Division of Biomedical

And Life Sciences

Lancaster University

MSc Biomedical Science (by research)

''Characterisation of Unusual DNA Glycosylases from *Trichomonas vaginalis*''

Amber Stephanie Hayes

Bsc (Hons) Biological Sciences

April 2019

I declare that this thesis is my own work and has not been submitted in part, or as a whole, for the award of a higher degree or qualification at this university or elsewhere.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Sarah Allinson for giving me the opportunity to work in her lab on this interesting project and for her guidance and feedback over the year. I would also like to thank the members of both the Allinson and Copeland labs for being so welcoming and answering any of my questions I had whilst working on my project. In particular I thank Daria Kania for her help at the start of my project.

My time at Lancaster would not have been what it was without all the amazing people I met in both my undergrad and my masters. I cannot begin to name them all, but I would particularly like to thank; Alice Padfield, Emily Bryson, Ellen Georgiou, Jessica Phoenix and Beth Smith for providing me with endless laughter, a love of darts and constant support whenever I needed it. Rosie Warburton and Jack Hill for three years of friendship and their words of encouragement this year. As well as Callum Ross and Danny Ward for their friendship during this year.

Lastly and most importantly, I would like to thank my mum for her unconditional love and support throughout my studies at Lancaster. Every trip to Lancaster was appreciated and this thesis would not have been possible without all she has done for me. I would not be where I am today without her.

Abstract

Trichomonas vaginalis is a parasitic protozoan responsible for 170 million new infections each year and is the most prevalent non-viral sexually transmitted infection worldwide according to The World Health Organisation. Research on the parasites pathogenicity and lifestyle has been limited until the publishing of the parasites' genome in 2007 and an emergence of studies documenting links between *T. vaginalis* infection and a number of adverse health outcomes, including an increased risk of HIV acquisition and cervical cancer. Research into DNA repair pathways within this organism have also been limited. In this research I describe two DNA 3-methyl adenine glycosylase homologues (TvAAG1 and TvAAG2) characterised via biochemical methods. In addition to the creation of an atlas of DNA repair proteins and enzymes present in *T. vaginalis,* and a close evolutionary relative, *Tritrichomonas foetus* via bioinformatical analysis. Biochemical analysis of TvAAG1 reveals it as the first bifunctional AAG type enzyme to be described. Bioinformatic analysis suggests TvAAG1 and TvAAG2 are most likely the result of horizontal gene transfer between an ancestor of both *T. vaginalis* and *T. foetus* and a donor belonging to the Bacteroidetes phylum. The DNA repair pathways of both parasites are reduced compared to humans. The parasites have the main components for base excision repair, mismatch repair and homologous recombination repair but not non-homologous end joining. In light of reports of antibiotic resistance further research should be done on the parasites DNA repair pathway to identify potential novel drug targets for treatment of *T. vaginalis* infection.

Contents Page

List of Figures

List of Tables and Appendices

List of Abbreviations

•OH – Hydroxyl radical

3'OH – (3'hydroxyl)

3'P – 3'phosphate

3meA – 3-methyl adenine

5'P – 5'phosphate

5meC – 5-methyl cytosine

7meG - 7-methylguanine

8oxoG - 8-oxoguanine

AAG – DNA 3-methyl adenine glycosylase

AGOG - *Pyrobaculum aerophilum* GO-glycosylase

Ala – Alanine

AlkA – *Eshcerichia coli* alkyladenine-DNA glycosylase

ALKs - alkylpurine DNA glycosylases

AP – Abasic (apurinic/apyrimidic) site

APE1 - Apurinic/apyrimidinic endonuclease 1

Arg – Arginine

BER – Base excision repair

BLAST - Basic Local Alignment Search Tool

BLM – Bloom syndrome protein

BRCA1 - Breast cancer predisposition gene 1

BRCA2 – Breast cancer predisposition gene 2

BSA – Bovine serum albumin

C. M. girerdii – Mycoplasma giredii

CB – Cyanoborohydride

CDC – The Centres for Disease Control

CDD – Conserved Domain Database

CDF – Cell detaching factor

dA – Deoxyadenosine

dC-Deoxycytidine

DDR – DNA damage response

dG – Deoxyguanosine

dH_2O – Distilled water

dI- Deoxyinosine

DMSO – Dimethyl sulfoxide

DNA – Deoxyribose nucleic acid

DNA-PKc - DNA-dependent protein kinase catalytic subunit

DNMT - DNA methyltransferase

dRP – 5'deoxyribosephosphate

DSB – Double strand break

DSBR – Double strand break repair

dT- Deoxythymidine

dU – Deoxyuridine

E value – Expected value

E. coli – Escherichia coli

E. cuniculi – Encephalitozoon cuniculi

E. histolytica – Entamoeba histolytica

ECM – Extra-cellular matrix

ERCC6 - Excision repair 6 chromatin remodelling factor

FCL – iron-sulfur cluster loop

FEN1 – Flap endonuclease 1

Fpg - Formamidopyrimidine-DNA glycosylase

H. sapiens – Homo sapiens

H2TH – Helix two turn-helix

HGT – Horizontal gene transfer

HhH – Helix – hairpin helix

HR – Homologous recombination

Hx – Hypoxanthine

IDL – insertion-deletion mismatches

IPTG - Isopropyl β-D-1-thiogalactopyranoside

K123A – TvAAG2 lysine 123 to alanine mutant

K123R – TvAAG2 lysing 123 to arginine mutant

K124A – TvAAG1 lysine 124 to alanine mutant

K124R – TvAAG1 lysine 124 to arginine mutant

L. infantum - Leishmania infantum

L. major - Leishmania major

LGT – Lateral gene transfer

LigI- DNA ligase I

LigIII – DNA ligase III

LP – BER – Long patch base excision repair pathway

M. hominis – Mycoplasma hominis

MAG - DNA-3-methyladenine glycosylase

MGMT - O^6-methylguanine DNA methyltransferase

MLC – Minimum lethal concentration

MMR – Mismatch repair

MPG - N-methylpurine DNA glycosylases

MpgII - N-methylpurine-DNA glycosylase II

MRN complex - MRE11-RAD50-NBS1 complex

MTZ – Metronidazole

MutSβ - MSH2-MSH3

MutSα - MSH2-MSH6

MutY/Mig - A/G-mismatch-specific adenine glycosylase

MYH – Adenine DNA glycosylase

NEB – New England Biolabs

Nei – Endonuclease VIII

NEIL1 – Nei-like DNA glycosylase 1

NEIL2 – Nei-like DNA glycosylase 2

NER – Nucleotide excision repair

NHEJ – Non-homologous end joining

Ni-NTA – Nickel – Nitrilotriacetic Acid

Nth – endonuclease III

OD – optical density

OGG1- 8-Oxoguanine DNA Glycosylase

OGG2 – 8-oxoguanine DNA glycosylase II

PCNA - proliferating cell nuclear antigen

PDB – Protein dialysis buffer

PNK – polynucleotide kinase

Pol δ - DNA polymerase δ

Pol β - DNA polymerase β

Polζ - DNA polymerase ζ

PUA – 3'phospho α,β-unsaturated aldehyde

pyr - pyrrolidine abasic nucleotide

RecQL5 – ATP-dependent DNA helicase Q5/recombinase

ROS – Reactive oxygen species

Rosetta™ - *Escherichia coli* Rosetta(DE3)pLysS cells

RPA – Replication protein A

S. cerevisiae – Saccharomyces cerevisiae

SAM - S-adenosylmethionine

SDM – site directed mutagenesis

SDS PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

SLS – Scientific Laboratory Supplies

SP – BER – Short patch base excision repair pathway

SSB – Single strand break

SSBR – Single strand break repair

T. foetus – Tritrichomonas foetus

T. vaginalis – Trichomonas vaginalis

T. brucei - Trypanosoma brucei

T. cruzi -Trypanosoma cruzi

TDG – Thymine DNA glycosylase

TMZ - Temozolomide

Turbo – *Escherichia coli* Turbo high efficiency cells

TvF – *Trichomonas vaginalis* factor

TvLF – *Trichomonas vaginalis* lateral gene transfer fragment

TVV- *T. vaginalis* virus

Tyr –Ttyrosine

UDG – Uracil-DNA glycosylase

Urea PAGE - denaturing urea polyacrylamide gel electrophoresis *

UV – ultraviolet radiation

VECs – Vaginal epithelial cells

WHO – World Health Organisation

XRCC1 – X-ray cross complementing protein 1

εA – 1-N^6-ethenoadenosine

1

Introduction

1.1 DNA Repair

Deoxyribonucleic acid (DNA) is the molecular carrier of genetic information in cellular organisms. The faithful replication and translation of the double helix structure, famously described by Watson and Crick in 1953, is a fundamental requirement in all organisms to ensure integrity of the genome. DNA consists of two polynucleotide strands linked together by hydrogen bonding between the bases of the nucleotides. Nucleotide monomers are made up of; one of four nitrogenous bases, a deoxyribose sugar, and a phosphate molecule. There are four types of bases; adenine, guanine, cytosine and thymine. Each one has a different chemical structure which allows them to be read by DNA processing enzymes. Despite DNA being a stable molecule, its bases are highly reactive (Lindahl, 1993). Spontaneous chemical reactions between the bases and other molecules within the cell can lead to base modifications, formation of abasic sites (AP sites) and single strand or double strand breaks. All of which can lead to mutations if not repaired and can be cytotoxic. These chemical alterations to the DNA are known as DNA lesions.

Up to 70,000 genomic lesions are formed in human cells per day (Lindahl and Barnes, 2000) due to constant attack by endogenous chemical reactions and exogenous DNA damaging agents. This means an efficient and effective DNA repair mechanism is essential in all organisms to prevent the effects of DNA damage and an increased mutational load. The variety of different DNA repair mechanisms that can combat these lesions and other types of DNA damage are collectively known as the DNA damage response (DDR) (and Elledge, 2007). The importance of DDR is highlighted by the evolutionary conservation of DNA repair. Defective repair systems can cause genomic instability leading to mutations and cell death or disease. Genomic instability is an enabling characteristic of the eight hallmarks of cancer (Hanahan and Weinberg, 2011) and mutations in specific DNA repair genes have been linked to an increased risk for several types of cancer; breast cancer predisposition gene 1 and 2 (BRCA1/BRCA2) mutations are associated with breast, ovarian and prostate cancer (Mavaddat et al., 2013; Pilié et al., 2017) adenine DNA glycosylase (MYH) mutations are linked to colorectal adenomas (Sieber et al., 2003) and ERCC2 mutations are associated with bladder cancer (Kim et al., 2016). This highlights the importance of the DDR system in respect to human disease as it has also been linked to Alzheimer's and ageing (Yu, Harrison and Xia, 2018; Hewitt et al., 2012)

1.1.1. Types of DNA Damage

DNA is subjected to various types of damage within the cell. Damage that occurs within the cell is known as endogenous DNA damage and includes base deamination, oxidative damage, depurination and methylation. Endogenous damage occurs when DNA reacts with naturally occurring molecules within the cell, such as water and reactive oxygen species (ROS), resulting in spontaneous DNA lesions. DNA is also damaged by environmental agents, such as radiation and alkylating agents, which is termed exogenous DNA damage.

1.1.1.1 Endogenous DNA Damage

Base deamination is a major source of endogenous damage to DNA. Within the cell, naturally present H_2O molecules participate in spontaneous hydrolysis reactions with DNA bases which results in a loss of their exocyclic amine (NH_2) (Lindahl, 1993). Adenine, guanine, cytosine and 5-methyl cytosine bases can be deaminated. Their deamination products are hypoxanthine (Hx), xanthine, uracil and thymine respectively (Figure 1.1). When bound to a deoxyribose sugar deoxyadenosine is deaminated to deoxyinosine (Hx is the corresponding base). Deoxyinosine formation leads to transition mutations as it is incorrectly read by DNA replicative machinery as deoxygaunosine due to its structure. This leads to the insertion of deoxycytidine in place of deoxythymidine (Yasui et al., 2008). Cytosine and 5-methyl cytosine (5meC) show the highest rates of deamination, with 5meC being more frequently deaminated than cytosine (Lindahl, 1979). Deamination of both cytosine and 5meC is repaired via the

base excision repair pathway (BER). Deamination of 5meC leads to G:T mismatches which are efficiently excised via BER with thymine DNA glycosylase (TDG). Deamination of cytosine causes the formation of uracil and leads to GC →TA mutations if not recognised and excised by uracil DNA glycosylase (UDG).

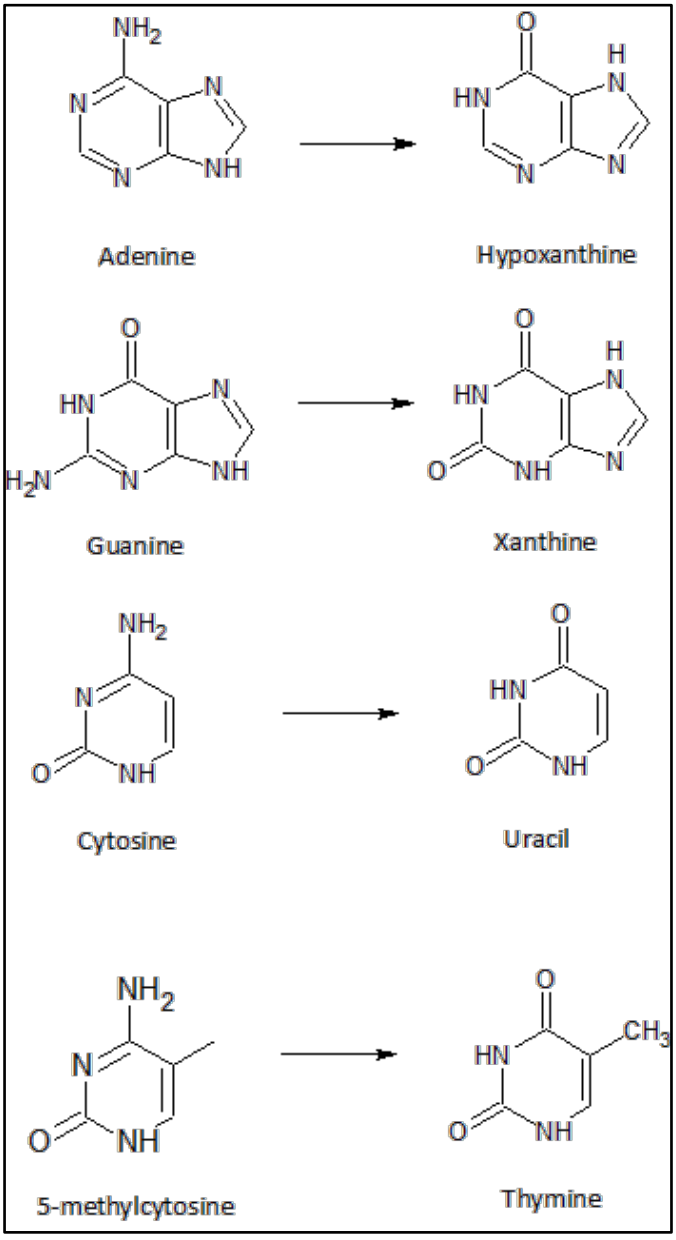


Figure 1.1: The Deamination of Bases with Exocyclic Amines via Hydrolysis.

Hydrolysis reactions within the cell result in the loss of the exocyclic amine (NH_2) group from the base. Adenine is deaminated to hypoxanthine, guanine is deaminated to xanthine, cytosine is deaminated to uracil and 5-methyl cytosine is deaminated to thymine.

Oxidative damage via endogenous attack by reactive oxygen species (ROS) causes a variety of DNA damage. ROS are highly electrophilic, and the most reactive species is the hydroxyl radical (•OH) (Tropp, 2011). Hydroxyl radicals react with DNA by attacking the sugar residue leading to single stranded DNA breaks (Giloni et al., 1981), removing hydrogen atoms from methyl groups and by adding to their double bonds both of which cause oxidised base damage.

ROS are normally produced within the mitochondria as a by-product of cellular respiration and peroxisomal metabolism (Henle and Linn, 1997). They are essential as cellular messengers in redox signalling and in signalling defence responses against pathogens (Errol C. Friedberg, 2005; Segal, 2005) However, excess ROS levels have been linked to diseases including cancer (Liou and Storz, 2010) and Alzheimer's (Mohsenzadegan and Mirshafiey 2012) and can cause over 100 different types of oxidative base lesions. The saturated imidazole ring 8-oxoguanine (8oxoG) is one of the most frequent lesions formed by oxidative damage (Ames, 1989). The guanine base is the most readily oxidised of all the DNA bases. 8oxoG incorrectly pairs with adenine in place of cytosine leading to mutations (Figure 2).

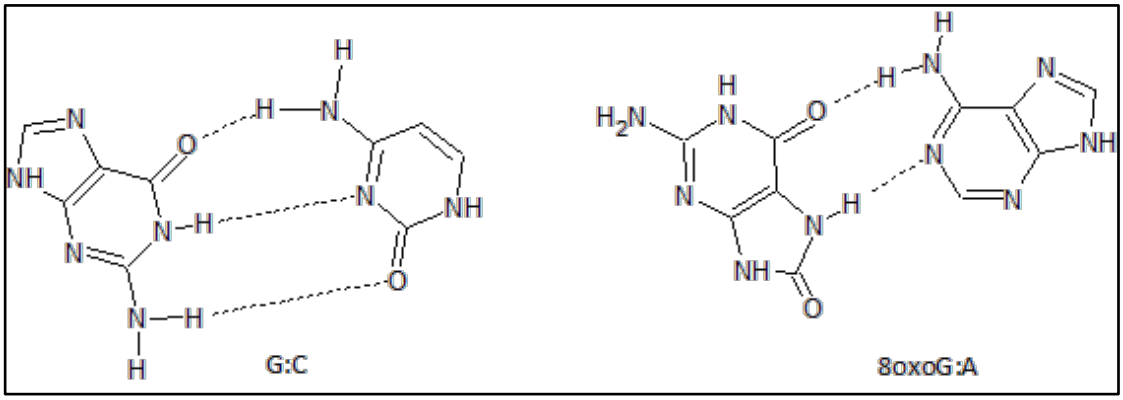


Figure 1.2: Correct and Mismatch Pairing of Guanine and 8oxoG. a) guanine and cytosine b) 8oxoG and adenine mismatch. 8oxoG – 8-oxoguanine.

