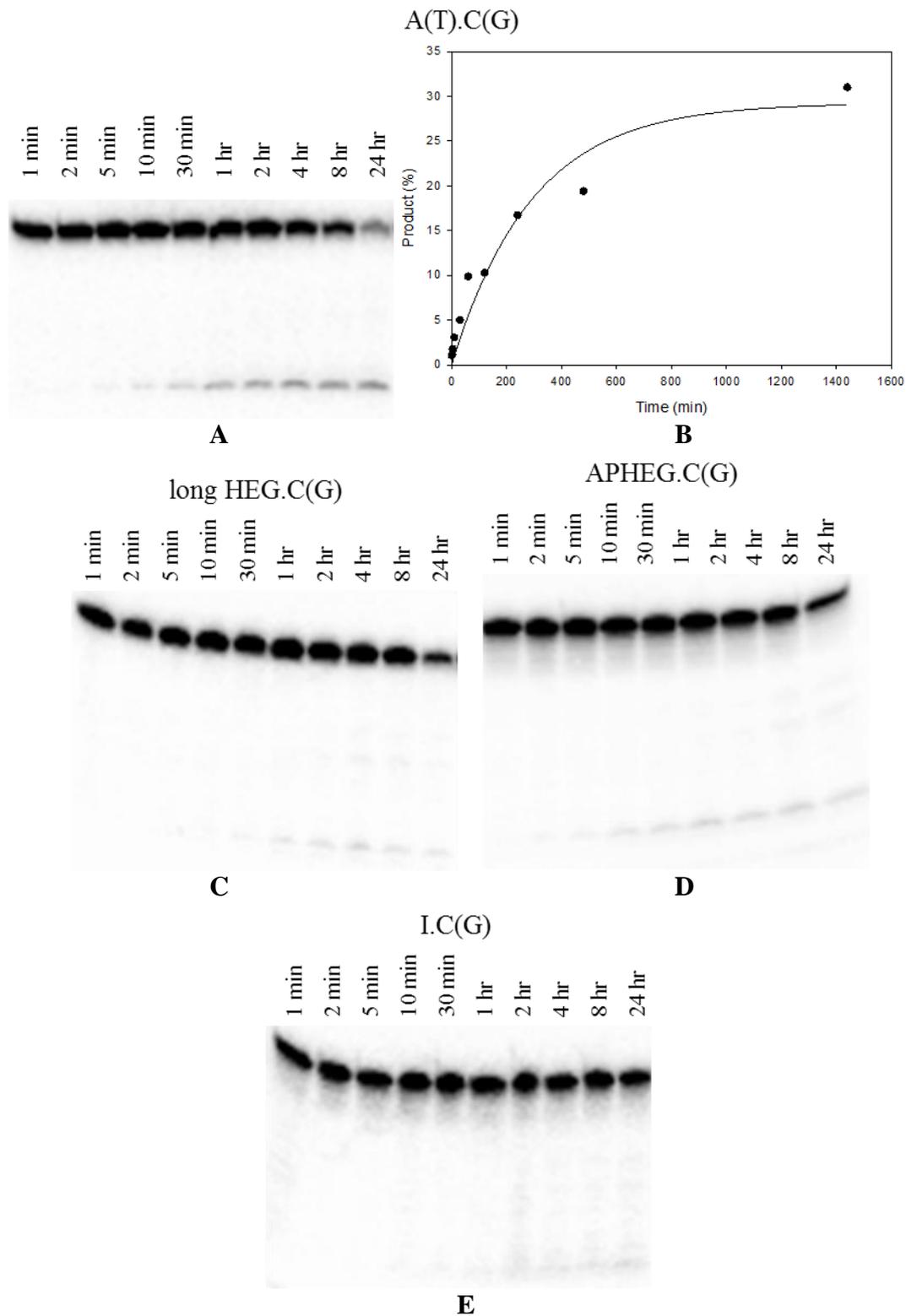


Figure 3.28 Kinetics of *e*CYDG cleavage of the 31mer substrates A(T).C(G), long HEG.C(G), APHEG.C(G), and I.C(G). The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *e*CYDG for up to 24 hours followed by heating in at 95°C 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, C, D, and E). The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. This is fitted with a single exponential curve (B). long HEG: hexaethylene glycol linker spanning three nucleotides, I: inosine.

3.3.11.5 The effect of pH on A.C excision

It has previously been shown that *e*CYDG is pH sensitive and shows optimum activity at pH 6.2, allowing correct protonation of the active site aspartates (Kwon *et al.*, 2003). We therefore performed cleavage assays with A.C at pH 7.4 (A.C pH 7.4; Figure 3.29A), a more physiological pH, in contrast to all previous *e*CYDG cleavage assays that were performed at pH 6.2. Surprisingly we found that the rate ($0.01 \pm 0.002 \text{ min}^{-1}$) increased by approximately 50% compared to A.C at pH 6.2.

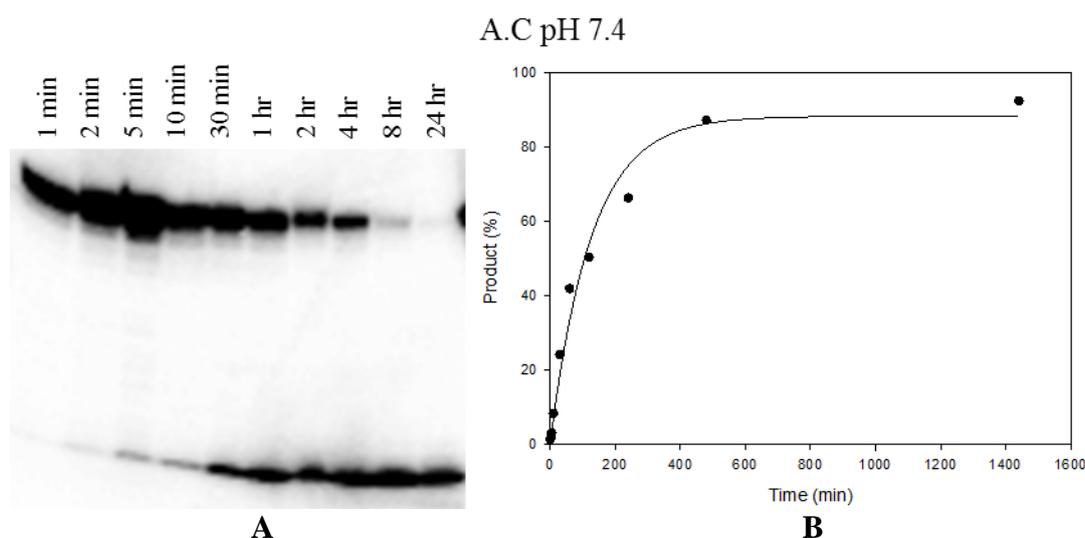


Figure 3.29 Kinetics of *e*CYDG cleavage the 31mer substrate A.C pH 7.4. The ^{32}P labelled 31mer duplex substrate ($\sim 50 \text{ nM}$) was incubated with $\sim 1.25 \mu\text{M}$ *e*CYDG for up to 24 hours followed by heating in at 95°C 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging (**A**). The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. This is fitted with a single exponential curve (**B**).

3.3.12 Mutating F77 and Y66

Other mutants of *e*UYDG and *e*CYDG were also investigated to examine cleavage of U and C. Phenylalanine 77 was mutated to tyrosine, tryptophan or histidine, since these residues are involved in base stacking interactions. We were most interested in histidine to investigate whether the nitrogens in its ring could interact with other residues or with the 2-amino of cytosine, and alter the stability and activity. Initial experiments performed with *e*UYDG showed no significant reduction in activity at G.U for F77Y/W but showed lower activity for F77H. We therefore investigated the effect of introducing the F77H mutation into *e*CYDG, in anticipation that the reduction with uracil in *e*UYDG might be accompanied by an increase with cytosine. Unfortunately preliminary results showed that

*e*CYDG F77H had lower excision activity towards C and U than *e*CYDG and *e*CYDG/*e*UYDG respectively.

As mutant CDGs may also be developed for other cytosine modifications, *i.e.* 5-hydroxymethylcytosine (^{HM}C), we investigated the effect of mutating Tyr66 to smaller residues, so as to open up the active site allowing entry of larger cytosine analogues. This has already been shown for UDG in which the Y66A mutation allows the accommodation of thymine and thus its excision. We investigated the Y66A mutation within the context of *e*UDG, *e*UYDG and *e*CYDG generating *e*TDG, *e*TYDG and *e*CTYDG respectively. We were unable to express *e*TDG suggesting that like *e*CDG it was cytotoxic. Cleavage assays to determine *e*TYDG and *e*CTYDGs activity were performed over 24 hours and both showed residual U excision activity, with only slight T excision activity observed with *e*TYDG when opposite the anthraquinone pyrrolidine; consistent with previous studies (Kwon *et al.*, 2003). Neither thymine, cytosine or 5-methylcytosine release was detected with *e*CTYDG suggesting that a triple mutation severely reduces the enzyme's activity. We also mutated Y66 to threonine, leucine and serine to see whether their side groups would be able to form any favourable interactions that would increase activity, while causing discrimination between ^MC and ^{HM}C. Preliminary results showed that all three mutations caused a loss of C excision activity while decreasing U excision activity compared to standard *e*CYDG.

3.4 Discussion

The main role of *e*UDG is to excise uracil in the context of a G.U base pair to prevent transition mutations that result from cytosine deamination. This base pair combination has not previously been investigated for *e*UYDG and *e*CYDG. The results show that both *e*UDG and *e*UYDG are able to excise uracil in any sequence context and show no activity towards any other base. It was surprising that *e*UYDG was able to excise uracil from A.U with such efficiency, as the L191A mutation has been reported to severely reduce the activity of the enzyme (Jiang *et al.*, 2001, Jiang *et al.*, 2002b), and L191A mutants can only cause base excision if the base is flipped into an extrahelical conformation by other means (such as pyrene). It is therefore not surprising that *e*UYDG is able to excise uracil from a G.U or AP.U base pair. This is because these base pairs are less stable than A.U, being a “wobble” base pair (G.U) and an unpaired (AP.U) respectively. The uracil is therefore more likely to become extrahelical through DNA breathing or by DNA distortion

caused by enzyme binding (Werner *et al.*, 2000). The ability of *eCYDG* to excise all uracil-containing substrates was also not expected, especially since previous studies have shown a large decrease in activity towards A.U (Kwon *et al.*, 2003). The difference may be due to different reaction conditions; 25°C in the previous study compared to 37°C in this work.

These results show that as expected *eCYDG* is able to discriminate between cytosine and 5-methylcytosine, since Y66 is positioned so as to cause a steric clash with the 5-methyl group of a pyrimidine base. The N123D mutation changes the specificity of *eUDG* enabling recognition of cytosine. This mutation does not affect its ability to discriminate between a pyrimidine and a 5-methylpyrimidine. The mutant enzyme is able to excise cytosine when paired with anything other than guanine, while retaining its ability to excise uracil (Kwon *et al.*, 2003), and no 5-methylcytosine activity was observed in any context. It is clear from these results that base pair stability is crucial for determination of the rate of excision (Krosky *et al.*, 2004, Krosky *et al.*, 2005). This is highlighted by *eCYDG* excising cytosine from Z.C faster than uracil from A.U. It is likely that the anthraquinone forces the cytosine out of the duplex into an extrahelical conformation more readily than uracil, which is held in a normal Watson-Crick A.U base pair. The faster rates observed with AP.C, gap.C and HEG.C (with the latter being equal to or greater than Z.C) than A.C, are probably because there is no opposing base to the cytosine.

The rate of cleavage of A.C was not significantly affected by a changing the pH to 7.4, in contrast to previous reports (Kwon *et al.*, 2003), though our experiments also showed that pH had no significant effect on excision of uracil. This difference could be due to the higher temperature used in the present work. It is interesting that only 50% of the substrate is cleaved with gap.C where the cytosine was completely unpaired. *eUDG* is known to bind strongly to AP sites (Parikh *et al.*, 1998) and it is conceivable that *eCYDG* may bind even more strongly to an unpaired site. The 50% maximal cleavage of gap.C could arise if *eCYDG* binds to the target cytosine and unpaired site with equal affinity. However, this would not explain why the same effect is not seen for AP.C. An alternative explanation is that the T_m of one half of the duplex (39°C) is very close to the reaction temperature (37°C) and may therefore have melted leaving a single stranded region that prevented binding. This explanation seems more plausible, and is consistent with the weak activity seen with ssC substrates. This could account for the increase in maximal cleavage from 50% (gap.C) to 80% (long gap.C) by working with a longer, more stable, gapped duplex. Similarly the

gap could generate a hinge region within the duplex, with the C acting as if single stranded, limiting enzyme binding and the maximal excision rate. We suggest that the fast rate seen with HEG.C is because the cytosine is unpaired; with a greater helical distortion than with AP.C. Furthermore, the HEG linker may generate a flexible region opposite cytosine allowing it to become extrahelical more easily upon enzyme binding (serine pinching; see Section 1.3.1.3.1) (Werner *et al.*, 2000).

We assumed that ssC would produce the fastest rate for *e*CYDG, based on the substrate preference of *e*UDG (ssU > G.U > A.U) (Panayotou *et al.*, 1998); however, this was not observed and very low single stranded activity was detected. This may be because the ssDNA can enter and quickly leave the active site as there is no leucine (L191A) to maintain the target base in the active site. Furthermore, since *e*CYDG prefers uracil over cytosine, it is possible that aspartate (N123D) adopts the optimal conformation for uracil recognition and it is not until cytosine enters the active site that it rotates to form the favourable interactions to allow excision of this base (Pearl, 2000). In this case the base is not retained within the active site for long enough for recognition and subsequent excision.

The ability of *e*CYDG to excise uracil from A.U but not cytosine from G.C suggests that the activity of the enzyme is dependent on the stability of the base pair; G.C is more stable as there is one more hydrogen bond compared to an A.U base pair. In addition, the enzyme may have a higher affinity for uracil and more readily forms favourable bonds with this as a result of D123's conformation (as discussed above). These results also suggest that the major role of L191 is to plug the space left behind after base flipping (thereby increasing the time the base resides in the active site), rather than base flipping itself as suggested by previous studies (Jiang and Stivers, 2002). It therefore appears that base flipping may be initiated by duplex destabilisation as a result of enzyme binding, which is sufficient to destabilise A.U but not G.C (Parikh *et al.*, 1998, Werner *et al.*, 2000, Jiang and Stivers, 2002). The lack of activity of *e*CYDG towards I.C (even though this has two hydrogen bonds as with A.U) suggests that this base pair increases local stability more than A.U due to base stacking interactions (SantaLucia *et al.*, 1996, Watkins and SantaLucia, 2005). This limits the amount of time cytosine is extrahelical, reducing the opportunity for which the enzyme can bind and base excision to occur.

The rate of A.C cleavage is severely reduced if it is flanked by G.C base pairs (A.C(G)) that is most likely explained by an increase in local DNA stability (Seibert *et al.*, 2002) and

the enzyme's inability to base flip (Jiang *et al.*, 2001, Jiang and Stivers, 2002, Jiang *et al.*, 2002b). The increase in local stability is most likely due to base stacking interactions (SantaLucia *et al.*, 1996) in which the 5' base seems to have the greatest impact as seen by A.C(GA), for which the cleavage activity is only slightly greater than A.C(G). This is supported by the increased activity with A.C(AG), though this is still slower than A.C and only reaches a maximum cleavage efficiency of 53%. These variations in the efficiency of cleavage are important when considering cytosine in a CpG context. This is consistent with the decreased cleavage of A.U activity when this is flanked by G.Cs (A.U(G)), as noted in Table 3 in Eftedal *et al.*, 1993, in which greater local duplex stability produces a slower rate of excision. Placing HEG opposite a cytosine which is flanked by G.C base pairs only increased the cleavage rate by approximately 3-fold. In contrast no activity was detected with long HEG.C(G), which is consistent with the result for APHEG.C(G). As the HEG linker spans three bases in long HEG.C, this small region may act like single-stranded DNA and thereby reduce the rate of excision. It is possible that the DNA is now too flexible at the target region, preventing the enzyme from binding and further compounding its loss of activity. The use of a G.T mismatch to cause a destabilisation of the flanking base pairs again increased the amount cleaved and enhanced the rate by approximately 3-fold compared to A.C(G). These results suggest that π - π base stacking provides significant stability to prevent the cytosine from becoming extrahelical.

It was not surprising that the F77Y mutation caused no discernible change in activity, as tyrosine is similar to phenylalanine and is unlikely to alter the structure of the active site. Tryptophan would be expected to have a similar effect, though it has a large aromatic surface area that might be able to provide improved base stacking interactions with the target base, which could stabilise interactions in the active site. Crystal structures (*i.e.* PDB 1SSP (Parikh *et al.*, 1998)) show that the base of the active site is fairly open and suggest that it can comfortably accommodate a tryptophan without any structural distortion to the enzyme(s). The reduction in activity with F77H is most likely due to unfavourable electrostatic interactions caused by the additional nitrogens in ring. The reduced activity of Y66A has previously been reported as to have the observation that *e*TDG is cytotoxic and cannot be expressed (Kavli *et al.*, 1996, Handa *et al.*, 2002, Kwon *et al.*, 2003). Furthermore, *e*TYDG has been shown to be 100-fold less active than *e*CYDG, and its slight activity towards thymine in a Z.T context is in agreement with this previous study (Kwon *et al.*, 2003). As the activity of *e*CYDG is approximately three orders of magnitude

lower than *e*UDG (Handa *et al.*, 2002, Kwon *et al.*, 2003), the addition of a third mutation (*e*CYTDG), has clearly impacted the activity of the enzyme, resulting in the loss of cytosine activity. As this mutation has such a major effect on the activity of the enzyme another approach may be needed to gain specificity towards other cytosine modifications. The other Y66 mutations (Y66T/S/L) showed no activity towards thymine and a decrease in activity towards uracil. The exception to this is Y66F for which the activity is similar to that of the wild type (Handa *et al.*, 2002), though based on our F77 results, we would suggest that this would also be the same for a tryptophan substitution. Further experiments will be required to determine the catalytic rates of these various mutants and to assess the difference in their lower activity.

In summary we have shown that *e*CYDG is able to discriminate between cytosine and 5-methylcytosine in any active substrate context. These results showed that the rate of excision is determined by the stability of the target base pair, the flanking base pairs and base stacking. This is consistent with the properties of UDG, which has a preference for G.U over A.U (Panayotou *et al.*, 1998). The inability of *e*CYDG to excise cytosine from G.C base pairs, and its lack of ssC activity, provides the possibility of using the enzyme to probe the methylation status of any cytosine by generating a mismatch at a target cytosine to allow for excision if unmethylated. The proposed assay is discussed in Chapter 6. One way to overcome the low activity with the more stable substrates could be to exploit the more active *h*CDG, and this is discussed in Chapter 4.

Chapter 4: Exploring the inability to clone *e*CDG

4.1 Introduction

The results presented in Chapter 3 (section 3.3.1) showed that plasmids containing *e*CDG could not be generated in *E. coli* and suggested that this protein is exceptionally cytotoxic and may also be translated by an unusual mechanism. This sequence could not be generated within pUC19, even though it was out of frame with the *LacZ* gene, suggesting some form of leakage or alternative initiation. In contrast the mutation could be made using the same sequence in pUC18, placing it in the opposite orientation, which would lead to transcription of the non-coding strand. We therefore decided to investigate this further in order to understand the basis of this toxicity and to see whether this could be overcome.

It is known that the basal expression of proteins can be repressed when cells are grown in glucose-containing media (Grossman *et al.*, 1998, Pan and Malcolm, 2000). The presence of glucose keeps the levels of cyclic AMP (cAMP) low and so reduces transcription, as the cAMP receptor protein (CAP) requires cAMP in order to bind DNA at the *lac* promoter. This complex then promotes transcription by recruiting *E. coli* RNA polymerase. Another possible mechanism by which this toxic gene might be produced is *via* a form of prokaryotic internal ribosomal entry (IRES) or ribosome “slippage” to initiate translation at the *e*UDG start codon (which is out of frame with the *LacZ* gene). In this process the first AUG start codon is missed or skipped and translation occurs at the next (or further downstream) AUG; in this instance at the start of the *e*UDG/CDG gene. To determine how active *e*CDG might be produced, even though it is out of frame with *LacZ*, we therefore prepared constructs with different regions between the *LacZ* and *e*UDG start codons.

Another approach for reducing the cytotoxic effects of *e*CDG would be to co-express it with the uracil DNA glycosylase inhibitor (Ugi). This binds very tightly (effectively irreversibly) to UDG under physiological conditions and inhibits its action (Bennett and Mosbaugh, 1992). Examination of the structure of the complex suggests that CDG should also be inhibited by Ugi (Kavli *et al.*, 1996). We predict that although the affinity of *h*CDG for DNA (K_d) is reduced compared to that of *h*UDG (Kavli *et al.*, 1996), Ugi should still be able to bind tightly and to cause inhibition. It has also been shown that UDG and Ugi can be expressed from the same plasmid (Roy *et al.*, 1998, Acharya *et al.*, 2002). However, this will require the two proteins to be generated at equal rates. We therefore investigated whether a fusion between *e*UDG and Ugi might overcome these problems as the proteins

will be generated in 1:1 stoichiometry and the covalent attachment will allow rapid and direct inhibition of *e*UDG/CDG. Generation of a fusion protein may therefore provide a method for *e*CDG expression by direct inhibition of its cytotoxic properties.

We therefore attempted to address the problem of cytotoxicity through (i) reverse mutagenesis, (ii) reducing basal expression, (iii) examining the mechanisms of IRES and (iv) co-expression with an inhibitor. Unfortunately all of these attempts were unsuccessful.

4.2 Experimental Design

4.2.1 Reverse Mutagenesis

To determine whether the cytotoxicity of *e*CDG was directly related to the N123D mutation we attempted to reverse the L191A mutation in *e*CYDG and convert it back to *e*CDG. We used both pUC19*e*CYDG and pUC18*e*CYDG as a control as we have already shown that pUC18*e*CDG is a stable clone. The A191L mutation was introduced *via* site directed mutagenesis (section 2.3.1).

4.2.2 Using different Cell Types to Generate a Stable *e*CDG Clone

4.2.2.1 Recombination Deficient Cells

Sure cells (Agilent Technologies) are used in the cloning of unstable clones as they are deficient in specific repair genes and may be able to accommodate a cytotoxic repair protein. Therefore the pUC19*e*CDG mutagenesis product was transformed into these cells.

4.2.2.2 Glucose Supplementation

To establish whether glucose could repress any basal expression of *e*CDG, agar plates were prepared containing 1% glucose. The pET*e*CDG mutagenesis product was transformed directly into BL21(DE3)pLysS expression cells that are sensitive to glucose, whilst also having the protein expression regulatory secondary plasmid of pLysS.

4.2.3 Altering the Region Upstream of the *e*UDG Start Codon

We designed a pair of 50mer oligonucleotides to increase the distance between the start codons of the *LacZ* and *e*UDG within pUC19, while maintaining the original reading frame. The insert (Figure 4.1A) contained pairs of restriction sites (coloured) that could be used to shorten the linker if necessary, as shown in Figure 4.1B and C.

A
 5' -AGCTACGTAAGATCTTATACTCGAGCAAGGATATCGACTGCTCGAGCCAGATCTTCC-3'
 3' -TGCATTCTAGAATATGAGCTCGTTCCTATAGCTGACGAGCTCGGTCTAGAAGGTCGA-5'

B
 5' -AGCTACGTAAGATCTTATACTCGAGCCAGATCTTCC-3'
 3' -TGCATTCTAGAATATGAGCTCGGTCTAGAAGGTCGA-5'

C
 5' -AGCTACGTAAGATCTTCC-3'
 3' -TGCATTCTAGAAGGTCGA-5'

Figure 4.1 Sequences produced to vary the distance between the *LacZ* and *eUDG* start codons. A) The full length sequence designed with overhangs complementary to the *Hind*III restriction site (green) for original cloning, *Bgl*II (purple) and *Xho*I (red) for production of sequences B and C. The full length sequence was designed to maintain the reading frame, whilst an *EcoRV* (blue) site was included to act as a diagnostic for successful cloning prior to sequencing.

We also inserted a stop codon, both in frame and out of frame of the *eUDG* start codon (Figure 4.2), between the *LacZ* start and *eUDG* in pUC19*eUDG*. It was hoped that this would stop any premature translation of *eCDG* that had initiated from the *LacZ* start codon.

A
 5' -ATGACCATGATTACGCCAAGCTTCATATGGCT-3'

B
 5' -ATGACCATGATTACGCCAAGTAAACATATGGCT-3'

C
 5' -ATGACCATGATTACGCCATAATTCATATGGCT-3'

Figure 4.2 The DNA sequences of the region between the start codons of *LacZ* (5' blue) and *eUDG* (3' blue). A) Unmutated sequence. B) Incorporation of a stop codon (red) in frame with *UDG*. C) Incorporation of a stop codon (red) in frame with *LacZ*.

4.2.4 Co-transformation of Ugi

The products of pET*eUDG* N123D mutagenesis reactions were co-transformed with Ugi (contained within the vector pRSETB that has ampicillin resistance) or with Ugi-containing competent cells. The transformations were then spread onto agar plates containing carbenicillin (100 µg/ml) and kanamycin (30 µg/ml) to give selectivity towards cells that contained both plasmids.

4.2.5 Generating an *eUDGUgi* Fusion Protein

Oligonucleotides were designed to amplify Ugi, including restriction sites for cloning into pET*eUDG*. *eUDG* and Ugi have 4 and 14 residues at the C-terminus and N-terminus respectively, which appear to be unstructured, as they are not seen in crystal structures. It

should therefore be possible to directly link the two proteins together. We also engineered a thrombin cleavage site into the 5' oligonucleotide primer so as to allow for separation of the proteins during purification. The thrombin site also acts to extend the linker between the two proteins and should allow Ugi to fold easily around the DNA binding site of *e*UDG/CDG.

4.2.6 Further Mutagenesis of *e*UDGUgi

After cloning the Ugi sequence into pET*e*UDG, the *e*UDG stop codons and the upstream thrombin cleavage site had to be removed. The thrombin cleavage site is 18 bp long which provides enough base pairing between the primer and template to generate a stable duplex. Gaining specificity for mutating one thrombin site (upstream of *e*UDG) over the other (between *e*UDG and Ugi) may be difficult to achieve. To increase the specificity, and the chances of success, we designed 31mer oligonucleotides that included the flanking regions specific to the upstream thrombin site (Figure 4.3).

Finally the N123D mutation could be performed to see if pET*e*CDGUgi could be generated. All mutations were introduced *via* site directed mutagenesis and were confirmed by sequencing. The oligonucleotides used are shown in Table 2.2.

5' -AGCGGCCTGGT**GCCG**GGCGGCAGCCATATGG-3'

3' -TCGCCGG**ACCACGGC**CGCCGTTCGGTATACC-5'

Figure 4.3 Oligonucleotide primers designed for mutation of the upstream thrombin cleavage site. The thrombin site and the single base mutation highlighted in blue and red respectively.

4.3 Results

4.3.1 Restoration of *e*CDG Activity

Both pUC18*e*CYDG and pUC19*e*CYDG were subjected to further SDM to generate *e*CDG by reversing the L191A mutation. Both reactions produced successful colonies, which were subjected to Taq^oI digestion in order to confirm that they still contained the N123D mutation. All clones showed the correct three band pattern. These were then sequenced using the CEQ8000 genetic analysis system. The results were positive for pUC18, producing pUC18*e*CDG, but negative for pUC19 in which all the clones contained a secondary mutation.

4.3.2 Effects of Different Cell Types

4.3.2.1 Using Sure Cells

The use of Sure cells did not produce any positive pUC19*e*CDG clones, suggesting that cytotoxicity is directly related to *e*CDG and/or the BER pathway and not general cellular repair.

4.3.2.2 Effects of Glucose

No positive pET*e*CDG clones were obtained when using glucose to repress basal expression of *e*CDG. This suggests that either glucose is not able to completely repress basal expression, or suppress it to a level at which *e*CDG expression is low enough not to cause cytotoxicity.

4.3.3 Alternative Initiation

The linker oligonucleotide (Figure 4.1A) was cloned into the upstream region of pUC19*e*UDG by ligating it into the plasmids *via* a single HindIII site. Once the full length insert had been successfully cloned, this was truncated by selective cutting with BglII (Figure 4.1B) or XhoI (Figure 4.1C) and re-ligated to generate the shorter linkers. Restriction digests were performed with XhoI/EcoRV or EcoRV respectively to determine successful cloning, as correct constructs would not contain the respective restriction sites resulting in an undigested plasmid. Positive clones were then sent for sequencing for confirmation. These were then used for site directed mutagenesis to introduce the N123D mutation. As with previous experiments, the N123D mutation yielded no clones of the correct mutated sequence with any of the three constructs.

4.3.4 Addition of Stop Codons

After successful mutagenesis to incorporate stop codons into the upstream region of *e*UDG, mutagenesis was performed to attempt to generate an *e*CDG clone. This still produced no viable clones, suggesting that translation initiation was occurring from the start codon of *e*UDG and not from that of *LacZ*.

4.3.5 Introduction of Ugi

Co-transformation of Ugi with the mutagenesis reaction mixture for converting pET*e*UDG to pET*e*CDG yielded very few colonies. These colonies were subjected to Taq⁹¹I digestion

and were determined to be wild type pET_eUDG. As a control, pET_eUDG was also co-transformed with Ugi to determine dual transformation efficiency. The number of colonies yielded was not significantly greater than with pET_eCDG transformation. This then led us to make competent cells containing Ugi that could then be transformed with pET_eCDG. Ugi was successfully transformed into XL-1 Blue cells and viable competent cells made from these. The number of colonies significantly increased (approximately 10 fold compared to co-transformation) upon transformation with pET_eUDG, but there were still very few (≤ 5 colonies) after pET_eCDG transformation. These again proved to be pET_eUDG and no positive pET_eCDG clones were obtained.

Rather than relying on basal expression of Ugi to inhibit any *e*CDG protein produced we attempted to induce production of Ugi using agar plates containing 0.2 mM IPTG. It was hoped that Ugi would be produced at a stoichiometry of at least 1:1 with *e*CDG. Expression from pET28a is also under IPTG control and therefore *e*CDG and Ugi would be induced together. This too failed to yield any positive colonies. As pRSETB is also induced by IPTG we decided to clone Ugi into a vector that was under a different inducer. We therefore amplified Ugi using primers containing SacI and EcoRI restriction sites for cloning into the pBADA vector that is induced by arabinose (see Appendix I for oligonucleotide sequences: Ugi For and Ugi Rev). Ugi was cloned into pBADA and colony PCR was performed and positive clones sent for sequencing for confirmation; generating pBADAUgi. pBADAUgi was co-transformed with pET_eCDG onto agar plates containing 0.02 to 20% (w/v) arabinose (and kanamycin) so that only Ugi would be induced to inhibit any basal expression of *e*CDG. Again no positive clones were obtained.