Another source of DNA damage is depurination via hydrolysis. Depurination causes the formation of AP sites and single strand breaks within the DNA. The generation of AP sites is an important step during the BER pathway (Lindahl and Barnes, 2000) but can be spontaneously produced when the *N*-glycosyl bond linking the nitrogenous base and phosphate groups of the DNA monomer is hydrolysed. Spontaneous generation of AP sites is affected by extreme pH and temperature conditions within the cell and around 10,000 AP sites are created per day within human cells (Lindahl, 1993) These AP sites can readily convert into single strand breaks due to AP sites being inherently unstable.

Epigenetic modification refers to the heritable modification of the genome outside of changes to the DNA sequence. DNA methylation is a post-replication modification involved in regulation of gene expression which involves the use of DNA methyltransferases (DNMTs) and S-adenosylmethionine (SAM) as a methyl donor (Moore, Le and Fan, 2012). Outside of normal epigenetic modification SAM acts as a source of endogenous damage as it can spontaneously produce over 4000 lesions per cell per day in mammals (Rydberg and Lindahl 1982). The lesions produced by SAM in Rydberg and Lindahl's experiment included 7-methylguanine (7meG), 3-methyladenine (3meA) and O^6-methylguanine.

The formation of 7meG and 3meA by abnormal methylation is cytotoxic as these DNA adducts block replication or form mutagenic AP sites if not repaired. Destabilization of the glycosyl bond on 7meG can generate imidazole ring opening which blocks replication (Barbarella et al., 1991). 7meG accounts for around 75% of methylated lesions in DNA (Beranek, 1990). 3meA is cytotoxic due to its interference

with polymerases and the DNA minor groove during replication (Larson et al., 1985).

Endogenous alkylation can also form the highly mutagenic O^6-methylguanine O^4-methylthymine, O^4-ethylthymine lesions which cause GC → AT and TA → CG transition mutations if replicated before repair (Warren, Forsberg and Beese, 2006). Table 1.1 lists the all possible positions of base modifications caused by alkylating agents. Alkylated lesions can also be formed by the effects of exogenous agents, as discussed in section 1.1.1.2.

Modified base	**Position**
Adenine	N1, N3, N^6, N7
Cytosine	N3, N^4, and O^2
Guanine	N1, N^2, N3, N7, and O^6
Thymine	N3, O^2, *and* O^4

Table 1.1: Possible Positions of Base Modifications Caused by Alkylating Agents. Exocyclic positions on DNA bases italicized superscripts. (Singer and Kusmierek, 1982; Singer, 1986; Errol C. Friedberg, 2005).

	DNA lesion	DNA repair mechanism
Deamination	Hypoxanthine Xanthine Uracil Thymine	BER
Oxidation	Oxidised purines and pyrimidines e.g 8-oxoguanine	BER
	SSBs	SSBR pathways
	DSBs	DSBR- homologous recombination Non-homologous end joining
Depurination	AP sites	BER
	SSBs	SSBR
DNA Methylation	*O*^{6}-methylguanine *O*^{4}-methylthymine *O*^{4}-ethylthymine *N*^{3}-methyladenine *N*^{7}-methylguanine	MMR or direct reversal of damage via methyltransferases e.g MGMT BER
Ionising radiation	Oxidised purines and pyrimidines SSBs DSBs	BER SSBR DSBR
UV radiation	Pyrimidine dimers	NER
	Oxidised purines and pyrimidines	BER
	Single strand breaks	SSBR
	Double strand breaks	DSBR homologous recombination Non-homologous end joining
Alkylating agents	*O*^{6}-methylguanine 1,N^{6}-ethenoadenosine	MMR or direct reversal of damage via methyltransferases e.g BER

Table 1.2: Biologically Significant Lesions Produced via Exogenous and Endogenous DNA Damage and the DNA Repair Mechanisms that Recognise the Lesions. BER- base excision repair, DSB- double strand break, DSBR- double strand break repair, MGMT – O^{6}-methylguanine DNA methyltransferase, MMR- mismatch repair, SSB- single strand break, SSBR – single strand break repair.

1.1.1.2 Exogenous DNA Damage

Exogenous damage occurs when substances and chemical agents from outside the cell damage the DNA directly or indirectly. The main sources of endogenous damage are radiation sources and alkylating agents.

Ionizing radiation includes, alpha, beta and gamma radiation, neutrons and X-rays. It is produced from many sources within the environment including; X-rays from medical devices, cosmic radiation and soil and rocks (World Health Organisation, 2016). Ionising radiation can cause direct and indirect DNA damage. Direct damage produces oxidised bases or single strand breaks (SSB) with unique 3'-phosphoglycolate modified ends (Henner et al., 1982; Obe et al., 1992). Indirect damage via ionising radiation involves generation of hydroxyl radicals via radiolysis of water within the cell (Friedberg, 2005) Damage via hydroxyl radicals accounts for 65% of DNA damage caused by ionising radiation (Vignard et al., 2013) Ionising radiation also causes toxic double strand breaks through oxidation of base and sugar residues closely located on opposite strands of the double helix (Hutchinson, 1985; Iliakis, 1991)

Ultraviolet radiation (UV) is another significant source of DNA damage, this type of radiation is emitted from the sun and is the major cause of skin cancer in humans (Davies, 1995; Kiefer 2007) UV radiation can be categorized into three groups based on its wavelength (Table 1.3). DNA absorbs the maximum amount of UV radiation at 260 nm and then absorption drops. UV-C is the most dangerous as its wavelength range includes the maximum absorbance for DNA. Sunlight is composed of a small percentage of UV-A and UV-B, visible light and infrared. The UV-C is filtered out by

atmospheric oxygen (Yu and Lee, 2017). UV-C and UV-B damage DNA by creating covalent linkages between two pyrimidines leading to the production of cyclobutane pyrimidine dimers and pyrimidine (6-4) pyrimidinone photoproducts (Varghese, 1972; Mitchell and Nairn, 1989; Davies, 1995) These two photoproducts are bulky dimers which distort the DNA helix leading to cytotoxicity if not repaired by nucleotide excision repair (NER) (Yu and Lee, 2017). Other photoproducts generated include; pyrimidine hydrate, thymine glycols and dipurine adducts (Demple and Linn, 1982; Bose et al., 1983; Kumar et al., 1991; Mitchell et al., 1991). UV-B radiation also causes the formation of pyrimidine dimers but at less frequency than UV-C (You et al., 2000; Errol C. Friedberg, 2005; Rastogi et al., 2010). UV-A radiation causes the formation of 8-oxoG lesions and cyclobutane pyrimidine dimers (Epe, 1991; Rochette et al., 2003).UV-A can also cause DNA strand breaks (Peak and Peak, 1986; Errol C. Friedberg, 2005).

UV Class	**Wavelength Range (nm)**
UV-A	320-400
UV-B	290-320
UV-C	190-290

Table 1.3: Types of UV Radiation and their Respective Wavelength Ranges (nm).

DNA alkylation, including endogenous methylation as previously discussed, can also be caused by exogenous alkylating agents to form DNA lesions. Alkylating agents are commonly used as chemotherapeutic agents, such as temozolomide (TMZ) and cyclophosphamide (Kondo et al., 2010). Tobacco smoke contains nitrosaminoketone which reacts with DNA to produce O^{6}-methylguanine (Goodsell, 2004). Alkylating agents are electrophilic with high affinity for base ring nitrogen and oxygen atoms forming DNA adducts (Table 1.2) which are removed by direct demethylation, BER and mismatch repair (MMR). Urethane and vinyl chloride exposure causes 1,N^{6}-ethenoadenosine (εA) production. εA is a bulky alkylation adduct which is highly mutagenic (Pandya and Moriya, 1996).

1.1.2. Base Excision Repair

Base excision repair (BER) is one of many DNA repair pathways which make up the DDR. The general features of the base excision repair pathway were first outlined by Thomas Lindahl in 1973 with his discovery of the *Escherichia coli* (*E. coli*) uracil-DNA glycosylase. The general mechanism of BER involves initiation by a DNA glycosylase which removes the damaged base leaving an AP site. The AP site is removed by an AP lyase or AP endonuclease and downstream proteins, including DNA polymerases and DNA ligases, work to ensure the DNA is repaired.

DNA glycosylases can be monofunctional or bifunctional. Monofunctional glycosylases have intrinsic DNA glycosylase activity whereas bifunctional glycosylases have glycosylase and additional AP lyase activity. Monofunctional

glycosylases, such as N-methylpurine DNA glycosylase (MPG) or UDG, initiate BER after recognition of a specific lesion whilst scanning duplex DNA. The glycosylase 'pinches' the DNA whilst scanning causing helix distortion and allowing the damaged base to be flipped out of the major groove into a substrate binding pocket to form a protein-substrate complex (Huffman et al., 2005). Within the substrate binding pocket an activated water molecule is used to hydrolyse the *N*-glycosidic bond between the base and deoxyribose molecule which generates an AP site (Schermerhorn and Delaney, 2014). AP sites are removed by an AP endonuclease, APE1 in mammalian cells (Demple, Herman and Chen, 1991), which cleaves the DNA backbone 5' to the AP site which generates a strand break with a 3'hydroxyl (3'OH) and 5'deoxyribosephosphate (dRP) ends.

Bifunctional glycosylases, such as 8-Oxoguanine DNA Glycosylase (OGG1), use an amine nucleophile to catalyse the glycosidic bond and DNA backbone cleavage. Bifunctional glycosylases have intrinsic β-lyase or β/δ-lyase activity which cleaves DNA backbone 3' to the AP site. β-elimination proceeds via formation of a Schiff base, a covalent protein DNA intermediate, to generate a single strand break with 3'phospho α,β-unsaturated aldehyde (PUA) and 5'phosphate (5'P) termini (Schermerhorn and Delaney, 2014). Schiff base intermediates bases can be detected by reduction with borohydride to forms a covalent DNA-glycosylase adduct that's detectable in SDS PAGE experiments. If the bifunctional glycosylase has β/δ-lyase activity the DNA backbone is cleaved in two steps, β-elimination followed by δ-elimination to generate 3'-phosphate (3'P) (Wiederhold et al., 2004).

The non-conventional termini of the single strand breaks generated by AP endonuclease or AP lyase block DNA polymerase extension and DNA ligase nick ligation which require conventional 3'OH and 5'P ends (Wilson, 2007). These non-conventional blocking residues, 5'dRP, 3'PUA and 3'P, need to be removed by end-processing enzymes before DNA repair can be completed. DNA polymerase-β has intrinsic dRP excision activity in addition to its polymerase activity which allows it to remove the 5'dRP ends left by APE1 and generate the 5'P required for ligation (Prasad et al., 1998). DNA polymerase λ and DNA polymerase ι can also clean up dRP residues (Bebenek, 2001; Garcı´a-Dı´az et al., 2001). 3'PUA and 3'P ends can be cleaned up by APE1 to leave conventional 3'OH ends. Polynucleotide kinase (PNK) is the primary enzyme that removes 3'P blocking termini (Jilani et al., 1999).

After the action of end processing enzymes to generate conventional 3'OH and 5'P ends Polβ inserts a single nucleotide at the 3'OH end of the nick, followed by nick sealing by a DNA ligase which generates a covalent phosphodiester bond to join the upstream 3'OH end and downstream 5'P ends. This pathway of processing the AP site, via removing blocking termini and the insertion of a single nucleotide, is known as the short patch base excision repair pathway (SP-BER). DNA ligase IIIα is the DNA ligase associated with SP-BER, it acts in a complex with scaffold protein X-ray cross-complementing protein 1 (XRCC1) (Cappelli et al., 1997).

If 5' blocking residues are not removed BER proceeds down the long patch repair pathway (LP-BER) to process the AP site. In LP-BER the AP site is removed by an AP endonuclease leaving a conventional 3'OH terminus at the 5' side to the nick and a 5'dRP blocking terminus at the 3' side to the nick. DNA polymerase δ or DNA

polymerase ε bind to the 3'OH substrate and insert up to 12 bases, which includes the removed base plus several bases downstream. This strand displaces the strand at the 3' side of the nick containing the 5'dRP terminus to generate a flap intermediate. The displaced strand is removed by flap endonuclease 1 (FEN1) and LigI, the DNA ligase associated with LP-BER (Tomkinson et al., 2006), seals the nick between the 3'OH and 5'P ends. The accessory protein proliferating cell nuclear antigen (PCNA) facilitates DNA synthesis with Polδ and Polε by binding to FEN1 (Gary et al., 1999).

1.1.3. Glycosylases and DNA 3-methyladenine Glycosylase

DNA glycosylases are an essential component of the base excision repair pathway. This is highlighted by the fact they are evolutionary conserved throughout the kingdom of life. DNA glycosylases can be found in eukaryotes (*Homo sapiens* (*H. sapiens)* AAG, *Saccharomyces cerevisiae* (*S. cerevisiae*) DNA-3-methyladenine glycosylase (MAG)), prokaryotes (*E. coli* alkyladenine-DNA glycosylase (AlkA)), and archaea (*Methanococcus jannaschii* 8-Oxoguanine DNA glycosylase II (OGG2)). Glycosylases are diverse and can excise a wide variety of DNA lesions.

Glycosylases can be sub-classified into six super families based on their structures. The four super families found within mammalian species include; the helix-hairpin-helix (HhH) glycosylases, the uracil glycosylases, the endonuclease VIII-like glycosylases and the 3-methylpurine glycosylases (Jacobs and Schär, 2011). The remaining two super families are found within bacterial and archaic genomes; helix-two turn-helix glycosylases (H2TH) and the alkylpurine DNA glycosylases (ALKs)

(Brooks et al., 2013) All classes of glycosylases have evolved a base-flipping mechanism of recognition except from the ALKs (Brooks et al 2013).

The HhH super family, also known as endonuclease III Nth-like superfamily, consists of glycosylases which share a HhH motif and catalytic aspartate residue and have similar 3-dimensional structures. The prototype for this family is *E. coli* endonuclease III (Nth) of which the crystal structure was used to identify the characteristic features of the HhH superfamily (Thayer et al., 1995). Most endonuclease III homologs also have an iron-sulfur cluster loop (FCL). The HhH super family can be further divided into six distinct families based on phylogenetic analysis of 234 HhH glycosylases from 94 genomes by Denver, Swenson and Lynch (2003), which named the families based on functionally characterised glycosylases which were included in the specific group. These six sub-families include Nth, OggI, A/G-mismatch-specific adenine glycosylase (MutY/Mig), AlkA, N-methylpurine-DNA glycosylase II (MpgII) and OggII. An additional archaeal sub-family was identified in 2005 by Lingaraju et al. through the crystallisation of *Pyrobaculum aerophilum* GO-glycosylase (AGOG)which contains the HhH motif, founding the AGOG family. Analysis of the crystal structure of *E. coli* TAG showed it to be an additional member of the HhH super family (Drohat et al., 2002).

The UDG super family consists of monofunctional glycosylases which remove uracil from DNA. The superfamily can be subdivided into six subfamilies based on characteristic motifs. Families I-III are found within eukaryotes; Family I/UDGs, family II/MUGs, family III/SMUGs. Families IV, V and VI are found in thermophilic

and hyper thermophilic bacteria and archaea (Schormann, Ricciardi and Chattopadhyay, 2014).

The endonuclease VII-like glycosylase superfamily also known as the Fpg/Nei superfamily includes bifunctional glycosylases which share a highly conserved C-terminus containing a helix-two-turn helix motif and a conserved zinc-finger motif (Boiteux, Coste and Castaing, 2017). This group includes *E. coli* formamidopyrimidine-DNA glycosylase (Fpg) (Boiteux, O'Connor and Laval, 1987), *E. coli* endonuclease VIII (Nei) (Saito et al., 1997), and mammalian homologs Nei-like (NEIL) NEIL1 (Hazra et al., 2002a) and NEIL2 (Hazra et al., 2002b).