4.3.6 Cloning and Mutagenesis of UDG-Ugi Constructs

4.3.6.1 Generating a pET_eUDGUgi Fusion Construct

The bacteriophage PBS1 Ugi was amplified by PCR to generate a product that contained an EcoRI restriction site and a thrombin cleavage site at its N-terminus, and a HindIII site at its C-terminus (see Appendix I for oligonucleotide sequences: UDGUgi For and UDGUgi Rev). The product (approximately 300 bp) generated was confirmed by agarose gel electrophoresis and cloned into pET_eUDG *via* the restrictions sites EcoRI and HindIII (Figure 4.4).



Figure 4.4 Cartoon of the pETeUDGUgi construct.

The cloning of Ugi into pETeUDG should generate an eUDGUgi construct of 972 bp. Upon transformation, colonies were subjected to colony PCR using primers designed to flank the construct and thus generate a product of 1102 bp. This can clearly be seen in Figure 4.5 (Lanes 2 – 5 and 10 – 12) where positive clones of pETeUDGUgi are seen at approximately the 1000 bp marker. The identity of the clones in lanes 2 – 4 was then confirmed by sequencing.

pETeUDGUgi was then subjected to two rounds of mutagenesis to remove the upstream thrombin cleavage site and the stop codons of eUDG. The stop codons were mutated to two glycine residues, to allow for continual translation into the Ugi gene. This construct was denoted pETeUDGUgiS. Secondly the upstream thrombin site was mutated so that Ugi could be released from the hybrid protein after thrombin cleavage, leaving the eUDG/CDG bound to the nickel column. The sequencing results show a cytosine to guanine mutation, which is sufficient to prevent thrombin cleavage, in the upstream thrombin site but not in the thrombin site between eUDG and Ugi, as indicated by the arrows. This final construct was denoted pETeUDGUgiST and its identity was confirmed by sequencing (Figure 4.6).

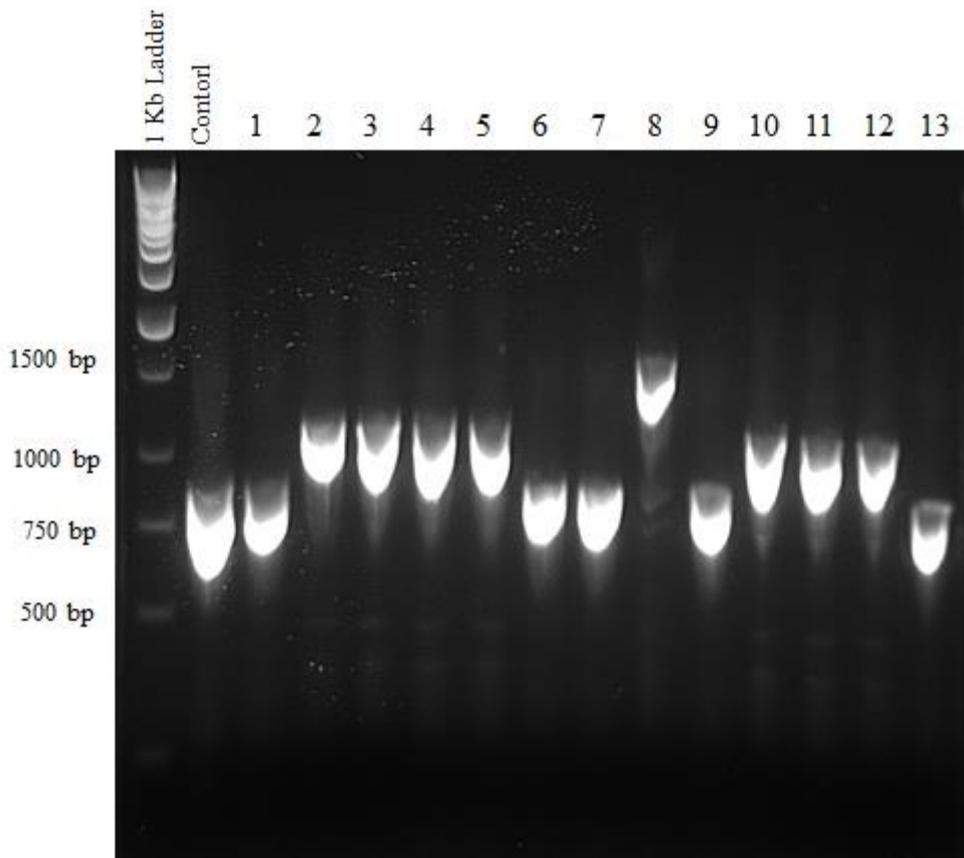


Figure 4.5 Colony PCR to determine positive pET*e*UDGUgi clones. Positive clones are shown in lanes 2 – 5 and 10 – 12; clones 2 – 4 sent for sequencing. 0.7% agarose gel ran in 1x TBE at 15 Vcm⁻¹ for one hour.

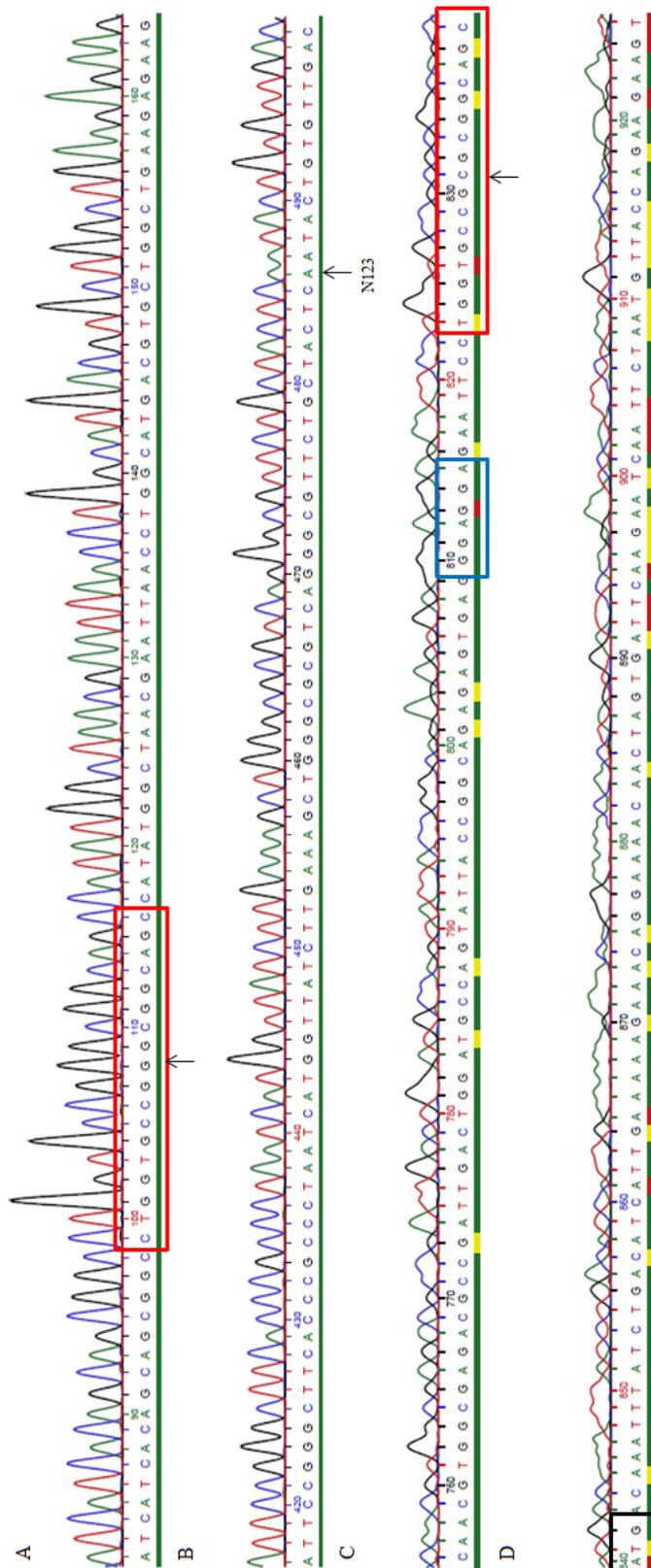


Figure 4.6 Sequencing analysis of pETεUDGiST. A) The upstream thrombin cleavage site (red box) with the C to G single base mutation indicated by an arrow. B) The base required for the N123D mutation. C) The mutation of the stop codons to glycine residues (GGA; blue box) and the internal unmutated thrombin cleavage site (red box). The equivalent base mutated in the upstream site is indicated by an arrow and remains as a cytosine. D) The start codon (ATG; black box) of Ugi.

4.3.6.2 Generation of pETHUDGUgi via complete gene synthesis

The bacteriophage *ugi* gene is very AT-rich and we were therefore concerned that its unusual codon usage might not facilitate expression in *E. coli*. We therefore prepared a synthetic version of this gene (*eUgi*) which, though still very AT-rich, was optimised for expression in *E. coli* (Figure 4.7). We combined this with the synthetic version of *hUDGΔ81* (which is described in Chapter 5) to generate the construct *hUDGΔ81Ugi*. This was prepared from two rounds of PCR (Figure 4.8A), as described for *hUDGΔ81* in chapter 5.

```

PBS1Ugi  ATGACCAATTTATCTGACATCATTGAAAAAGAAACAGGAAAACAACAGTGATTCAAG
          M T N L S D I I E K E T G K Q L V I Q
eUgi     ATGACCAACTTGCCGACATCATCGAAAAAGAGACCGGCAAGCAACTGGTTATTCAAG
          M T N L S D I I E K E T G K Q L V I Q

AATCAATCTAATGTTACCGAAGAAGTAGAGGAAGTAATTGGGAATAAACAGAAAGTGATATTT
E S I L M L P E E V E E V I G N K P E S D I
AATCTATCTTGATGCTCCCTGAAGAAGTAGAAGAGGTATATCGGTAATAAGCCGGAGTCCGACATTC
E S I L M L P E E V E E V I G N K P E S D I

TAGTTCATACTGCTTATGATGAAAGTACAGATGAAAATGTAATGCTATTAACCTCAGATGCTCCAG
L V H T A Y D E S T D E N V M L L T S D A P
TGGTGCACACTGCTTATGACGAATCTACTGATGAGAACGTAATGCTGCTGACTAGCGACGCTCCGG
L V H T A Y D E S T D E N V M L L T S D A P

AATATAAACCTTGGGCTTTAGTAATTCAAGACAGTAATGGAGAAAATAAAATTAAAATGTTATAA
E Y K P W A L V I Q D S N G E N K I K M L
AATACAAACCGTGGGCTCTGGTAATTCAAGACTCTAACGGCGAAAACAAAATCAAGATGCTGTAA
E Y K P W A L V I Q D S N G E N K I K M L

```

Figure 4.7 DNA and protein sequence alignments of bacteriophage PBS1 and *E. coli* to show optimisation for expression in *E. coli*. Base, but not amino acid, changes highlighted in red.

As with *hUDGΔ81* the first round of PCR generated a mixture of species (shown as a smear in lane 2) containing the full length product, which was selectively amplified by a second round of PCR using excess of the terminal primers, to produce an amplified product of 973 bp (lane 3). As *hCDG* can be produced in *E. coli* we had no need to use the clone for the potential production of *hCDG*.

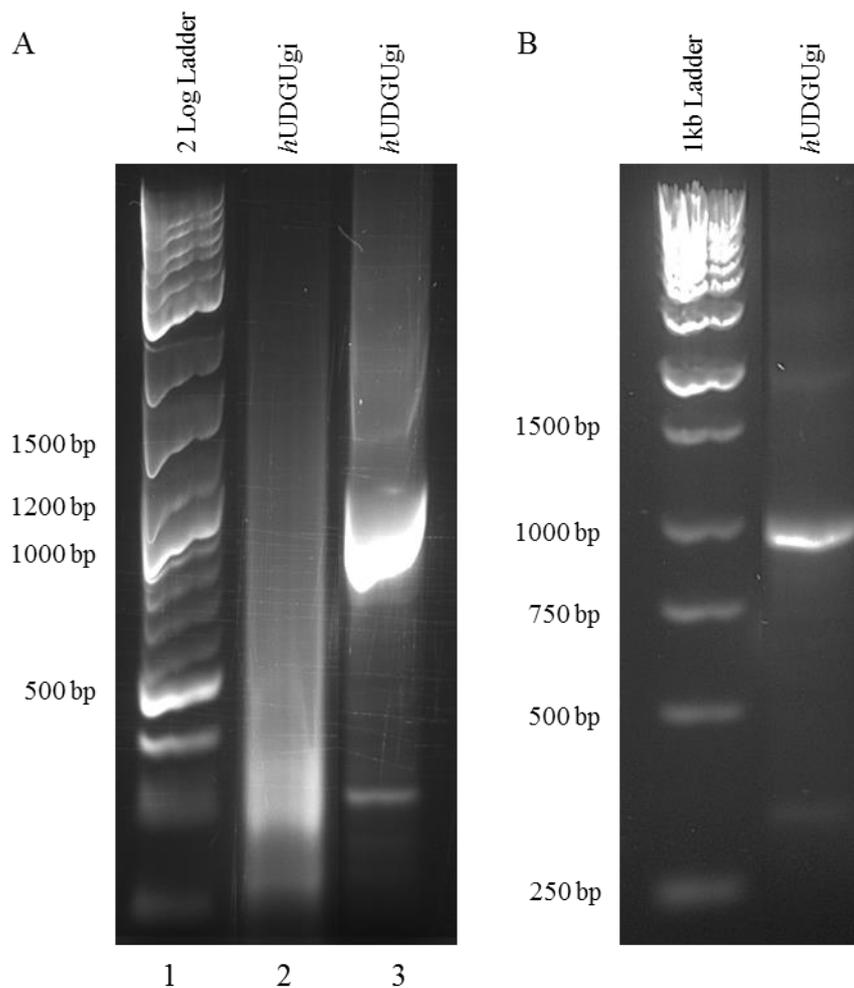


Figure 4.8 Complete gene synthesis of *hUDGΔ81Ugi*. A) Products after first round (lane 2) and second rounds (lane 3) of PCR amplification. B) PCR purification of lane 3 of A. 0.7% agarose gel run in 1x TBE at 15 Vcm^{-1} for one hour.

4.3.7 Expression of pET*eUDGUgiST*

As positive pET*eCDGUgiST* clones could not be generated, it suggested that the fusion protein was still cytotoxic or that the Ugi was not able to inhibit the *eCDG* protein. To investigate this we attempted to express *eUDGUgi*, with the aim of testing its activity against a G.U substrate. If the fusion protein still has UDG activity it would suggest that the Ugi is not inhibiting *eUDG* and therefore explain why a positive *eCDGUgi* construct cannot be made. Alternatively if the fusion protein is inactive it would suggest that the Ugi was causing inhibition and that it should be possible to generate an *eCDGUgi* construct. Surprisingly we were unable to express and purify *eUDGUgiST* as shown by Figure 4.9A and B. Figure 4.9A a potential band of *eUDGUgiST* (~ 37 kDa) in fraction 1, though no induction band, or a corresponding band in the sonication sample can be seen, suggesting this is an artefact at the expected MW. The purification was repeated to investigate this

further (Figure 4.9B) and again no clear induction can be seen and this time a darker band can be seen in fraction 1 and 2 at approximately 27 kDa. This is indicative of *eUDG* expression and not the *eUDG_{UGiST}* construct. This was not pursued further.

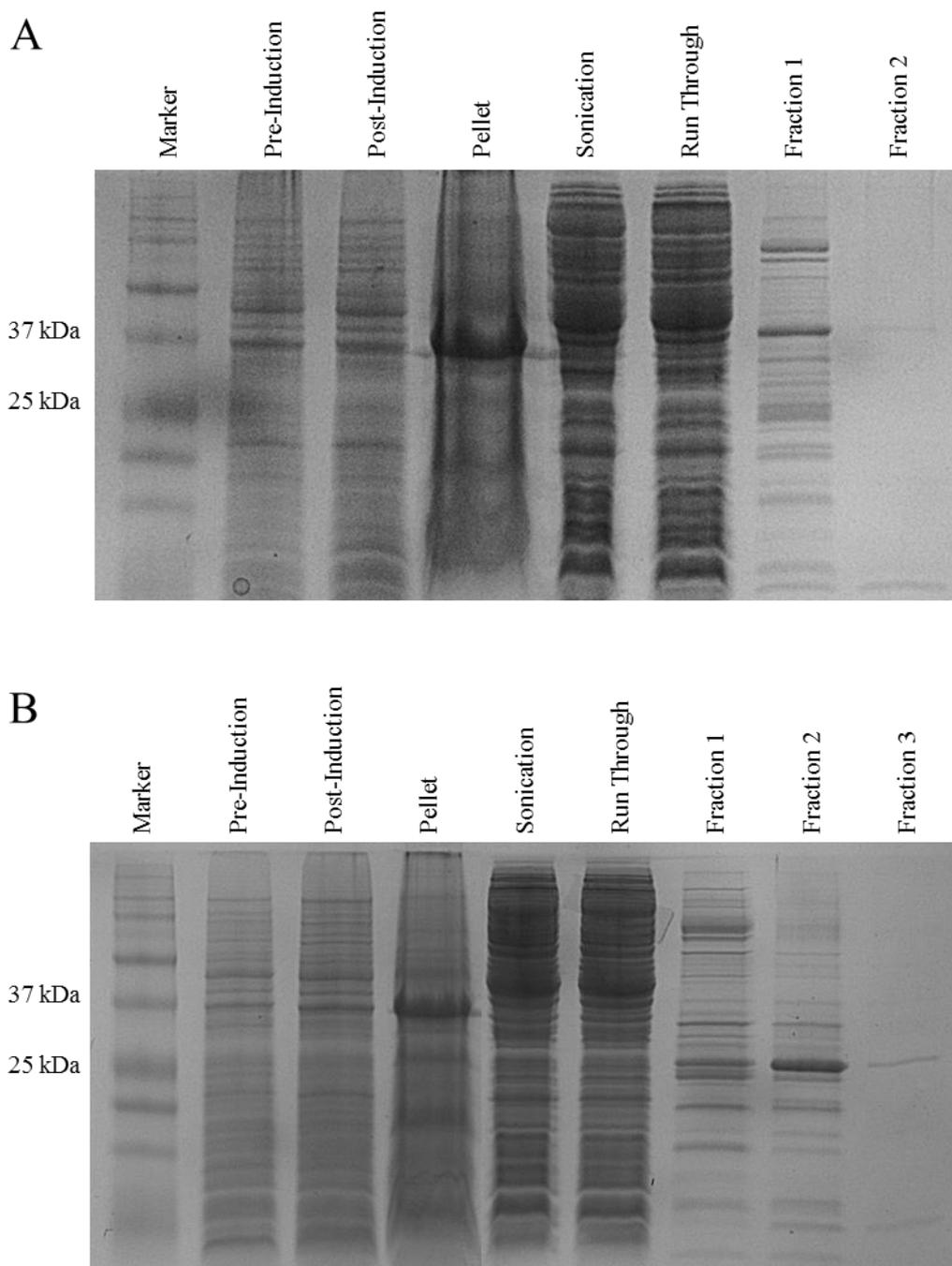


Figure 4.9 SDS-PAGE purification of *eUDG_{UGiST}*. Samples were taken at different purification stages to determine enzyme expression and purity.

4.4 Discussion

We assume that *e*CDG is very cytotoxic, as expression of the enzyme will lead to excision of cytosine from plasmid and genomic DNA and result in fragmentation of the genome. If this excision occurs at a faster rate than the ability of the *E. coli* cells to repair the damage, then cell death will be inevitable. This appeared to be the case as clones for *e*CDG could not be obtained in either pUC19 or pET28a. We assumed that this occurs through “leaky” expression in the pET vector, but is more puzzling for the pUC clone. The mRNA produced by reading from the *LacZ* promoter, would contain the *e*CDG sequence, but this will be out of frame with the *LacZ* sequence and therefore produce a nonsense protein. Therefore any *e*CDG produced must have occurred through translation from its own start codon arising from ribosomal scanning and/or slippage missing the initial *LacZ* start codon *via* a form of alternative initiation.

This alternative initiation is most likely to occur through ribosomal binding at Shine-Dalgarno sequences (AGGAGG) within the *LacZ* and *e*CDG start codons. Upon review of this region, though there is no exact consensus match, there are two positions that correspond to the sequence AXXAXG. It has been reported that GAGG is only required for translation of early genes in phages (Malys, 2012). It is therefore conceivable that other partial sequences can be tolerated, allowing ribosomal binding, and in our case translation from the *e*CDG start codon. Any expression of *e*CDG in this way is likely to be very low but these results suggest that only a small amount of *e*CDG is enough for it to be cytotoxic to the cell. This is consistent with the observation that this sequence could be stably cloned in pUC18, in which it is inserted in the opposite orientation, so that transcription from the *LacZ* promoter would lead to production of a completely different nonsense protein. It is of note that even using different cell types that are under tighter control of expression (BL21DE3pLysS cells) or are repair deficient (Sure cells), we were still unable to produce an *e*CDG clone.

All the clones that were obtained when attempting to produce the N123D mutation of *e*UDG, in pUC19 or pET28a, were found to be incorrect upon sequencing. These had each undergone a secondary mutation generating either a deletion or insertion of a nucleotide (or nucleotides). As a result the rest of the protein would be out of frame and would produce a nonsense protein. These secondary mutations therefore produced a protein without CDG activity and so enabled the host and the plasmid to survive. This was also

shown by the reverse mutagenesis of *eCYDG* to *eCDG*, in which the A191L mutation was successful for pUC18*eCYDG* (generating pUC18*eCDG*) but not for pUC19*eCYDG*, for which the N123D was only obtained in conjunction with a secondary mutation(s).

The failure of Ugi to allow stable *eCDG* expression was not completely unexpected. As the two proteins are made from separate plasmids it is highly likely that they are produced in different amounts and at different rates. If *eCDG* is produced at a greater rate than Ugi then the Ugi concentration will not be sufficient to prevent the cytotoxicity. Secondly, as the proteins are made separately, Ugi may not inhibit *eCDG* fast enough. There may be sufficient time for *eCDG* to locate and start excising cytosine from plasmid and genomic bacterial DNA, leading to DNA damage and cell death. Therefore, for Ugi to be effective, even if it was being produced at a greater rate, it would still have to locate and bind to the *eCDG*. Finally, Ugi is a phage protein and the gene for this is extremely AT rich, meaning the codon usage is different to *E. coli* and could reduce the rate of Ugi translation. To overcome these problems we attempted to generate a fusion protein between *eUDG* and Ugi, whereby this would guarantee a 1:1 stoichiometry while keeping the proteins in close proximity to allow for rapid inhibition upon expression. Though we successfully generated a full fusion construct of *eUDG* and Ugi (pET*eUDGUgiST*), we were still unable to obtain any clones with the N123D mutation without a secondary mutation. Furthermore, we were unable to express *eUDGUgiST* and our results showed inconsistent purification in the various fractions. Further investigation is therefore required to understand this and to ascertain whether a fusion protein can allow the expression and purification of an *eCDG*.

The investigation of cytotoxicity from a variety of different methods strongly suggest that *eCDG* is extremely cytotoxic to *E. coli* cells as the enzyme is unable to be produced except when 'inactivated', *i.e.* through secondary mutations. Therefore its production may only be possible by *in vitro* transcription translation systems. We suggest that apart from *eCDG* being more active (as our results in Chapter 3 would suggest), *eCDG* is cytotoxic compared to *hCDG* in *E. coli* as *hCDG* is unable to recruit the other *E. coli* proteins in the BER pathway for DNA repair. Based on this assumption it would be interesting to attempt to express both CDGs in other cell lines (*e.g.* HeLa cells, yeast or insect cells) to see if the effect is reversed and cell type dependent.

Chapter 5: Excision Properties of *h*CDGs

5.1 Introduction

As previously reported (Kavli *et al.*, 1996, Handa *et al.*, 2002), *e*CDG is toxic and cannot be expressed in *E. coli*. This was confirmed in Chapter 3 of this thesis, which therefore examined the properties of the double mutant N123D, L191A (*e*CYDG). It would have been preferable to study *e*CDG itself, though this enzyme has previously only been prepared using an *in vitro* transcription-translation system (Handa *et al.*, 2002). The yield of enzyme from this system was much lower than that of direct expression of UDG in *E. coli*. Another disadvantage with this system for generating *e*CDG is that the transcription and translation steps have to be performed separately rather in a one-pot dual assay. As soon as *e*CDG is produced it will begin to degrade the template DNA preventing any further transcription. Therefore the amount of *e*CDG mRNA is limited and decreases over time through degradation of the DNA template. Therefore finding a way to produce a CDG *in vivo* would be highly advantageous. This has previously been achieved for the human enzyme *h*CDG using *recA*⁻ strains of *E. coli* as it this enzyme appears to have lower cytotoxicity than the equivalent *E. coli* enzyme (Kavli *et al.*, 1996). The difference in toxicity might be because the human and *E. coli* enzymes have different amino acid codon usage, resulting in a lower rate of *h*CDG expression in *E. coli*, allowing repair of any lesions to keep pace with the damage. Alternatively it is possible that the toxic effects of *e*CDG result from other *E. coli* proteins that are recruited to the sites, which are not able to interact with *h*UDG. We therefore looked to exploit this as a means for expressing *h*CDG in *E. coli* and assessing its activity towards a variety of cytosine-containing substrates. This chapter describes the preparation and properties of *h*CDG, as well as *h*CYDG, to allow direct comparison with *e*CYDG.

5.2 Experimental Design

5.2.1 Complete Gene Synthesis of *h*UDG

The sequence of human placental UDG cDNA (Olsen *et al.*, 1989) was used to generate the *h*UDG gene by total gene synthesis. Slupphaug *et al.* (1995) have shown that *h*UDG is fully active with a truncated N-terminus, as this signal sequence is only required for translocation to the nucleus/mitochondria. This truncated sequence was divided into ~60mers that allowed for approximately 20 base overlap between each of the

complementary strands. The central unpaired regions were then filled in and the whole gene amplified using a high fidelity polymerase. An additional benefit of this method is that mutations can be introduced made by simply changing one of the cassette oligonucleotides. The sequences of the oligonucleotides that were used to generate this clone are shown in Appendix II. As described below, this generated the sequence coding for *hUDGΔ81* (*hUDG* with residues 3 - 84 removed).

5.2.2 Mutagenesis of *hUDG*

The N204D and L191A mutations were introduced into *hUDGΔ81* *via* site directed mutagenesis as in described in Section 2.3.1. The oligonucleotides used for this are shown in the Materials and Methods; Table 2.2.

5.2.3 Excision Properties

As with *eCYDG*, DNA cleavage assays were used to assess the excision properties of the *hCDGs*. These assays were performed in triplicate; the rate constants were calculated and averaged to determine the rate of reaction of the enzyme against a variety of different substrates. The assays were performed over 24, 4 or 1 hour time courses, depending on the activity of the enzyme towards each particular substrate. The oligonucleotides containing the target base were radiolabelled with ^{32}P at the 5' end and annealed to a complementary strand to form duplex substrates for the enzyme. The substrates generated contained target U, T, C or $^{\text{M}}\text{C}$ paired opposite a G, A, AP (abasic site), Z (anthraquinone pyrrolidine) and HEG (hexaethylene glycol). As in the previous chapter anthraquinone pyrrolidine was used as a bulky synthetic nucleoside instead of pyrene, as the pyrene nucleoside is not commercially available. After incubation with *hCDG* the DNA was treated with 10% (v/v) piperidine to cause specific cleavage at the abasic site, generating a single product band. The resulting products were run on denaturing PAGE, phosphorimaged and analysed.

5.3 Results

5.3.1 Complete Gene Synthesis of *hUDGΔ81*

hUDGΔ81 was generated through complete gene synthesis from two rounds of PCR (Figure 5.1). The first round generated the full length construct from a series of partially overlapping oligonucleotides, optimising the PCR reaction as detailed in Materials and Methods 2.3.7. The resulting product seen as a smear in Figure 5.1A (lane 2), which

contained the full length product as well as many other partial sequences. This was further amplified using the terminal primers to generate a band of 697 bp (lane 3). This can be seen more clearly after purification of the PCR reaction using a PCR product purification kit (QIAGEN) (Figure 5.1B).

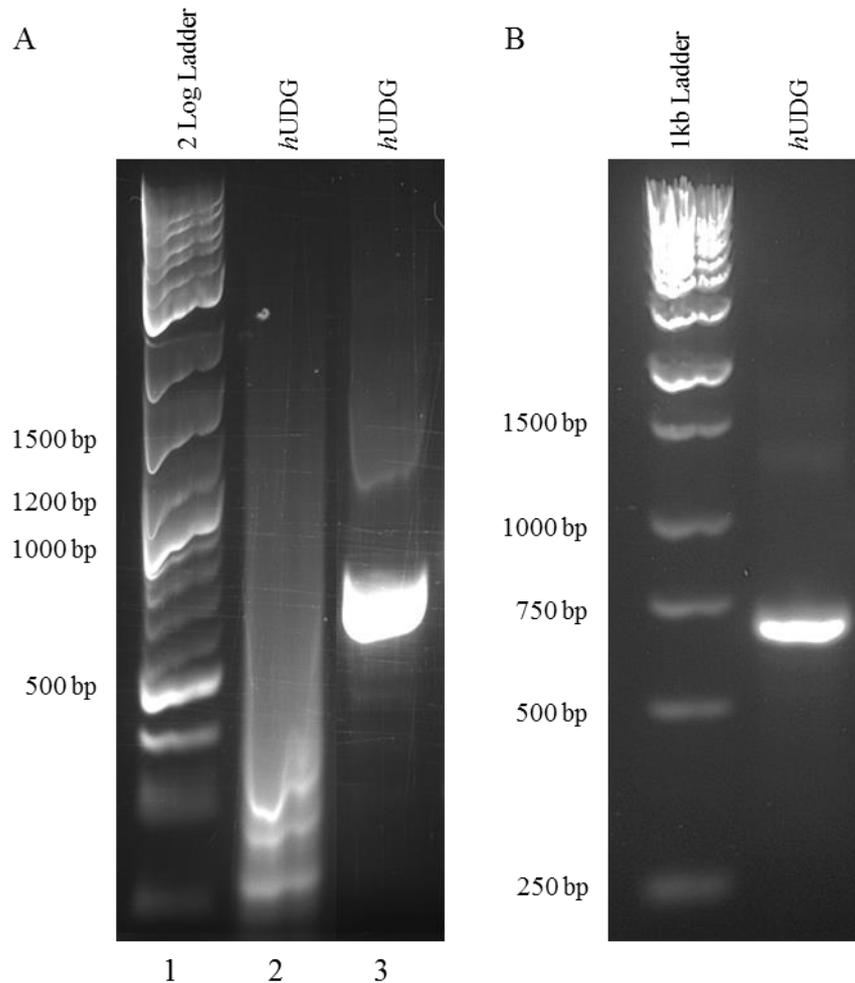


Figure 5.1 Complete gene synthesis of *hUDG*. A) Products after the first (lane 2) and second (lane 3) rounds of PCR amplification. B) PCR purification of lane 3 of A. These 0.7% agarose gels were run in 1x TBE at 15 Vcm^{-1} for one hour.