The fourth super family of glycosylases found in mammalian cells is the N-methylpurine DNA glycosylases that do not contain HhH motifs and so form a separate structural class of N-methylpurine glycosylases. Human MPG also known as DNA 3-methyladenine glycosylase (AAG) repairs alkylation and oxidative damage by excising a wide range of lesions including; 3meA, 7meG, εA and Hx (Engelward et al., 1997). The glycosylases in this super family share similarity with the protein core or active site of human AAG. The structure of AAG consists of a single domain of mixed α/β helices and strands. The first MPG was originally identified in rat hepatoma cells (O'Connor and Laval, 1990) and other glycosylases within this class can be found in; mouse, *Arabidopsis thaliana* (Santerre and Britt, 1994), *Bacillus subtilis* (Morohoshi et al., 1993), *Borrelia burgdorferi* (Fraser et al., 1997), *Mycobacterium tuberculosis* (Cole et al., 1998) and humans. AAG is the only enzyme to be identified in humans that repairs alkylation damage. Work by O'Brien and Ellenberger (2003) suggests that human AAG is predominantly a hypoxanthine DNA glycosylase due to the large difference in rate

enhancement for substrate excision between Hx and alkylated purines. Like most other class of glycosylase AAG utilizes a base flipping mechanism to flip the base into the enzyme active site. Lau and Scharer et al (1998) used the crystal structure of human AAG in complex to DNA, containing a pyrrolidine abasic nucleotide (*pyr)*, to investigate how AAG was able to recognise and excise so many substrates that are not structurally similar and do not all distort the DNA helix. Their work revealed that AAG binds to the DNA minor groove and flips out the base into its substrate binding pocket. AAG has a mixed α/β structure with a protruding β hairpin which inserts into the minor groove and displaces the nucleotide for excision. A Tyrosine (Tyr)-162 residue inserts into the space left by the flipped-out base and the DNA is bent by about 22° at this site. The *pyr* is held in the substrate binding pocket, next to the enzymes active site, by protein contact with flanking AAG surface phosphates. In the pocket the base stacks between aromatic side chains Tyr-127, His-136 and Tyr 159. A water molecule within the active site then displaces the glycosidic bond between the base and deoxyribose sugar. Further studies with AAG in complex with DNA containing εA confirm the base flipping mechanism for base excision by AAG is similar for different substrates. Base excision of both *pyr* and εA involve the same conformation of AAG, a similar bent angle of DNA at Tyr 162 and similar binding of the substrates within the substrate binding pocket including the side chain orientation where the base is stacked (Lau et al., 2000).

An additional fifth glycosylase superfamily has recently been described. The ALKs were first identified as a glycosylase superfamily after the purification of AlkC and AlkD from *Bacillus cereus* (Dalhus et al., 2007). The two alkylpurine glycosylases repair alkylation damage and are specific for 3meA and 7meG DNA adducts (Alseth et

al., 2006), ALK glycosylases contain characteristic HEAT repeats, a motif consisting of two alpha helices linked by a short loop (Brooks et al., 2013).

1.2. The Protozoan Parasite *Trichomonas vaginalis*

Trichomonas vaginalis (*T. vaginalis*) is an early diverging eukaryotic protozoan which belongs to the Parabasalia class of the Excavata supergroup (Adl et al., 2012). The classification of *T. vagianlis* is shown in Figure1.3. *T. vaginalis* is the causative agent of the sexually transmitted infection trichomoniasis. Trichomoniasis was long regarded as a 'self-clearing nuisance' and there was little research regarding its pathogenicity and virulence or control efforts despite it being the most commonly transmitted, non-viral sexual infection worldwide according to the World Health Organisation (WHO, 2012). The Centres for Disease Control and Prevention (CDC, 2018) have marked trichomoniasis as one of five neglected parasitic diseases prioritised for public health action. Studies associating *T. vaginalis* infection with an increased risk of HIV acquisition (Kissinger and Adamski, 2013) cervical cancer (Tao et al., 2014; Yang et al., 2018), in addition to reports of antibiotic resistance (Schwebke and Barrientes, 2006; Upcroft et al., 2009) have led to an increased importance of research into this parasite.

Tritrichomonas foetus (*T. foetus*) is an evolutionary relative to *T. vaginalis* belonging to the Tritrichomonadida order of the Parabasalia class (Figure 1.3). It infects the reproductive system of cattle causing bovine trichomoniasis (Rae et al., 2004; Mardones et al., 2008) and the intestinal tract of cats (Levy et al., 2003; Gunn-Moore

et al., 2007; Lim et al., 2010). Bovine trichomoniasis is of global importance due to its detrimental effect on beef cattle in the livestock industry.

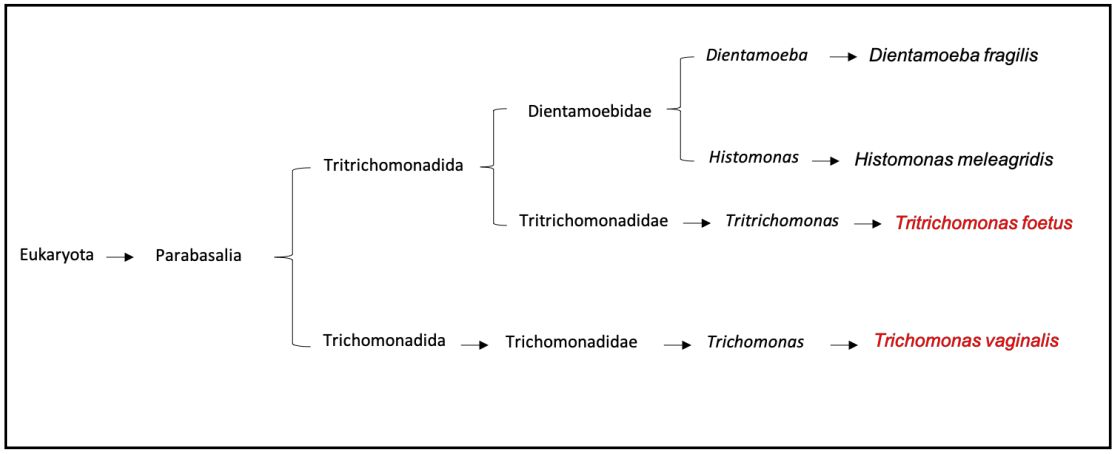


Figure1.3 Classification of Trichomonads. Classification of Trichomonads including *Trichomonas vaginalis and Tritrichomonas foetus*. Adapted from Barratt et al., (2016)

1.2.1. Epidemiology

The World Health Organisation survey data from 2008 estimated a total of 276.4 million new cases of *T. vaginalis* infection in humans between the age of 15-49 years, making it the most common STI worldwide, with more than 100 million more new cases compared to *Chlamydia trachomatis* infection which was recorded at 105.7 million new cases. Global distribution of *T. vaginalis* cases is listed in Table 1.4.

Region	Incidence (million)	Prevalence (million)	Estimated population
Africa	59.7	42.8	384.4
Americas	85.4	57.8	125.7
South-East Asia	42.9	28.7	945.2
Europe	22.6	14.3	450.8
East Mediterranean	20.2	13.2	309.6
Western Pacific	45.7	30.1	986.7

Table 1.4: Incidence and Prevalence Data for *T. vaginalis* Cases by Region in Adults Between 15 and 49 Years of Age. Data from The World Health Organisation Global incidence and prevalence of selected curable sexually transmitted infections 2008 report (accessed at: https://www.who.int/reproductivehealth/publications/rtis/stisestimates/en/)

Humans are the only natural host of *T. vaginalis*. The parasite infects the squamous epithelium of the lower urogenital tract of women (vulva, vagina and urethra) and men (urethra) after transmission via sexual contact. Common symptoms of trichomoniasis, as listed on the CDC website, include; irritation, redness and soreness of genitals, unusual vaginal discharge, irritation and discharge from the penis and dysuria. Vaginal discharge is a clinical symptom of trichomoniasis described as frothy

and yellow or green in colour (Swygard, 2004). It is caused by an influx in polymorphonuclear leukocytes in response to inflammation caused by *T. vaginalis* and possibly by the parasite releasing a chemotactic factor (Demirezen, Safi and Beksaç, 2000). Strawberry cervix or *colipitis macularis* is another specific symptom for trichomoniasis where the urogenital mucosa develops haemorrhagic spots and, although this is the most indicative symptom of *T. vaginalis* infection, it is only diagnosed in 2-5% of patients (Lehker and Alderete, 2000). Trichomoniasis is more common in women than men (CDC) and around 80% of cases are asymptomatic (Allsworth, Ratner and Peipert, 2009).

The general symptoms of trichomoniasis are easily treatable with antibiotics whereas the complications associated with trichomoniasis are more severe. *T. vaginalis* infection has been associated with adverse pregnancy outcomes such as low birth weight and pre-term birth (Cotch et al 1997; Silver et al., 2014), infertility, an increased susceptibility to HIV infection (Quinlivan et al., 2012; Kissinger and Adamski, 2013) and an increased risk of cervical cancer (Tao et al., 2014; Yang et al., 2018). The recommended treatment for *T. vaginalis* infection is the use of 5'nitromidiazoles. Metronidazole (MTZ) is the FDA approved drug for the treatment on trichomoniasis. After its synthesis in 1959 (Cosar and Julou 1959) it was approved in 1960 as the main course of treatment for trichomoniasis. Tinidazole can also be used to treat trichomoniasis (CDC, 2017). In regard to vaginal trichomoniasis, MTZ resistance occurs in 4-10% of cases and tinidazole resistance occurs in 1% of cases in the US (Kirkcaldy et al., 2012; Schwebke and Barrientes, 2006) There have been reports of increased prevalence of *Trichomonas vaginalis* resistance in Papua New Guinea ~ 17% (Upcroft et al., 2009).

1.2.2. Cell Organisation

Discovered in 1836 by Alfredo Francois Donné *T. vaginalis* is a flagellated microaerophilic protozoan with the typical structure and organisation of a trichomonad. Its size ranges between 7 and 9 μM and free-swimming trophozoites are pyriform in shape with four anterior flagella and one recurrent flagellum that runs down the axis of the trichomonad cell and is attached to form the undulating membrane. Upon cytoadhesion to epithelial cells the trophozoite becomes amoeboid in shape (Arroyo et al., 1993).

The parasites nucleus is located at the anterior portion of the cell and is contained within a porous nuclear envelope (Honigberg and King, 1964). The main cytoskeleton structures within the trichomonad cell are the pelta-axostylar complex and the costa (Benchimol, 2004). The axostyle is a string of microtubules, arranged longitudinally, that runs from the posterior to the anterior end of the cell (Benchimol et. Al., 2000). The pelta is located in the anterior part of the cell and is also made up of microtubules (Ribeiro et al., 2000). The region where the two sheets of microtubules overlap and are connected via bridges is known as the pelta-axostylar junction (Honigberg et al., 1971; Benchimol et al., 1993). The costa cytoskeletal complex is a striated fibre, made up of proteins organised into filamentous structures, that supports the undulating membrane structures (Honigberg, Mattern and Daniel, 1971; Benchimol, 2004)

T. vaginalis lack mitochondria and instead have anaerobic hydrogenosomes, double-membraned organelles which produce ATP through substrate level phosphorylation with pyruvate and malate and generate hydrogen as an end product (Muller et al., 2012). Figure 1.4 shows the structure of a *T. vaginalis* cell.

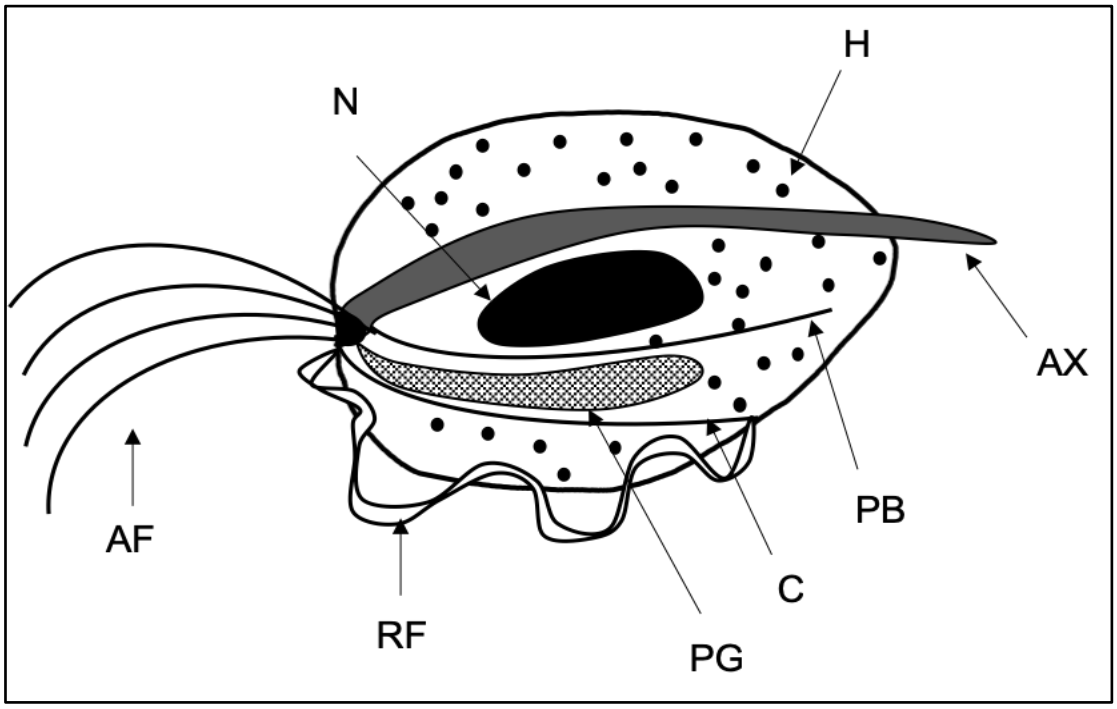


Figure 1.4 *Trichomonas vaginalis* Structure. Drawing highlighting the main components of a pyriform *T. vaginalis* cell. AF - anterior flagella; AX - axostyle; C - costa; H - hydrogenosome; N - nucleus; PB - parabasal body; PG - parabasal body and golgi apparatus; RF - recurrent flagellum. Adapted from Barratt et al., (2016)

1.2.3. Life Cycle and Population Structure

T. vaginalis exists only in a trophozoite stage and is transmitted via sexual contact. The parasite divides by binary fission within the urogenital tract (Figure 1.5) and is transmitted by vaginal or prostatic secretions. Pseudocyst forms of the parasite have

been observed and are formed when the parasite is under stressful conditions (Pereira-Neves, Ribeiro and Benchimol, 2003). Isolates have also been taken from the respiratory tract in a small number of cases with *T. vaginalis* identified as the cause of respiratory tract symptoms (Carter and Whithaus, 2008) (McLaren LC et al., 1983) (Press et al., 2001) (Duboucher et al., 2003). Rare transmission of *T. vaginalis* via fomites has been described in one study in Zambia (Crucitti et al., 2011). Upon examination of *T. vaginalis* genome genes for meiotic recombination machinery have been found (Malik et al 2008) but meiosis or has never been observed within *T. vaginalis*.

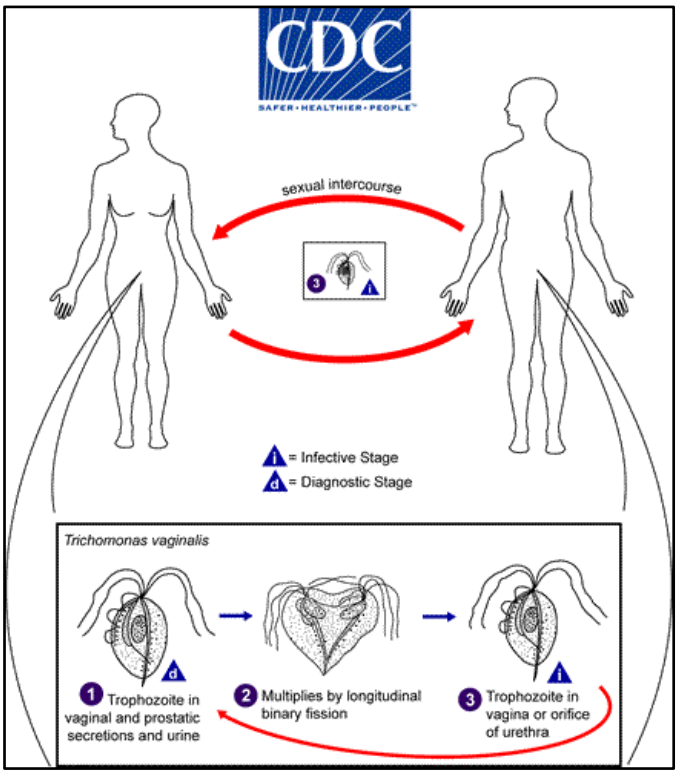


Figure 1.5: *Trichomonas vaginalis* Transmission and Life Cycle (CDC available at: https://www.cdc.gov/dpdx/trichomoniasis/index.html)

Clinical isolates of *T. vaginalis* show a high frequency of variation in pathogenicity, virulence and antibiotic resistance and studies have used a variety of methods to examine the genetic diversity of the parasite as an explanation to the

differences between isolates. Earlier work on genetic diversity of *T. vaginalis* involved isoenzyme analysis and antigen characterisation (Soliman et al., 1982; Krieger et al., 1985). Later studies utilised large samples and random amplified polymorphic DNA (Snipes et al., 2000), single nucleotide polymorphisms and microsatellite genotyping (Conrad et al., 2012) with statistical programmes to analyse phylogeny and population genetics to determine a 'two type' population structure in *T. vaginalis*. The 2012 study by Conrad et al., examined 235 *T. vaginalis* isolates from across the globe to confirm the presence of a 'two population' structure within *Trichomonas vaginalis*. The study used genetic markers to conclusively demonstrate a global distribution of 'type 1' and 'type 2' *T. vaginalis* population types. The types were found to be equally distributed worldwide, with the exception of South Africa and Mexico where a higher frequency of type 1 was calculated. The differences found between the two isolates involved the minimum lethal concentration (MLC) of antibiotic required to kill isolates and infection with the *Trichomonas vaginalis* virus (TVV). Type 1 isolates required a lower concentration of MTZ to be killed with a mean of 76.6 μg/ml compared to type 2 which had a mean MLC of 228.4 μg/ml. Type 1 isolates were more frequently infected with TVV compared to Type 2. These findings correlate with findings from Snipes et al., (2000) and Cornelius et al., (2012). Phylogenetic analysis suggests type 1 is the ancestral type which type 2 diverged from (Conrad et al., 2012).