5.3.2 *hCDG* Mutagenesis

The PCR mixture from Figure 5.1B underwent a PCR clean-up and the construct was cloned into the vector pET28a. Colonies were screened by colony PCR (Materials and Methods 2.3.9) and a positive clone was obtained (Figure 5.2A, *hUDG* Δ 81 2). Plasmid DNA from the positive colony (pETHUDG Δ 81 2) was prepared and the sequence confirmed by sequencing. This clone was then subjected to further site directed mutagenesis, to introduce the N204D mutation, generating human CDG (pETHCDG Δ 81).

After transformation plasmids were prepared from two of the successful colonies and were subjected to digestion by Taq^{qI} . The N204D mutation creates an extra restriction site for Taq^{qI} , which then generates a different digestion pattern. Though there is a difference in the cleavage patterns produced with $hUDG\Delta81$ and $hCDG\Delta81$ (Figure 5.2B) the pattern is not as predicted from a theoretical digest. The largest fragment should be 1420 bp, while the additional restriction site should remove a 627 bp fragment, which should be cleaved into two fragments of 526 and 101 bp. This may be due to secondary structure forming preventing restriction enzyme binding. Despite this, the two clones were sent for sequencing and confirmed the N204D mutation had been successful. The ability to obtain a pETHCDG clone suggests that $hCDG$ does not have the same downstream cytotoxic properties as $eCDG$ or it may be less active.

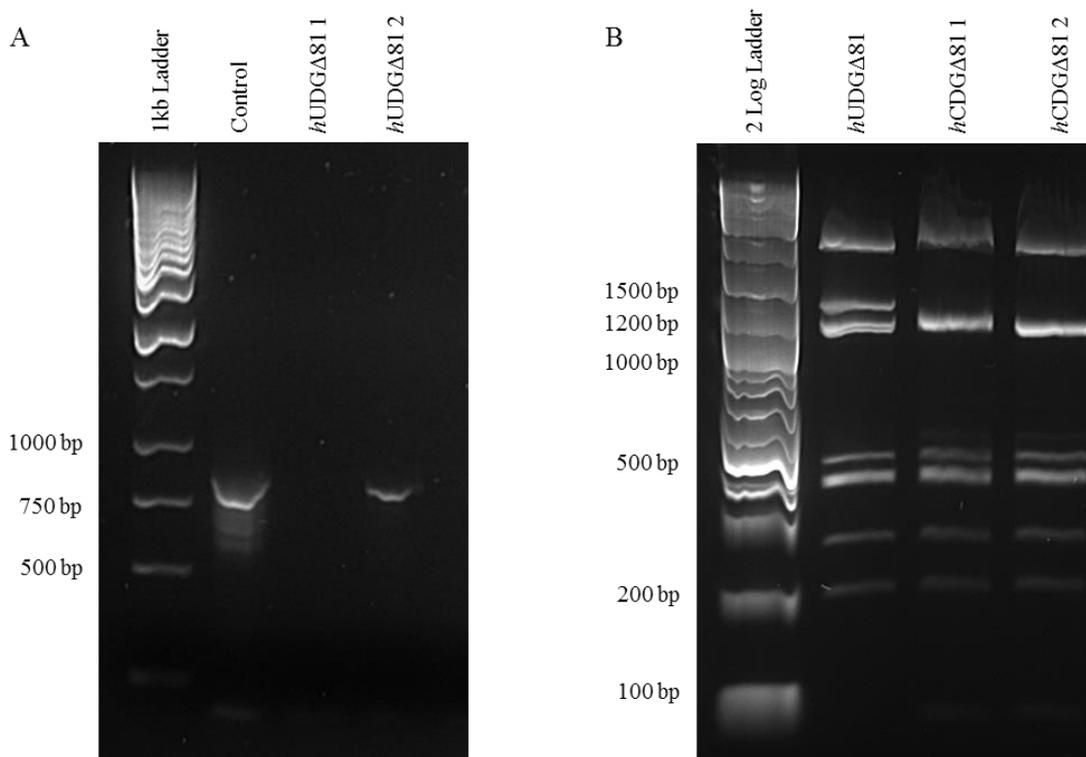


Figure 5.2 Generation of $hCDG\Delta81$. A) Colony PCR of the $hUDG$ gene of two potential clones. Control relates to the amplification of pUC19 $eUDG$ for size reference. B) Taq^{qI} digestion of pETHCDG $\Delta81$ clones. A different digest pattern can be seen in the region of 1200 to 1500 bp. 0.7% agarose gel ran in 1x TBE at 15 Vcm^{-1} for one hour.

5.3.3 Expression of $hCDG\Delta81$

$hCDG\Delta81$ was expressed and purified (Figure 5.3) in the same manner as for all other UDG variants (Chapter 3). The yield obtained was 0.45 mg from 0.5 L of culture, which is significantly less than the for the other mutant *E. coli* proteins that varied between 0.75 and

1.5 mg. Despite the low yield, Figure 5.3 shows an enriched protein of the correct molecular weight (approximately 28 kDa) in the final two lanes.

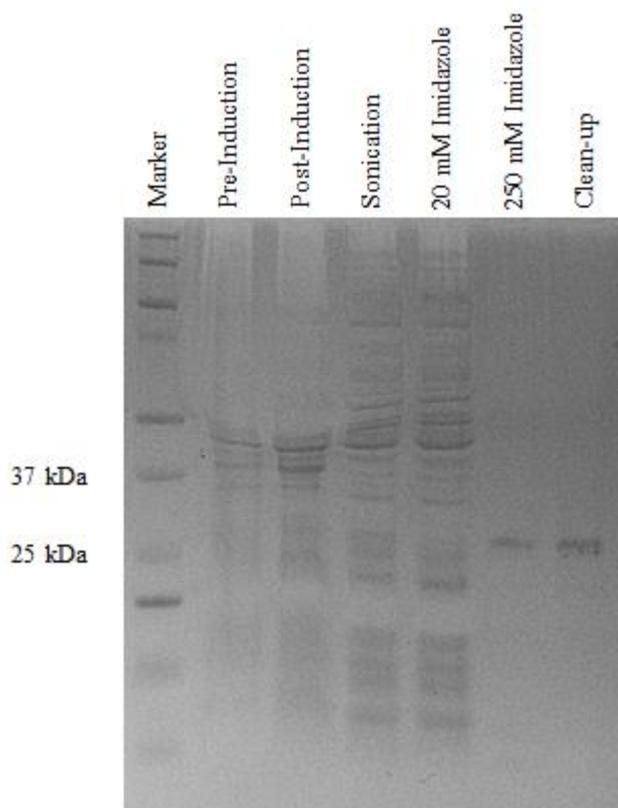


Figure 5.3 SDS-PAGE showing the purification of *hCDG*. Samples were taken at different purification stages to assess enzyme expression and purity.

5.3.4 Excision Activity Determination

To determine the substrate preference of *hCDG*, initial cleavage assays were performed by incubating *hCDG* with a variety of substrates for 24 hours and analysing these by denaturing PAGE. The results are presented in Figure 5.4. As observed with *eCYDG*, *hCDG* retains activity towards the U-containing substrates (G.U, A.U and AP.U), which show the expected single cleavage product. As expected, no activity is seen against T (G.T and Z.T). The enzyme also cleaves the C-containing substrates A.C, AP.C and Z.C, but shows no activity against G.C. As anticipated, the enzyme shows no activity against any of the substrates containing ^MC. There is also some activity against the single stranded substrate (ssC: see below for ssC(GAT)), though only about 50% of this DNA is cleaved within the 24 hour incubation. Once again, no cleavage is seen with the single-strand substrate containing ^MC.

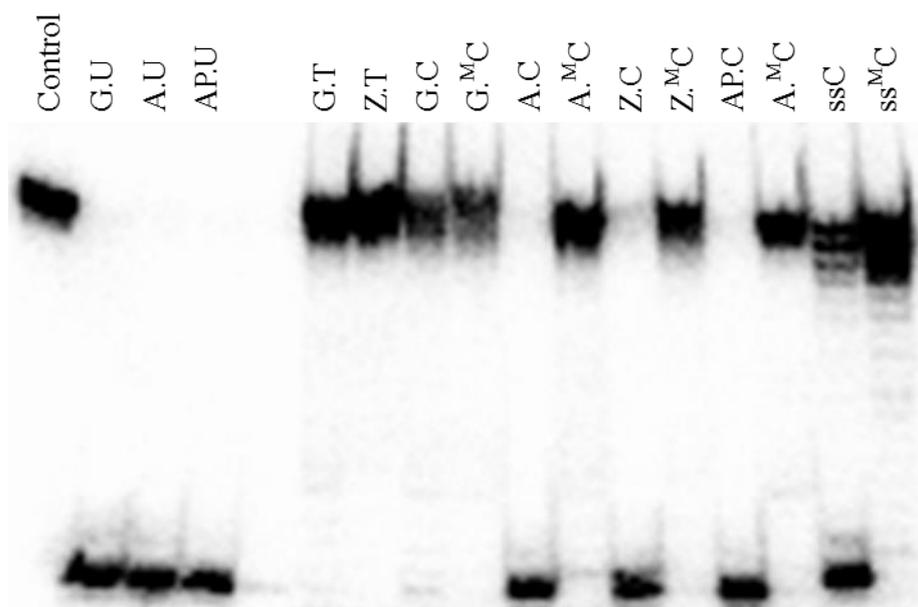


Figure 5.4 *hCDG* excision of uracil, thymine, cytosine and 5-methylcytosine substrates. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *hCDG* for 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The top and bottom bands correspond to the uncleaved 31mer or 15mer cleaved product respectively. AP; abasic site, Z; anthraquinone pyrrolidine.

Although *hCDG* showed no activity against the substrate with a central G.C base pair, we also examined its activity against a substrate containing a central G.C base pair that is flanked by blocks of A and T (sequence G.C(AT); Table 5.1) and the results are shown in Figure 5.5, and can be compared with equivalent experiments with *eCYDG* (Figure 3.20). Surprisingly it can be seen that *hCDG* is able to cleave the central G.C base pair in this sequence, albeit with low efficiency, though again no activity was observed with $^{\text{M}}\text{C.G}$.

Oligonucleotide	Sequence
G.C	5' -CCGAATCAGTGC GCA CAGTCGGTATTTAGCC-3' 3' -GGCTTAGTCACGCGT G TCAGCCATAAATCGG-5'
G.C(AT)	5' -CGAATAATTATATA A CATATATATATTTAGC-3' 3' -GCTTATTAATATATT G TATATATATAAATCG-5'

Table 5.1 G.C oligonucleotides used in excision assays. Central G.C base pair highlighted in bold.



Figure 5.5 *hCDG* excision of G.C^MC. The ³²P labelled 31mer duplex substrates (~50 nM) were incubated with ~1.25 μM *hCDG* for 24 hours followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. Top and bottom bands correspond to a 31mer unexcised or 15mer excised product respectively.

5.3.5 Rate of Reaction values for *hCDG* with different DNA substrates

To assess the best substrate for *hCDG* for discriminating between C and ^MC, we performed a series of cleavage assays to determine the rate of reaction against a range of different substrates (similar to the experiments with *eCYDG* that are described in Chapter 3). Again, the rate refers to the initial velocity of the reaction by assumption that all substrate is bound by enzyme. The results are summarised in Table 5.2. Figure 5.6 shows the cleavage profiles of substrates containing central A.C, AP.C and Z.C pairs. The rates with A.C ($0.165 \pm 0.015 \text{ min}^{-1}$) and AP.C ($0.038 \pm 0.004 \text{ min}^{-1}$) were faster than *eCYDG* by approximately 28 and 3-fold respectively, demonstrating that *hCDG* has greater activity. However, the rate at Z.C was $0.0051 \pm 0.0008 \text{ min}^{-1}$, which is approximately 20-fold lower than *eCYDG*. This was surprising since this substrate produced one of the fastest rates with *eCYDG*. The rate for Z.C was calculated from a single time point at 2 hours, assuming a simple exponential.

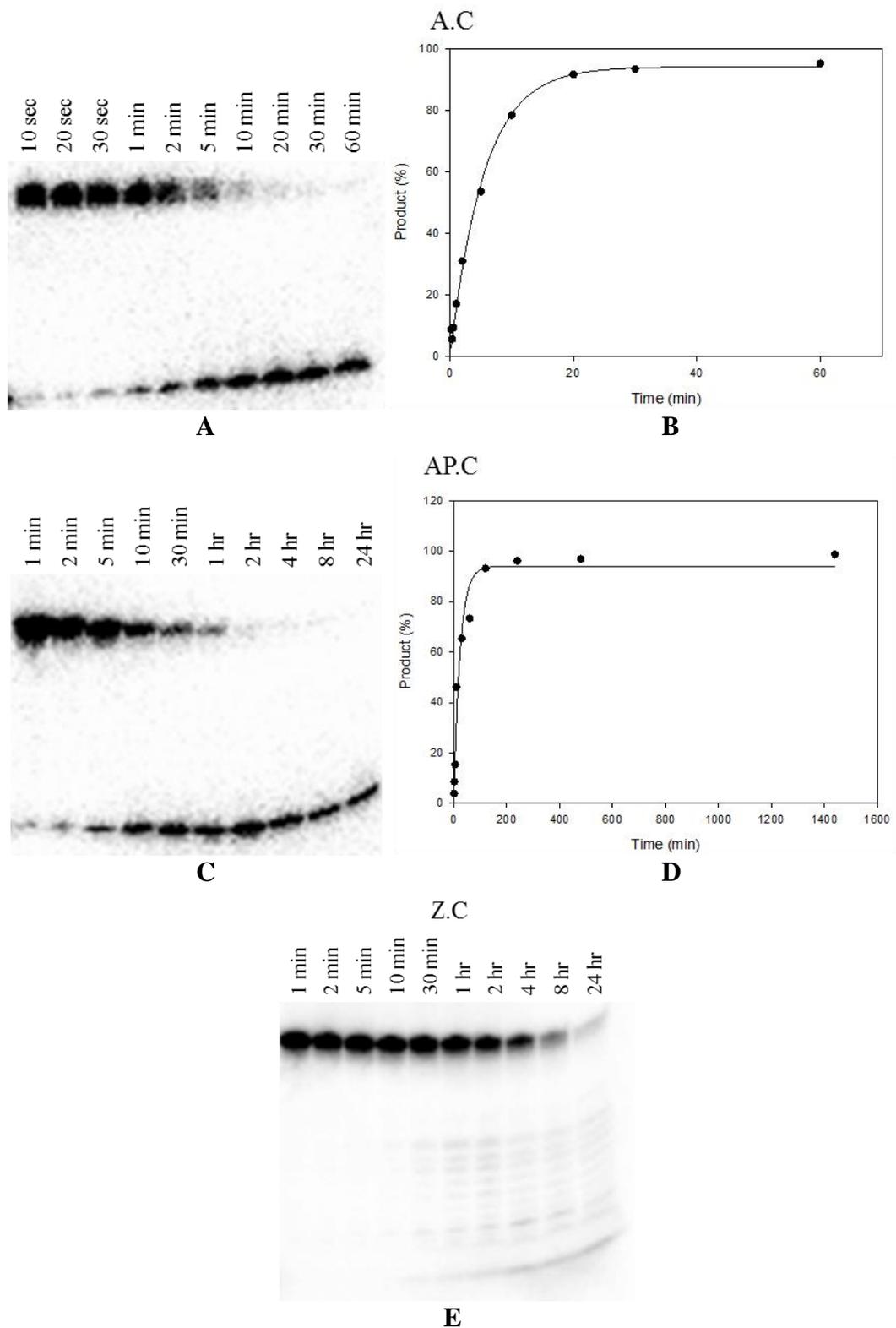


Figure 5.6 Kinetics of *h*CDG cleavage of the 31mer substrates A.C, AP.C and Z.C. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *h*CDG for up to 24 hours and followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, C and E). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B and D). AP: abasic site, Z: anthraquinone pyrrolidine.

<i>h</i> CDG	Rate of Reaction (min ⁻¹)	Relative Activity (%)
G.U	3.8 ± 0.7	100
A.U	1.27 ± 0.08	33
HEG.C	0.26 ± 0.02	6.8
A.C	0.17 ± 0.02	4.3
AP.C	0.04 ± 0.004	0.99
long gap.C	0.03 ± 0.007	0.67
ssC(polyA)	0.03 ± 0.005	0.78
gap.C	0.02 ± 0.003	0.58
ssC(GAT)	0.02 ± 0.002	0.46
A.C pH 7.4	0.02 ± 0.002	0.43
A(T).C(G)	0.006 ± 0.001	0.15
Z.C ²	0.005 ± 0.001	0.13
A.C(G) ¹	0.0012	0.03
G.C(AT) ²	0.0006	0.09
I.C ³	0.0004	0.01

Table 5.2 *h*CDG reaction rates. Relative activity is compared to the G.U substrate. The rate of reaction was determined as an average of three experiments. Rate values were estimated from a single time point at 8 hrs¹, 2 hrs² and 4 hrs³ assuming a simple exponential. Z, anthraquinone pyrrolidine; AP, abasic site; I, inosine; gap, unpaired C; long gap, unpaired C in a 41mer duplex; HEG, hexaethylene glycol; ND, not detectable.

Figure 5.7 shows the activity of *h*CDG against G.U and A.U which are cut with rates of 3.8 ± 0.7 min⁻¹ and 1.27 ± 0.08 min⁻¹ respectively, which are approximately 10- and 6-fold faster than with *e*CYDG. This Figure also shows the results for *h*CDG cleavage of HEG.C, which is cut with a rate of 0.26 ± 0.02 min⁻¹. HEG.C is the best C-containing substrate for *h*CDG, as with *e*CYDG. However *e*CYDG cut Z.C and HEG.C at similar rates, while *h*CDG has only very weak activity against Z.C. This suggests that the low rate with Z.C is a combination of the properties of both the mutant enzyme and the substrate. It is also worth noting that *e*CYDG cut Z.C and HEG.C faster than A.U, while the activity of *h*CDG at HEG.C is only about one fifth of that at A.U.

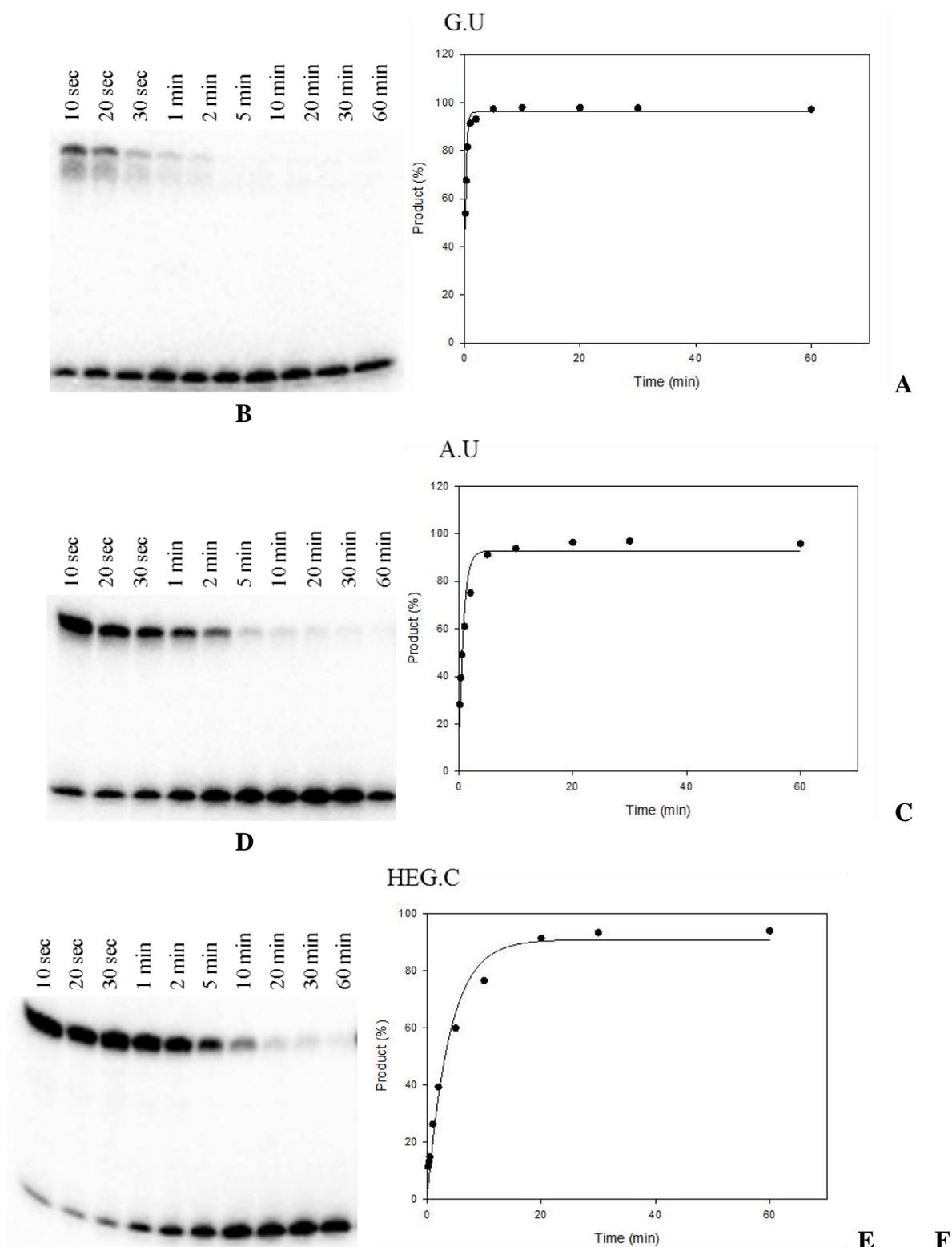


Figure 5.7 Kinetics of *hCDG* cleavage of the 31mer substrates G.U, A.U and HEG.C. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *hCDG* for up to 1 hour, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, C and E). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B, D and F). HEG: hexaethylene glycol.

5.3.5.1 Excision of ssC substrates with *hCDG*

The activity of *h*CDG against single-stranded substrates was also investigated (Figure 5.8); *e*CYDG had weak activity towards ssC substrates. These were both excised to completion; the rate with ssC(polyA) ($0.03 \pm 0.005 \text{ min}^{-1}$) was nearly twice as fast as ssC(GAT) ($0.02 \pm 0.002 \text{ min}^{-1}$); showing how local sequence context can affect the rate of excision, even for these single stranded substrates. Additional bands can be seen at the longer time points for both substrates.

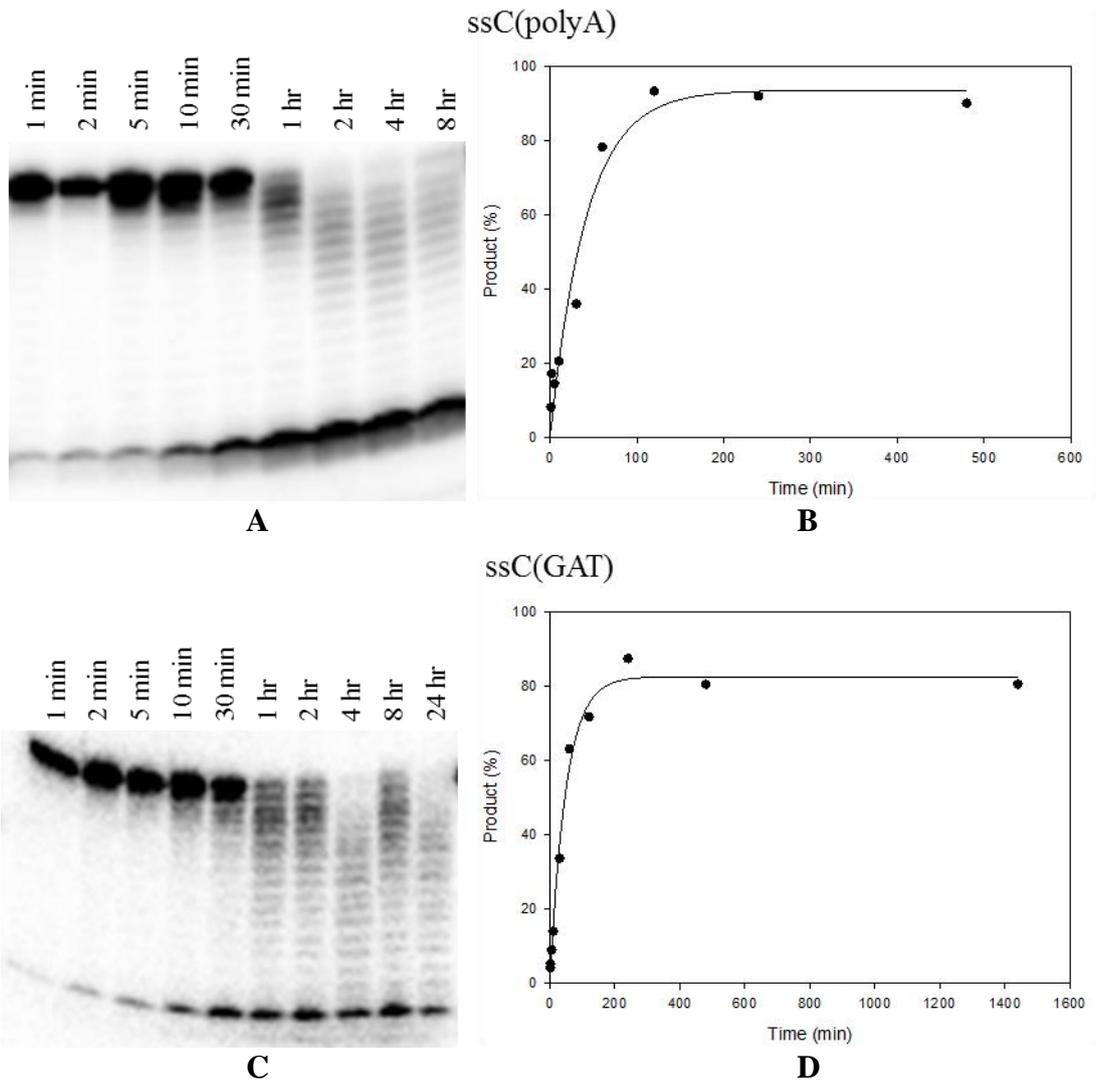


Figure 5.8 Kinetics of *h*CDG cleavage of the 31mer substrates ssC(polyA) and ssC(GAT). The ^{32}P labelled 31mer duplex substrates ($\sim 50 \text{ nM}$) were incubated with $\sim 1.25 \mu\text{M}$ *h*CDG for up to 1 hour, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A and C). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B and D).

These are clearly not specific cleavage products and are most likely generated by some contaminating 3'-exonuclease activity (Hodskinson *et al.*, 2007), which is able to act on ssDNA rather than dsDNA substrates.

5.3.5.2 Examining the Effect of Flanking Regions on Excision

As previously noted (Figure 5.5) *hCDG* is able to excise C from a G.C base pair when surrounded by long blocks of A.T base pairs (sequence G.C(AT)). We therefore further examined the activity of *hCDG* against this substrate and a number of other related substrates, assessing the effect of flanking base pairs. G.C(AT) produced a rate of $0.0006 \pm 0.0001 \text{ min}^{-1}$ (Figure 5.9 and Table 5.2), though a maximal cleavage of only 35% was obtained. The rate against A.C(G) (a central A.C mismatch flanked by G.C base pairs) was $0.0012 \pm 0.003 \text{ min}^{-1}$. While this is slow (and only about 40% of the substrate is cleaved) it is twice as fast as against G.C(AT) and is very different to the result with *eCYDG*, which showed no activity against this substrate. The ability of *hCDG* to cleave at I.C was also tested, as this base pair only contains two hydrogens bonds, like A.U. A similar slow rate, compared to G.C(AT), of $0.0006 \pm 0.0001 \text{ min}^{-1}$ was observed and the cleavage only reached approximately 20%. These results suggest that the loss of a hydrogen bond between a G.C and I.C base pair does not affect the rate of excision.

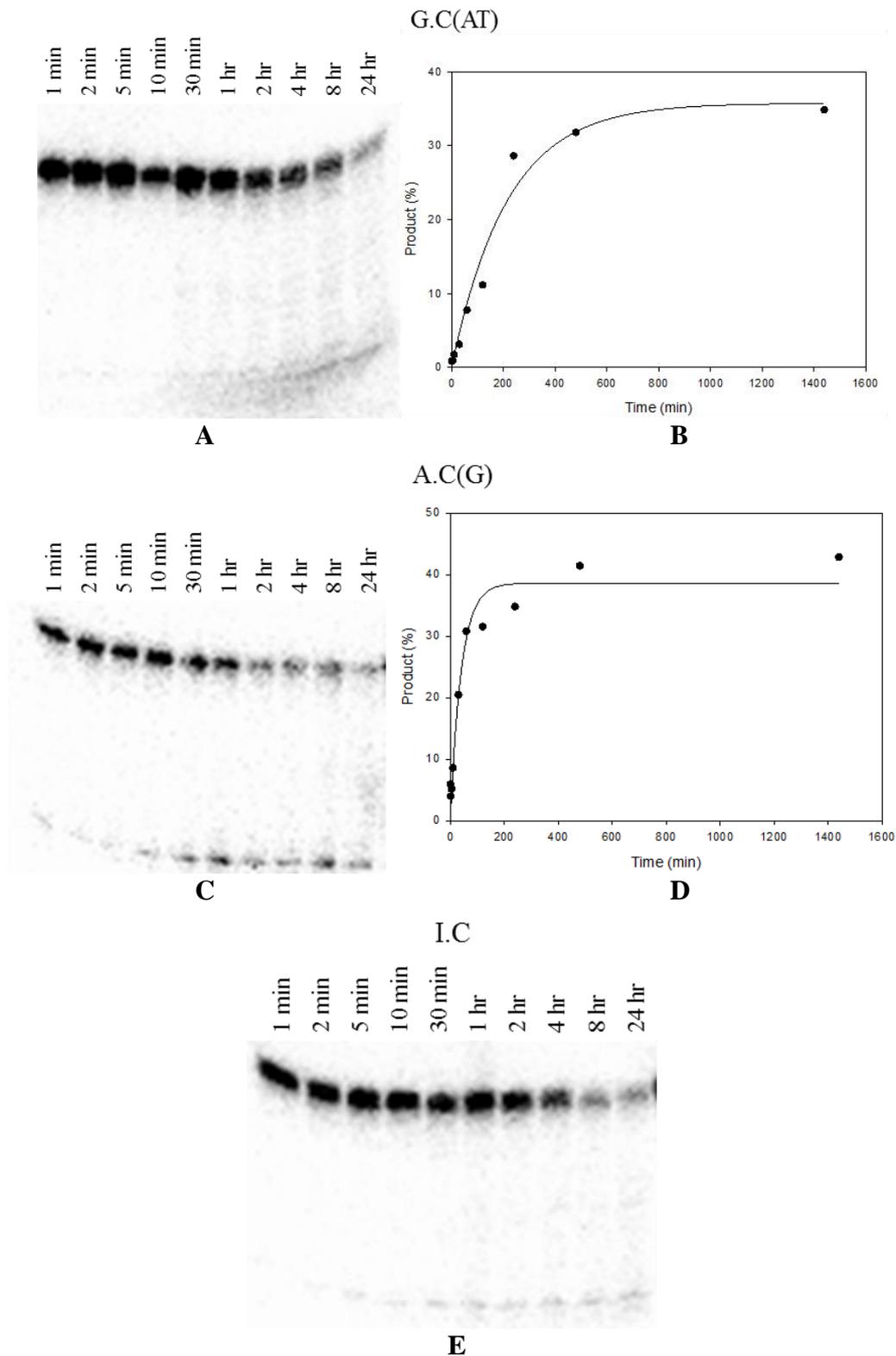


Figure 5.9 Kinetics of *h*CDG cleavage of the 31mer substrates G.C(AT), A.C(G) and I.C. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *h*CDG for up to 1 hour, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (**A**, **C** and **E**). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (**B** and **D**). I: inosine.