1.2.4. Pathogenesis

Trichomonas vaginalis is a microaerophilic parasite which utilises the human urogenital tract as its niche where it exists extracellularly. *T. vaginalis* deploys a number of virulence mechanisms to colonise the human urogenital tract. Differences in the

microenvironments of female and male urogenital tracts may affect the parasites ability to colonise and explain the difference between men and women's ability to clear *T. vaginalis* infection. The vaginal mucosa is rich in iron required for the parasite's survival (Lehker, 1990) whereas the prostatic glands of the penis have a hostile zinc rich environment. Zinc is antimicrobial and concentrations of zinc salts, similar to concentration found in the prostatic fluid, were found to kill trichomonads *in vitro* (Fair, Couch and Wehner, 1976; Krieger and Rein, 1982)

Once transmitted the free swimming trophozoites remain extracellular and adhere to squamous epithelial cells. Cytoadherence is essential for establishment of infection, and during adhesion the parasite undergoes a conformational change from pyriform to amoeboid and adhesion expression is increased (Arroyo et al., 1993). The process of cytoadherence and involves the surface proteins adhesins and glycoconjugates as well as cytoskeletal proteins and extra-cellular matrix (ECM) protein receptors (Figueroa-Angulo et al., 2012).

The parasite overcomes the host's mucous layer, an innate immune defence, by binding and releasing mucinases that degrade the mucin glycoprotein (Alderete et al., 2002) After successful adherence colonisation of vaginal epithelial cells (VECs) is initiated by the parasite via a multistep process that results in cytolysis, phagocytosis and disintegration of cell monolayers (Figueroa-Angulo et al., 2012). Contact-independent mechanisms also damage host tissues, *T. vaginalis* secretes cytolytic factors to aid in colonisation including; *Trichomonas vaginalis* factor (TvF) and cell detaching factor (CDF). TvF induces clumping of target cells and CDF promotes cell detachment (Lushbaugh et al., 1989).

Some *T. vaginalis* isolates harbour a double stranded RNA virus. The *T. vaginalis* virus is a member of the *Totiviridae* virus family which parasitizes *T. vaginalis* (Goodman et al., 2010). Four species of TVV have been reported and multiple strains can infect one single isolate causing a mixed infection (Flegr, Čerkasov and Štokrová, 1988; Goodman et al., 2011). As previously mentioned, Type 1 isolates appear to be more frequently infected with TVV, Conrad et al., (2012) analysis of 153 isolates showed 75% of type 1 were infected compared to only 3% of type 2. TVV has been linked to upregulation of cysteine proteases, virulence factors which degrade proteins involved in the immune response and immunoglobulins (Alderete, Provenzano and Lehker, 1995; Provenzano and Alderete, 1995). Cysteine proteases are also involved in cytoadherence (Arroyo and Alderete, 1989). Upregulation of cysteine proteases could confer advantage for the parasites' survival within the human urogenital tract.

In addition to TVV *T. vaginalis* isolates can also harbour the bacterium *Mycoplasma hominis* (*M. hominis)* (Rappelli et al., 1998; Vancini and Benchimol, 2007) and *Candidatus Mycoplasma girerdii (C. M. giredii)* (Martin et al., 2013; Fettweis et al., 2014). *M. hominis* is a *T. vaginalis* symbiont which has been reported to contribute to the pathogenicity of *T. vaginalis* by increasing growth rate and ATP production (Margarita et al., 2016). *In vitro* studies have shown *M. hominis* infected *T. vaginalis* strains to have an enhanced rate of cytopathogenicity of epithelial cells and haemolysis of red blood cells compared to non-infected strains (Vancini et al., 2007; (Margarita et al., 2016). *M. hominis* has been linked to bacterial vaginosis, adverse pregnancy outcomes and pelvic inflammatory disease (Larsen and Hwang, 2010).

1.2.5. Genome

The draft genome for the *Trichomonas vaginalis* G3 lab isolate was first published in 2007 allowing more research to be done on the parasites' genetic diversity, population structure and pathogenesis. The draft genome sequence, originally stated as ~160 Mb in size with more recent data stating ~176 Mb (TrichDB.org), contains around 46,000 protein coding genes on six haploid chromosomes (Carlton et al.,2007) Around 65% of the genome is composed of transposable elements. The *T. vaginalis* genome is considerably large compared to other parasitic protists genomes; *Leishmania major* (32 Mb) (Ivens et al., 2005), *Toxoplasma gondii* (65 Mb) (Lau et al., 2016) and *Plasmodium falciparum* (22 Mb) (Gardner et al., 2002). This larger size of the genome could be explained by genome duplication events (Carlton et al., 2007).

1.2.6. Horizontal Gene Transfer

Horizontal gene transfer (HGT), or lateral gene transfer (LGT), refers to the non-sexual transmission of genes between two organisms. It is widely documented in prokaryotes via transformation, conjugation and transduction (von Wintersdorff et al., 2016). Instances of horizontal gene transfer between prokaryotes and eukaryotes are less well understood. However, phylogenomic studies have estimated 0.25% of *T. vaginalis* genes are acquired by LGT (Alsmark et al., 2013). *T. vaginalis* lives within the human urogenital tract, an ecological niche inhabited by several species of bacteria, viruses and fungi, which are part of the normal microbiota of humans (Pruski et al., 2018). Within this niche HGT can occur and has been widely documented in *T. vaginalis*. Strese, Backlund and Alsmark (2014) analysed a 34 kbp region of the *T. vaginalis* genome, the

Trichomonas vaginalis lateral gene transfer fragment (TvLF), and concluded it was probably a result of a single lateral gene transfer event from a *Peptoniphilus harei* donor.

Horizontal gene transfer is a way of picking up advantageous genes to enrich survival, and in *T. vaginalis'* case, to increase pathogenicity. Work by Alsmark et al., (2013) identified nine *T. vaginalis* enzymes involved in the degradation of glycans in host cell membranes as products of gene transfers, highlighting the significance HGT has had within the parasite. The BSPA-like proteins are also likely to have been acquired by *T. vaginalis* via horizontal gene transfer from ancient prokaryotes (Noël et al., 2010). *Trichomonas vaginalis* has also been documented to be involved in eukaryote to eukaryote HGT. Several HGTs have been identified between the human parasite *Entamoeba histolytica* (*E. histolytica*) and *T. vaginalis*. Grant and Katz (2014) propose an ancestral relationship between the two human parasites which allowed them to take part in horizontal gene transfer.

One hypothesis on the mechanism of HGT between prokaryotes and eukaryotes is that gene transfer is mediated by phagocytosis, the 'you are what you eat' hypothesis (Doolittle, 1998). *T. vaginalis* uses phagocytosis to ingest necrotic tissue after the break down of host cell membrane during infection and it has been shown that *T. vaginalis* is able to phagocytose *Saccharomyces cerevisiae, lactobacilli,* leukocytes and erythrocytes *in-vitro* (Midlej and Benchimol, 2009; Pereira-Neves and Benchimol, 2007; Rendón-Maldonado et al., 1998). The 'you are what you eat' hypothesis could explain the high percentage of HGT within the *T. vaginalis* genome.

2

Materials and Methods

2.1. Materials

Synthetic genes and oligonucleotide primers were obtained from Thermo Fisher Scientific. Fluorescently labelled substrate oligonucleotides, deoxyinosine (dI), deoxythymidine (dT), deoxycytidine (dC) and 8oxoG, were obtained from Sigma or from Eurogentec (deoxyuridine (dU)) All other chemicals were purchased from Sigma unless otherwise stated. *Escherichia coli* Competent cells were obtained from Novagen, *Escherichia coli* Rosetta(DE3)pLysS cells (Rosetta™) and New England BioLabs *Escherichia coli* Turbo high efficiency cells (Turbo) genotypes for both strains can be found in Table 2.1.

Strain	Genotype
Rosetta™	F- ompT hsdSB(rB- mB-) gal dcm pRARE (CamR)
Turbo	F' proA+B+ lacIq ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)TetS endA1 thi-1 Δ(hsdS-mcrB)5

Table 2.1 *E. coli* Competent Cell Genotypes. The genotypes of Rosetta (DE3)pLysS cells obtained from Novagen and Turbo competent high efficiency cells obtained from New England Biolabs.

2.2. Methods

2.2.1. Molecular Cloning

2.2.1.1 Preparation of Competent Cells

Escherichia coli Rosetta(DE3)pLysS cells (Novagen) were spread on LB agar plates (Table 2.2) with chloramphenicol (34 μg/ml) and incubated at 37°C overnight. A single colony was used to inoculate 5 ml of LB broth (Table 2.2) which was incubated with shaking overnight or for 8 hours at 37°C. After incubation, 0.5 ml of the culture was used to inoculate 50 ml LB broth which was incubated with shaking until an optical density (OD) of 0.6 was reached. OD was measured using a spectrophotometer with distilled water (dH_2O) water as a blank measurement. The culture was transferred into 50 ml Oakridge tubes, chilled on ice and centrifuged in an Avanti-JX centrifuge using a JA 25.50 rotor (Beckman Coulter Inc.) at 3000 g, 4°C for 5 minutes. After centrifuging, the supernatant was discarded, and tubes were inverted to ensure all traces of supernatant were removed. The pellet was re-suspended into 0.4 volumes ice cold 1 M Calcium Chloride ($CaCl_2$). The culture was left to incubate on ice for 30 minutes and cells were harvested by centrifugation with the same settings as previously described. After removal of the supernatant the cells were re-suspended in 3 ml ice cold $CaCl_2$ and 1 ml 50% glycerol. Newly prepared competent cells were snap-frozen and stored in aliquots at -80°C.

2.2.1.2 Transformation of Bacterial Cells

New England BioLabs (NEB) Turbo competent high efficiency *E. coli* cells were used for the propagation of plasmids and *E. coli* Rosetta (DE3)pLysS cells were used for protein expression. Cells were thawed and 2 µl of plasmid, ligation mixture or site directed mutagenesis reaction mixture were added. The culture was heat shocked at 42°C for 30 seconds and left on ice for 2 minutes. 200 µl of S.O.C medium (NEB) was added and the culture was incubated with shaking at 160 rpm 37°C for 1 hour. After incubation, 50 µl was spread on LB agar plates (Table 2.2) containing the appropriate antibiotics (Table 2.3) and incubated at 37°C overnight. Plates were wrapped with parafilm and stored in the 4°C fridge until use.

Name	Composition	Manufacturer
LB broth	0.5% Yeast Extract 1% Tryptone 1% NaCl	Scientific Laboratory Supplies (SLS) Ltd SLS Ltd Sigma Aldrich
LB agar	0.5% Yeast Extract 1% Tryptone 1% NaCl 1.5% Agar	SLS Ltd SLS Ltd Sigma Aldrich SLS Ltd
SOC media	2% Vegetable Peptone 0.5% Yeast Extract 10 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 20 mM Glucose	New England Biolabs

Table 2.2 Composition of Media used for Bacterial Cell Growth.

Antibiotic	mg/ml
Chloramphenicol	34 in DMSO
Kanamycin	30 in dH_2O

Table 2.3: Details of Stock Solutions of Antibiotics used for Selection of Resistant *E. coli* Colonies and Cultures. Stock solutions were added to medium at a dilution of 1:1000. DMSO - Dimethyl sulfoxide

2.2.1.3 Plasmid DNA Preparation

A single colony of previously transformed NEB Turbo cells was used to inoculate 5 ml of LB broth with appropriate antibiotics (Table 2.3) and incubated with shaking at 37°C 160 rpm for 8 hours or overnight. After incubation, the culture was centrifuged at 5000 rpm for 15 minutes at 4°C and the resulting pellet was frozen at -80°C or purified using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific) according to manufacturer instructions. The concentration of the purified plasmid DNA was measured using a NanoDrop UV-Vis spectrophotometer (Thermo Fisher Scientific) at 280 nm (A_{280}) wavelength.

2.2.1.4 Restriction Digest

The plasmid containing the artificially synthesised TVAAG2 insert and pET28a expression vector were each cut with NdeI and XhoI restriction enzymes (NEB). The restriction digest mixture (Table 2.4) was incubated at 37°C for 2 hours. Samples were covered with tinfoil in the dryblock incubator during incubation to prevent evaporation.

	Restriction mixture composition	**Final concentration**
TvAAG2	TvAAG2 vector NdeI XhoI Cutsmart buffer dH_2O	2 µg 1.5 µl 1.5 µl 10 x Up to 20 µl
pET28a	pET28a expression vector NdeI XhoI Cutsmart buffer SAP dH_2O	2 µg 1.5 µl 1.5 µl 10 x 1 µl Up to 30 µl

Table 2.4 Composition of Restriction Digest Mixture. Components sourced from New England Biolabs.

2.2.1.5 Agarose Gel Electrophoresis

Electrophoresis analysis of the digested plasmids from section 2.2.14 was carried out on agarose gel with 0.5 X TBE buffer (Table 2.5) at 90-120 V for 1 hour to determine if the restriction digest was successful. Plasmid DNA samples were prepared by adding DNA loading dye and ran alongside a DNA molecular marker (Table 2.5). The gel was then imaged using a UV transilluminator.

Name	**Composition**	
Agarose gel	TBE buffer Agarose GelRed	0.5 X 0.24 g 1:10000
0.5 x TBE buffer	TRIS Boric acid EDTA	41.6 mM 44.5 mM 1.4 mM
6 x DNA loading buffer	Tris-HCl Bromophenol blue, Xylene cyanol FF Glycerol EDTA.	10 mM 0.03% 0.03% 60% 60 mM
DNA molecular marker	100 bp DNA Ladder	New England Biolabs

Table 2.5 Components used for Agarose Gel Electrophoresis.

2.2.1.6 Purification of DNA Fragments

Appropriate bands identified with the UV transilluminator in section 2.2.1.5 were cut out with a razor blade and purified using the QIAEX II Agarose gel extraction kit (QIAGEN). According to manufacturer instructions.

2.2.1.7 Ligation of DNA into Vector

The ligation reaction mixture for the TvAAG2 insert and pET28a expression vector (Table 2.6) was incubated for 15 minutes at room temperature. 2 µl of the reaction mixture was used to transform Turbo competent cells which were grown on LB agar plates.

Single colonies from these plates were used to inoculate 5 ml cultures and plasmids prepared as described in Section 2.2.1.3. Plasmids were screened by restriction digestion as described in Section 2.2.1.4 and 2.2.1.5. If a band at the right size was visualised from restriction digestion the same plasmid was sent for Sanger sequencing at Source BioScience to confirm successful ligation of insert.

Ligation mixture composition	Final concentration
Quick ligase reaction buffer	10 µl
pET28a Vector	50 ng
TVAAG2 Insert	37.5 ng
dH_2O	Up to 20 µl
Quick ligase	1 µl

Table 2.6 Composition of Ligation Reaction Mixture. Components from New England Biolabs.

2.2.2. Protein Expression and Purification

2.2.2.1 Test Expression

5 ml LB broth 'starter' cultures were inoculated with a single colony from previously transformed Rosetta 2 (DE3) pLysS cells (as described in 2.2.1.2) with the appropriate antibiotics and left to grow for 8 hours or overnight. These starter cultures were used to inoculate 50 ml of LB with the appropriate antibiotics and incubated with shaking at 160 rpm at 37°C for 3 hours until mid-log phase which was confirmed by OD measurement of 0.6 or above using UV-vis spectrophotometry. 1 ml of the culture was taken to be processed for subsequent analysis by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the remaining culture to a final concentration of 1 mM to induce protein expression. The culture was incubated for a further 3 hours with 1 ml samples taken at hour intervals and processed for SDS-PAGE. This involved samples being spun down in a microcentrifuge at 600 rpm, supernatant discarded, and the pellet resuspended in 200 µl of 1 X SDS-PAGE loading buffer before heating in a 95°C heat block for 3 minutes. Samples were stored at -20°C. 10 µl of each

of the thawed samples were loaded onto a 10% SDS-PAGE gel (described later in Section 2.2.2.6) alongside 5 µl of molecular weight marker.

2.2.2.2 Large Scale Expression

5 ml starter cultures were grown as previously described (Section 2.2.2.1) and added to 750 ml LB cultures with the appropriate antibiotics and incubated with shaking 160 rpm at 37°C for 3 hours until mid-log phase was reached. IPTG was added to a final concentration of 1 mM and the culture was left to grow for a further 3 hours. After growth the cultures were transferred to 1 L polypropylene centrifuge bottles (Beckman Coulter Inc) and centrifuged in an Avanti-JX centrifuge using the JLA 8.1000 rotor from Beckman Coulter Inc at 3000 rpm for 20 minutes, 4°C. The supernatant was discarded, and the cell lysate pellet was frozen at -80°C for later use.

2.2.2.3 Preparation of the Cell Lysate

The frozen cell lysate pellet was left to thaw on ice. One tablet of protease inhibitor (cOmplete ULTRA Tablets Mini EDTA Free, Sigma Aldrich) was dissolved in 10 ml of cold buffer A (Table 2.7) and left on ice for 15 minutes. This buffer was used to resuspend the cell lysate pellet. The resuspended sample was sonicated to lyse the bacterial cells. Sonication was done 5 times for a total of 15 seconds with a 15 second break between each sonication. The sample was kept in an ice-water bath during sonication to ensure no degradation of protein. The sonicated sample was transferred to Oakridge tubes and centrifuged in an Avanti JX centrifuge using a JLA 25.25 rotor at

20,000 rpm for 30 minutes, 4°C. The cell lysate was used in subsequent purification steps and the pelleted cell debris discarded.

2.2.2.4 Nickel – Nitrilotriacetic Acid Affinity Column Chromatography

The resin used for the Nickel - Nitrilotriacetic acid (Ni-NTA) purification column was made by adding 300 µl resin (QIAGEN) to the cell lysate (step 2.2.2.3) which was incubated on a rotator for 30 minutes, 4°C. The resin-cell lysate mixture was passed through the Ni-NTA column and unbound protein was collected as the ''flow through'' fraction. 1.2 ml of Buffer A was used for the first wash (Wash 1) and 1.2 ml of 10% buffer B (Table 2.7) in A was used for the second wash (Wash 2). 600 µl of buffer B was used to elute the purified protein (Elute). The fractions were prepared for SDS-PAGE (Section 2.2.2.6) to analyse the purification of protein. All steps were done at 4°C to prevent protein degradation.

Name	Composition
Buffer A	50 mM NaH_2PO_4 1 M NaCl 10 mM Imidazole
Buffer B	50 mM NaH_2PO_4 1 M NaCl 250 mM Imidazole
Protein Dialysis Buffer	50 mM HEPES 100 mM NaCl 1 mM DTT 0.1 mM EDTA 10% Glycerol

Table 2.7 Composition of Buffers used in Ni-NTA Chromatography

2.2.2.5 Dialysis

If a protein band could be seen at the appropriate molecular weight following SDS-PAGE the final eluted fraction was dialysed in pre-cooled protein dialysis buffer (PDB) (Table 2.7). The fraction was inserted into a dialysis cassette from Invitrogen using a fresh syringe. The cassette was left in a beaker of PDB on a magnetic spinner overnight. After overnight dialysis the protein was removed from the dialysis cassette and centrifuged at 13,000 rpm for 10 minutes at 4°C. The concentration of the protein was calculated using a UV-vis spectrophotometer.

2.2.2.6 SDS-PAGE

SDS PAGE gels were hand cast using the Mini-PROTEAN Tetra Cell casting stand system and glass plates from Biorad or precast Novex Tris-Glycine midi protein gels were used. For hand cast gels 10% resolving gel (Table 2.8) was poured between the glass plates. 1-butanol was poured onto the top of the 10% resolving gel to prevent quenching of polymerisation by air and to ensure no bubbles were present before pouring of stacking gel. After gel polymerisation 1-butanol was washed off the top of the gel and 4% stacking gel (Table 2.8) was poured onto the top. A 1 mm 12 or 15 well comb was inserted, and the gel was left to polymerise. Samples for SDS-PAGE were mixed with 3 X or 2 X SDS loading buffer, heated for 5 minutes at 95°C and loaded onto the gels alongside a protein ladder. Hand cast gels were ran for 1 hour 30 minutes in a Mini-PROTEAN Tetra vertical electrophoresis cell at 200 V. Precast gels were ran with the mini gel tank (Thermo Fisher) at 225 V for 45 minutes. Both gels used 1 X TGS running buffer and were stained with Coomassie Blue dye and left overnight to

stain. Gels were imaged using a ChemiDoc XRS+ from Biorad with the Fluro orange option on Image Lab Software.

Name	Composition
10% Resolving gel	4% Acrylamide 0.375 M Tris pH 8.8 0.1% SDS 0.06% APS 0.20% TEMED dH_20
4% Stacking gel	10% Acrylamide 0.125 M Tris pH 6.8 0.1% SDS 0.05% APS 0.20% TEMED dH_2O
1 X Tris Glycine SDS buffer	25 mM Tris, 192 mM Glycine, 0.1% SDS dH_2O

Table 2.8 Composition of SDS PAGE Gel and Buffers

2.2.3. Preparation of TvAAG1 and TvAAG2 Mutants

2.2.3.1 Designing Site Directed Mutagenesis Primers

Amino acids of interest were identified using BLAST search to find closely related species to *T. vaginalis* with TvAAG homologs. These organisms were used in a Clustal Omega Alignment to identify conserved regions of the proteins. The only lysine conserved between all of the identified homologs was chosen as a mutation target. For TvAAG1 and TvAAG2 there would be a Lysine → Arginine (Arg) mutant and Lysine → Alanine (Ala) mutant. Site directed mutagenesis primers (Table 2.9) were designed using Quikchange Primer Design Program at Aglient genomics website.

Primers	Sequence	Type of Mutation
TVAAG1 K124A	5'-gatcaatgccatacagatctgcgctaatttccagacggctac-3' 5'-gtagccgtctggaaattagcgcagatctgtatggcattgatc-3'	Point mutation of lysine to alanine
TVAAG1 K124R	5'-gccatacagatctctgctaatttccagacggctacc-3' 5'-ggtagccgtctggaaattagcagagatctgtatggc-3'	Point mutation lysine to arginine
TVAAG2 K123A	5'-gatcaataccatgcatgcttgcgctaacttccagacgtgcac-3' 5'-gtgcacgtctggaagttagcgcaagcatgcatggtattgatc-3	Point mutation lysine to alanine
TVAAG2 K123R	5'-tcaataccatgcatgcttctgctaacttccagacgtg-3' 5'-cacgtctggaagttagcagaagcatgcatggtattga-3'	Point mutation lysine to arginine

Table 2.9 Primers Provided by Invitrogen for use in Site Directed Mutagenesis

2.2.3.2 Phusion Flash Site Directed Mutagenesis

Site directed mutagenesis was carried out in a thermal cycler. The reaction mixture (Table 2.10) was put in a thermal cycler to run with the Phusion flash programme (Table 2.11). Following completion of this step, restriction enzyme DpnI (10 U) was added and the reaction incubated at 37°C for a further 1 hour. After site directed mutagenesis the reaction mixture was used to transform Turbo cells as described in part 2.2.2.1.

Composition	Volume
2 X Phusion Flash PCR Master Mix (Thermo Fisher Scientific)	25 µl
Forward Primer (Invitrogen)	2 µl
Reverse Primer (Invitrogen)	2 µl
10 mM dNTP	1 µl
Template	1 µl
dH_2O	20 µl

Table 2.10 Composition of the Site Directed Mutagenesis Reaction Mixture.

Step	Temperature (°C)	Time (s)	Cycles (no.)
Initial denaturation	98	10	1
Denaturation	98	1	30
Extension	72	15 s / 1 kb	30
Final Extension	72 4	60 hold	1

Table 2.11 Phusion Flash PCR Cycling Instructions.

2.2.4. Glycosylase Activity Assays

Glycosylase activity was assessed by measuring the intensity of fragments produced by cleavage of fluorescently-labelled oligonucleotide substrates by glycosylases resolved by denaturing PAGE. The resulting fragments of cleavage would be shorter in length, 18 oligomer (mer) compared to the original 39 mer substrate and will migrate further in the denaturing urea polyacrylamide gel according to its molecular weight. The substrate oligomer was labelled with a Cy5 fluorescent label to allow imaging.

2.2.4.1 Enzyme Activity Assay Samples

Enzyme dilutions of 1 μM - 1 nM were made from previously purified protein stock (Section 2.2.2) using 10% bovine serum albumin (BSA) in PDB as dilution buffer. 2 μl of each dilution and 2 μl of PDB/BSA as a negative control were added to 18 μl of master mix, containing the substrate to be tested, in separate Eppendorf tubes (Table 2.12) and heated at 37°C for 15 minutes. Tinfoil was used to cover the Eppendorf's to prevent condensation. FA/NaOH was added to stop the reaction and samples were either frozen in -20°C or processed straight away. Samples were heated at 95°C for 5 minutes and then loaded onto a urea polyacrylamide gel.

2.2.4.2 Denaturing Urea Polyacrylamide Gel Electrophoresis

Urea polyacrylamide gel was hand cast using the Mini-PROTEAN Tetra Cell casting stand system and glass plates from Biorad and prepared according to the composition in Table 2.13 The gel was poured between the glass plates on the casting stand system then a 1 mm 12 or 15 well comb was inserted and the gel was left to polymerise. The gel was pre-ran at 200 V for 30 minutes in the BioRad Mini-PROTEAN Tetra vertical electrophoresis cell before loading the samples which ran at 200 V for 1 hour. After electrophoresis the glass plates were imaged using the ChemiDoc XRS+ Cy5 blot option.

Enzyme Activity Assay Master Mix	**Concentration**
Glycosylase buffer	50 mM HEPES 50 mM NaCl 10 mM DTT 1 mM EDTA
DNA substrate	50 nM
BSA	0.1 mg/ml

Table 2.12 Composition of Master Mix for Enzyme Activity Assays

Name	Composition	
Urea gel	Urea gel	79.53%
	5x Urea buffer	14.88%
	APS	0.53%
	TEMED	0.53%
Urea buffer	Urea	
	10x TBE	
	MilliQ water	
FA/NaOH	Formamide	
	1 M NaOH	
	Blue dextran	
TBE buffer	10x TBE buffer (Severn Biotech Ltd)	
	dH_2O	

Table 2.13 Composition of Gel and Buffers used in Denaturing Urea-PAGE

2.2.5. Schiff Base Assay

The Schiff base assay was used to assess the glycosylases to determine if they were monofunctional or bifunctional. A master mix for the assay was made up as according to Table 2.12 and 1 µl of the protein and either 1 µl of dH_20 or 25 mM cyanoborohydride was added. Samples were run on SDS-PAGE gels as described in 2.2.2.6 and imaged using the ChemiDoc XRS+ Cy5 blot option.

2.2.6. Bioinformatic Analysis

To generate the atlas of DNA repair enzymes found in *T. vaginalis* and *Tritrichomonas foetus* (*T. foetus)* the NCBI database was used to obtain the amino acid sequence for various human DNA repair enzymes which were used with the Basic Local Alignment Search Tool (BLAST) to identify homologs. Alignments were put together using the FASTA sequences of these homologs and Clustal OMEGA.

2.2.6.1 BLAST

BLAST search (available at the NCBI website) was used to identify homologs of common human DNA repair enzymes in *T. vaginalis* and *T. foetus*. Reliability of the results was assessed using the expected value (E value). The E value describes the number of hits that are expected to be found by chance, results with an E value less than 1e-4 were acceptable and results with a higher value discarded or analysed further using reciprocal BLAST the NCBI conserved domains database and Clustal OMEGA alignments.

2.2.6.2 Clustal OMEGA Alignments

Clustal OMEGA (accessed at: www.ebi.ac.uk/Tools/msa/clustalo/) was used to create multiple sequence alignments using FASTA sequences copied from NCBI. The resulting alignment was copied into BoxShade (accessed at: https://embnet.vital-it.ch/software/BOX_form.html) to colour the alignment according to conservation of amino acids.

3

Bioinformatic Analysis of DNA Repair Pathways in *Trichomonas vaginalis* and *Tritrichomonas foetus*

3.1. Introduction

The aim of this section of my work was to create an atlas of DNA repair enzymes and proteins for *Trichomonas vaginalis* and *Tritrichomonas foetus*. I performed bioinformatic analysis across four major DNA repair pathways, including the proteins named TvAAG1 and TvAAG2, which were to be characterised in the biochemistry arm of my project. I also determined the presence of AAG and DNA methyltransferase enzymes within a range of species from all kingdoms of life. Another aim of this section of my project was to identify the origin of the TvAAG genes within *T. vaginalis*.

3.2. Analysis of DNA Repair Proteins

Human DNA repair enzymes involved in base excision repair, mismatch repair, non-homologous end joining, and homologous recombination were identified and used to find possible homologs in *T. vaginalis* and *T. foetus*. BLAST search (Altschul et al., 1990) was used with each gene's nucleotide accession number as a query within the *T. vaginalis* G3 (taxid:412133) and *T. foetus* (taxid:1144522) protein database. Hits were deemed acceptable if they had an E value below 1e-4 and an alignment score above 80. Any hits with an acceptable E value were checked by doing a reciprocal BLAST search; using the resulting protein accession number as a query in a BLAST search against the *Homo sapiens* database to see if the original query gene was the top hit. If more than one query gene resulted in the same hit, or if the reciprocal BLAST showed a different gene, further analysis was carried out. Further analysis involved; looking for common conserved domains using the NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al., 2016) and analysis of sequences by multiple sequence alignment using the CLUSTAL Omega programme at The European Bioinformatics Institute website (https://www.ebi.ac.uk/Tools/msa/clustalo/).

3.2.1. Base Excision Repair

Base excision repair removes single base lesions caused by oxidative and alkylation damage as described in Section 1.1.1.1 and Section 1.1.1.2. The main DNA repair enzymes within the BER pathway are DNA glycosylases which recognise the damaged base and act to excise the base which generates an AP site (Lindahl, 1993). DNA glycosylases can be monofunctional or bifunctional. Bifunctional glycosylases have

intrinsic β-lyase or β/δ-lyase activity allowing the enzyme to cleave the DNA backbone after generation of AP site to leave a strand break. Following excision by monofunctional glycosylases, a class of enzymes known as AP endonucleases remove the AP site to generate a strand break. The strand breaks generated by bifunctional glycosylases or AP endonucleases are further processed by DNA polymerases and joined by DNA ligase. DNA glycosylases recognise different DNA lesions and twelve DNA glycosylases have been identified within mammals which are present in humans. Most notably OGG1 which removes the highly mutagenic 8-oxoguanine lesion and UDG which removes uracil.

I used a list of nucleotide accession numbers, for a total of 23 genes, involved in human BER to identify which were conserved within *T. vaginalis* and *T. foetus*. The list comprised of twelve glycosylases, three endonucleases, two DNA ligases, three DNA polymerases and four accessory proteins. Table 3.1 shows the genes involved in base excision repair within *T. vaginalis* and *T. foetus*. Six of the genes analysed encode proteins involved in replication and are highlighted in Table 3.1. Proteins involved DNA repair and replication would be expected to be conserved within the parasites due to their replicative function. This is true for both *T. vaginalis* and *T. foetus* as almost all of the genes that encode for proteins involved in replication are conserved. The only exception to this is that *T. vaginalis* does not have a Polε homolog and *T. foetus* does not have a Polδ homolog.

Out of the 23 genes analysed *T. vaginalis* had eleven significant hits: three glycosylases, two endonucleases (i.e. Apurinic/apyrimidinic endonuclease 1 (APE1)

and Flap endonuclease 1 (FEN1)), one DNA ligase, one polymerase, PCNA and RF-C. Two homologs were identified for MPG and APE1.

T. foetus also had eleven hits, three glycosylases, two endonucleases, one DNA ligase, one polymerase, PCNA and RF-C. *T. foetus* differs from *T. vaginalis* as it has two homologs for NTH1 and only one homolog for MPG. *T. foetus* also has two APE1 homologs.

Gene designation	***Homo sapiens***	***Trichomonas vaginalis***	***Tritrichomonas foetus***
UDG Uracil DNA *N*-glycosylase	NM_003362	XP_001296939.1	OHS99142.1
UDG2 Uracil DNA *N*-glycosylase 2	NM_080911	x	x
SMUG1 Single-strand-selective monofunctional uracil-DNA glycosylase	NM_014311	x	x
OGG1 8-Oxo-guanine glycosylase 1	NM_016821	x	x
TDG Thymine DNA glycosylase	NM_003211	x	x
MBD4 Methyl-CpG-binding domain 4	NM_003925	x	x
MYH Mut Y homolog DNA glycosylase	NM_012222	x	x
NTH1 Endonuclease three homolog 1 DNA glycosylase	NM_002528	XP_001329048.1	OHT05554.1 OHS93885.1
MPG Methyl purine DNA glycosylase	NM_002434	XP_001311901.1(TvAAG1) XP_001315594.1(TvAAG2)	OHT07844.1
NEIL1 Nei-like DNA glycosylase 1	NM_024608	x	x
NEIL2 Nei-like DNA glycosylase 2	NM_145043.3	x	x
NEIL3 Nei-like DNA glycosylase 3	NM_018248	x	x
APE1 Apurinic/apyrimidinic endonuclease 1	NM_001641	XP_001299171.1 XP_001324898.1	OHT00257.1 OHT05750.1
APE2 Apurinic/apyrimidinic endonuclease 2	NM_014481	x	x
LIG I DNA ligase 1	NM_000234	XP_001581589.1	OHT10360.1
LIG III DNA ligase 3	NM_013975	x	x
Polβ DNA polymerase β	NM_002690	x	x
Polε DNA polymerase ε	NM_006231	x	OHT04305.1
Polδ DNA polymerase δ	NM_002691	XP_001303643.1	x
XRCC1 X-ray repair cross-complementing protein 1	NM_006297	x	x
PCNA Proliferating cell nuclear antigen	NM_182649	XP_001329442.1	OHT09382.1
RF-C Replication factor C	NM_002913	XP_001322740.1	OHS93402.1

FEN1 Flap endonuclease 1	NM_004111	XP_001294157.1	OHS94785.1

Table 3.1: Genes Involved in Base Excision Repair in Humans and their Homologs in *T. vaginalis* and *T. foetus*. Gene designations and protein names of common proteins involved in BER are listed with the accession numbers for the corresponding human nucleotide sequences. BLAST search was used to identify homologs within *T. vaginalis* (taxid:412133) and *T. foetus* (taxid:1144522). Entries for proteins involved in DNA replication are shaded grey.

During initial analysis on *T. vaginalis* the same top hit, HhH-GPD superfamily base excision DNA repair protein (accession no. XP_001329048.1) was shown for both MYH and NTH1. To determine which of these two human glycosylases the *T. vaginalis* protein was most likely to be a homolog of, a reciprocal BLAST was carried out and CLUSTAL alignment was produced using the FASTA sequences for MYH, NTH1 and the *T. vaginalis* protein. The alignment is shown in Figure 3.1. Upon analysis of the conservation of key catalytic and substrate binding residues, XP_001329048.1 was identified as a homolog for NTHL1. This is consistent with the reciprocal BLAST search using XP_001329048.1 against the *Homo sapiens* (taxid:9606) database which lists NTH1 with 90% query cover and an expected value of 1e-50 compared to MYH which had 61% query cover and an expected value of 0.022.