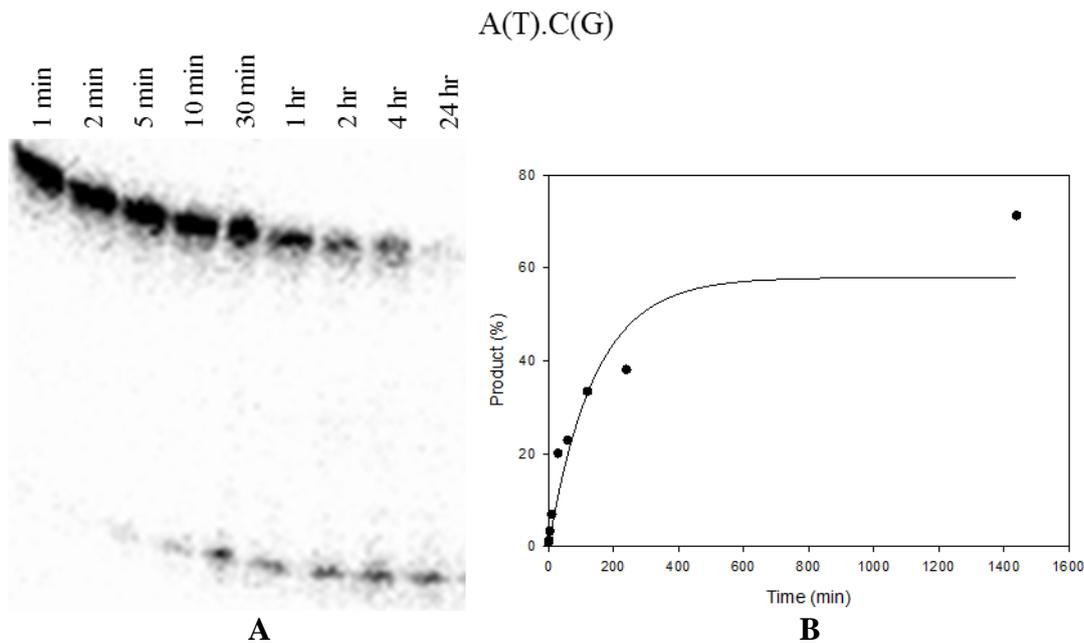


Figure 5.10 Kinetics of *h*CDG cleavage of the 31mer substrate A(T).C(G). The ^{32}P labelled 31mer duplex substrate (~ 50 nM) was incubated with ~ 1.25 μM *h*CDG for up to 1 hour followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging (A). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graph is derived from phosphorimage analysis of the gel and show the rate of formation of the cleavage product. This is fitted with a single exponential curve (B).

*h*CDG's greater activity led us to investigate whether it would be able to excise C more efficiently in a context in which the surrounding base pairs are GT mismatches using sequence A(T).C(G) (see Table 3.4) and the results are shown in Figure 5.10. With this sequence the rate of cleavage increased to $0.006 \pm 0.001 \text{ min}^{-1}$ and the maximum cleavage observed improved to 70%, suggesting that the destabilising properties of the G.T mismatches increased the excision efficiency (also seen with *e*CYDG). The targets with an unpaired central C (gap.C and long gap.C) produced similar results to *e*CYDG with maximal cleavage of 60 and 75%, though their rates were significantly faster ($0.022 \pm 0.003 \text{ min}^{-1}$ and $0.026 \pm 0.007 \text{ min}^{-1}$ respectively (Figure 5.11)). Interestingly a second longer but weaker cleavage product can be seen with gap.C (indicated by \blacktriangle in the cleavage plot (Figure 5.11B)). This probably corresponds to the region with four Cs on the 3'-side of the expected target C. Since these oligonucleotides have T_{ms} that are close to the reaction temperature it is possible that these regions temporarily melt and the additional *h*CDG cleavage products correspond to cleavage of Cs in this transient single stranded region. The rate calculated for the primary cleavage product with gap.C was $0.023 \pm 0.003 \text{ min}^{-1}$; \bullet as this was equivalent to the rate of the total product ($0.022 \pm 0.003 \text{ min}^{-1}$; \blacksquare).

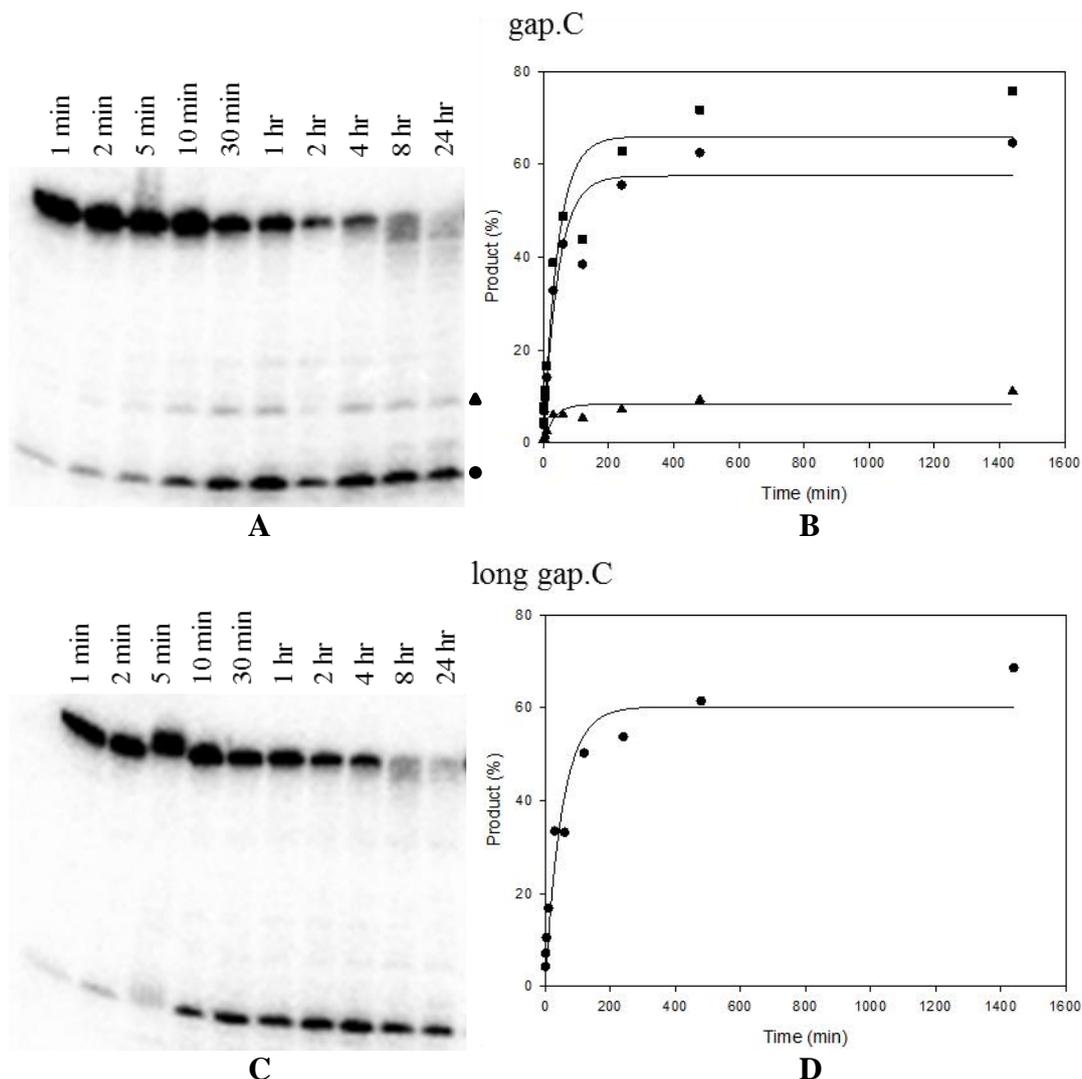


Figure 5.11 Kinetics of *h*CDG cleavage of the 31mer substrates gap.C and long gap.C. The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *h*CDG for up to 1 hour, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A and C). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B and D). For the gap.C graph; ● bottom/primary band, ■ middle band, ▲ top/total band. gap: unpaired C, long gap: unpaired C in a 41mer duplex.

5.3.5.3 The Effect of pH on A.C excision

Since *e*CYDG showed an unexpected increase in the rate of excision at higher pH we also investigate the effect of pH on the activity of *h*CDG. Unlike *e*CYDG, *h*CDG showed a decrease in the rate of excision at A.C (0.016 ± 0.002 min^{-1}) at pH 7.4 (Figure 5.12A), as would be expected, compared to A.C performed at pH 6.3 (the standard pH used for all other cleavage assays). This is different to the results with *e*CYDG (Section 3.3.12), but is consistent with (Kwon *et al.*, 2003), and is as would be expected from the protonation of both the aspartates in the active site.

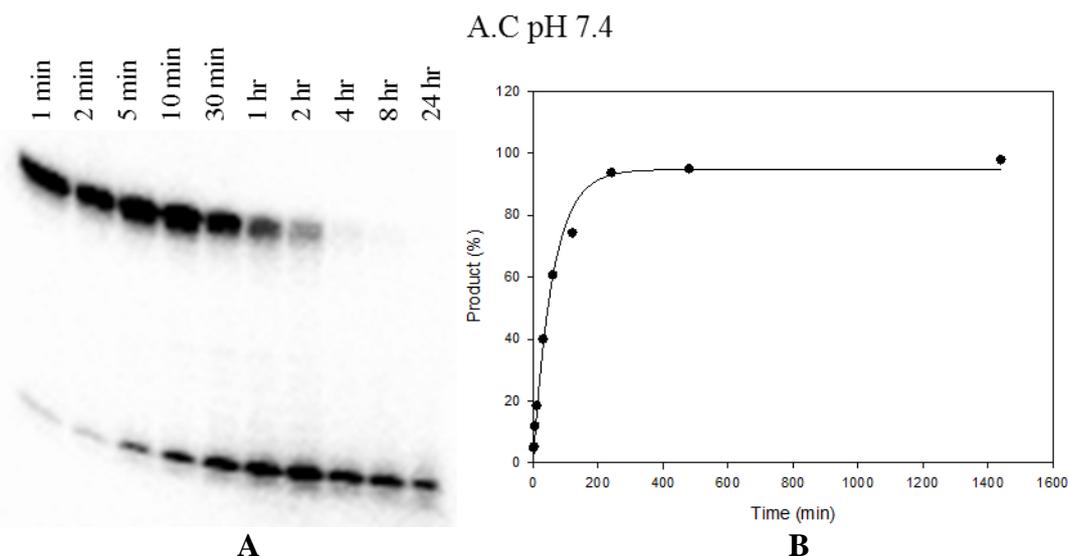


Figure 5.12 Kinetics of *hCDG* cleavage of the 31mer substrate A.C pH 7.4. The ^{32}P labelled 31mer duplex substrate (~ 50 nM) was incubated with ~ 1.25 μM *hCDG* for up to 1 hour, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging (A). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graph is derived from phosphorimage analysis of the gel and show the rate of formation of the cleavage product. This is fitted with a single exponential curve (B).

5.3.6 Rate of Reaction Determination of *hCYDG*

Since Chapter 3 examined the properties of the double mutant of the *E. coli* enzyme (*eCYDG*) we also examined the properties of the equivalent human enzyme (*hCYDG*) against a range of different substrates. The results are summarised in Table 5.3.

<i>hCYDG</i>	Rate of Reaction (min^{-1})	Relative Activity (%)
G.U	0.51 ± 0.01	100.00
Z.C	0.02 ± 0.005	3.6
HEG.C	0.013 ± 0.002	26
AC	0.013 ± 0.001	2.6
AP.C ¹	0.0001	0.7
A.C(G)	ND	0
ssC(GAT)	ND	
G.C(AT)	ND	

Table 5.3 *hCYDG* reaction rates. Relative activity is in relation to the G.U substrate. The rate of reaction was determined as an average of three rates. ¹The rate value was estimated from a single time point at 24 hrs assuming a simple exponential. Z; anthraquinone pyrrolidine, AP; abasic site, HEG; hexaethylene glycol, ND; not detectable.

Figure 5.13 shows the cleavage of A.C by *hCYDG* from which a rate constant of 0.013 ± 0.001 min^{-1} was determined. This is about 10-fold lower than *hCDG* with the substrate

confirming that, as expected, the double mutant *h*CYDG has a much lower activity than *h*CDG. This is also about 3-fold slower than the rate of *e*CYDG cleavage of this substrate; though it should be noted that only 30% of the substrate was cleaved.

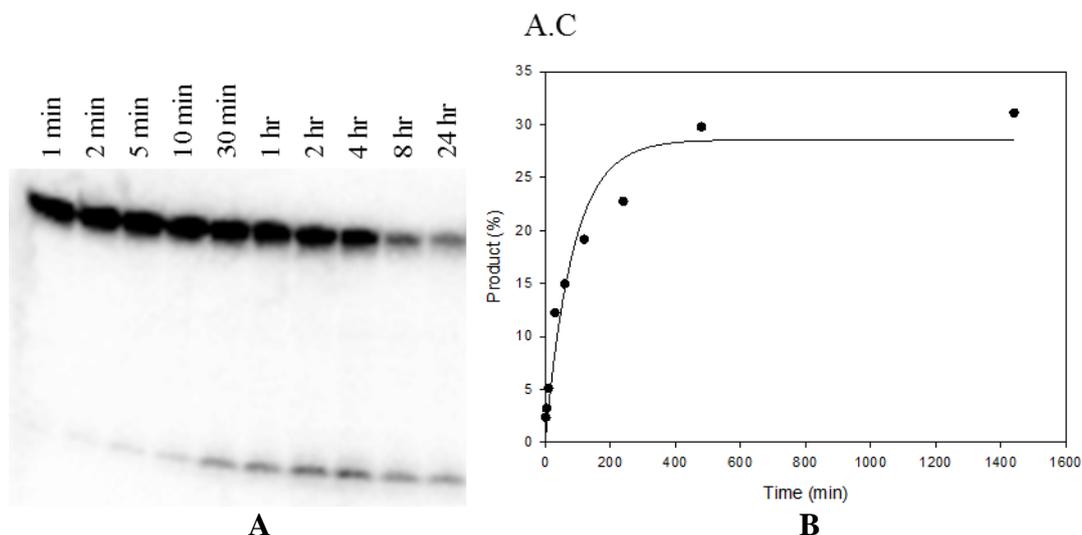


Figure 5.13 Kinetics of *h*CYDG cleavage of the 31mer substrate A.C. The ^{32}P labelled 31mer duplex substrate (~ 50 nM) was incubated with ~ 1.25 μM *h*CDG for up to 24 hours, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging (A). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graph is derived from phosphorimage analysis of the gel and show the rate of formation of the cleavage product. This is fitted with a single exponential curve (B).

Figure 5.14 shows the cleavage of Z.C, HEG.C and AP.C by *h*CYDG. Z.C (0.02 ± 0.005 min^{-1}) and HEG.C (0.013 ± 0.002 min^{-1}) are cut at comparable rates, though this is much slower than with *e*CYDG, while cleavage of AP.C was barely detectable and a rate (0.0001 min^{-1}) could only be estimated from a single time point at 24 hours. No detectable activity was observed with ssC(GAT) or the more stable duplex substrates G.C(AT) or A.C(G) (Figure 5.15). Though *h*CYDG has the weakest activity against C out of the three enzymes investigated, it showed greater activity against U than *e*CYDG with a rate of 0.5 ± 0.01 min^{-1} (Figure 5.16).

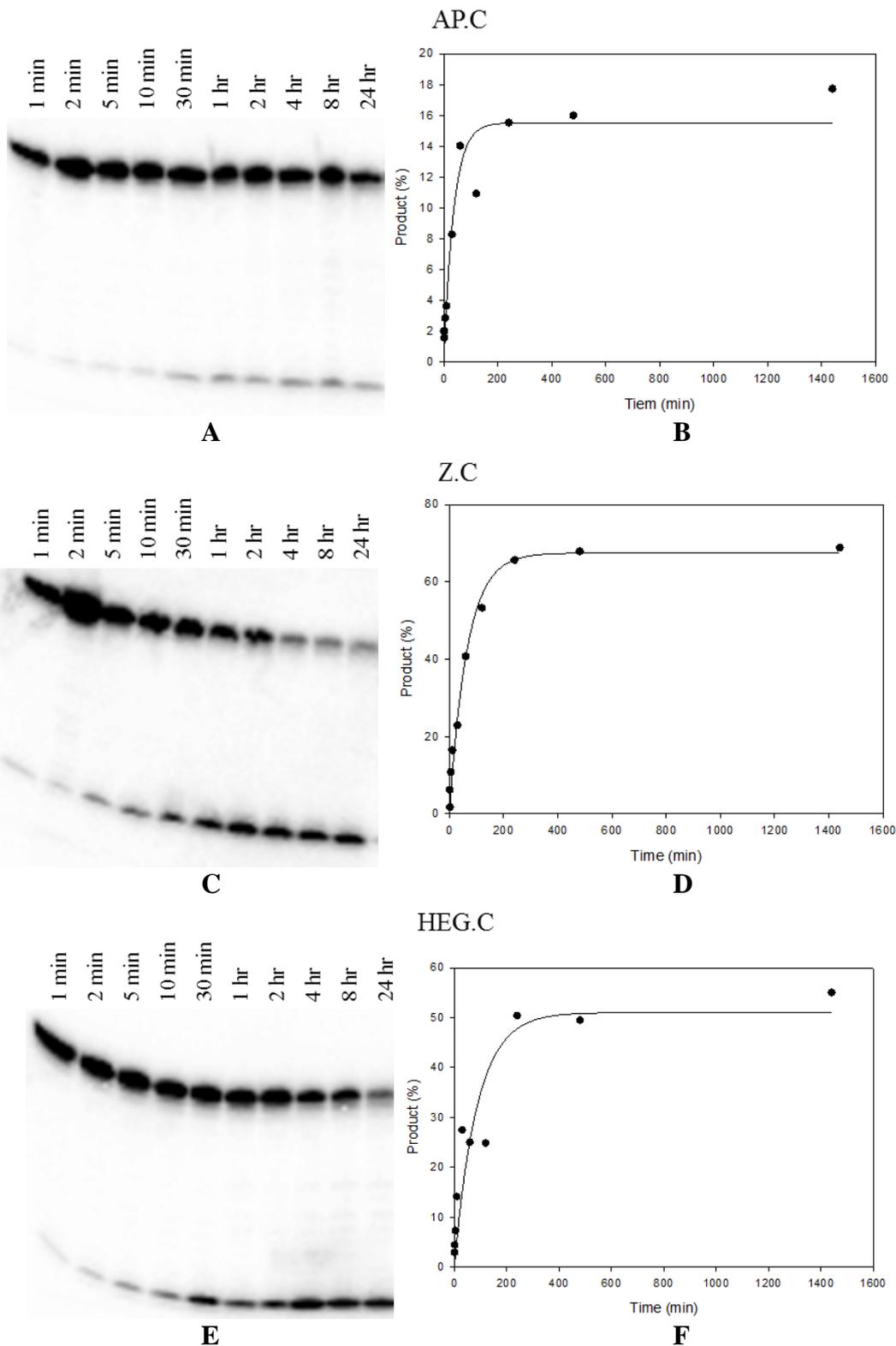


Figure 5.14 Kinetics of *h*CYDG cleavage of the 31mer substrates AP.C, Z.C and HEG.C. The ³²P labelled 31mer duplex substrates (~50 nM) were incubated with ~1.25 μM *h*CDG for up to 24 hours, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, C and E). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves (B, D and F). AP: abasic site, Z: anthraquinone pyrrolidine, HEG: hexaethylene glycol.

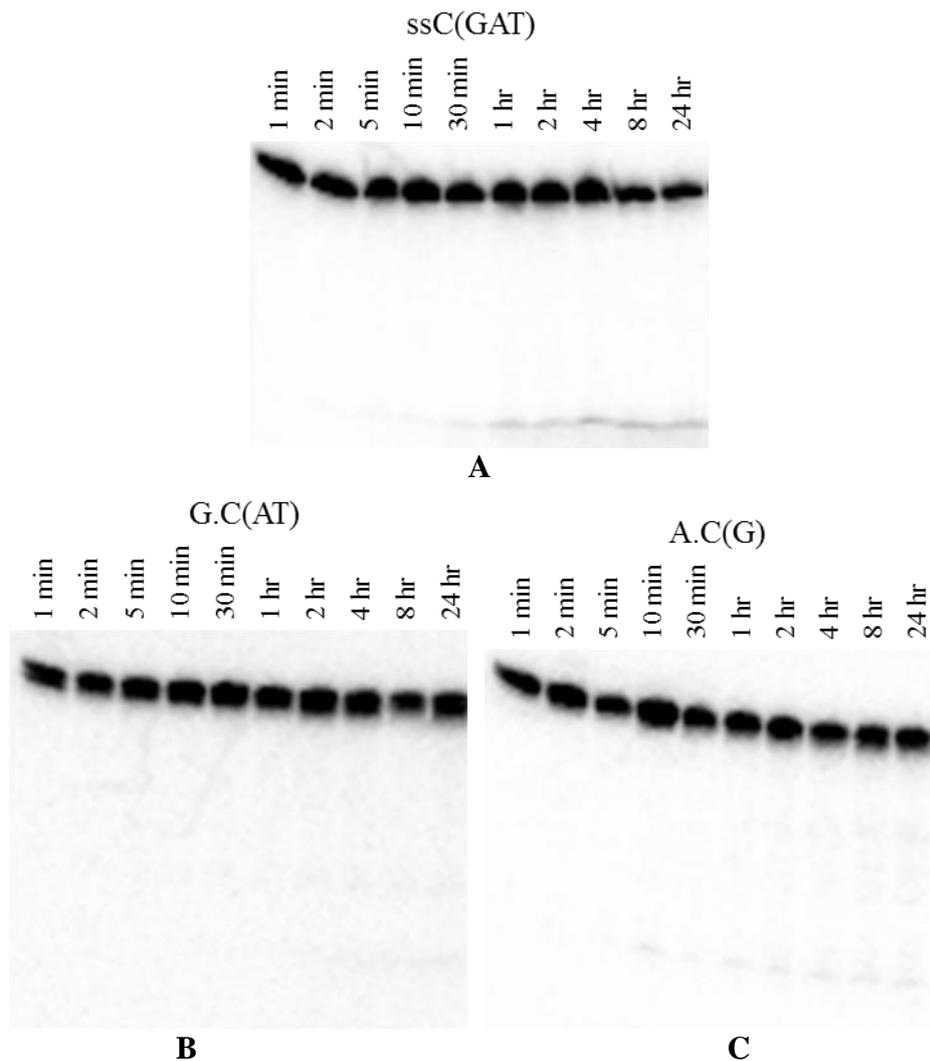


Figure 5.15 Kinetics of *h*CYDG cleavage of the 31mer substrates ssC(GAT), G.C and A.C(G). The ^{32}P labelled 31mer duplex substrates (~ 50 nM) were incubated with ~ 1.25 μM *h*CDG for up to 24 hours, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on 12.5% denaturing polyacrylamide gels and analysed by phosphorimaging (A, B and C). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively.

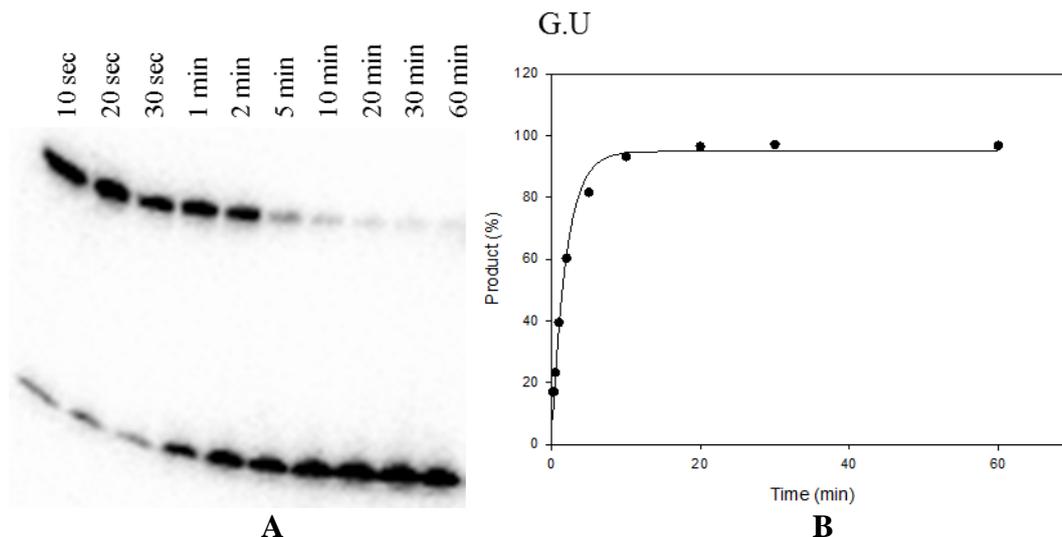


Figure 5.16 Kinetics of *hCYDG* cleavage of the 31mer substrate G.U. The ^{32}P labelled 31mer duplex substrate (~ 50 nM) was incubated with ~ 1.25 μM *hCDG* for up to 24 hours, followed by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging (A). Top and bottom bands correspond to a 31mer uncleaved or 15mer cleaved product respectively. The graph is derived from phosphorimage analysis of the gel and show the rate of formation of the cleavage product. This is fitted with a single exponential curve (B).

5.4 Discussion

By using complete gene synthesis we have generated two functional human glycosylases (*hCDG* and *hCYDG*) that are able to excise cytosine and show no activity towards 5-methylcytosine. As with *eCYDG*, they are most efficient when the cytosine is in an unstable base pair (*i.e.* not paired with guanine). The ability to express *hCDG* in *E. coli* suggests that it is not cytotoxic and therefore only has weak activity. This could be explained on account of the asparagine to aspartate mutation, which compromises the activity of the human variant more than that in the *E. coli* enzyme. Despite this expectation, *hCDG* still showed significant activity with a 25-fold and 10-fold greater activity for A.C and G.U compared to *eCYDG*. A possible explanation for the lack of cytotoxicity of *hCDG* could be that it is unable to recruit other *E. coli* proteins in the BER pathway, while the equivalent *E. coli* protein will be able to do so as it is expressed in its normal host organism. It is not surprising that *hCYDG* has weaker activity than *hCDG* as it is a double mutant. *hCYDG* has the weakest activity of the three enzymes investigated and suggest that the human variants have weaker intrinsic activity than their *E. coli* counterparts. These results, along with the cytotoxicity of *eCDG*, strongly suggests that if an *eCDG* could be expressed it would be the most active enzyme: $eCDG > hCDG > eCYDG > hCYDG$. This is supported by the observation that *hCDG* has greater activity against cytosine than *eCYDG* (*i.e.* at A.C and AP.C). This suggests that an *eCDG* might have greater activity

than *h*CDG, and could be cytotoxic through excision of G.C base pairs within the *E. coli* DNA, for which *h*CDG has very low activity (and *e*CYDG has none).

As G.C activity was not seen initially, but was seen when put into the context of flanking A and T bases (G.C(AT)), strongly suggests that the G.C duplex was being cleaved into smaller fragments due to excision at other G.C sites within the duplex. Alternatively, binding of the enzyme to other G.C sites could prevent binding and therefore cleavage at the central G.C, reducing observed activity. Therefore the ability of *h*CDG to excise cytosine from a G.C base pair, albeit at a low efficiency, supports the suggestion that the role of leucine 191 is to “plug” rather than to “push”. This is because the leucine, which is still present in *h*CDG, is unable to force the cytosine into an extrahelical conformation. This is energetically less likely at a G.C pair, as a result of its greater stability compared to A.U. However, *h*CDG has weak activity towards I.C, even though this base pair only contains two hydrogen bonds, as with A.U, compared to three with G.C. This suggests firstly that *h*CDG has greater affinity towards uracil than cytosine, due to the greater activity at A.U than I.C. Secondly this suggests that base stacking plays an important role in base flipping, and it is not merely dependent on the stability of the base pair itself (SantaLucia *et al.*, 1996, Watkins and SantaLucia, 2005). This is consistent with the observation that the rate at I.C is similar to that of A.C(G) and G.C(AT). This is further supported by the increased rate of cleavage at A(T).C(G), in which the flanking guanines are mismatched with thymine. This suggests that the G.T mismatches not only destabilise the flanking base pairs but also affect the base stacking due to their “wobble” base pairs. The role of leucine as a “plugger” is also supported by the greater ssC activity of *h*CDG compared to *e*CYDG, suggesting that the leucine is able to hold the ssDNA in the enzyme’s active site for longer. This allows for the rotation of the aspartate so that the correct hydrogen bonds are formed with the base that is to be excised (as previously explained in the discussion of Chapter 3). The increase in ssC activity is also evident with gap.C, for which a second product band is observed, and which may arise from local duplex melting around the gap, exposing a 3' ssC region that exposes a cytosine four bases 3' of the target cytosine, allowing its excision.

As with *e*CYDG, excision of cytosine by *h*CDG is most efficient when it is placed opposite the hexaethylene glycol linker, presumably because this easily allows the cytosine to become extrahelical. Although *e*CYDG cuts Z.C and HEG.C at comparable rates Z.C is a very poor substrate for *h*CDG. This may be due to steric clash between leucine and the

anthraquinone when leucine intercalates into the duplex DNA, occupying the space left by the extrahelical base. This cannot happen when an anthraquinone is present as it occupies the space of both base pairs, as with pyrene (Jiang *et al.*, 2001), causing steric hindrance and preventing enzyme binding and therefore hindering base docking and excision. This is supported by the much greater activity of *hCYDG* at Z.C, compared to *hCDG*, though this still occurs at a slow rate and does not go to completion within a 24 hour period. In contrast to the other enzymes *hCYDG* appears to have very little activity at AP.C. It is not clear why this occurs as both *eCYDG* and *hCDG* show AP.C activity. Although HEG and AP are similar, in that they do not form a base pair with the target cytosine, AP is shorter by ~18 Å and still contains a ribose sugar that makes it more rigid and would likely produce greater local stability that is only noticeable with the weak *hCYDG* enzyme. Though the A.C pair has one hydrogen bond and would be expected to be more stable than AP.C, it seems to have a greater effect on the cytosine base stacking interactions, allowing it to become extrahelical and thus providing A.C activity. This also suggests that base stacking is important in determining the rate of excision.

In contrast to *eCYDG*, *hCDG* shows the expected decrease in rate at pH 7.4 compared to when the assay was performed at pH 6.3. This suggests that the human enzyme may be more susceptible to changes in pH. This is consistent with the notion that CDGs are pH dependent as a result of the protonation state of the critical aspartate residue, and that activity should decrease at pH 7.4 (Kwon *et al.*, 2003).

In summary *hCDG* is able to discriminate between cytosine and 5-methylcytosine and shows cytosine excision in any base pair combination, though the rate is greatly reduced in more stable duplex contexts, while showing no 5-methylcytosine in any context. The data for both *hCDG* and *hCYDG* provide further evidence that base stacking is a major player in determining variations in cleavage efficiency. In view of its greater activity *hCDG* might be a better enzyme than *eCYDG* for use in developing of an assay for detection of 5-methylcytosine (Chapter 6).

Chapter 6: Developing an Assay for the Detection of 5-methylcytosine

6.1 Introduction

The detection and quantification of 5-methylcytosine is important in epigenetics and for understanding its effects on gene regulation. Many different methods are now available for detecting the location of 5-methylcytosines (Shapiro *et al.*, 1973, Cedar *et al.*, 1979, McClelland *et al.*, 1994, Weber *et al.*, 2005, Tahiliani *et al.*, 2009, Manrao *et al.*, 2011), with the gold standard being bisulphite sequencing. However, bisulphite sequencing requires μg s of DNA as 80 – 95% degradation usually occurs during the reaction process. It is therefore difficult to investigate the methylation status of cytosines in single cells or in a small sample size. The average methylation state of each cytosine is therefore generated, preventing the detection of weakly or highly regulated/modified methylated sites. All of the methods have their disadvantages (as discussed in section 1.4) with another common one being the sequence dependency of the detection techniques. One example is the use of methylation sensitive restriction enzymes (Cedar *et al.*, 1979, McClelland *et al.*, 1994) which are necessarily limited to probe the recognition sequence of the methylation sensitive restriction enzyme(s), *e.g.* HpaII for CCGG. This has been useful for investigation of CpG methylation sites (the main sites for cytosine methylation relating to gene expression), though CpG islands are generally unmethylated. Cytosine methylation can also occur at other (non-CpG) sites, *i.e.* CpN (Grafstrom *et al.*, 1985, Ramsahoye *et al.*, 2000, Guo *et al.*, 2014a, Guo *et al.*, 2014b), which may be of unrealised importance in gene regulation, and therefore cannot be investigated by sequence dependant methods.

Here we propose a new method for detecting the presence of 5-methylcytosine that should enable its detection in both fully methylated or hemi-methylated DNA sites, as well as being applicable to non-CpG sequences. By utilising the abilities of CDGs to distinguish between cytosine and 5-methylcytosine, combined with real-time PCR, the methylation (or other modified) state of any cytosine can be determined, regardless of its sequence context. It is hoped that this could be developed into a simple detection assay.

6.2 Experimental Design

We propose using CDG to discriminate between cytosine and 5-methylcytosine with a simple assay in which the cleavage (at C) or no cleavage (at ^MC) of a target sequence is detected by real-time PCR. No PCR amplification will occur if the template DNA has been cleaved.

To test the viability of the proposed assay we designed an 80mer oligonucleotide containing a central cytosine or 5-methylcytosine in a C/^MCpG context, flanked by sequences that are similar to those in the A.C(AG) substrate that was used for determining the cleavage rates in Chapter 3. This synthetic template also contained sites for PCR primers at the 5' and 3' primer ends. This 80mer is hereafter known as the 'target' oligonucleotide. A complementary 20mer oligonucleotide (hereafter known as the 'probe' oligonucleotide) is annealed to the centre of this target, covering the cytosine to be probed, and generating a C.X mismatch at this location. This generates a dsDNA substrate for efficient cleavage by CDG if the target base is unmethylated cytosine. The sequences of the oligonucleotides are shown in Table 6.1.

Oligonucleotide	Sequence
Target	5' -TAGCGTAGGGAGATATAACCATGGGCAGCAGCCAGTGCGCAC/ ^M CG GTCGGTATTCTGAGATCCGGCTGCTAACAAAGTTGATCAT-3'
Probe	5' -GGCTAAATACCGACTHTGCGCACTGATTCGG-propyl-3'

Table 6.1 Oligonucleotides used in the cytosine detection assay. H; HEG.

The target-probe complex (with or without annealing) was incubated with *e*CYDG for 24 hours to allow cleavage to occur, followed by cleavage of the backbone with hot alkali or AP endonuclease. The resulting products were subjected to PCR; amplification should only occur if the substrate contained 5-methylcytosine, but not with cytosine at this position as the template will have been cleaved in two (Figure 6.1).

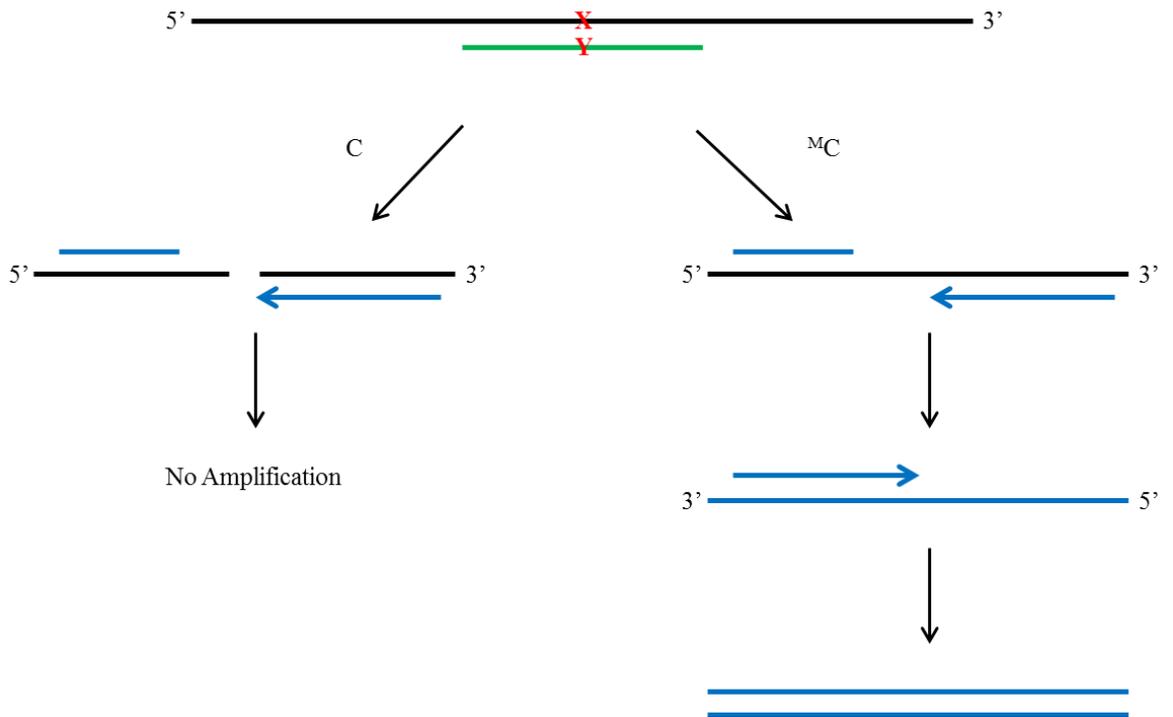


Figure 6.1 Proposed method for detecting 5-methylcytosine using *e*CYDG. The target DNA (black) and probe oligonucleotide (green) are reacted with *e*CYDG. The target base (X) will be excised if it is a cytosine, but not if it is 5-methylcytosine. The reaction products are then treated with Apurinic Endonuclease 1 (APE1) or alkali conditions in order to break the phosphodiester backbone at the abasic site. The products are then subjected to PCR using flanking primers (blue). A PCR product will only be generated if the target contained 5-methylcytosine, since the cleaved cytosine-containing substrate will not be amplified.

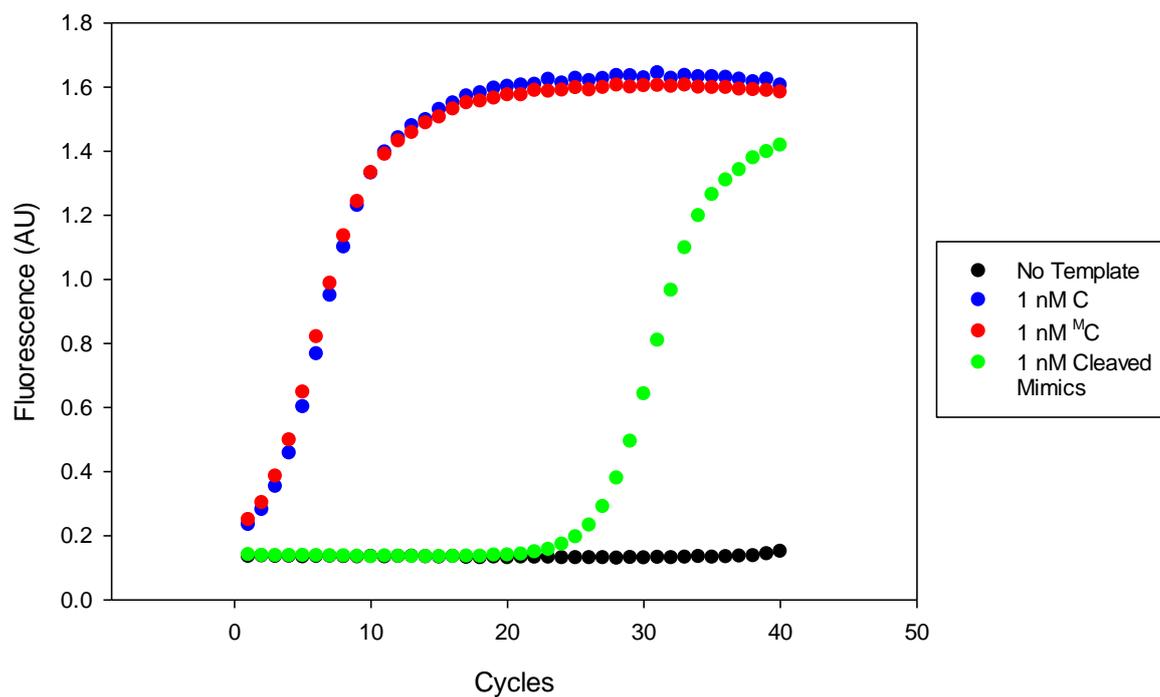


Figure 6.3 Real-time PCR analysis of a 1 nM C, ^MC, and a cleaved C mimic, target oligonucleotide with a 10 nM HEG probe.

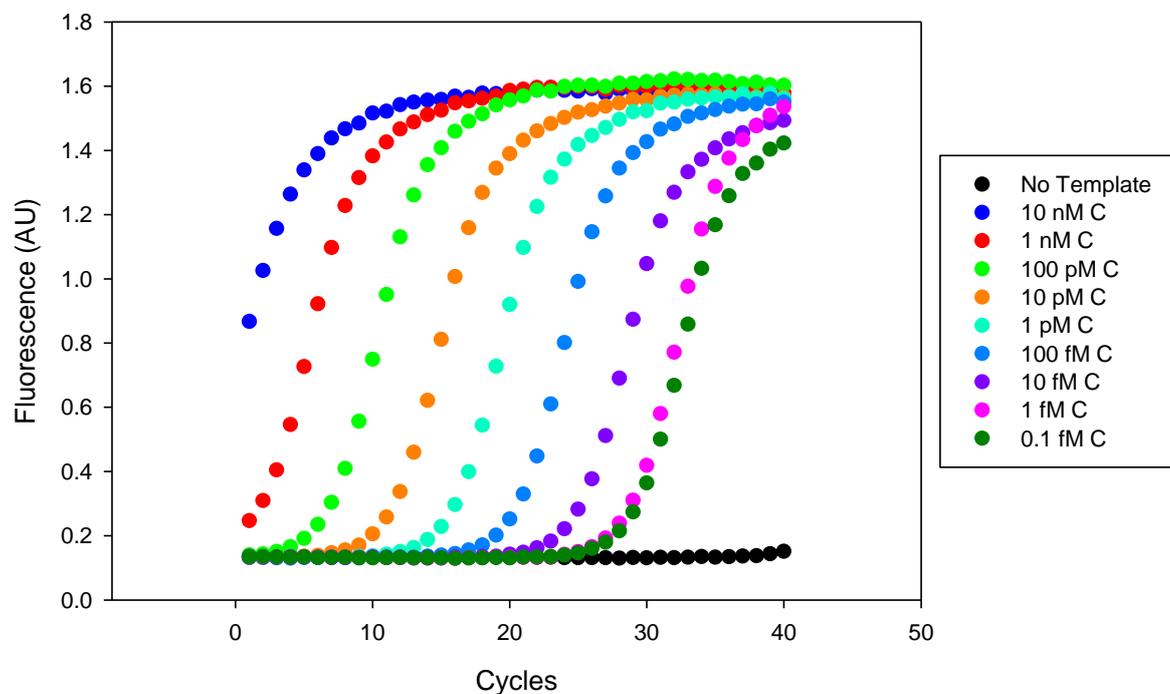


Figure 6.4 Real-time PCR concentration gradient analysis of 10 nM to 0.1 fM C.

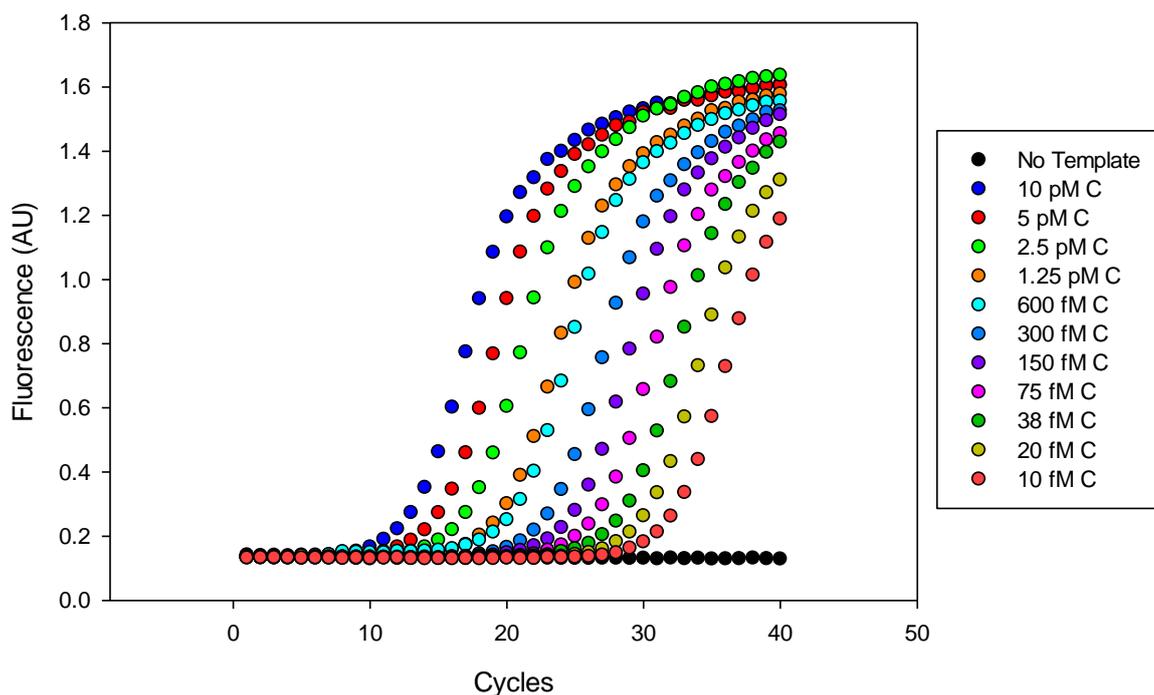


Figure 6.5 Real-time PCR concentration gradient analysis of 10 pM to 10 fM C.

Though the lower concentrations used in Figure 6.5 are slightly more right shifted than comparable concentrations in Figure 6.4 (e.g. 10 pM, 100/150 fM and 10 fM), there is still a consistent right shift for each 2-fold dilution, showing the assay's sensitivity to small differences in concentration.

6.3.2 *e*CYDG:PCR Assay

As there was discrimination between the synthetic cleaved mimics and the full length substrates, we proceeded to test the assay with cleavage by *e*CYDG. The template-probe mixture was incubated with *e*CYDG for 24 hours and then subjected to PCR. We expected amplification with the 5-methylcytosine-containing template, and at least a shift to the right with the cytosine-containing template, similar to the curves shown in Figure 6.3. However, initial experiments using piperidine to cleave at the basic site, showed no clear discrimination between amplification of the cytosine and 5-methylcytosine-containing templates. This was also attempted with the more reactive *h*CDG but still no discrimination was observed.

6.3.3 *e*CYDG Reaction Analysis

As the PCR showed no clear discrimination between the C and ^MC containing targets after reaction with *e*CYDG, we decided to analyse the cleavage reaction itself, using the standard radiolabelled cleavage assays (Chapters 3 and 5) with these substrates. These assays were also used to investigate the effect of annealing conditions on the reaction and to examine different methods of backbone cleavage at the basic site. As shown in Figure 6.7 *e*CYDG efficiently cleaves the cytosine-containing template and very little intact substrate is left after 24 hours incubation. The reaction does not depend on whether the probe and template are properly annealed (slow cooling from 95°C) or simply mixed. This suggests that under the reaction conditions with *e*CYDG, the target and probe are able to hybridise to allow for efficient C excision. Most importantly no ^MC activity was detected. We can assume that activity is due to cleavage of duplex DNA, rather than unannealed single strands as the previous cleavage assays (Figure 3.21) showed there was little or no ssC activity with *e*CYDG. Despite this, other cleavage products were generated for both the C and ^MC containing substrates (Figure 6.7) showing low ssC activity in the single stranded flanking regions of the duplex. This is most apparent with the 5-methylcytosine-containing substrate as the reactivity of cytosine-containing substrate only generates products of half length (from the position of the target C) or smaller. Products can also be seen within the duplex region at Cs (where the probe is annealed) suggesting G.C activity, but previous results discussed in Chapter 3 clearly show that *e*CYDG has no G.C activity. Therefore, these products must be due to ssC activity from a small amount of an unannealed or partially annealed probe:target substrate.

To investigate this further, the PCR assay was also designed so the primers could also act as protection oligonucleotides on either side of the probe to reduce the likelihood of any single-stranded activity (Table 6.2 and as illustrated in Figure 6.6).

Description	Sequence
For Primer	5' -GGAGATATACCATGGGCAGCAGC-3'
Rev Primer/Protection	5' -CTTTGTTAGCAGCCGGATCTCAG-3'
Rev Protection	5' -GCTGCTGCCCATGGTATATCTCC-3'

Table 6.2 Primers and protection oligonucleotides used in the cytosine detection assay.



Figure 6.6 Binding of protection oligonucleotides to the target DNA. The target DNA (black) hybridised with the probe oligonucleotide (green), protection oligonucleotide (red) and by the PCR primer (blue).

This should minimise the production of non-specific products, and provide the greatest difference between C and ^MC for discrimination by PCR. Figure 6.8 shows that single stranded activity, with *e*CYDG incubation, is reduced by the addition of the protection oligonucleotides; as with the previous results activity is seen with the cytosine-containing substrate (arrow) but not with the 5-methylcytosine-containing substrate.

Finally we investigated the effects of different backbone cleavage agents as shown in Figure 6.9. As expected piperidine, NaOH (10 mM) and the use of APE1 (1 µl; approx. 10 u/µl) all acted to cleave the phosphodiester backbone. However, this Figure also shows efficient cleavage after simply heating the enzyme cleavage products at 95°C (Figure 6.9; arrow). This was surprising as backbone cleavage is expected to require alkaline conditions and the sample is in a slightly acidic reaction buffer (pH 6.3). However, this would make any PCR-based assay simpler as this can simply be included as the first heating step of the reaction. Again, no activity was observed at the central 5-methylcytosine.

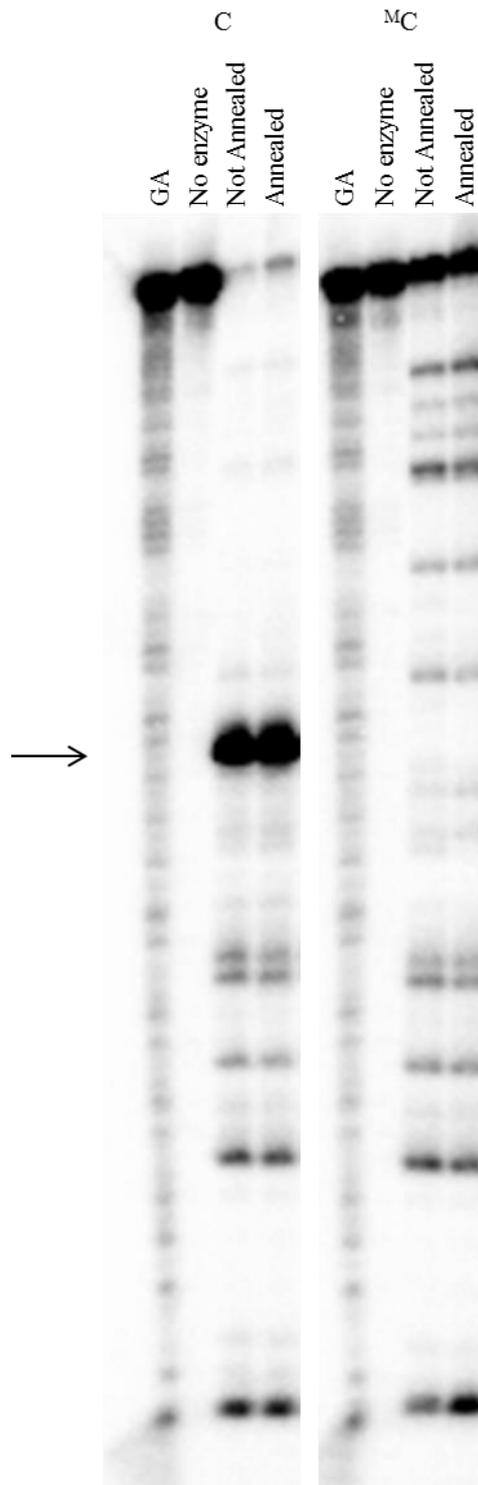


Figure 6.7 *e*CYDG cleavage of the C- and ^MC-containing substrates. The ³²P labelled target (~50 nM) was incubated with ~1.25 μM *e*CYDG for 24 hours and the backbone cleaved by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The arrow indicates the 39mer product produced by cleavage at the target cytosine. GA corresponds to a GA tract for sequence determination. No enzyme is a control without incubation with *e*CYDG.

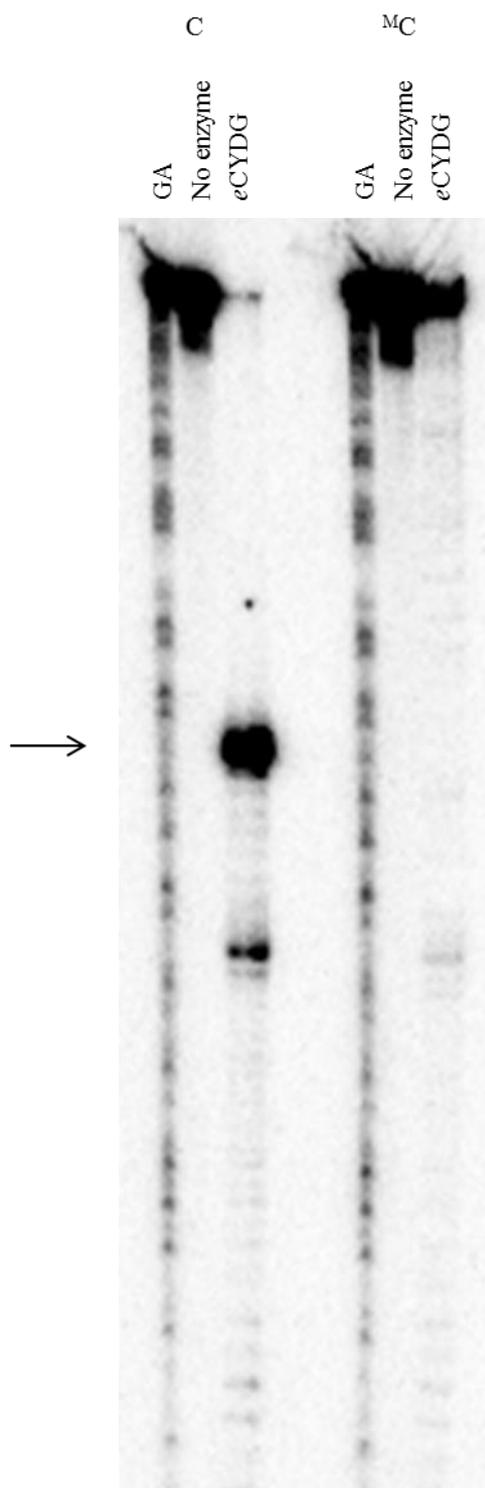


Figure 6.8 *eCYDG* cleavage of C- and ^MC-containing substrates in the presence of protecting oligonucleotides. The ³²P labelled target (~50 nM) was incubated with ~1.25 μM *eCYDG* for 24 hours and the backbone cleaved by heating at 95°C in 10% piperidine for 20 minutes. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The arrow indicates the 39mer product produced by cleavage at the target cytosine. GA corresponds to a GA tract for sequence determination. No enzyme is a control without incubation with *eCYDG*.

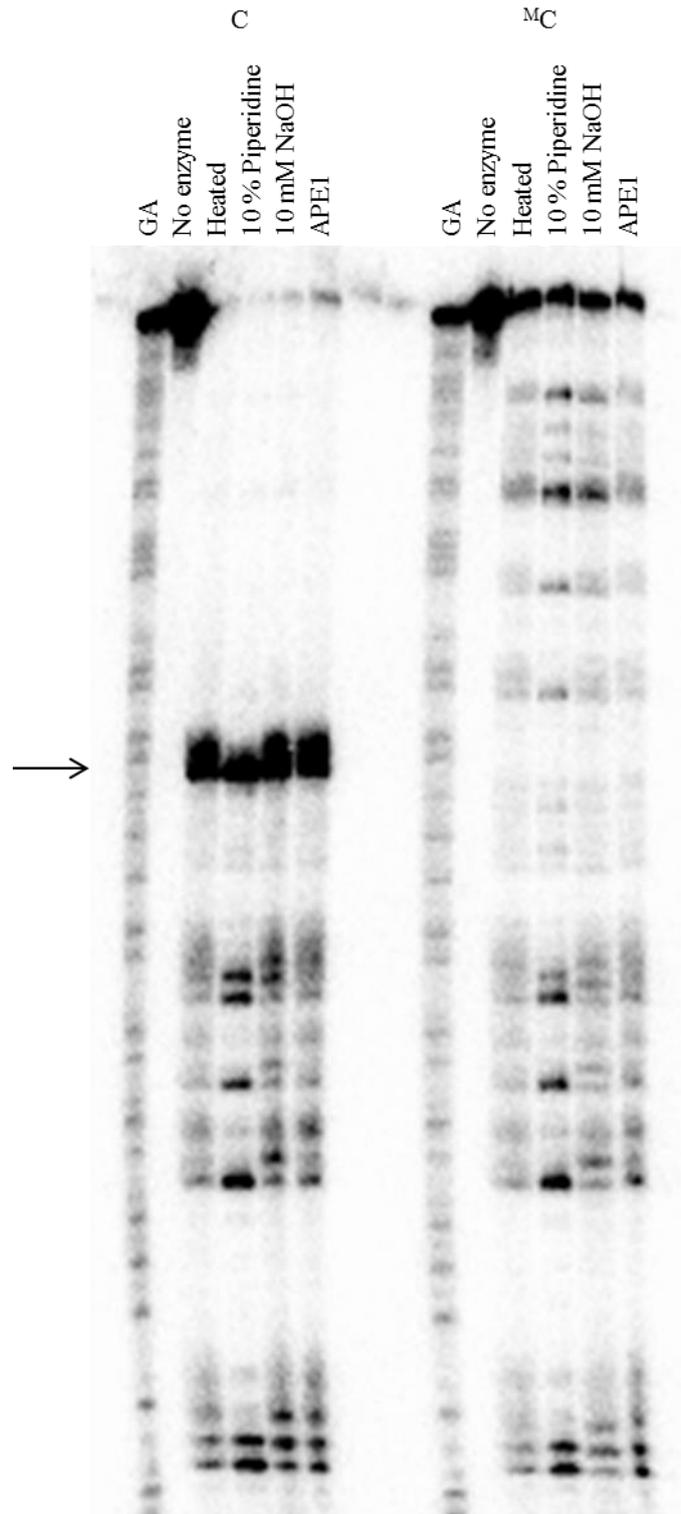


Figure 6.9 *e*CYDG cleavage of C- and ^MC-containing oligonucleotide templates with different backbone cleavage agents. The ³²P labelled target (~50 nM) was incubated with ~1.25 μM *e*CYDG for 24 hours and the backbone cleaved by heating at 95°C without and with 10% (v/v) piperidine or 10 mM NaOH for 20 minutes, or reacted with APE1 (1 μl; approx. 10 u/μl) for 1 hour. The products were resolved on a 12.5% denaturing polyacrylamide gel and analysed by phosphorimaging. The arrow indicates the 39mer product produced by cleavage at the target cytosine. GA corresponds to a GA tract for sequence determination. No enzyme is a control without incubation with *e*CYDG.

6.4 Discussion

The use of synthetic oligonucleotides to mimic the cleavage reaction of target DNA by *e*CYDG clearly shows that real time PCR should be able to distinguish between cleaved (cytosine-containing) and uncleaved (5-methylcytosine-containing) oligonucleotides. The concentration dependence of the signal confirms, as expected, that the assay should be able to detect small differences in DNA concentration. This is important since the concentration of DNA obtained from cellular extracts may be very low, and the differences between samples when analysing their cytosine methylation state may be very small. These controls highlight that the proposed assay could be used effectively for the detection of 5-methylcytosine, despite the lack of discrimination seen when the target DNA was incubated with *e*CYDG followed by PCR analysis. This requires further investigation and optimisation.

The experiments in Chapter 3 showed that a HEG.C pair is one of the best substrates for *e*CYDG and allowed excision of cytosine within the sequence context ACG. This is superior to that seen with a simple AC mismatch at the same position. Although the use of HEG overcomes the limitations of *e*CYDG within the sequence context ACG, further experiments will be needed for optimisation and to increase the reaction efficiency in other stable sequence combinations, such as GCA and GCG. This may be overcome by the use of longer incubation times with *e*CYDG. Alternatively *h*CDG might be used due to its greater activity, though this will be accompanied by other disadvantages. *h*CDG has greater activity at G.C and ssC that could lead to fragmentation of the DNA, reducing the ability to discriminate *via* PCR. However, this might allow the detection of 5-methylcytosine in any sequence context. The single-stranded cleavage activity can be reduced by using protection oligonucleotides (that may also serve as primers as with our assay), increasing the amount of full length target (for ^MC) that is available for amplification and enhancing the assay's discriminatory ability between cytosine and 5-methylcytosine containing DNA.