The results of the BLAST search for MYH and NTH1 within *T. foetus* resulted in the same hits with acceptable E values; the putative endonuclease III like protein (OHS93885.1) and the HhH-GPD superfamily base excision DNA repair protein (OHT05554.1). Reciprocal searches using both proteins against the *H. sapiens* database resulted in NTH1 as the top hit. A multiple sequence alignment with OHS93885.1 confirms it is more similar to NTH1 (Figure 3.2). The multiple sequence alignment with OHT05554.1 also shows it to be more similar to NTH1 than MYH (Figure 3.3) suggesting *T. foetus* has two NTH1 homologs and no MYH.

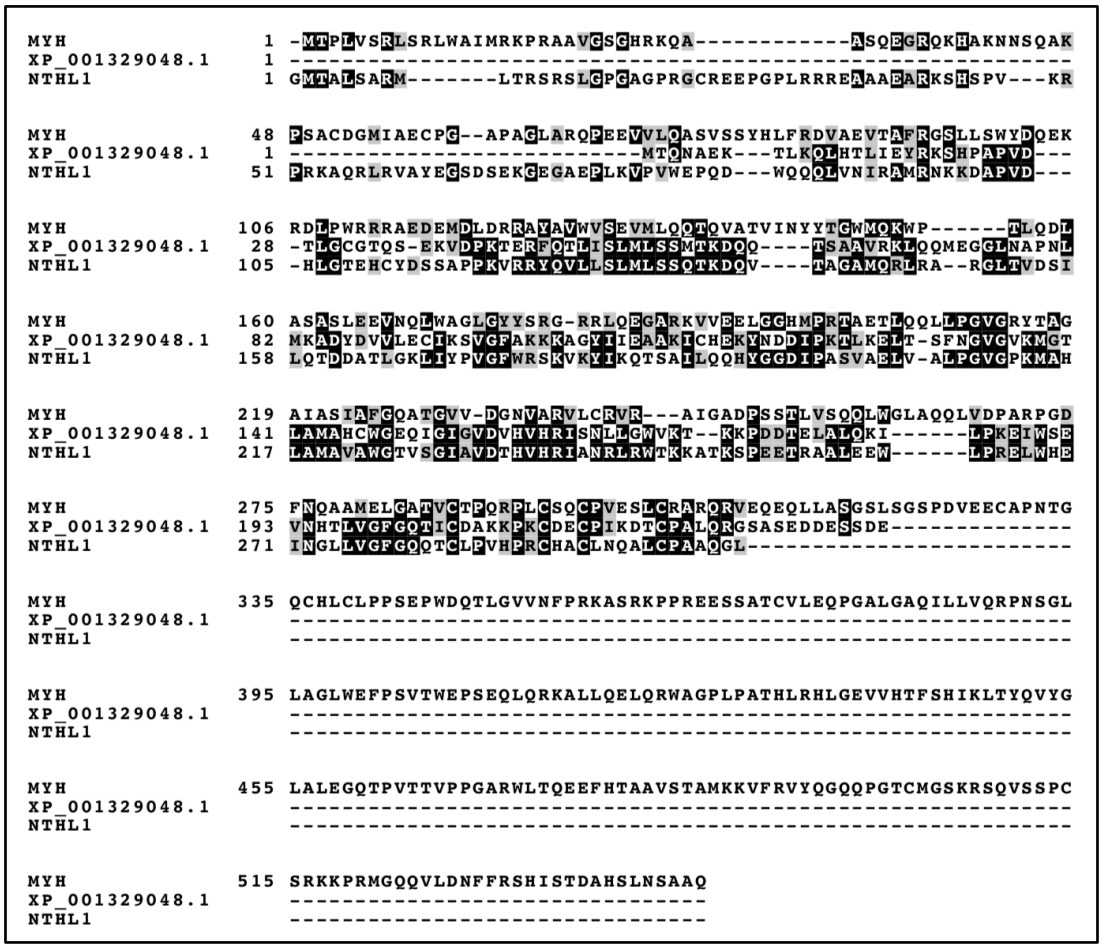


Figure 3.1: Multiple Sequence Alignment of Sequences for Human MYH, NTH1 and *Trichomonas vaginalis* XP_001329048.1. The FASTA sequences for the genes were found on the NCBI website and used with the Clustal Omega tool on the European Bioinformatics Institute website to produce the Clustal alignment. Box shade (ExPASy) was used to highlight identical amino acids in black and similar amino acids in grey.

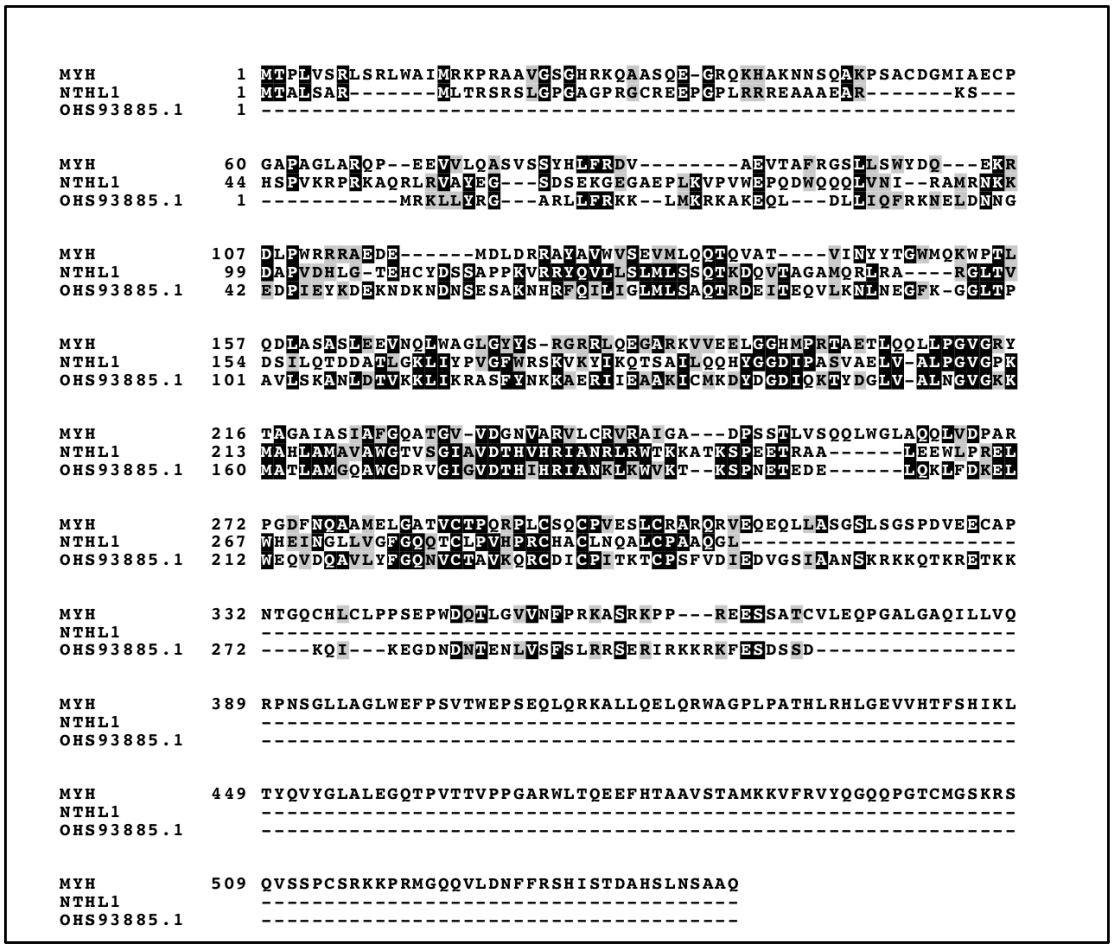


Figure 3.2 Multiple Sequence Alignment of Human MYH, NTH1 and *Tritrichomonas foetus* OHS93885.1. FASTA sequences for the genes were found on the NCBI website and used with the Clustal Omega tool on the European Bioinformatics Institute website to produce the alignment. Box shade (ExPASy) was used to highlight identical amino acids in black and similar amino acids in grey.

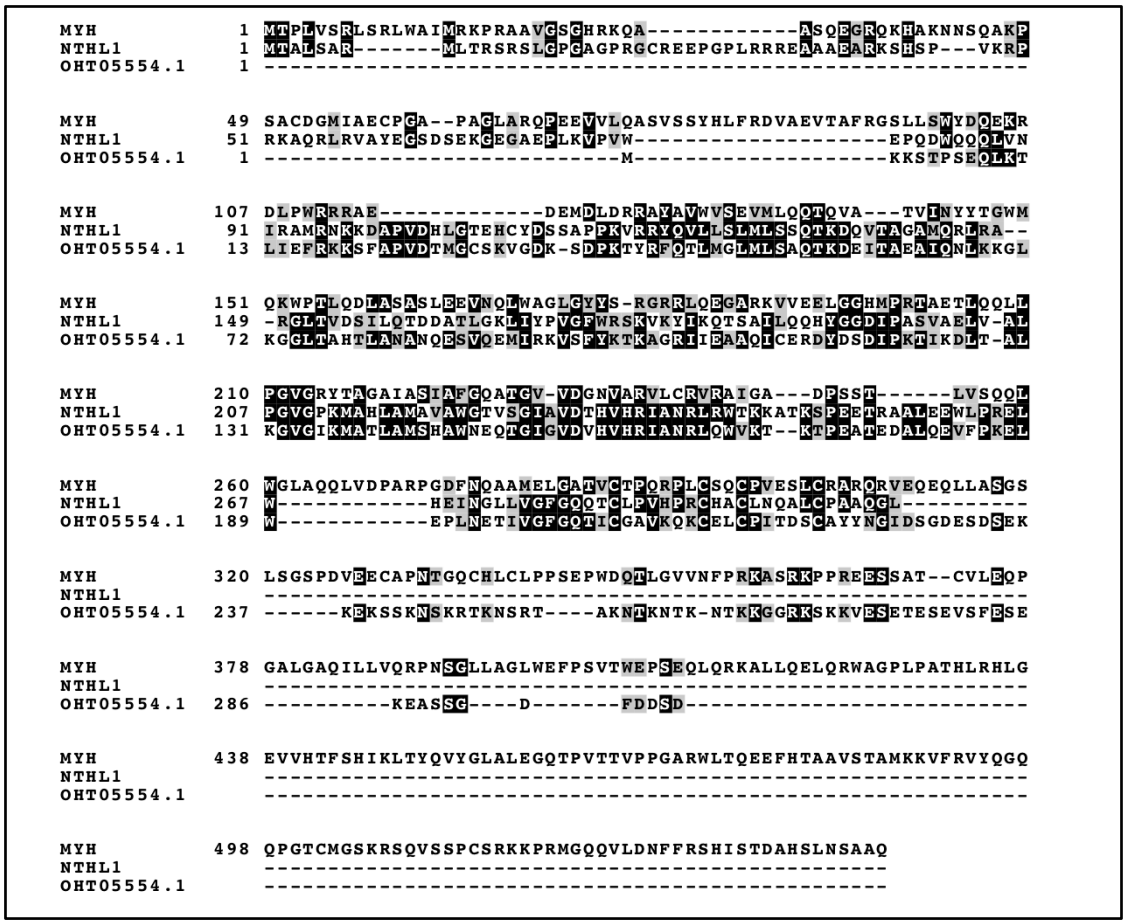


Figure 3.3: Multiple Sequence Alignment with Human MYH, NTHL1 and *Tritrichomonas foetus* OHT05554.1. FASTA sequences for the genes were found on the NCBI website and used with the Clustal Omega tool on the European Bioinformatics Institute website to produce the alignment. Box shade (ExPASy) was used to highlight identical amino acids in black and similar amino acids in grey.

The BLAST search for APE2 in both *T. vaginalis* and *T. foetus* resulted in hits (XP_001324898.1 and OHT05750.1 respectively) which, when used in a reciprocal blast search, did not come up with the original APE2 query but instead showed APE1. A multiple sequence alignment was created using APE1, APE2 and XP_001324898.1 or OHT05750.1 to test if the proteins were APE1 or APE2 homologs. Both XP_001324898.1 (Figure 3.4) and OHT05750.1 (Figure 3.5) are more similar to APE1 on the multiple sequence alignments. The original BLAST search with APE1 as a query

did show XP_001324898.1 and OHT05750.1 as hits in addition to the top hits of XP_001299171.1 (*T. vaginalis)* and OHT00257.1 (*T. foetus)*. XP_001299171.1 and OHT00257.1 showed APE1 when used in reciprocal BLAST searches. Suggesting both *T. vaginalis* and *T. foetus* have two APE1 homologs.

The BLAST search for DNA polymerase δ (NM_002691) in *T. foetus* listed DNA polymerase family B containing protein (OHT08034.1) as a hit with an acceptable E value but the reciprocal blast shows it is a polymerase ζ (NM_002912.4) homolog so was not listed in the table. The same occurred when searching for DNA polymerase δ homologs with *T. vaginalis*. A hit with an accepted E value (XP_001303643.1) was listed that, when used as the query for a reciprocal blast, is actually a polymerase ζ (Polζ) homolog. Pol ζ is present in mammals and yeast and is involved in translesion synthesis, a repair mechanism used to overcome bulky lesions that stall the replication fork (Gibbs et al., 1998; Johnson et al., 2000).

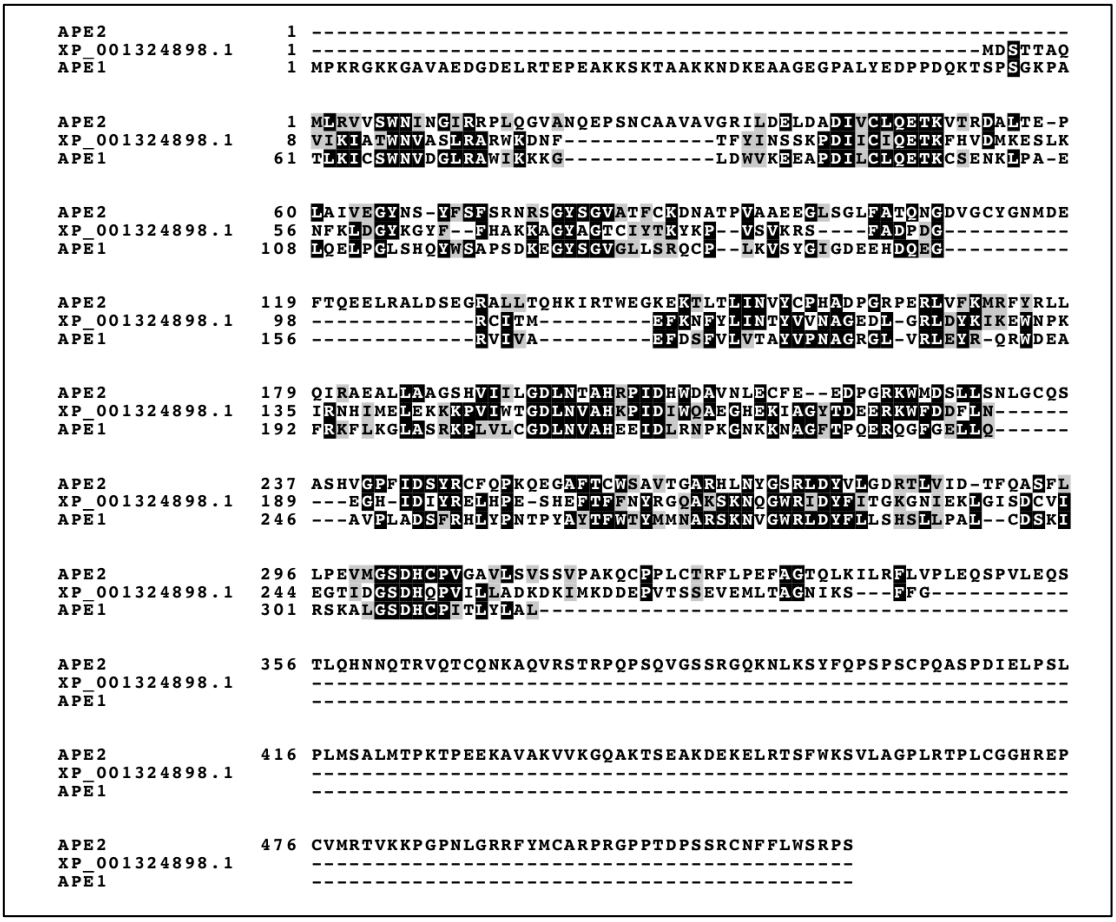


Figure 3.4 Multiple Sequence Alignment with Human APE1, APE2 and *Trichomonas vaginalis* XP_001324898.1. FASTA sequences for the genes were found on the NCBI website and used with the Clustal Omega tool on the European Bioinformatics Institute website to produce the alignment. Box shade (ExPASy) was used to highlight identical amino acids in black and similar amino acids in grey.