These results also showed that it is possible to obtain full phosphodiester backbone cleavage merely by heating the sample for 20 minutes at 95°C. This will be a real advantage for development of a simple 5-methylcytosine detection assay, as the initial heating step can be programmed into the PCR protocol.

We anticipate that if the problems described above can be overcome then the assay could be developed into a 96-well plate format to allow for higher throughput. This would involve each well containing a different probe, providing the ability to examine the methylation state of different cytosines over a larger region. Alternatively, the methylation status of the complementary strand could also be examined to determine whether any CpG site is fully or hemi-methylated; such detection of hemi-methylated DNA is often not possible within other assays for ^MC. We therefore hope that this assay can be developed to produce an accurate and quick way of determining the methylation state of cytosine.

Chapter 7: General Discussion

This thesis has investigated the ability of an enzyme to discriminate between cytosine and 5-methylcytosine and to assess the feasibility of using them in an assay for 5-methylcytosine detection. *h*UDG was generated by complete gene synthesis, using an oligonucleotide-stitching-PCR based method, followed by successful N204D mutagenesis to generate *h*CDG. This also highlights the viability of this method to generate whole genes. The equivalent mutation in *E. coli* UDG (N123D) generated a highly cytotoxic protein that couldn't be expressed, even in the cloning vector pUC19 and when out of frame with the *LacZ* gene. Since L191 aids base flipping by forcing the target base into an extrahelical conformation, while also preventing the base from reinserting into the duplex (Krokan *et al.*, 1997, Parikh *et al.*, 1998), mutation to alanine (L191A) significantly reduces the enzyme's activity (Jiang *et al.*, 2001, Jiang and Stivers, 2002). This effect can be rescued by placing a large bulky synthetic nucleoside opposite the target base that occupies the space of the base pair forcing the base into a permanent extrahelical formation. This mutation was performed first to generate *e*UYDG (Jiang *et al.*, 2001, Jiang *et al.*, 2002b) followed by N123D to generate *e*CYDG (Kwon *et al.*, 2003), a functional *e*CDG.

By exploiting UDG's natural ability to discriminate between uracil and thymine, due to the presence of the 5-methyl group being sterically excluded by Y66, we have shown that these CDGs have the ability to excise cytosine when it is unpaired, mispaired or placed opposite a non-nucleosidic linker, while showing no activity towards 5-methylcytosine in any context. The excision efficiency of CDGs is affected by the stability of the base pair. This is shown by faster rates of excision being obtained with HEG or anthraquinone pyrrolidine opposite cytosine compared to an A.C mismatch or a normal G.C base pair. Investigation into the effect of the flanking base pairs led us to conclude that base stacking interactions are more important in base stability, and have a greater effect on excision efficiency. To reduce this effect the flanking bases were mismatched or left unpaired to attempt to reduce base stacking interactions with the target cytosine. This proved unsuccessful though further investigation with the use of DMSO or DNA destabilising compounds (*e.g.* betaine) may be able to reduce the local stability, providing more applicable cleavage rates for use in 5-methylcytosine detection assay.

The extremely slow cleavage rates of ssDNA substrates strongly suggest that the major role of leucine is to "plug" the space left by the target base becoming extrahelical, rather

than base flipping (Krokan *et al.*, 1997, Parikh *et al.*, 1998). This would suggest that the destabilisation of the base to allow it to become extrahelical occurs by the distortion of the DNA by 45° upon enzyme binding, with base flipping only being aided by leucine. The activity of the enzymes was determined as *h*CDG > *e*CYDG > *h*CYDG, and therefore it would be reasonable to predict that an *e*CDG would be even more active. Experiments into producing *e*CDG would be of potential benefit to the 5-methylcytosine detection assay by reducing the cleavage reaction time. Though this can be prepared *via in vitro* transcription translation systems (Handa *et al.*, 2002), it would more efficient if the enzyme could be produced using *E. coli* (or other amenable cell types such as yeast), as with our investigation into generating a fusion enzyme with Ugi. As Ugi binds strongly to UDG and is even stable in 8 M urea (Acharya *et al.*, 2002), a strategy would need to be developed to release the Ugi from the UDG/CDG if an expressible fusion construct was produced.

Analysis of 5-methylcytosine is important in the understanding of gene regulation and is an area of great interest. All current methods suffer from two main limitations of having to undergo a reaction with bisulphite and/or are dependent on sequence context (*e.g.* restriction enzymes). A method that overcomes these limitations while being fast, reliable, accurate and easy to perform would be of great benefit. Real-time PCR is one such technique that satisfies these requirements. We proposed an assay whereby DNA would first be reacted with a CDG, with the use of a probing oligonucleotide to form a duplex with the target cytosine, followed by real-time PCR to detect the presence or absence of a product indicating the presence of 5-methylcytosine or cytosine respectively. CDG is able to fully excise cytosine in this context as shown by PAGE assays with radiolabelled substrates, however, this proved unsuccessful when analysed by PCR and reproducible and significant discrimination was not observed. Further investigation will be required for potential assay commercialisation. One solution may be to use molecular HyBeacons with melt analysis to detect the difference in one base pair. This is because upon amplification a guanine will be incorporated opposite a 5-methylcytosine, while any base, most likely adenine, will be incorporated opposite an AP site produced by excision of cytosine by CDG.

Another possibility as to why the *e*CYDG:PCR assay failed to produce any clear discrimination between a cytosine and 5-methylcytosine substrate could be associated with the purity of the enzyme, due to contaminants during the purification. As no DNases were added and since our enzymes bind DNA, it is conceivable that the purified enzymes

contained traces of DNA. Furthermore, Figure 5.4 shows possible exonuclease (most likely exonuclease III) due to the laddering effect seen for the ssC and ss^MC substrates. Figure 5.6E also shows potential phosphatase activity due to a radioactive signal, most likely from free radiolabelled phosphate, being apparent at the dye front at the bottom of the gel. It would therefore be appropriate for future work to include further purification steps, *e.g.* the addition of DNases and ion exchange, to yield purer enzymes.

Initial experiments have been performed to assess the activity of Y66 UDG and CDG mutants in their ability to accept 5-methylcytosine into their active site. Further studies will be required to fully determine their activity, including their activity against 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Further Y66 mutants will need to be generated to alter the activity and/or specificity, which may be aided by other active site mutations. The same mutations will also be compared to and examined in *h*CDG; *h*UDG Y147A has already been shown to excise thymine and it is effectively a TDG (Kavli *et al.*, 1996). Obtaining crystal structures of the CDGs bound to a DNA substrate will show what residues form interactions, or where favourable interactions could form with the substrate, allowing better determination of residues for mutagenesis. Random mutagenesis may also be used and active enzymes characterised to determine favourable mutations. These studies would further develop the 5-methylcytosine detection assay for detection of hmC, fC and caC.

In conclusion we have successfully generated fully functional CDGs that are able to discriminate between cytosine and 5-methylcytosine, providing a potential new method for 5-methylcytosine detection.

References

- Abner, C. W., Lau, A. Y., Ellenberger, T. & Bloom, L. B. 2001. Base excision and DNA binding activities of human alkyladenine DNA glycosylase are sensitive to the base paired with a lesion. *J Biol Chem.* 276, 13379-87.
- Abu, M. & Waters, T. R. 2003. The main role of human thymine-DNA glycosylase is removal of thymine produced by deamination of 5-methylcytosine and not removal of ethenocytosine. *J Biol Chem.* 278, 8739-8744.
- Acharya, N., Roy, S. & Varshney, U. 2002. Mutational analysis of the uracil DNA glycosylase inhibitor protein and its interaction with Escherichia coli uracil DNA glycosylase. *J Mol Biol.* 321, 579-90.
- Acharya, N., Talawar, R. K., Saikrishnan, K., Vijayan, M. & Varshney, U. 2003. Substitutions at tyrosine 66 of Escherichia coli uracil DNA glycosylase lead to characterization of an efficient enzyme that is recalcitrant to product inhibition. *Nucleic Acids Res.* 31, 7216-26.
- Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A. R., Gutig, D., Grabs, G., Howe, A., Kursar, M., Lesche, R., Leu, E., Lewin, A., Maier, S., Muller, V., Otto, T., Scholz, C., Schulz, W. A., Seifert, H. H., Schwöpe, I., Ziebarth, H., Berlin, K., Piepenbrock, C. & Olek, A. 2002. Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* 30, e21.
- Arand, J., Spieler, D., Karius, T., Branco, M. R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V. & Walter, J. 2012. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* 8, e1002750.
- Barrett, T. E., Savva, R., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. 1998a. Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct. *Nat Struct Biol.* 5, 697-701.
- Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. 1998b. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell.* 92, 117-29.
- Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L. & Pearl, L. H. 1999. Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J.* 18, 6599-609.
- Baylin, S. B., Esteller, M., Rountree, M. R., Bachman, K. E., Schuebel, K. & Herman, J. G. 2001. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet.* 10, 687-92.
- Bennett, M. T., Rodgers, M. T., Hebert, A. S., Ruslander, L. E., Eisele, L. & Drohat, A. C. 2006. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J Am Chem Soc.* 128, 12510-9.
- Bennett, S. E. & Mosbaugh, D. W. 1992. Characterization of the Escherichia coli uracil-DNA glycosylase.inhibitor protein complex. *J Biol Chem.* 267, 22512-21.
- Berdal, K. G., Johansen, R. F. & Seeberg, E. 1998. Release of normal bases from intact DNA by a native DNA repair enzyme. *EMBO J.* 17, 363-7.
- Bharti, S. K. & Varshney, U. 2010. Analysis of the impact of a uracil DNA glycosylase attenuated in AP-DNA binding in maintenance of the genomic integrity in Escherichia coli. *Nucleic Acids Res.* 38, 2291-301.
- Bianco, T., Hussey, D. & Dobrovic, A. 1999. Methylation-sensitive, single-strand conformation analysis (MS-SSCA): A rapid method to screen for and analyze methylation. *Hum Mutat.* 14, 289-93.
- Bjelland, S., Birkeland, N. K., Benneche, T., Volden, G. & Seeberg, E. 1994. DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in Escherichia coli. *J Biol Chem.* 269, 30489-95.

- Boiteux, S., O'Connor, T. R., Lederer, F., Gouyette, A. & Laval, J. 1990. Homogeneous Escherichia coli FPG protein. A DNA glycosylase which excises imidazole ring-opened purines and nicks DNA at apurinic/apyrimidinic sites. *J Biol Chem.* 265, 3916-22.
- Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W. & Balasubramanian, S. 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science.* 336, 934-7.
- Borys-Brzywczy, E., Arczewska, K. D., Saparbaev, M., Hardeland, U., Schar, P. & Kusmierk, J. T. 2005. Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts. *Acta Biochim Pol.* 52, 149-65.
- Cedar, H., Solage, A., Glaser, G. & Razin, A. 1979. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res.* 6, 2125-32.
- Clark, S. J., Harrison, J., Paul, C. L. & Frommer, M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22, 2990-7.
- Colella, S., Shen, L., Baggerly, K. A., Issa, J. P. & Krahe, R. 2003. Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques.* 35, 146-50.
- Cullinan, D., Johnson, F., Grollman, A. P., Eisenberg, M. & De Los Santos, C. 1997. Solution structure of a DNA duplex containing the exocyclic lesion 3,N4-etheno-2'-deoxycytidine opposite 2'-deoxyguanosine. *Biochemistry.* 36, 11933-43.
- Dinner, A. R., Blackburn, G. M. & Karplus, M. 2001. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature.* 413, 752-5.
- Dodson, M. L., Michaels, M. L. & Lloyd, R. S. 1994. Unified catalytic mechanism for DNA glycosylases. *J Biol Chem.* 269, 32709-12.
- Dong, J., Drohat, A. C., Stivers, J. T., Pankiewicz, K. W. & Carey, P. R. 2000. Raman spectroscopy of uracil DNA glycosylase-DNA complexes: Insights into DNA damage recognition and catalysis. *Biochemistry.* 39, 13241-13250.
- Doseth, B., Ekre, C., Slupphaug, G., Krokan, H. E. & Kavli, B. 2012. Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil. *DNA Repair (Amst).* 11, 587-93.
- Down, T. A., Rakyar, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E. M., Thorne, N. P., Backdahl, L., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Marioni, J. C., Birney, E., Hubbard, T. J., Durbin, R., Tavare, S. & Beck, S. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol.* 26, 779-85.
- Drohat, A. C., Jagadeesh, J., Ferguson, E. & Stivers, J. T. 1999a. Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry.* 38, 11866-75.
- Drohat, A. C., Kwon, K., Krosky, D. J. & Stivers, J. T. 2002. 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. *Nat Struct Biol.* 9, 659-64.
- Drohat, A. C. & Stivers, J. T. 2000. Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry.* 39, 11865-75.
- Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L. & Stivers, J. T. 1999b. Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry.* 38, 11876-86.
- Eftedal, I., Guddal, P. H., Slupphaug, G., Volden, G. & Krokan, H. E. 1993. Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase. *Nucleic Acids Res.* 21, 2095-101.
- Engstrom, L. M., Partington, O. A. & David, S. S. 2012. An iron-sulfur cluster loop motif in the Archaeoglobus fulgidus uracil-DNA glycosylase mediates efficient uracil recognition and removal. *Biochemistry.* 51, 5187-97.

- Fitzgerald, M. E. & Drohat, A. C. 2008. Coordinating the initial steps of base excision repair. Apurinic/aprimidinic endonuclease 1 actively stimulates thymine DNA glycosylase by disrupting the product complex. *J Biol Chem.* 283, 32680-90.
- Friedman, J. I. & Stivers, J. T. 2010. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry.* 49, 4957-67.
- Gallinari, P. & Jiricny, J. 1996. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature.* 383, 735-8.
- Grafstrom, R. H., Yuan, R. & Hamilton, D. L. 1985. The characteristics of DNA methylation in an in vitro DNA synthesizing system from mouse fibroblasts. *Nucleic Acids Res.* 13, 2827-42.
- Grippon, S., Zhao, Q., Robinson, T., Marshall, J. J., O'Neill, R. J., Manning, H., Kennedy, G., Dunsby, C., Neil, M., Halford, S. E., French, P. M. & Baldwin, G. S. 2011. Differential modes of DNA binding by mismatch uracil DNA glycosylase from *Escherichia coli*: implications for abasic lesion processing and enzyme communication in the base excision repair pathway. *Nucleic Acids Res.* 39, 2593-603.
- Grossman, T. H., Kawasaki, E. S., Punreddy, S. R. & Osburne, M. S. 1998. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene.* 209, 95-103.
- Grunau, C., Clark, S. J. & Rosenthal, A. 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* 29, E65-5.
- Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G. L. & Song, H. 2014a. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci.* 17, 215-22.
- Guo, W., Chung, W. Y., Qian, M., Pellegrini, M. & Zhang, M. Q. 2014b. Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic Acids Res.* 42, 3009-16.
- Hackett, J. A., Reddington, J. P., Nestor, C. E., Dunican, D. S., Branco, M. R., Reichmann, J., Reik, W., Surani, M. A., Adams, I. R. & Meehan, R. R. 2012. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development.* 139, 3623-32.
- Handa, P., Acharya, N. & Varshney, U. 2001. Chimeras between single-stranded DNA-binding proteins from *Escherichia coli* and *Mycobacterium tuberculosis* reveal that their C-terminal domains interact with uracil DNA glycosylases. *J Biol Chem.* 276, 16992-16997.
- Handa, P., Acharya, N. & Varshney, U. 2002. Effects of mutations at tyrosine 66 and asparagine 123 in the active site pocket of *Escherichia coli* uracil DNA glycosylase on uracil excision from synthetic DNA oligomers: evidence for the occurrence of long-range interactions between the enzyme and substrate. *Nucleic Acids Research.* 30, 3086-3095.
- Hang, B., Medina, M., Fraenkel-Conrat, H. & Singer, B. 1998. A 55-kDa protein isolated from human cells shows DNA glycosylase activity toward 3,N4-ethenocytosine and the G/T mismatch. *PNAS.* 95, 13561-13566.
- Hardeland, U., Bentele, M., Jiricny, J. & Schar, P. 2000. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem.* 275, 33449-56.
- Hashimoto, H., Zhang, X. & Cheng, X. 2012. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* 40, 8276-84.
- Haug, T., Skorpen, F., Aas, P. A., Malm, V., Skjelbred, C. & Krokan, H. E. 1998. Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase. *Nucleic Acids Res.* 26, 1449-57.
- Haushalter, K. A., Stukenberg, P. T., Kirschner, M. W. & Verdine, G. L. 1999. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Current Biology.* 9, 174-185.

- He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C. & Xu, G. L. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 333, 1303-7.
- Higley, M. & Lloyd, R. S. 1993. Processivity of uracil DNA glycosylase. *Mutat Res*. 294, 109-16.
- Hodskinson, M. R., Allen, L. M., Thomson, D. P. & Sayers, J. R. 2007. Molecular interactions of *Escherichia coli* ExoIX and identification of its associated 3'-5' exonuclease activity. *Nucleic Acids Res*. 35, 4094-102.
- Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S. & Kuramitsu, S. 2003. Crystal structure of a family 4 uracil-DNA glycosylase from *Thermus thermophilus* HB8. *J Mol Biol*. 333, 515-26.
- Huang, R., Wang, J., Mao, W., Fu, B., Xing, X., Guo, G. & Zhou, X. 2013. A novel combined bisulfite UDG assay for selective 5-methylcytosine detection. *Talanta*. 117, 445-8.
- Jiang, Y. L., Drohat, A. C., Ichikawa, Y. & Stivers, J. T. 2002a. Probing the limits of electrostatic catalysis by uracil DNA glycosylase using transition state mimicry and mutagenesis. *J Biol Chem*. 277, 15385-15392.
- Jiang, Y. L., Ichikawa, Y., Song, F. & Stivers, J. T. 2003. Powering DNA repair through substrate electrostatic interactions. *Biochemistry*. 42, 1922-9.
- Jiang, Y. L., Kwon, K. & Stivers, J. T. 2001. Turning on uracil-DNA glycosylase using a pyrene nucleotide switch. *J Biol Chem*. 276, 42347-42354.
- Jiang, Y. L. & Stivers, J. T. 2001. Reconstructing the substrate for uracil DNA glycosylase: tracking the transmission of binding energy in catalysis. *Biochemistry*. 40, 7710-9.
- Jiang, Y. L. & Stivers, J. T. 2002. Mutational analysis of the base-flipping mechanism of uracil DNA glycosylase. *Biochemistry*. 41, 11236-47.
- Jiang, Y. L., Stivers, J. T. & Song, F. H. 2002b. Base-flipping mutations of uracil DNA glycosylase: substrate rescue using a pyrene nucleotide wedge. *Biochemistry*. 41, 11248-11254.
- Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M. R., Barton, M. C. & Issa, J. P. 2014. TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res*.
- Johannsen, M. W., Gerrard, S. R., Melvin, T. & Brown, T. 2014. Triplex-mediated analysis of cytosine methylation at CpA sites in DNA. *Chem Commun (Camb)*. 50, 551-3.
- Jones, P. A. & Baylin, S. B. 2002. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 3, 415-28.
- Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J. & Wolffe, A. P. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*. 19, 187-91.
- Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A*. 93, 13770-3.
- Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A. & Krokan, H. E. 1996. Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J*. 15, 3442-7.
- Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell*. 76, 357-69.
- Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S. & Masui, R. 2007. Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. *J Mol Biol*. 373, 839-50.
- Kriaucionis, S. & Heintz, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 324, 929-30.
- Krokan, H. & Wittwer, C. U. 1981. Uracil DNA-glycosylase from HeLa cells: general properties, substrate specificity and effect of uracil analogs. *Nucleic Acids Res*. 9, 2599-613.

- Krokan, H. E., Otterlei, M., Nilsen, H., Kavli, B., Skorpen, F., Andersen, S., Skjelbred, C., Akbari, M., Aas, P. A. & Slupphaug, G. 2001. Properties and functions of human uracil-DNA glycosylase from the UNG gene. *Prog Nucleic Acid Res Mol Biol.* 68, 365-86.
- Krokan, H. E., Standal, R. & Slupphaug, G. 1997. DNA glycosylases in the base excision repair of DNA. *Biochem J.* 325 (Pt 1), 1-16.
- Krosky, D. J., Schwarz, F. P. & Stivers, J. T. 2004. Linear free energy correlations for enzymatic base flipping: how do damaged base pairs facilitate specific recognition? *Biochemistry.* 43, 4188-95.
- Krosky, D. J., Song, F. & Stivers, J. T. 2005. The origins of high-affinity enzyme binding to an extrahelical DNA base. *Biochemistry.* 44, 5949-59.
- Kwon, K., Jiang, Y. L. & Stivers, J. T. 2003. Rational engineering of a DNA glycosylase specific for an unnatural cytosine:pyrene base pair. *Chem Biol.* 10, 351-9.
- Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M., Bartlett, J. J., Pavlenok, M. & Gundlach, J. H. 2013. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A.* 110, 18904-9.
- Leithauser, M. T., Liem, A., Stewart, B. C., Miller, E. C. & Miller, J. A. 1990. 1,N6-ethenoadenosine formation, mutagenicity and murine tumor induction as indicators of the generation of an electrophilic epoxide metabolite of the closely related carcinogens ethyl carbamate (urethane) and vinyl carbamate. *Carcinogenesis.* 11, 463-473.
- Lindahl, T. 1974. An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc Natl Acad Sci USA.* 71, 3649-3653.
- Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature.* 362, 709-15.
- Lindahl, T. & Nyberg, B. 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry.* 13, 3405-10.
- Liu, P. F., Theruvathu, J. A., Darwanto, A., Lao, V. V., Pascal, T., Goddard, W. A. & Sowers, L. C. 2008. Mechanisms of base selection by the *Escherichia coli* mispaired uracil glycosylase. *J Biol Chem.* 283, 8829-8836.
- Liu, X. & Liu, J. 2004. Cloning, expression, and characterization of uracil-DNA glycosylase of *Chlamydia pneumoniae* in *Escherichia coli*. *Protein Expr Purif.* 35, 46-53.
- Lutsenko, E. & Bhagwat, A. S. 1999. The role of the *Escherichia coli* Mug protein in the removal of uracil and 3,N4-ethenocytosine from DNA. *J Biol Chem.* 274, 31034-31038.
- Maiti, A., Morgan, M. T., Pozharski, E. & Drohat, A. C. 2008. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci U S A.* 105, 8890-5.
- Malys, N. 2012. Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes. *Mol Biol Rep.* 39, 33-9.
- Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., Pavlenok, M., Niederweis, M. & Gundlach, J. H. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol.* 30, 349-53.
- Manrao, E. A., Derrington, I. M., Pavlenok, M., Niederweis, M. & Gundlach, J. H. 2011. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS One.* 6, e25723.
- McClelland, M., Nelson, M. & Raschke, E. 1994. Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* 22, 3640-59.
- Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N. & Cao, W. 2009. Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. *J Mol Biol.* 385, 761-78.
- Moe, E., Leiros, I., Smalas, A. O. & McSweeney, S. 2006. The crystal structure of mismatch-specific uracil-DNA glycosylase (MUG) from *Deinococcus radiodurans* reveals a novel catalytic residue and broad substrate specificity. *J Biol Chem.* 281, 569-77.

- Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. 1995a. Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*. 82, 701-8.
- Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. 1995b. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*. 80, 869-78.
- Morera, S., Grin, I., Vigouroux, A., Couve, S., Henriot, V., Saparbaev, M. & Ishchenko, A. A. 2012. Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Res.* 40, 9917-26.
- Morgan, M. T., Maiti, A., Fitzgerald, M. E. & Drohat, A. C. 2011. Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. *Nucleic Acids Res.* 39, 2319-29.
- Mosbaugh, D. W. 1988. Enzymology of uracil-DNA repair in mammalian cells. *Rev. Biochem. Tox.* 9, 69-130.
- Muller, U., Bauer, C., Siegl, M., Rottach, A. & Leonhardt, H. 2014. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.*
- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. & Bird, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*. 393, 386-9.
- Neddermann, P. & Jiricny, J. 1994. Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci U S A*. 91, 1642-6.
- Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T. A., Skorpen, F. & Krokan, H. E. 1997. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res.* 25, 750-5.
- O'Neill, R. J., Vorob'eva, O. V., Shahbakhti, H., Zmuda, E., Bhagwat, A. S. & Baldwin, G. S. 2003. Mismatch uracil glycosylase from *Escherichia coli*: a general mismatch or a specific DNA glycosylase? *J Biol Chem*. 278, 20526-32.
- Oka, Y., Peng, T., Takei, F. & Nakatani, K. 2008. The reaction of cytosine with bisulfite by base flipping from the duplex. *Nucleic Acids Symp Ser (Oxf)*. 435-6.
- Oka, Y., Peng, T., Takei, F. & Nakatani, K. 2009. Synthesis and reaction of DNA oligomers containing modified cytosines related to bisulfite sequencing. *Org Lett*. 11, 1377-9.
- Olsen, L. C., Aasland, R., Wittwer, C. U., Krokan, H. E. & Helland, D. E. 1989. Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme. *EMBO J*. 8, 3121-5.
- Pan, S. H. & Malcolm, B. A. 2000. Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *Biotechniques*. 29, 1234-8.
- Panayotou, G., Brown, T., Barlow, T., Pearl, L. H. & Savva, R. 1998. Direct measurement of the substrate preference of uracil-DNA glycosylase. *J Biol Chem*. 273, 45-50.
- Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. 1998. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J*. 17, 5214-26.
- Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M. & Tainer, J. A. 2000. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc Natl Acad Sci U S A*. 97, 5083-8.
- Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M. & Stivers, J. T. 2007. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*. 449, 433-7.
- Pearl, L. H. 2000. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat Res*. 460, 165-81.

- Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E. & Kavli, B. 2007. Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res.* 35, 3879-92.
- Purmal, A. A., Lampman, G. W., Pourmal, E. I., Melamede, R. J., Wallace, S. S. & Kow, Y. W. 1994. Uracil DNA N-glycosylase distributively interacts with duplex polynucleotides containing repeating units of either TGGCCAAGCU or TGGCCAAGCTTGGCCAAGCU. *J Biol Chem.* 269, 22046-22053.
- Ramsahoye, B. H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A. P. & Jaenisch, R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A.* 97, 5237-42.
- Robins, P., Jones, C. J., Biggerstaff, M., Lindahl, T. & Wood, R. D. 1991. Complementation of DNA repair in xeroderma pigmentosum group A cell extracts by a protein with affinity for damaged DNA. *EMBO J.* 10, 3913-21.
- Rowland, M. M., Schonhoft, J. D., Mckibbin, P. L., David, S. S. & Stivers, J. T. 2014. Microscopic mechanism of DNA damage searching by hOGG1. *Nucleic Acids Res.*
- Roy, R., Brooks, C. & Mitra, S. 1994. Purification and biochemical characterization of recombinant N-methylpurine-DNA glycosylase of the mouse. *Biochemistry.* 33, 15131-40.
- Roy, S., Purnapatre, K., Handa, P., Boyanapalli, M. & Varshney, U. 1998. Use of a coupled transcriptional system for consistent overexpression and purification of UDG-Ugi complex and Ugi from Escherichia coli. *Protein Expr Purif.* 13, 155-62.
- Rusling, D. A., Powers, V. E., Ranasinghe, R. T., Wang, Y., Osborne, S. D., Brown, T. & Fox, K. R. 2005. Four base recognition by triplex-forming oligonucleotides at physiological pH. *Nucleic Acids Res.* 33, 3025-32.
- Sagi, J., Hang, B. & Singer, B. 1999. Sequence-dependent repair of synthetic AP sites in 15-mer and 35-mer oligonucleotides: role of thermodynamic stability imposed by neighbor bases. *Chem Res Toxicol.* 12, 917-23.
- Sagi, J., Perry, A., Hang, B. & Singer, B. 2000. Differential destabilization of the DNA oligonucleotide double helix by a T:G mismatch, 3,N4-ethenocytosine, 3,N4-ethanocytosine, or an 8-(hydroxymethyl)-3,N4-ethenocytosine adduct incorporated into the same sequence contexts. *Chem Res Toxicol.* 13, 839-845.
- Sandigursky, M., Faje, A. & Franklin, W. A. 2001. Characterization of the full length uracil-DNA glycosylase in the extreme thermophile *Thermotoga maritima*. *Mutat Res.* 485, 187-95.
- Sandigursky, M. & Franklin, W. A. 1999. Thermostable uracil-DNA glycosylase from *Thermotoga maritima* a member of a novel class of DNA repair enzymes. *Curr Biol.* 9, 531-4.
- Sandigursky, M. & Franklin, W. A. 2000. Uracil-DNA glycosylase in the extreme thermophile *Archaeoglobus fulgidus*. *J Biol Chem.* 275, 19146-9.
- Sano, H., Imokawa, M. & Sager, R. 1988. Detection of heavy methylation in human repetitive DNA subsets by a monoclonal antibody against 5-methylcytosine. *Biochim Biophys Acta.* 951, 157-65.
- Sano, H., Royer, H. D. & Sager, R. 1980. Identification of 5-methylcytosine in DNA fragments immobilized on nitrocellulose paper. *Proc Natl Acad Sci U S A.* 77, 3581-5.
- Santalucia, J., Jr., Allawi, H. T. & Seneviratne, P. A. 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry.* 35, 3555-62.
- Saparbaev, M. & Laval, J. 1998. 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. *PNAS.* 95, 8508-8513.
- Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H. & Jiricny, J. 2002. A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J.* 21, 3182-91.
- Savva, R., Mcauley-Hecht, K., Brown, T. & Pearl, L. 1995. The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature.* 373, 487-93.

- Savva, R. & Pearl, L. H. 1995. Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nat Struct Biol.* 2, 752-7.
- Schonhoft, J. D., Kosowicz, J. G. & Stivers, J. T. 2013. DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry.* 52, 2526-35.
- Schonhoft, J. D. & Stivers, J. T. 2013. DNA translocation by human uracil DNA glycosylase: the case of single-stranded DNA and clustered uracils. *Biochemistry.* 52, 2536-44.
- Seibert, E., Ross, J. B. & Osman, R. 2002. Role of DNA flexibility in sequence-dependent activity of uracil DNA glycosylase. *Biochemistry.* 41, 10976-84.
- Shapiro, R., Braverman, B., Louis, J. B. & Servis, R. E. 1973. Nucleic acid reactivity and conformation. II. Reaction of cytosine and uracil with sodium bisulfite. *J Biol Chem.* 248, 4060-4.
- Shapiro, R. & Danzig, M. 1972. Acidic hydrolysis of deoxycytidine and deoxyuridine derivatives. The general mechanism of deoxyribonucleoside hydrolysis. *Biochemistry.* 11, 23-9.
- Shen, J. C., Rideout, W. M., 3rd & Jones, P. A. 1992. High frequency mutagenesis by a DNA methyltransferase. *Cell.* 71, 1073-80.
- Singer, B. & Hang, B. 1997. What structural features determine repair enzyme specificity and mechanism in chemically modified DNA? *Chem Res Toxicol.* 10, 713-32.
- Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N. M., Haug, T., Levine, D. W. & Krokan, H. E. 1995. Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry.* 34, 128-38.
- Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. 1996. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature.* 384, 87-92.
- Sponer, J., Jurecka, P. & Hobza, P. 2004. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J Am Chem Soc.* 126, 10142-10151.
- Starkuviene, V. & Fritz, H. J. 2002. A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium *Thermus thermophilus*. *Nucleic Acids Res.* 30, 2097-102.
- Stivers, J. T., Pankiewicz, K. W. & Watanabe, K. A. 1999. Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry.* 38, 952-63.
- Stratagene 2004. QuikChange site-directed mutagenesis kit instruction manual. 1-13.
- Su, J., Yan, H., Wei, Y., Liu, H., Liu, H., Wang, F., Lv, J., Wu, Q. & Zhang, Y. 2013. CpG_MPs: identification of CpG methylation patterns of genomic regions from high-throughput bisulfite sequencing data. *Nucleic Acids Res.* 41, e4.
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. & Rao, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* 324, 930-5.
- Talhaoui, I., Couve, S., Ishchenko, A. A., Kunz, C., Schar, P. & Saparbaev, M. 2013. 7,8-Dihydro-8-oxoadenine, a highly mutagenic adduct, is repaired by *Escherichia coli* and human mismatch-specific uracil/thymine-DNA glycosylases. *Nucleic Acids Res.* 41, 912-23.
- Van Aalten, D. M., Erlanson, D. A., Verdine, G. L. & Joshua-Tor, L. 1999. A structural snapshot of base-pair opening in DNA. *Proc Natl Acad Sci U S A.* 96, 11809-14.
- Varshney, U. & Van De Sande, J. H. 1991. Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by *Escherichia coli* uracil DNA glycosylase. *Biochemistry.* 30, 4055-61.
- Volkert, M. R. 1988. Adaptive response of *Escherichia coli* to alkylation damage. *Environmental and Molecular Mutagenesis.* 11, 241-255.
- Waters, T. R., Gallinari, P., Jiricny, J. & Swann, P. F. 1999. Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J Biol Chem.* 274, 67-74.

- Waters, T. R. & Swann, P. F. 1998. Kinetics of the action of thymine DNA glycosylase. *J Biol Chem.* 273, 20007-14.
- Watkins, N. E., Jr. & Santalucia, J., Jr. 2005. Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.* 33, 6258-67.
- Watt, F. & Molloy, P. L. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* 2, 1136-43.
- Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. & Schubeler, D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet.* 37, 853-62.
- Werner, R. M., Jiang, Y. L., Gordley, R. G., Jagadeesh, G. J., Ladner, J. E., Xiao, G., Tordova, M., Gilliland, G. L. & Stivers, J. T. 2000. Stressing-out DNA? The contribution of serine-phosphodiester interactions in catalysis by uracil DNA glycosylase. *Biochemistry.* 39, 12585-94.
- Werner, R. M. & Stivers, J. T. 2000. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: evidence for an oxocarbenium ion-uracil anion intermediate. *Biochemistry.* 39, 14054-64.
- Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L. & Pearl, L. H. 2003. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol Cell.* 11, 1647-59.
- Wojdacz, T. K. & Dobrovic, A. 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res.* 35, e41.
- Wong, I., Lundquist, A. J., Bernards, A. S. & Mosbaugh, D. W. 2002. Presteady-state analysis of a single catalytic turnover by Escherichia coli uracil-DNA glycosylase reveals a "pinch-pull-push" mechanism. *J Biol Chem.* 277, 19424-32.
- Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. 1999. Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited. *Proteins.* 35, 13-24.
- Ye, Y., Stahley, M. R., Xu, J., Friedman, J. I., Sun, Y., Mcknight, J. N., Gray, J. J., Bowman, G. D. & Stivers, J. T. 2012. Enzymatic excision of uracil residues in nucleosomes depends on the local DNA structure and dynamics. *Biochemistry.* 51, 6028-38.
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J. H., Jin, P., Ren, B. & He, C. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 149, 1368-80.
- Zharkov, D. O., Mechetin, G. V. & Nevinsky, G. A. 2010. Uracil-DNA glycosylase: Structural, thermodynamic and kinetic aspects of lesion search and recognition. *Mutat Res.* 685, 11-20.

Appendix I - List of General Oligonucleotides used

Description	Sequence
N123D For	5' -GCGTCAGGGCGTTCTGCTACTCGATACTGTGTTGACGGTACGCGC-3'
N123D Rev	5' -GCGCGTACCGTCAACACAGTATCGAGTAGCAGAACGCCCTGACGC-3'
L191A For	5' -GCACCGCATCCGTCGCCGCGTCGGCGCATCGTGGATTC-3'
L191A Rev	5' -GAATCCACGATGCGCCGACGCCGCGACGGATGCGGTGC-3'
A191L For	5' -GCACCGCATCCGTCGCCGCTTTCGGCGCATCGTGGATTC-3'
A191L Rev	5' GAATCCACGATGCGCCGAAAGCGGCGACGGATGCGGTGC 3'
pET28a For	5' -GGAGATATACCATGGGCAGCAGC-3'
pET28a Rev	5' -CTTTGTTAGCAGCCGGATCTCAG-3'
N204D For	5' -GGTGTTCCTTCTCGACGCTGTCCTCACGG-3'
N204D Rev	5' -CCGTGAGGACAGCGTCGAGAAGGAGAACC-3'
L272A For	5' -GCTCATCCCTCCCCTGCCTCAGTGTATAGAGGG-3'
L272A Rev	5' -CCCTCTATACTGAGGCAGGGGAGGGATGAGC-3'
A	5' -GGCTAAATACCGACTATGCGCACTGATTCGG-3'
G	5' -GGCTAAATACCGACTGTGCGCACTGATTCGG-3'
AP	5' -GGCTAAATACCGACTAPTGCGCACTGATTCGG-3'
U	5' -CCGAATCAGTGCGCAUAGTCGGTATTTAGCC-3'
T	5' -CCGAATCAGTGCGCATAGTCGGTATTTAGCC-3'
C	5' -CCGAATCAGTGCGCACAGTCGGTATTTAGCC-3'
^M C	5' -CCGAATCAGTGCGCA ^M CAGTCGGTATTTAGCC-3'
A2	5' -GGCTAAATACCGACCACGCGCACTGATTCGG-3'
C3	5' -CCGAATCAGTGCGCGCGGTTCGGTATTTAGCC-3'
Z	5' -GGCTAAATACCGACTZTGCGCACTGATTCGG-3' where Z = anthraquinone pyrrolidine
ssC(GAT)	5' -GGATAAATAGGGAGTCTGAGAAGTGATTAGG-3'
ss ^M C	5' -GGATAAATAGGGAGT ^M CTGAGAAGTGATTAGG-3'
ssC(polyA)	5' -AAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAA-3'
G2	5' -GCTAAATATATATATGTTATATAATTATTCG-3'
C2	5' -CGAATAATTATATAACATATATATATTTAGC-3'
Long C	5' -CCGTACTGAATCAGTGCGCACAGTCGGTATTTACGATAGCC-3'
Long Gap 1	5' -GGCTATCGTAAATACCGACT-3'
Long Gap 2	5' -TGCGCACTGATTCAGTACGG-3'
I	5' -GGCTAAATACCGACTITGCGCACTGATTCGG-3' where I = Inosine
Gap 1	5' -GGCTAAATACCGACTG-3'
Gap 2	5' -GCGCACTGATTCGG-3'
HEG	5' -GGCTAAATACCGACTHTGCGCACTGATTCGG-3' where H = Hexaethylene glycol
Long HEG	5' -GGCTAAATACCGACHGCGCACTGATTCGG-3' where H = Hexaethylene glycol
pBADA For	5' -GATCTGTACGACGATGACGATAAGG-3'
pBADA Rev	5' -GGCTGAAAATCTTCTCTCATCCGCC-3'
IRES Linker 1	5' -AGCTACGTAAGATCTTATACTCGAGCAAGGATATCGACTGCTCGA GCCAGATCTTCC-3'
IRES Linker 2	5' -AGCTGGAAGATCTGGCTCGAGCAGTCGATATCCTTGATCGAGTAT AAGATCTTACGT-3'
Stop Codon	5' -CATGATTACGCCAAGTAACATATGGCTAACGAA-3'

(UDG)	
Stop Codon (UDG)	5' -TTCGTTAGCCATATGTTACTTGGCGTAATCATG-3'
Stop Codon (LacZ)	5' -ACCATGATTACGCCATAATTCATATGGCTAACG-3'
Stop Codon (LacZ)	5' -CGTTAGCCATATGAATTATGGCGTAATCATGGT-3'
Ugi For	5' -CTGTAGAGCTCCATATGACAAATTTATCTGACAT-3'
Ugi Rev	5' -AATCGAATTCAAGCTTATAACATTTTAATTTTATT-3'
UDGUgi For	5' -TCAGCTGAATTCCTGGTGCCGCGCGGCAGCATGACAAATTTATCTGACATCATTG-3'
UDGUgi Rev	5' -TAGTACAAGCTTATAACATTTTAATT-3'
Stop Codons For	5' -TTACCGGCAGAGAGTGAGGGAGGAGAATTCCTGGTGCCGCGC-3'
Stop Codons Rev	5' -GCGCGGCACCAGGAATTCTCCTCCCTCACTCTCTGCCGGTAA-3'
Thrombin For	5' -AGCGGCCTGGTGCCGGGCGGCAGCCATATGG-3'
Thrombin Rev	5' -CCATATGGCTGCCGCCCGGCACCAGGCCGCT-3'

Appendix II - List of Oligonucleotides used for the Generation of *hUDGΔ81* and *hUDGΔ81Ugi*

Description	Sequence
<i>hUDG1</i>	5' -AGCTCAGTCATATGGGCGGCTTTGGAGAGAGCTGGAAGAA-3'
<i>hUDG2</i>	5' -CTTGATAAAATACGGTTTCCCGAACTCCCCGCTGAGGTGCTTCTTCCAGCTCTCTCCAAA-3'
<i>hUDG3</i>	5' -GGAAACCGTATTTTATCAAGCTAATGGGATTTGTTGCAGAAGAAAGAA GCATTACTG-3'
<i>hUDG4</i>	5' -ATCTGGGTCCAGGTGAAGACTTGGTGTGGGGGTGGATAAACAGTGTAATGCTTTCTTTCT-3'
<i>hUDG5</i>	5' -GTCTTCACCTGGACCCAGATGTGTGACATAAAAAGATGTGAAGGTTGTCA TCCTGGGACAG-3'
<i>hUDG6</i>	5' -GCAGAGCCCGTGAGCTTGATTAGGTCCATGATATGGATCCTGTCCAGGATGACAACCT-3'
<i>hUDG7</i>	5' -TCAAGCTCACGGGCTCTGCTTTAGTGTTCAAAGGCCTGTTCCGCTCCGCCAGTTTGGA-3'
<i>hUDG8</i>	5' -AAAATCCTCTATGTCTGTAGACAACTCTTTATAAATGTTCTCCAACTGGCGGAGGCGG-3'
<i>hUDG9</i>	5' -CTACAGACATAGAGGATTTTGTTCATCCTGGCCATGGAGATTTATCTGGGTGGGCCAAGC-3'
<i>hUDG10</i>	5' -GCACGAACCGTGAGGACAGCGTTGAGAAGGAGAACACCTTGCTTGGCCC ACCCAGATAAA-3'
<i>hUDG11</i>	5' -GCTGTCTCACGGTTCGTGCCATCAAGCCAACCTCTCATAAGGAGCGAGGCTGGGAGCAG-3'
<i>hUDG12</i>	5' -TCGAGTTCTGATTTAGCCAGGACACAACCTGCATCAGTGAAGTCTCCCA GCCTCGCTCCT-3'
<i>hUDG13</i>	5' -CTGGCTAAATCAGAACTCGAATGGCCTTGTTTTCTTGCTCTGGGGCTCTTATGCTCAGAA-3'
<i>hUDG14</i>	5' -TAGTACATGGTGCCGCTTCCCTATCAATGGCACTGCCCTTCTTCTGAGCA TAAGAGCCCCA-3'
<i>hUDG15</i>	5' -GGAAGCGGCACCATGTACTACAGACGGCTCATCCCTCCCCTTTGTGAGTGTATAGAGGGTTC-3'
<i>hUDG16</i>	5' -AGCAGCTCATTGGTCTTTGAAAAGTGTCTACATCCAAAGAACCCTCTATACACTGA-3'
<i>hUDG17</i>	5' -TCAAAGACCAATGAGCTGCTGCAGAAGTCTGGCAAGAAGCCCATTGACTGGAAGGAGCTG-3'
<i>hUDG18</i>	5' -TCCGAGTCGAATTCTCACAGCTCCTTCCAGTCAATGG-3'
Ugi1	5' -CGGACAAGTTGGTCATGCTGCCGCGGGCACCAGTTCTTCCAGCTCCTTCCAGTCAATGG-3'
Ugi2	5' -CAGCATGACCAACTTGTCCGACATCATCGAAAAAGAGACCGGCAAGCAACTGGTT-3'
Ugi3	5' -ACTTCTTCAGGGAGCATCAAGATAGATTCTTGAATAACCAGTTGCTTGC CGGTCT-3'
Ugi4	5' -TTGATGCTCCCTGAAGAAGTAGAAGAGGTTATCGGTAATAAGCCGGAGTCCGACA-3'
Ugi5	5' -ATCAGTAGATTTCGTCATAAGCAGTGTGCACCAGAATGTCGGACTCCGGC TTATTA-3'
Ugi6	5' -CTTATGACGAATCTACTGATGAGAACGTAATGCTGCTGACTAGCGACGC

	TCCGGA-3'
Ugi7	5'-TAGAGTCTTGAATTACCAGAGCCCACGGTTTGTATTCCGGAGCGTCGCT AGTCAG-3'
Ugi8	5'-TCTGGTAATTCAAGACTCTAACGGCGAAAACAAAATCAAGATGCTGTAA GAATTC-3'
Ugi9	5'-TCCGAGTCGAATTCTTACAGCATCTTGA-3'

Appendix IIIa - List of Oligonucleotides used for the PCR assay

Description	Sequence
Target	5' -TAGCGTAGGGAGATATACCATGGGCAGCAGCCAGTGCGCA [C/ ^M C] GGTCGGTATTCTGAGATCCGGCTGCTAACAAAGTTGATCAT-3'
Probe	5' -GGCTAAATACCGACTATGCGCACTGATTCGG-3'
C Mimic 1	5' -TAGCGTAGGGAGATATACCATGGGCAGCAGCCAGTGCGCA-3'
C Mimic 2	5' -GGTCGGTATTCTGAGATCCGGCTGCTAACAAAGTTGATCAT-3'
HEG Probe	5' -GGCTAAATACCGACTHTGCGCACTGATTCGG-Propyl-3' Where H = Hexaethylene glycol

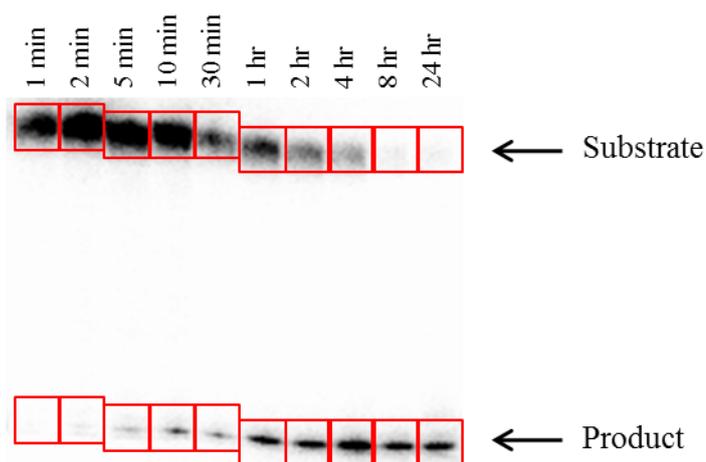
Appendix IIIb - List of Primers/Protection Oligonucleotides used for the eCYDG:rtPCR Assay

Description	Sequence
For Primer/Protection	5' -GGAGATATACCATGGGCAGCAGC-3'
Rev Primer/Protection	5' -CTTTGTTAGCAGCCGGATCTCAG-3'
Rev Protection	5' -GCTGCTGCCCATGGTATATCTCC-3'
For Protection	5' -CTGAGATCCGGCTGCTAACAAAG-3'

Appendix IV – Cleavage Assay Interpretation

To determine the rate of reaction for the enzymes investigated against a range of different substrates the following procedure was used, using *e*CYDG A.C as an example.

The cleavage assay was performed and the samples were run on a denaturing polyacrylamide gel to separate out substrate from product. The gel can be seen below. Boxes were then drawn around each band (Red Boxes) to analyse the intensity of each band. This was performed using the analysis software ImageQuantTL.



The percentage product for each time point was calculated using the equation:

$$\% P = \frac{PI}{PI + SI} \times 100$$

% P = Percentage Product

PI = Product Intensity

SI = Substrate Intensity

The percentage product was plotted against time to generate a graph (using SigmaPlot) allowing the rate of reaction to be calculated (0.0072 min^{-1}) by fitting a curve (see graph below). This was calculated by using the “exponential rise to maximum” equation given as:

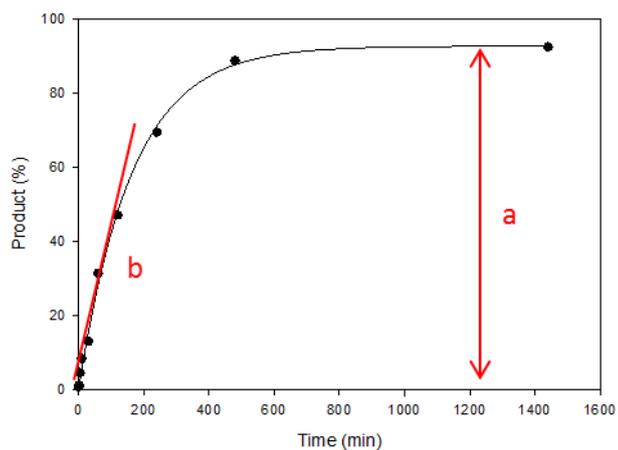
$$y = a(1 - e^{-bx})$$

y = y axis value

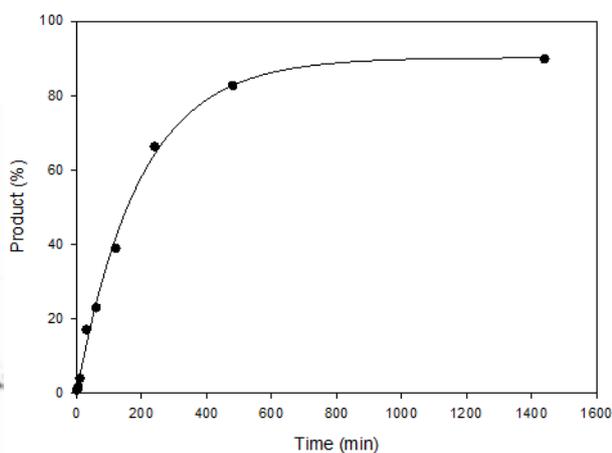
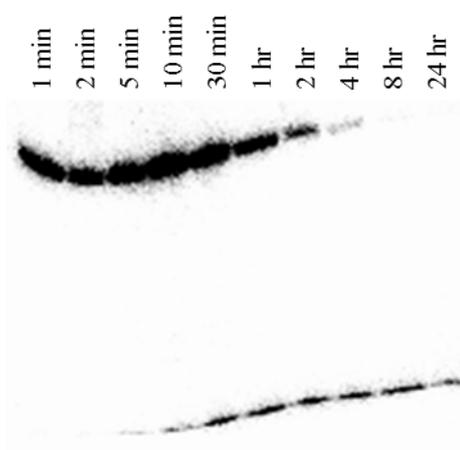
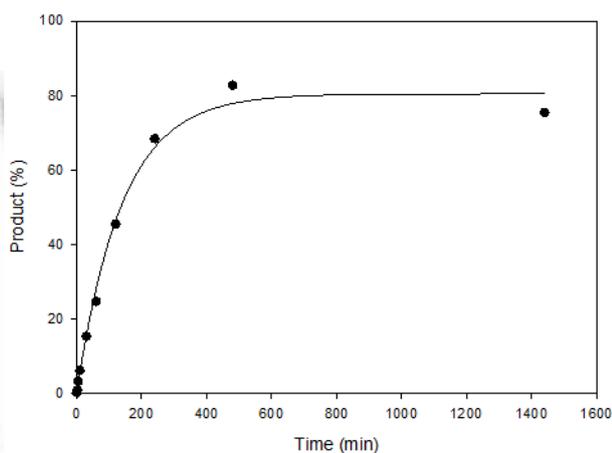
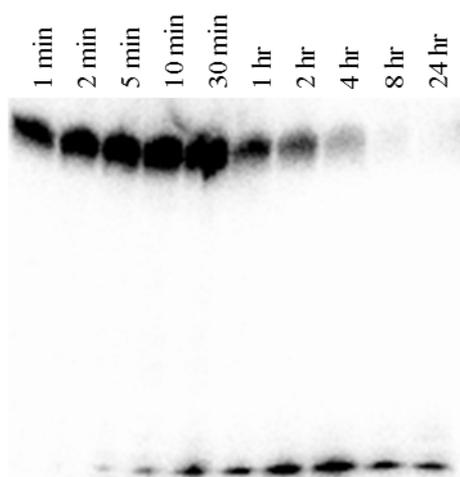
x = x axis value

a = maximum amplitude of the curve (highlighted on the graph)

b = initial velocity/rate of reaction (highlighted on the graph)



This process was repeated for a further two gels and graphs (see below) to give rates of 0.0061 and 0.0052 min^{-1} . This generates a triplicate of rates that were then averaged and the error calculated (by standard deviation) giving $0.0062 \pm 0.001 \text{ min}^{-1}$. Rates were then compiled into Table 3.2, Table 5.2 and Table 5.3.



Appendix V – DNA Sequences of the ORFs of *eUDG* and *hUDGΔ81*

eUDG

ATGGCTAACGAATTAACCTGGCATGACGTGCTGGCTGAAGAGAAGCAGCAACCCCTATTTTC
TTAATACCCCTTCAGACCGTCGCCAGCGAGCGGCAGTCCGGCGTCACTATCTACCCACCACA
AAAAGATGTCTTTAACCGTTCGCTTTACAGAGTTGGGTGACGTTAAAGTGGTGATTCTC
GGCCAGGATCCTTATCACGGACCGGGACAGGCGCATGGTCTGGCATTTTCCGTTTCGTCCCG
GCATTGCCATTCTCCGTCAATTATTGAATATGTATAAAGAGCTGGAAAATACTATTCCGGG
CTTACCCCGCCCTAATCATGGTTATCTTGAAAGCTGGGCGCGTCAGGGCGTTCTGCTACTC
AATACTGTGTTGACGGTACGCGCAGGTCAGGCGCATTCACGACCCAGCCTCGGCTGGGAAA
CCTTACCGATAAAGTGATCAGCCTGATTAACCAGCATCGCGAAGGCGTGGTGTTTTTTGT
GTGGGGATCGCATGCGCAAAAGAAAGGGGCGATTATAGATAAGCAACGCCATCATGTACTG
AAAGCACCGCATCCGTGCGCGCTTTCGGCGCATCGTGGATTCTTTGGCTGCAACCATTTTG
TGCTGGCAAATCAGTGGCTGGAACAACGTGGCGAGACGCCGATTGACTGGATGCCAGTATT
ACCGGCAGAGAGTGAGTAA

Y66A/L/T/W/H/S/C: TAT → GCG/CTG/ACC/TGG/CAT/AGC/TGC

F77W/Y/H: TTT → TGG/TAT/CAT

N123D: AAT → GAT

L191A: CTT → GCG

hUDG

ATGGGCCTCTTCTGCCTTGGGCCGTGGGGGTTGGGCCGGAAGCTGCGGACGCCTGGGAAGG
GGCCGCTGCAGCTCTTGAGCCGCCTCTGCGGGGACCCTTGCAGGCCATCCCAGCCAAGAA
GGCCCCGGCTGGGCAGGAGGAGCCTGGGACGCCGCCCTCCTCGCCGCTGAGTGCCGAGCAG
TTGGACCGGATCCAGAGGAACAAGGCCGCGGCCCTGCTCAGACTCGCGGCCCGCAACGTGC
CCGTGGGCTTTGGAGAGAGCTGGAAGAAGCACCTCAGCGGGGAGTTCGGGAAACCGTATTT
TATCAAGCTAATGGGATTTGTTGCAGAAGAAAGAAAGCATTACACTGTTTATCCACCCCCA
CACCAAGTCTTACCTGGACCCAGATGTGTGACATAAAAGATGTGAAGGTTGTCATCCTGG
GACAGGATCCATATCATGGACCTAATCAAGCTCACGGGCTCTGCTTTAGTGTTCAAAGGCC
TGTTCCGCCTCCGCCAGTTTGGAGAACATTTATAAAGAGTTGTCTACAGACATAGAGGAT
TTGTTTCATCCTGGCCATGGAGATTTATCTGGGTGGGCCAAGCAAGGTGTTCTCCTTCTCA
ACGCTGTCCTCACGGTTCGTGCCCATCAAGCCAACCTCTATAAGGAGCGAGGCTGGGAGCA
GTTCACTGATGCAGTTGTGTCTGGCTAAATCAGAACTCGAATGGCCTTGTTTTCTTGCTC
TGGGGCTCTTATGCTCAGAAGAAGGGCAGTGCCATTGATAGGAAGCGGCACCATGTACTAC
AGACGGCTCATCCCTCCCCTTTGTCAGTGTATAGAGGGTTCTTTGGATGTAGACACTTTTC
AAAGACCAATGAGCTGCTGCAGAAGTCTGGCAAGAAGCCATTGACTGGAAGGAGCTGTGA

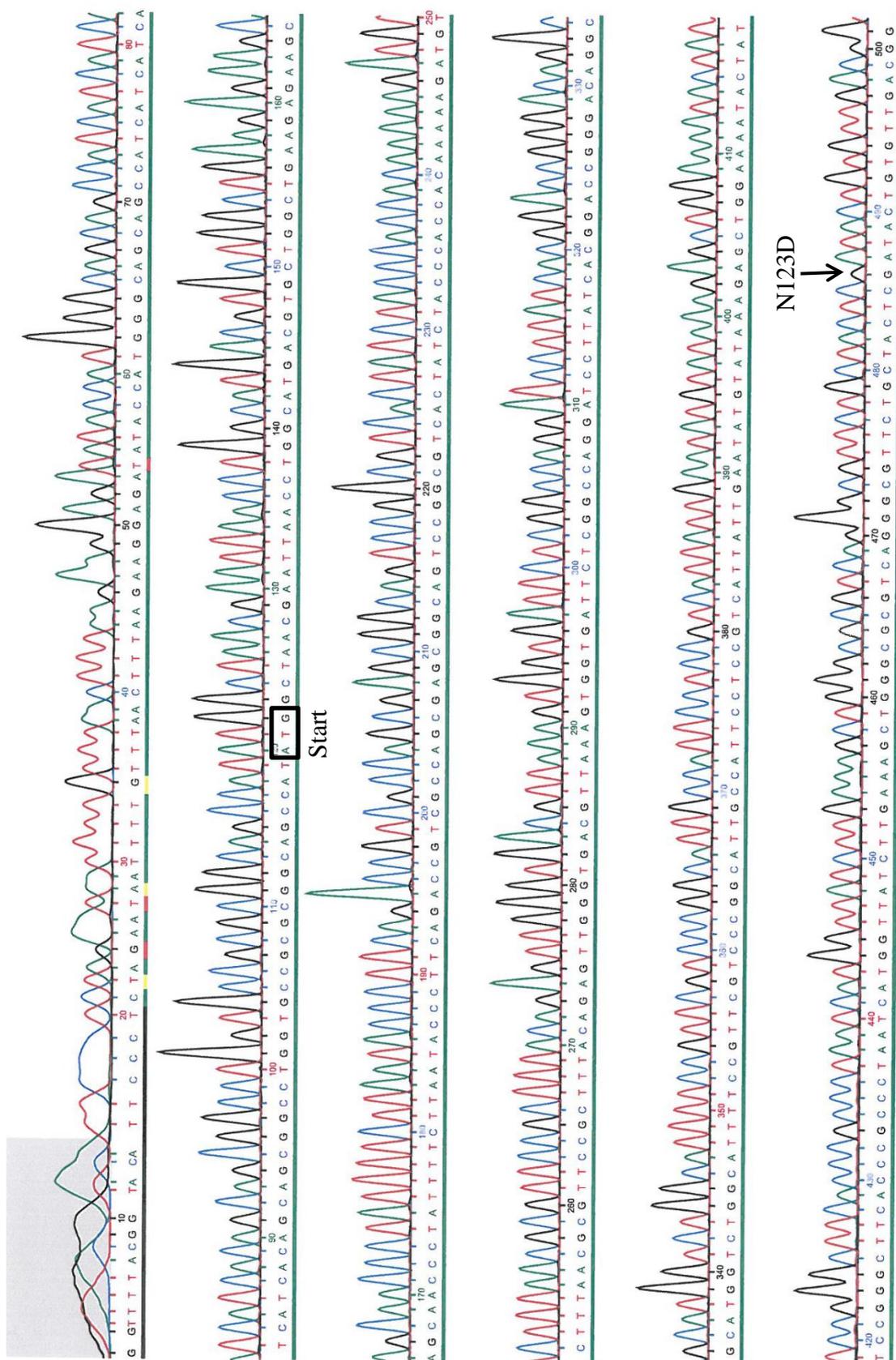
N204D: AAC → GAC

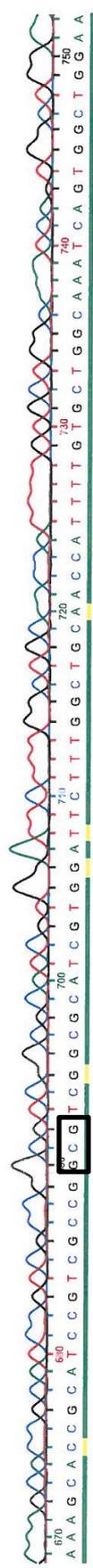
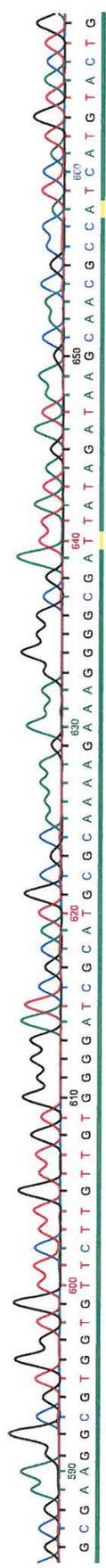
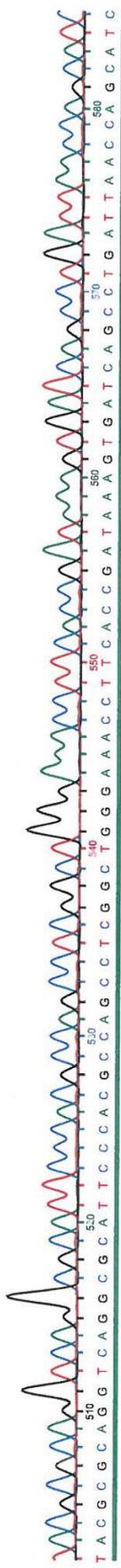
L272A: TTG → GCC