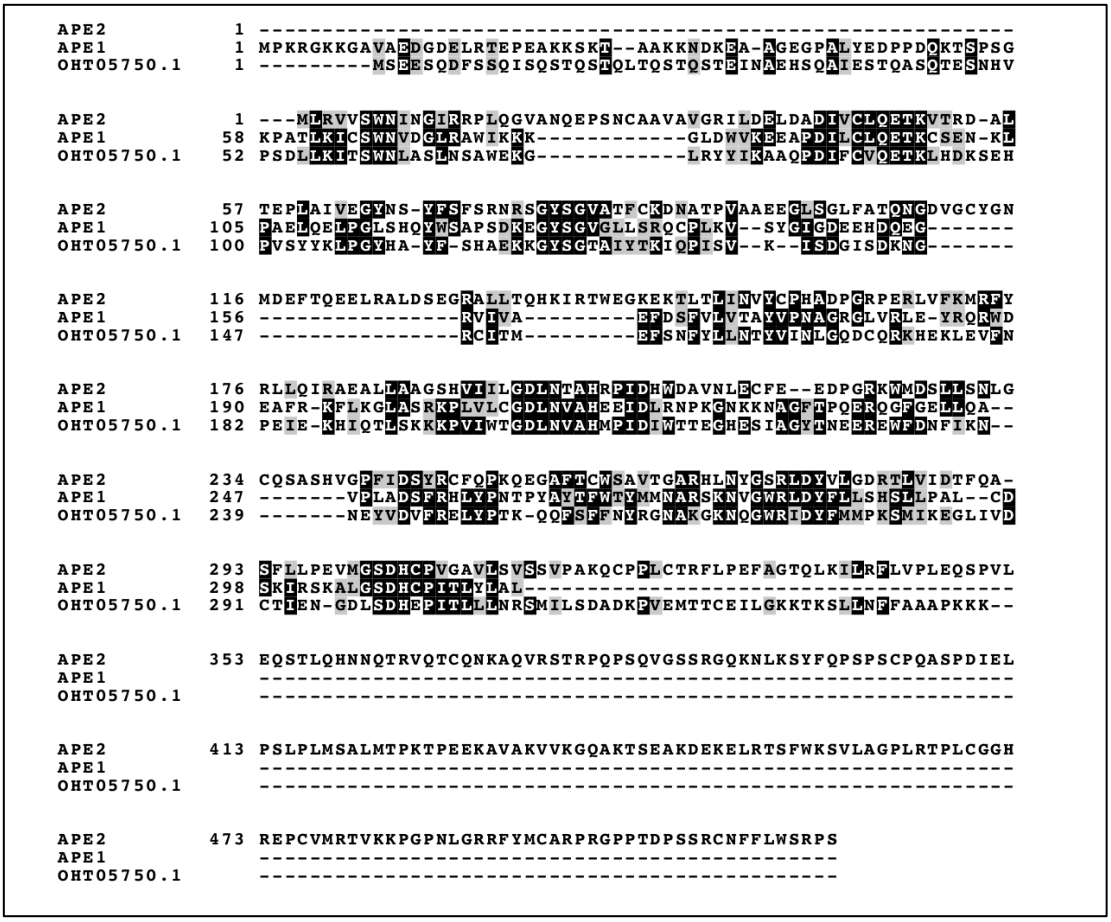


Figure 3.5 Multiple Sequence Alignment for Human APE1, APE2 and *Tritrichomonas foetus* OHT05750.1. FASTA sequences for the genes were found on the NCBI website and used with the Clustal Omega tool on the European Bioinformatics Institute website to produce this Clustal alignment. Box shade (ExPASy) was used to highlight identical amino acids in black and similar amino acids in grey.

3.2.2. Mismatch Repair

Mismatch repair corrects DNA base substitution mismatches and insertion-deletion mismatches (IDLs), that escape the proofreading ability of DNA polymerases generated during replication. Studies on MMR in *E. coli* were the first to establish the basic mechanism (Wildenberg and Meselson, 1975; Glickman and Radman, 1980). This mechanism involves recognition of the mismatch, excision of newly synthesised strand containing the erroneous base and gap filling by re-synthesis of the strand. Prokaryotic

MMR uses the mismatch recognition protein MutS to recognise base mismatches and to recruit MutL and MutH. The MutS-MutL-MutH complex slides along the template DNA strand to find a hemi-methylated GATC sequence a site where the unmethylated daughter strand is currently being synthesised. MutH incises at the GATC site on the newly synthesised strand, allowing excision of the strand by DNA exonuclease and DNA helicase (Lahue and Modrich, 1988., Burdett et al., 2001., Mechanic, Frankel and Matson, 2000). DNA polymerase III carries out repair of the single stranded gap and then DNA ligase works to ligate the nick.

MMR in eukaryotes is more complex than in prokaryotes and less is known about the mechanism. However the main steps of recognition, excision and gap filling are conserved across species. In mammals, mismatches are recognised by two MutS homologs. Unlike the prokaryotic MutS homodimer (Bjornson et al., 2003) eukaryotic homologs of MutS form a heterodimer of two distinct proteins allowing for different substrate specificities depending on the subunit protein. MSH2-MSH6 (MutSα) heterodimer recognises base to base mismatches and IDLs up to three nucleotides long. MSH2-MSH3 (MutSβ) recognises IDLs up to 16 nucleotideslong. (Liu, Keijzers and Rasmussen, 2017., McCulloch, Gu and Li, 2002). There is no MutH homolog in eukaryotes, but the MutL homolog MLH1/PMS2 heterodimer has latent endonuclease activity which is thought to carry out the incision of the daughter strand (Kadyrov et al., 2006). After incision MSH2 and MLH1 recruit exonuclease 1 which excises the daughter strand, DNA is re-synthesised by DNA polymerase δ (Pol δ), and DNA ligase I (LigI) ligates the remaining nick (Nielsen et al., 2003., Longley, Pierce and Modrich, 1997).

Fifteen genes involved in human MMR were identified and used to search for within *T. vaginalis* and *T. foetus*. The results are shown in Table 4.2. Note that some of the homologs, particularly those involved in DNA replication, had previously been identified in Table 3.1 due to their involvement in BER as well. BLAST searches with all six MutS homologs in *T. vaginalis* produced the same hits: XP_001322723.1, XP_001301506.1, XP_001317815 and XP_001306678. Except for MSH3 and MSH4 which did not include XP_001306678. Due to this ambiguity, results from the reciprocal BLAST searches using the *T. vaginalis* proteins as queries against the *H. sapiens* database were used in the construction of the results table (Table 3.2). The same hits were also listed in searches with the MutS homologs in *T. foetus*: OHT00741.1, OHT07430.1, OHT11453.1, OHS97192.1. Once again, results from reciprocal BLAST searches using *T. foetus* proteins as queries against the *H. sapiens* database were therefore used in the construction of the table.

Gene designation	*Homo sapiens*	*Trichomonas vaginalis*	*Tritrichomonas foetus*
MSH2 MutS homolog 2	NM_000251	XP_001317815.1	OHT07430.1
MSH3 MutS homolog 3	NM_002439	x	x
MSH4 MutS homolog 4	NM_002440	XP_001306678.1	OHT11453.1
MSH5 MutS homolog 5	NM_002441	XP_001301506.1	OHS97192.1
MSH6 (GTBP, MSH8) MutS homolog 6	NM_000179	XP_001322723.1	OHT00741.1
MLH1 MutK homolog 1	NM_000249	XP_001317043.1	OHT10723.1
MLH2* MutL homolog 2	x	x	x
MLH3 MutL homolog 3	NM_014381	x	x
PMS1	NM_000534	XP_001322219.1	OHT04625.1
PMS2	NM_000535.7	XP_001301639.1	x
LIGI (CDC9) DNA ligase I	NM_000234	XP_001581589.1	OHT10360.1
Polε DNA polymerase	NM_006231	x	OHT04305.1
Polδ DNA polymerase	NM_002691	XP_001303643.1	x
Polβ DNA polymerase	NM_002690	x	x
EXO 1 (HEX1) Exonuclease 1	NM_003686	XP_001325288.1	OHT16064.1

Table 3.2: Genes Involved in Mismatch Repair in Humans and their Homologs in *T. vaginalis* and *T. foetus*. Gene designations and protein names of common proteins involved in MMR are listed with the accession numbers for the corresponding human nucleotide sequences. BLAST search was used to identify homologs within *T. vaginalis* (taxid:412133) and *T. foetus* (taxid:1144522). * The accession number for MLH2 from *S. cerevisiae* (NP_013135) was used as a query term as there is no MLH2 homolog in *H. sapiens*.

3.2.3. Non-homologous End Joining

Non-homologous end joining (NHEJ) repairs double strand breaks (DSB) in DNA. DSBs can be caused by ROS action, ionising radiation (Costes et al., 2010), Type 2 topoisomerase stalling and chemotherapeutic agents. This DNA repair mechanism requires the recognition of DNA ends by the Ku heterodimer, made up of Ku70 and Ku80 proteins, which binds to the DSB (Davis and Chen, 2013). After binding, the Ku heterodimer recruits the other main enzymes involved in NHEJ to the DNA ends, which subsequently repair the DSB. These are: DNA-dependent protein kinase catalytic subunit (DNA-PKc) (Uematsu et al., 2007), XLF (Tsai, Kim and Chu, 2007), X-ray cross complementing protein 4 (Yano et al., 2007), and DNA ligase IV (Costantini et al., 2007). DNA-PKcs bind to Ku80 and the phosphorylation of DNA-PKcs causes a conformational change allowing the activation of end processing nucleases, such as PNKP, and dissociation of DNA-Pkcs (Gottlieb and Jackson, 1993., (Goodarzi et al., 2006). XRCC4-Ligase IV complex and XLF then take part in ligation of the DNA strand (Gu et al., 2007).

Fifteen genes identified in human NHEJ were used to search for in *T. vaginalis* and *T. foetus* (Table 3.3). Both *T. vaginalis* and *T. foetus* only carry a single gene encoding a protein that participates in human NHEJ: MRE11. Notably, MRE11 also participates in homologous recombination (see next section). The results of the BLAST searches are shown in Table 3.3.

Gene designation	***Homo sapiens***	***Trichomonas vaginalis***	***Tritrichomonas foetus***
MRE11 (Rad32)	NM_005590	XP_001321917.1	OHT17588.1
XRCC1 X-ray repair cross complementing protein 1	NM_006297	x	x
XRCC4 (Lif4) X-ray repair cross complementing protein 4	NM_003401	x	x
XRCC5 (Ku80 Pku80 Yku80) X-ray repair cross complementing protein 5	NM_021141	x	x
XRCC6 (Ku70, Pku70, Yku70) X-ray repair cross complementing protein 6	NM_001469	x	x
DNA-PK (XRCC7) DNA dependant protein kinase	NM_006904	x	x
Artemis (DCLRE1C)	NM_022487	x	x
XLF (NHEJ1) XRCC4-like factor	NM_024782	x	x
APTX Aprataxin	NM_175073	x	x
APLF/PALF Aparataxin and PNKP-like factor	NM_173545	x	x
CtIP (Sae2, RIM, RBBP8) Carboxy-terminal BRCA1-interacting protein	NM_002894	x	x
PNKP Polynucleotide kinase-phosphatase	NM_007254	x	x
LIG III DNA ligase	NM_013975	x	x
LIG IV (Lig4p) DNA ligase	NM_002312	x	x
Polλ DNA polymerase	NM_001174084	x	x
Pol μ DNA polymerase	NM_013284	x	x

Table 3.3: Genes Involved in Non-Homologous End Joining in Humans and their Homologs in *T. vaginalis* and *T. foetus*. Gene designations and protein names of common proteins involved in NHEJ are listed with the accession numbers for the corresponding human nucleotide sequences. BLAST search was used to identify homologs within *T. vaginalis* (taxid:412133) and *T. foetus* (taxid:1144522).

The BLAST search for DNA dependant protein kinase (NM_006904) in both *T. vaginalis* and *T. foetus* lists hits with acceptable E values; XP_001317016.1 and OHT01329.1 respectively. However, when used in a reciprocal BLAST as the query against the *H. sapiens* database both result in FKBP-rapamycin associated protein as the top hit and so were not counted as DNA dependant protein kinase homologs.

3.2.4. Homologous Recombination

Homologous recombination (HR) also repairs DSB and interstrand crosslinks and stalled replication forks. Unlike NHEJ HR repairs DSBs by using the sister chromatid sequence as a template strand for high fidelity repair. Due to the use of sister chromatids HR is restricted to S and G2 phases of the cell cycle whereas NHEJ can occur in any phase. The first step in HR known as 'pre-synapsis' is the processing of the DSB to a 3' overhang tail. MRE11-RAD50-NBS1 complex (MRN complex) recognises the DSBs which are processed by helicases and nucleases including; CtIP, BLM, EXO1, and DNA2 (Sartori et al., 2007., Nimonkar et al., 2011). The exposed 3' single stranded overhangs are coated by replication protein A (RPA) to prevent nuclease digestion and secondary structure formation. RPA is then removed by recombination mediators, such as BRCA2, which allow Rad51 to bind to ssDNA forming the 'pre-synaptic filament' Rad51 nucleoprotein filament (Sung et al., 2003). The second step consists of homology search and DNA strand invasion together known as synapsis. Rad51 nucleoprotein filament catalyses homology search and DNA strand exchange into the

sister chromatid, forming a joint molecules known as D-loops. The third HR step is the postsynaptic phase where the 3' end of the invading strand is extended by DNA synthesis. RAD54 enables the disassociation of RAD51 to allow DNA synthesis mainly by polymerase eta in mammals (McIlwraith et al., 2005).

39 genes identified within homologous recombination in humans and 4 from *S. cerevisiae* were analysed and searched for within the *T. vaginalis* and *T. foetus* databases. The results are shown in Table 3.4 and Table 3.5 respectively.

Gene designation	**Protein**	***Homo sapiens***	***Trichomonas vaginalis***	***Tritrichomonas foetus***
Ctip (Sae2, RIM, RBBP8)	Carboxy-terminal BRCA1 interacting protein	NM_002894	x	x
MRE11 (Rad32)	Meiotic recombination 11	NM_005590	XP_001321917.1	OHT17588.1
RAD50	MRN complex protein	NM_005732	ABC61989.1	OHS93514.1
NBS1 (Xrs2)	Nijmegen breakage syndrome 1	NM_002485	x	x
BLM (Rqh1, Sgs1)	Bloom syndrome helicase	NM_000057	XP_001326459.1	OHS96980.1
RMI1	RecQ-mediated genome instability protein 1	NM_024945	x	x
RMI2	RecQ-mediated genome instability protein 2	NM_152308	x	x
Top3α	DNA topoisomerase 3α	NM_004618	XP_001308043.1	OHT04075.1
RecQL5 (FBH1, Srs2)	Antirecombinase	NM_001003715	XP_001307056.1	x
DNA2	ssDNA endonuclease	NM_001080449		OHT11922.1
EXO1 (HEX1	Exonuclease 1	NM_003686	XP_001325288.1	OHT16064.1
RPA1 (p70)	Replication protein A 1	NM_002945	XP_001311988.1	OHT09105.1
RPA2 (p32)	Replication protein A 2	NM_002946	x	x
RPA3 (p11)	Replication protein A 3	NM_002947	x	x
RAD51 (Rhp51)	RecA homolog	NM_002875	XP_001303202.1	OHT07838.1 OHT00703.1
RAD51B (RAD51L1)	RAD51 paralog	NM_002877	x	x
RAD51C (RAD51L2 ,FANCO)	RAD51 paralog C	NM_002876	x	x
RAD51D (RAD51L3)	RAD51 paralog D	NM_002878	x	x
XRCC2	X-ray repair cross complementing protein 2	NM_005431	x	x
XRCC3	X-ray repair cross complementing protein 3	NM_005432	XP_001301981.1	OHT11229.1
RAD52	Recombination mediator	NM_134424	x	x

RAD54L	RAD54-like	NM_003579	x	OHT17419.1
RAD54B (Rdh54)	RAD54 homolog B	NM_012415	XP_001304972.1	x
BRCA1	Breast cancer type 1	NM_007294	x	x
BRCA2 (FANCD1	Breast cancer type 2	NM_000059	XP_001316845.1	OHT06102.1
DSS1 (SHFM1)	Deleted in split hand/split foot 1	NM_006304	x	x
MUS81	Methyl methanesulfonate and UV sensitive 81	NM_025128	x	x
EME1 (Mms4)	Essential meiotic endonuclease 1	NM_152463	x	x
SLX1 (GIYD1)	Synthetic lethal X 1	NM_001014999	XP_001313690.1	OHT01848.1
SLX4 (BTBD12, FANCP	Synthetic lethal X 4	NM_032444	x	x
GEN1 (Yen1)	XPG-like endonuclease 1	NM_182625	x	x
PARI (Srs2)	PCNA-associated recombination inhibitor	NM_017915	x	x
PALB2 (FANCN)	Partner and localizer of BRCA2	NM_024675	x	x
WRN	Werner syndrome	NM_000553	XP_001307056.1	OHS96980.1
Polα	DNA polymerase alpha	NM_002689	XP_001316810.1	x
Polδ	DNA polymerase delta	NM_002691	XP_001303643.1	x
Polε	DNA polymerase epsilon	NM_006231	x	OHT04305.1
Polζ	DNA polymerase zeta	NM_002912	XP_001303643.1	OHT08034.1

Table 3.4: Genes Involved in Homologous Recombination in Humans and their Homologs in *T. vaginalis* and *T. foetus*. Gene designations and protein names of common proteins involved in HR are listed with the accession numbers for the corresponding human nucleotide sequences. BLAST search was used to identify homologs within *T. vaginalis* (taxid:412133) and *T. foetus* (taxid:1144522).

Gene designation	*Saccharomyces cerevisiae*	*Trichomonas vaginalis*	*Tritrichomonas foetus*
Shu1	YHL006	x	x
Psy3	YKR376C	x	x
Rad55	YDR076W	x	x
Rad57	YDR004W	x	x

Table 3.5 Genes Involved in Homologous Recombination in *S. cerevisiae* and their homologs in *T. vaginalis* and *T. foetus*. Gene designations and protein names of common proteins involved in HR are listed with the accession numbers for the corresponding yeast protein sequences. BLAST search was used to identify homologs within *T. vaginalis* (taxid:412133) and *T. foetus* (taxid:1144522).