Orange: Deleted sequence of residues 3 - 83

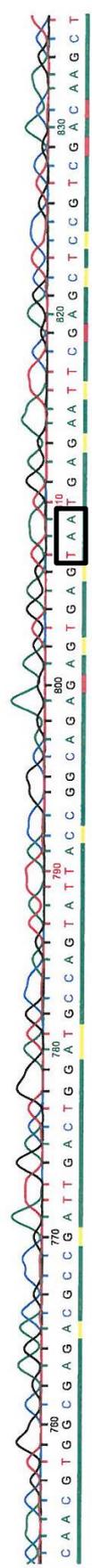
Appendix VI - Sequencing Chromatograms

pET₇CYDG

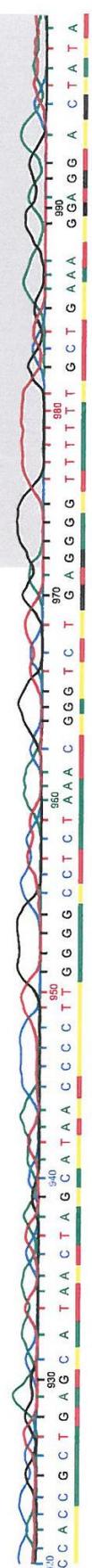
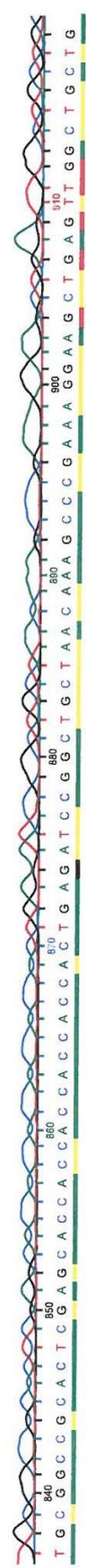


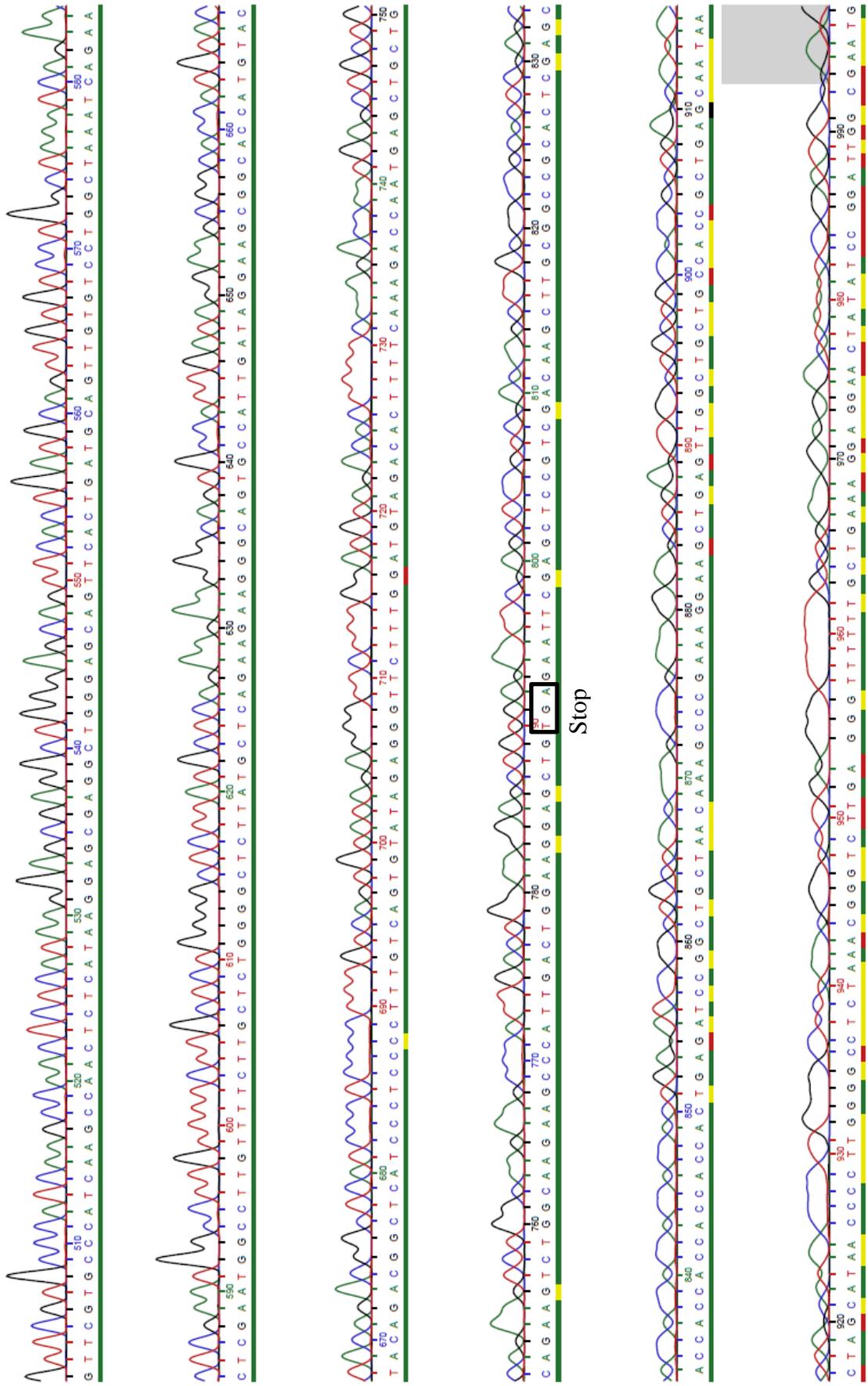


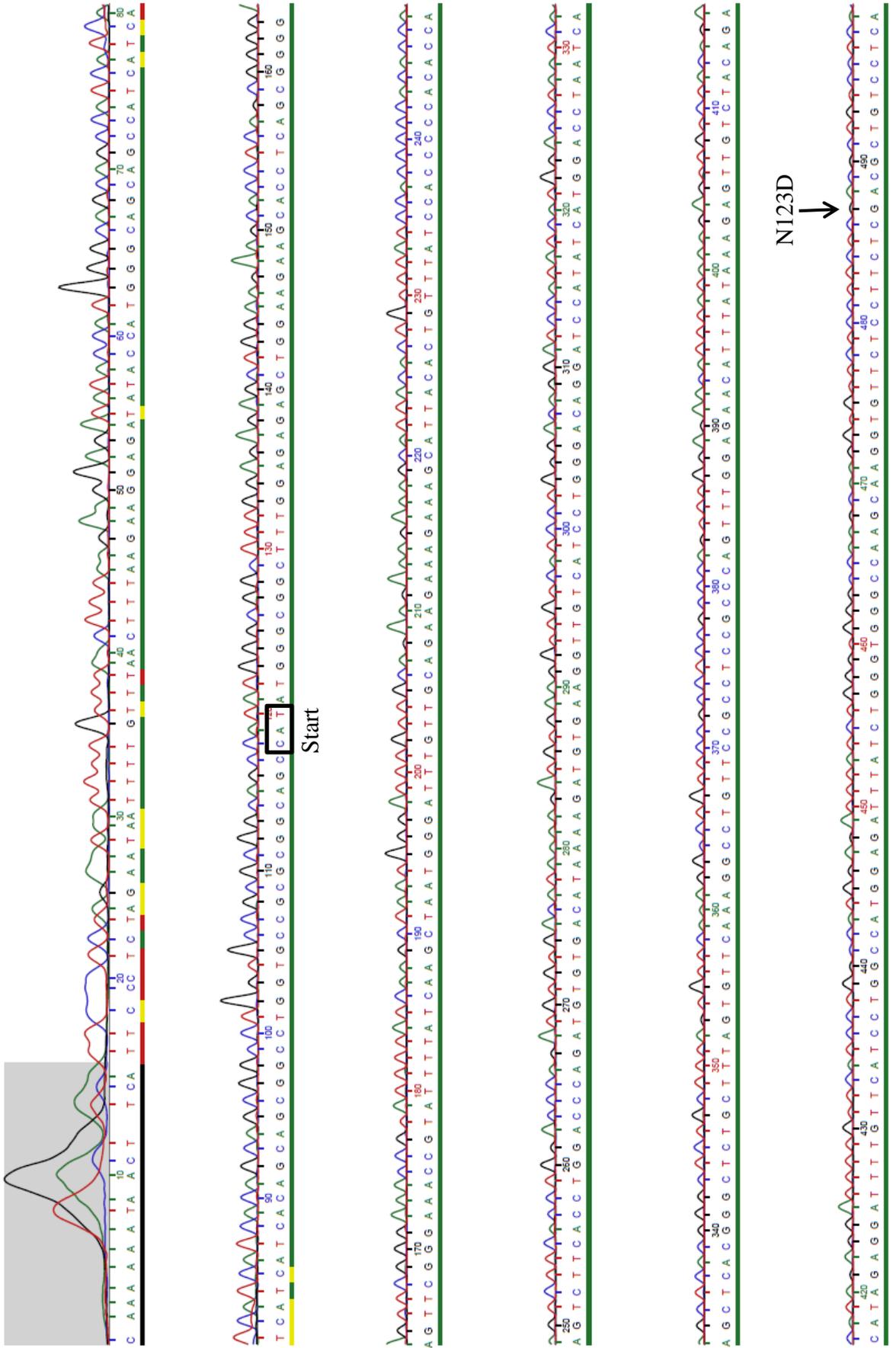
L191A



Stop







CGGTTCTGTCCTCAAGCCAACTCTCATAAAGGAGCGAGGCTGGAGCAGTTTCACTGATGCCAGTTGTCTCCTGGCTAAATCAGA
500 510 520 530 540 550 560 570 580

ACTCGAATGGCCCTTGTCTTCTTGGCTCTGGGCTTATGCTCAGAAAGGCGAGTGCATTGATAGGAAAGCGGCACCATGTA
590 600 610 620 630 640 650 660

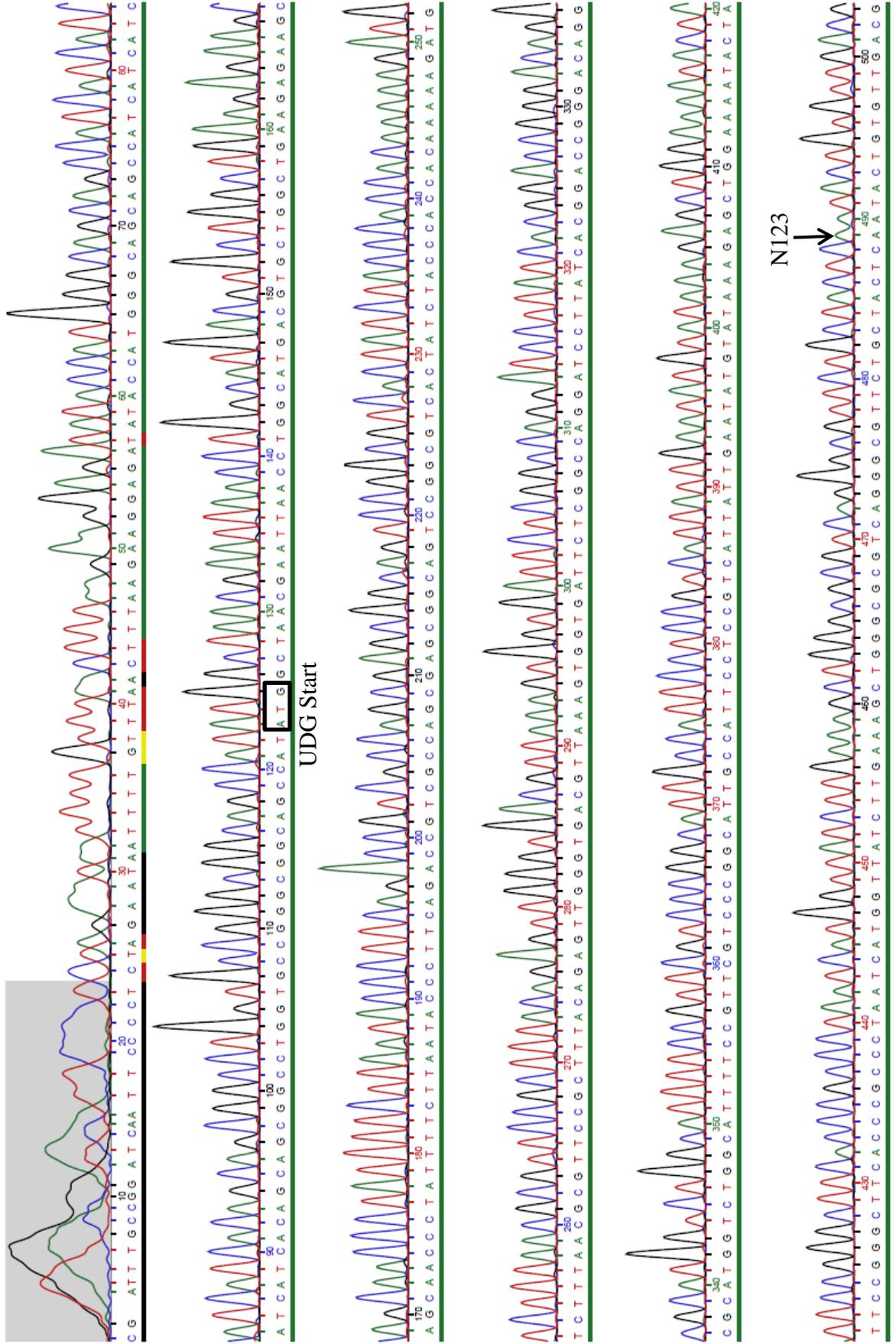
CTACGACGGCTCATCCCTCCCTTTGTCAGTATAGAGGTTCTTTGGATGTAGACACTTTCAAAGACCAATGAGCTGCTG
670 680 690 700 710 720 730 740 750

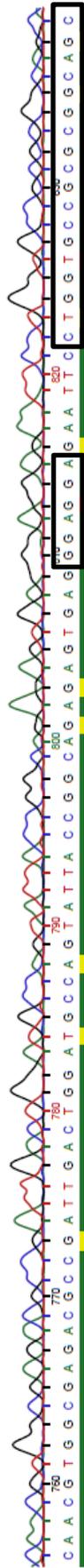
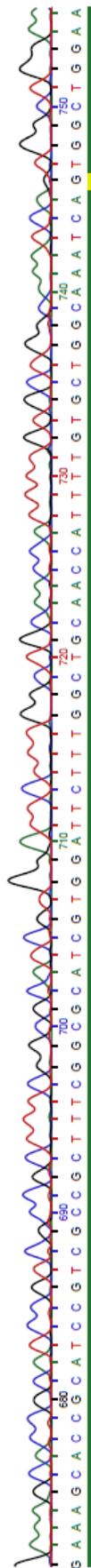
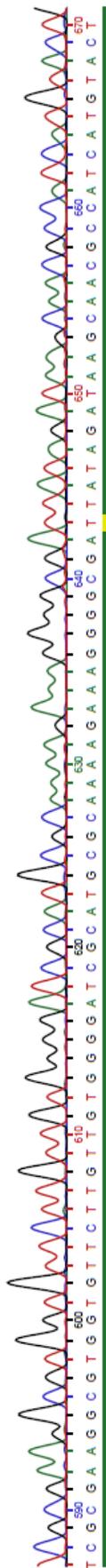
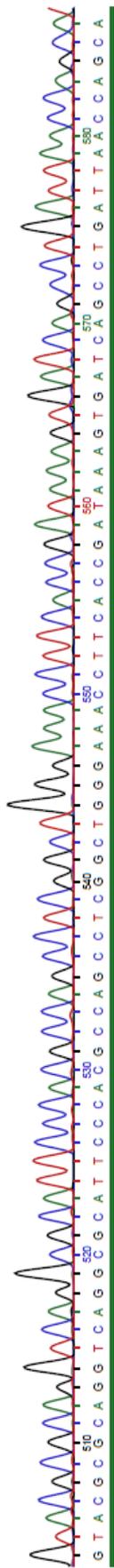
CAGAACTCTGGCAAGAAGCTTCACTTGAC TGG AAGGAGCTG TGA AATTCGAGCTCCGTCGACAAAGCTTGC GCCGCACTCGAGCA
760 770 780 790 800 810 820 830

Stop

CCACCACCACACTGAGATCCGGCTGCTGCTAACAA GCCCGAAGGAAAGCTGAGTTGGCTGCTGCCACC GCTGAGCAA TAACT
840 850 860 870 880 890 900 910

AGCA TAA CCCC TTGGG CCTCTAAC GGG TC TTGAGGGTTTTTTCCTGAAGGAGAACTATCCGGA TTGGC GAAT GGG AC
920 930 940 950 960 970 980 990 1000



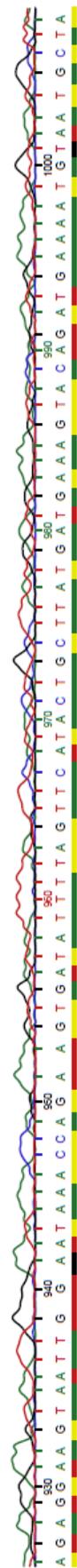


2 x Glycines

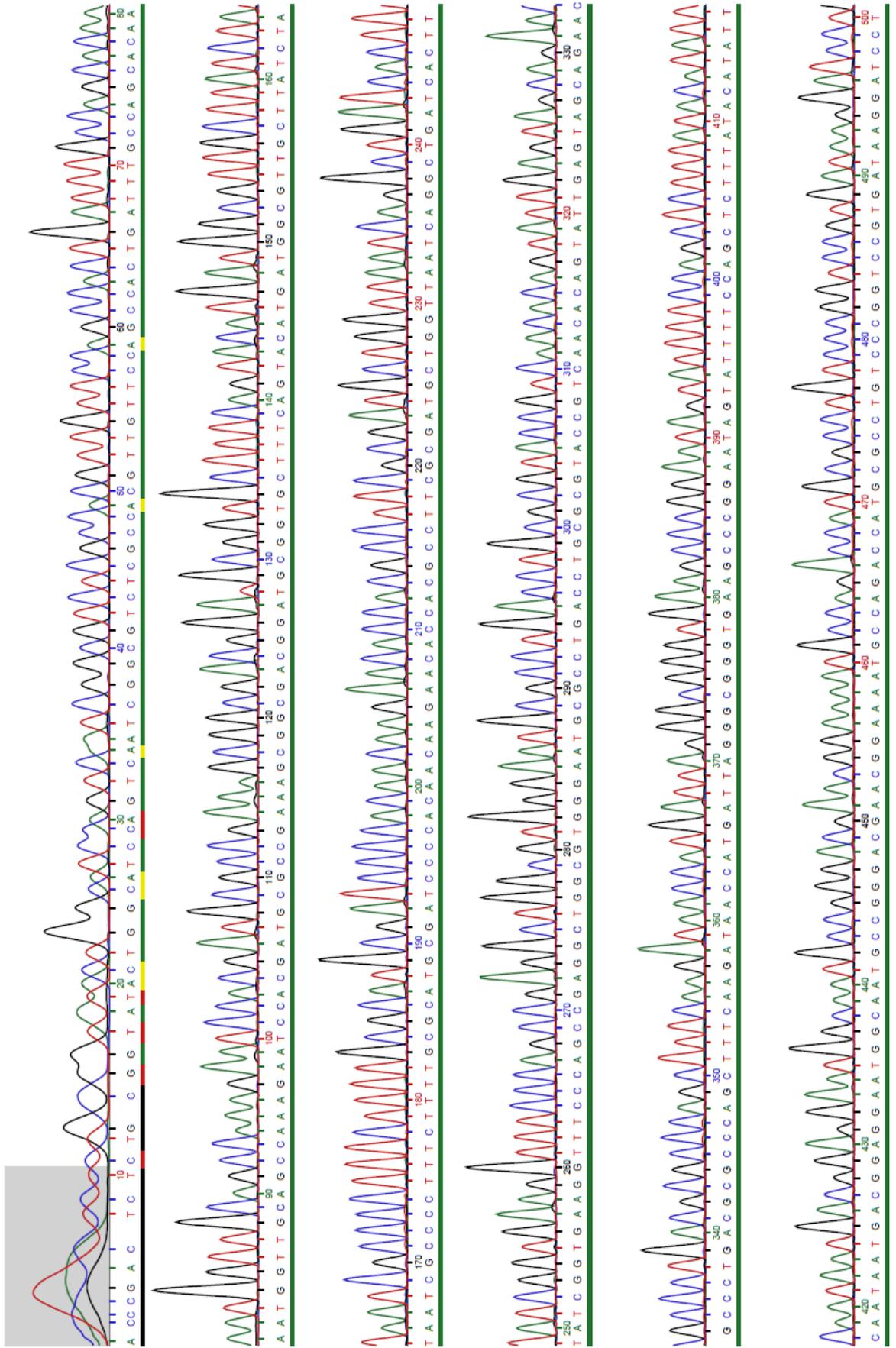
Thrombin Site

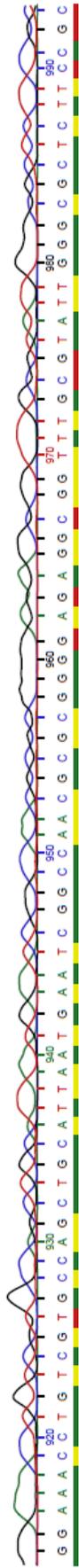
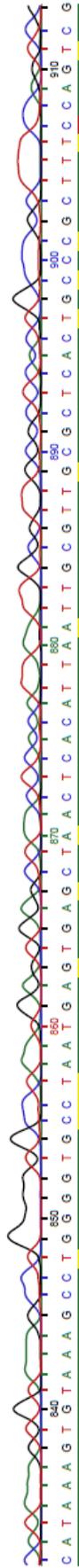
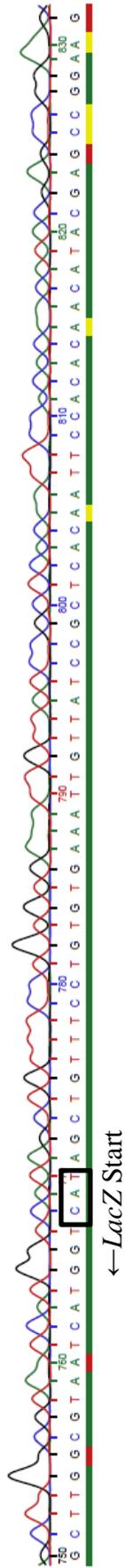
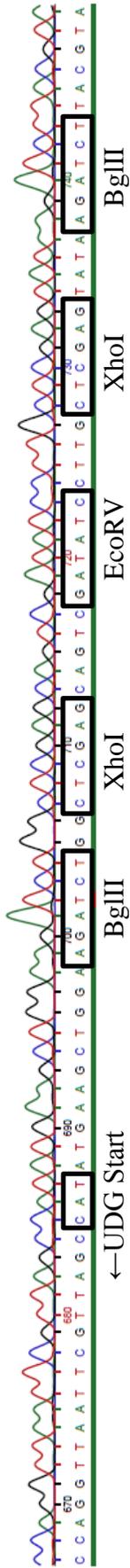
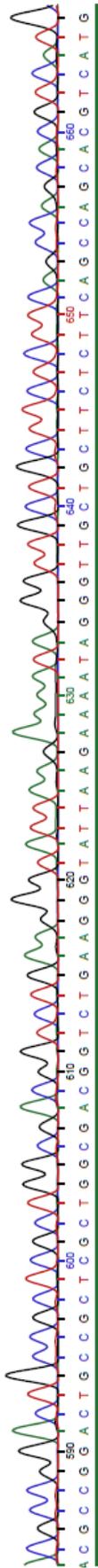
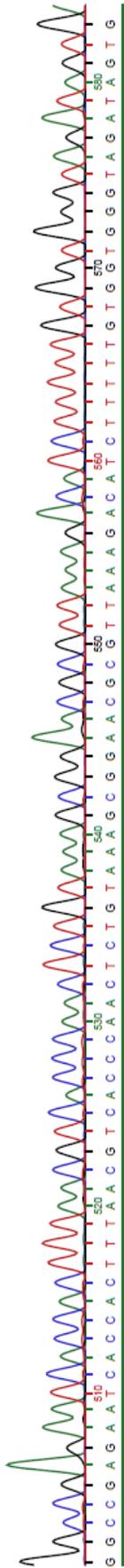


Ugi Start

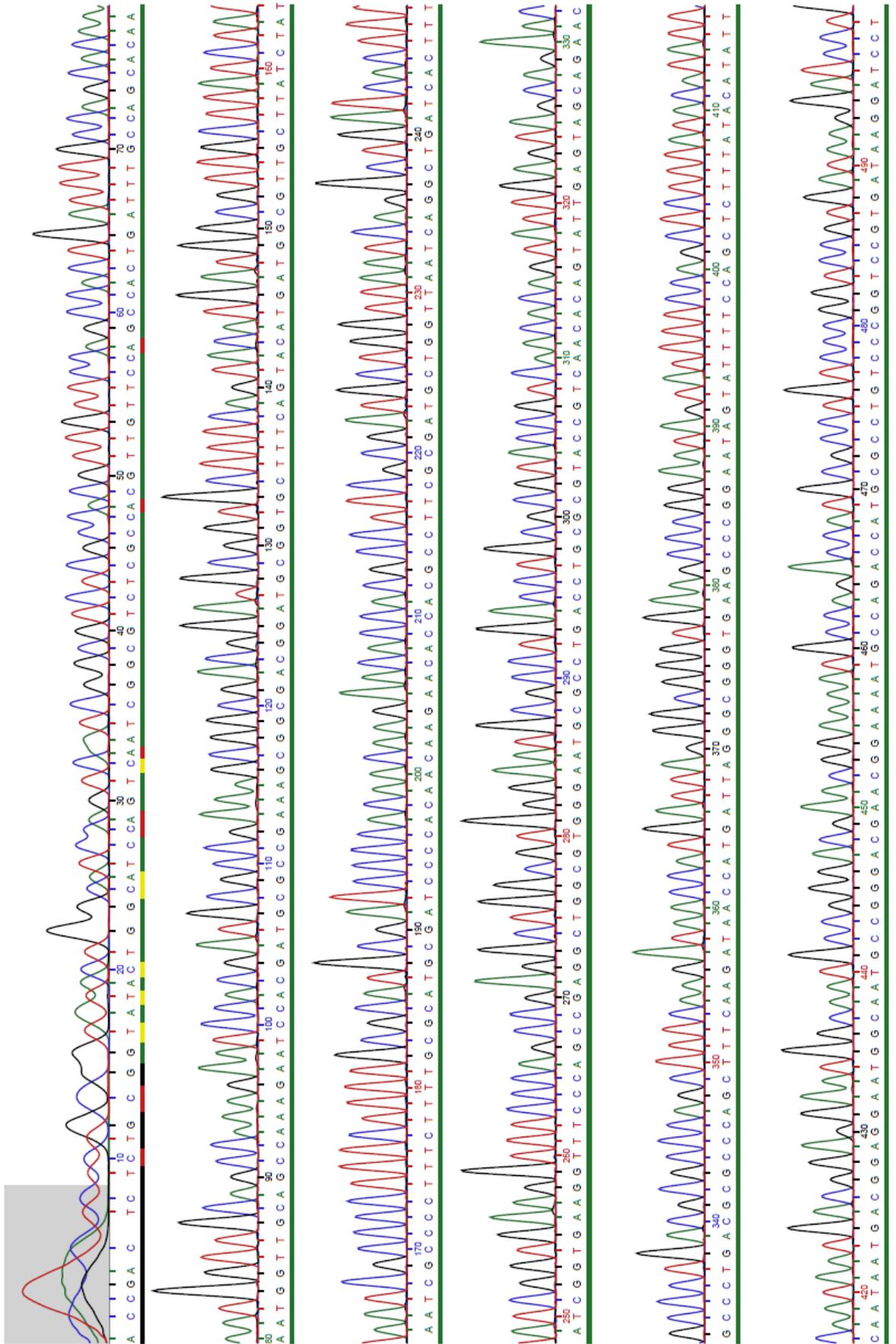


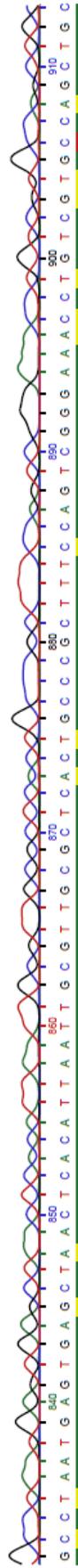
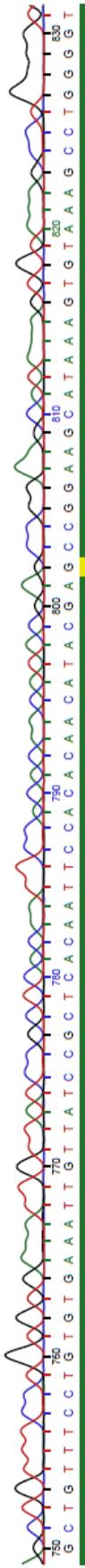
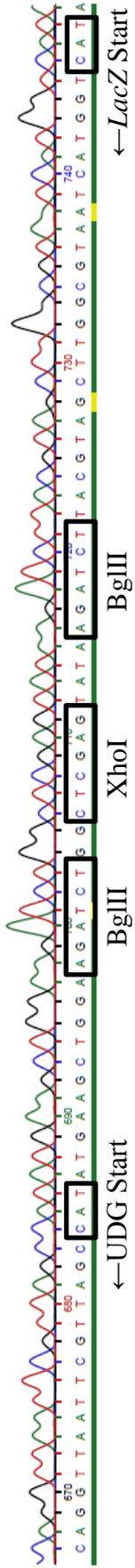
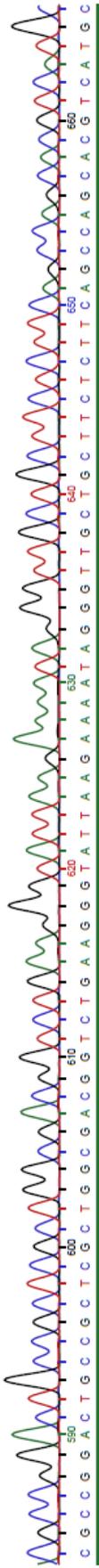
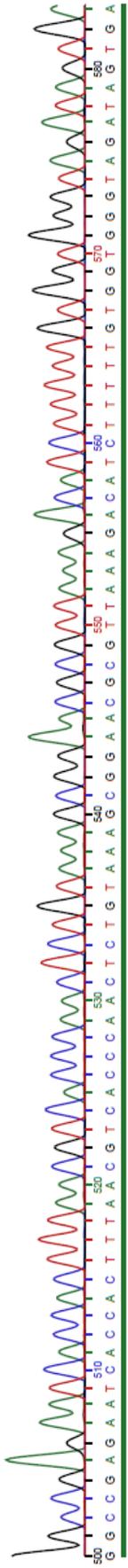
pUC19eUDG IRES





pUC19eUDG IRES X





pUC19eUDG IRES B