T. foetus ATP-dependent DNA helicase, RecQ family protein (OHS96980.1) came up as a top hit for both Bloom syndrome protein (BLM) (NM_000057) and ATP-dependent DNA helicase Q5/recombinase (RecQL5) (NM_001003715). To determine which gene OHS96980.1 was most similar to, the conserved domains of each protein were compared using NCBI CDD. All three proteins have a specific hit for the RecQ family domain. A reciprocal BLAST search was used to further analyse the protein which showed BLM as the top hit. OHS96980.1 was listed as BLM homolog and the other hits listed in the BLAST search were analysed to determine if there is a RecQL5 homolog in *T. foetus*. The other hits with acceptable e values were used in reciprocal BLASTs; OHT16234.1, OHS98960.1 and OHT08502.1 which all came up with BLM as the top hit so were deemed not RecQ5 homologs. Four more hits with acceptable E values were ruled out as possible homologs due to low alignment scores <50.

T. vaginalis DNA repair protein RAD51 homolog (XP_001303202.1) was shown as the top hit for both RAD51 (NM_002875) and RAD51B (NM_002877). A reciprocal BLAST search was carried out, which revealed XP_001303202.1 to be a

RAD51 homolog and it was therefore listed in the table as such. The second hit with an acceptable E value for RAD51B was the meiotic recombination protein DMC1/LIM15 homolog (XP_001303137.1). This protein was further analysed to determine if there is a RAD51B homolog within *T. vaginalis*. This reciprocal BLAST showed DNA meiotic recombinase 1 (NM_007068.4) as the top hit: therefore *T. vaginalis* does not have a RAD51B homolog. In *T. foetus* RAD51B BLAST search shows a protein annotated in the database as DNA repair protein RAD51 1 (OHT00703.1) as a result, but reciprocal BLAST shows it to be more similar to RAD51A.

When searching for RAD54L homologs within *T. vaginalis*, a protein annotated as Type III restriction enzyme, res subunit family protein (XP_001330700.1) came up as a top hit with an acceptable E value but upon further analysis with reciprocal BLAST it was found to be more similar to *H. sapiens* excision repair 6 chromatin remodelling factor (ERCC6) (NM_000124.4) and so it was discounted as a possible RAD54L homolog.

The GEN1 blast search within *T. vaginalis* and *T. foetus* resulted in top hits with acceptable E values; XP_001294157.1 and OHS94785.1 respectively. These results were ruled out as possible GEN1 homologs due to the results of reciprocal BLASTs with the proteins against the *H.* sapiens database not showing GEN1 as the top hit. Instead, XP_001294157.1 and OHS94785.1 were found to be most similar to *H. sapiens* FEN1 (NM_004111.6).

The blast search for Polδ in *T. foetus* showed a top hit with an acceptable E value, OHT08034.1, which was ruled out as discussed in Section 3.2.1. The same hit

was shown when searching for a Polymerase ζ homologue for homologous recombination in *T. foetus*. The reciprocal BLAST search using OHT08034.1 as query returned Polζ as the top hit therefore OHT08034.1 was listed as a Polζ homologue.

Four genes involved in homologous recombination in *S. cerevisiae* were included in the analysis of *T. vaginalis* and *T. foetus*. Shu1, Psy3, Rad55, and Rad57 play important roles in yeast homologous recombination and are not present in humans. These four genes are also missing from both *T. vaginalis* and *T. foetus* (Table 3.5). This is consistent with findings from other excavates including; *Leishmania infantum* (*L. infantum*), *Leishmania major* (*L. major*), *Trypanosoma brucei* (*T. brucei*) and *Trypanosoma cruzi* (*T. cruzi*) (Genois et al., 2014).

3.3. AAG Type Enzymes Throughout the Kingdom of Life

I wanted to explore the spread of AAG within other organisms, including other parasitic protists. To do this I used NCBI's BLAST search to search for AAG homologs in several species from different kingdoms, including other parasitic members of the Excavata. I also searched for the DNA methyltransferases; DNMT1, DNMT3A and DNMT3B. DNA methyltransferases are a family of enzymes which catalyse the transfer of a methyl group to DNA. The rationale for this is that it has been suggested that AAG may play an important role in countering the pro-mutagenic effect of aberrant methylation.

The results identified six species that possessed all three DNMTs in addition to AAG: *Homo sapiens, Xenopus laevis, Danio rerio, Nematostella vectenis, Ciona intestinalis* and *Amphimedon queenslandica*, all of which belong to the Animalia kingdom. There was no otherwise obvious correlation in the conservation of any of the DNA methyl transferases and having a homologue of AAG (Table 3.6).

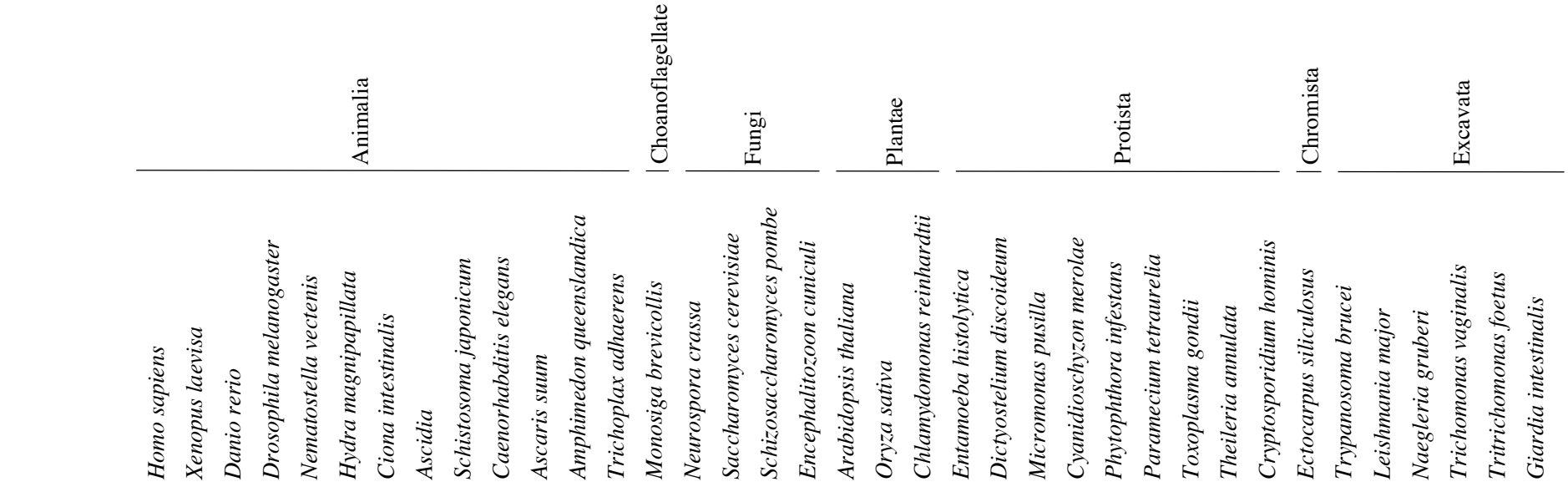


Table 3.6: Analysis of Presence of AAG, DNMT1, DNMT3A, DNMT3B in a Range of Eukaryotic Species. The accession numbers for the human nucleotide sequence for each enzyme (AAG: NP_002425.2) (DNMT1: NP_001370.1) (DNMT3A: NP_072046.2) (DNMT3B: NP_008823.1) were used in BLAST search against the database for each species. Results were accepted if the expected value was less than 1e-4. Blue shaded boxes show the presence of the enzyme.

T. vaginalis and the related organism *T. foetus* are the only members of the Excavata that we have found to have one or more homologue of AAG, I was interested in how it acquired the gene as its evolutionary relatives do not have a copy. To begin to answer this question I did an unrestricted BLAST search with *T. vaginalis* TvAAG1 (XP_001311901.1) TvAAG2 (XP_016784459.1) and the *T. foetus* AAG (OHT07844.1*) to* identify the organismal source for the top five hits for each gene (Table 3.7). In each case, the top hits were not derived from eukaryotes but were instead prokaryotic homologues. The only exception was for TvAAG1, where the second most similar protein in the NCBI database was derived from *T. foetus*.

TvAAG1	**TvAAG2**	**OHT07844.1**
Mariniphaga sediminis	*Mariniphaga sediminis*	*Parabacteroides chartae*
Tritrichomonas foetus (OHTO07844.1)	*Draconibacterium orientale*	*Macellibacteroides sp. HH-ZS*
Draconibacterium orientale	*Draconibacterium sediminis*	*Prolixibacteraceae bacterium XSD2*
Parabacteroides chartae	*Parabacteroides chartae*	*Dysgonomonas capnocytophagoides*
Macellibacteroides sp. HH-ZS	*Macellibacteroides sp. HH-ZS*	*Psychromonas aquimarina*

Table 3.7 Top Five Most Similar Proteins Identified by an Unrestricted BLAST Search using TvAAG1 and TvAAG2 as a Search Term. Presented in the order as shown in the BLAST significant alignments list. TvAAG1 shows TvAAG2 as the top hit and vice versa but were ruled out as the purpose of this table is to identify AAGs in other species. Hits that were present in all three searches are highlighted.

3.4. Discussion

Analysis of DNA repair pathways within *T. vaginalis* and *T. foetus* shows that both parasites have a greatly reduced number of DNA repair enzymes compared to humans. In comparison to each other the parasites have similar conserved DNA repair enzymes, which is to be expected as the two parasites are closely related in evolutionary terms (Adl et al., 2012).

In BER the parasites are missing some enzymes and proteins essential to the human pathway, XRCC1, DNA polymerase β (Polβ) and DNA ligase III (LigIII). However, they do have copies of certain DNA glycosylases and other BER components such as APE1, FEN1 and PCNA. The parasites have a reduced library of glycosylases yet the three that are present, UDG, NTHL1 and MPG recognise the majority of lesions repaired by BER. Surprisingly the parasites do not have a copy of an 8oxoG specific glycosylase such as OGG1. 8oxoG is one of the major oxidative lesions in mammals so it would be expected that *T. vaginalis* and *T. foetus* have a way of repairing it. The only reported members of the Excavata to possess copies of OGG1 are *L. major* and *T. brucei* (Genois et al., 2014). *T. vaginalis* and *T. foetus* lack LigIII which was previously thought of as the main ligase involved in short-patch BER, however studies have since shown LigIII is not essential in short patch XRCC mediated BER repair (Simsek et al., 2011 and Gao et al., 2011) suggesting LigIII can be substituted by another ligase during BER. Indeed, LigI has previously been shown to substitute for LigIII/XRCC *in vivo* (Sleeth, Robson and Dianov, 2004) *T. vaginalis* and *T. foetus* do possess a copy of LigI which likely replaces LigIII. This finding is similar to analysis of DNA repair in the

parasitic protist *Entamoeba histolytica* which lacks LigIII and contains a copy of LigI for sealing of DNA during DNA repair (Azuara-Liceaga et al., 2018). *S. cerevisiae* also lacks a LigIII homologue and possesses a LigI homologue CDC9 (Tomkinson, Tappe and Friedberg, 1992). LigI is known to participate in the long patch repair pathway of BER. In humans, DNA Polδ and DNA Polε are used for synthesis of DNA as an alternative to DNA Polβ in the long patch repair pathway of BER (Fortini et al., 1998). *T. vaginalis* has a copy of DNA Polδ and *T. foetus* has a copy of DNA Polε suggesting BER in the parasites takes place via the long patch repair pathway. This is supported by both parasites having a copy of PCNA and RF-C proteins essential for LP-BER (Gary et al., 1999; Levin et al., 2000). It is therefore likely that *T. vaginalis* and *T. foetus* have a preference for LP BER.

Throughout the DNA repair pathways there is some overlap of proteins involved in more than one repair pathway. The majority of these proteins are also involved in replication. Almost all of the genes encoding replicative proteins are conserved in *T. vaginalis* and *T. foetus*. LigI is the main ligase involved in the joining of Okazaki fragments during lagging strand synthesis of the lagging DNA strand in replication (Okazaki et al., 1968). Studies have shown in LigI null mutants are viable which suggests LigIII or LigIV can act as alternate ligases during Okazaki fragment ligation (Han et al., 2014) LigI is still considered the primary replicative ligase due to the fact it is conserved throughout all mammals (Petrini, Xiao and Weaver, 1995) and this holds true in *T. vaginalis* and *T. foetus*. In eukaryotes DNA Polα, Polδ and Polε are considered as the three main replicative polymerases. Whilst there are three different models of the roles of polymerases at the replication fork the main consensus is that DNA Polα initiates DNA synthesis, DNA Polε continues synthesis of the leading strand whilst Polδ

synthesises the lagging strand ((Stodola and Burgers, 2016). Out of the 7 DNA polymerases analysed *T. vaginalis* contains three, Polα, Polδ and Polζ, and *T. foetus* contains two, Polε and Pol ζ. This is a drastic reduction compared to other excavates, *Leishmania major, Leishmania infantum, T. cruzi* and *T. brucei* all possess seven polymerases (Genois et al., 2013). *T. vaginalis* DNA replication likely takes place in the same manner as other eukaryotes with respect to DNA Polα initiating replication however, as the parasite lacks DNA Polε, DNA Polδ may synthesise both the leading and lagging DNA strands. Polδ has been shown to synthesise both strands in *Schizosaccharomyces pombe* by Miyabe et al. (2015) and by Johnson et al 2015. *T. foetus* only possesses Polε out of the three main replicative polymerases, suggesting that it is solely responsible for DNA replication within the parasite. Human and *S. cerevisiae* Polε has been shown to participate in leading strand synthesis (Shinbrot et al., 2014) but not lagging strand synthesis which leads to the question of how *T. foetus* would replicate the lagging strand. Other replicative proteins that the parasites possess include the PCNA sliding clamp. The flap endonuclease FEN-1 and RF-C.

Mismatch repair works to repair base substitutions and insertion-deletion loops. *T. vaginalis* and *T. foetus* are missing some proteins and enzymes involved in mismatch repair, although they do possess the majority of the components essential to the pathway. In eukaryotes mismatch repair is initiated by recognition of base substitution or insertion-deletion mismatches by the MutSα heterodimer or the MutSβ heterodimer. The two heterodimers differ in their recognition of substrate. Both parasites lack MSH3 part of the MutSβ heterodimer, this suggests that they can only repair insertion-deletion mismatches of up to 3 nucleotides long via MutSα as MutSβ recognises longer IDLs. Other species have been found to lack a MSH3 homologue including, *Drosophila*

(Sekelsky et al. 2000), *Caenorhabditis elegans* (Marti, Kunz and Fleck, 2002) and *Encephalitozoon cuniculi (E. cuniculi)* (Gill and Fast, 2007*)*. After recognition the MLH1/PMS2 heterodimer is thought to carry out the incision of the daughter strand, *T. vaginalis* has copies of MLH1/PMS2 however, *T. foetus* is lacking a PMS2 homologue. It could be the case that PMS1, which *T. foetus* does have a copy of, replaces the function of PMS2, or that the daughter strand is incised by a separate endonuclease. MSH2 and MLH1 recruit exonuclease 1 in eukaryotic MMR. Both parasites have copies of MSH2, MLH1 and Exo1 which, with DNA Polδ (*T. vaginalis*) and DNA Polε (*T. foetus)*, can work to complete the repair of base substitution mismatches or insertion-deletion loops.

Homologous recombination repair is one of two pathways responsible for repairing double strand breaks in humans, the second pathway is non-homologous end joining. Out of the 39 HR components analysed *T. vaginalis* possesses 15 and *T. foetus* possesses 16, which includes two homologs of Rad51. MRE11 and RAD50 make up the essential MRN complex which acts as a DSB sensor. In mammals the Nijmegen breakage syndrome 1 protein (NHS1), or Xrs2 in yeast, is also part of the MRN complex. In humans, mutations in the NHS1 gene result in Nijmegen breakage syndrome, a condition which causes microcephaly and cancer predisposition (Carney et al., 1998; Antoccia et al., 2006). T. *vaginalis* and *T. foetus* contain copies of MRE11 and RAD50 but surprisingly do not contain a homologue for the NHS1 protein. This finding is similar to analysis of HR in *Plasmodium falciprium* and *Toxoplasma gondii* which also lack a NHS1/Xrs2 homologue (Fenoy et al., 2016; Badugu et al., 2015). The parasites may contain another gene that can take over the role of NHS1 in the MRN complex or contain such highly divergent copies that bioinformatic analysis has not

identified them. *T. vaginalis* and *T. foetus* contain homologs for the main enzymes and proteins involved in homologous recombination; Mre11, Rad51 , Rad 50, RPA1, BLM, BRCA1, Exo1 and polymerase δ or ε (as previously discussed) so most likely use solely this pathway to repair DSBs.

In respect to non-homologous end joining both *T. vaginalis* and *T. foetus* have only one enzyme, the MRE11 homolog, out of the 15 analysed. MRE11 is also involved in homologous recombination. It is therefore highly likely that both species are unable to repair DSBs by non-homologous end-joining. This is consistent with other studies on Excavata. Analysis on the *Plasmodium falciparum* genome sequence did not find any proteins involved in NHEJ, except from MRE11, whilst the main proteins involved in HR; MRE11, DMC1, Rad50 and Rad51 were present (Gardner et al., 2002). There is also evidence of HR in *Giardia intestinalis* (Sandoval-Cabrera et al., 2015; Ordoñez-Quiroz et al., 2018) In *Trypanosoma brucei* HR is utilised for Variant Surface Glycogen (VSG) antigenic coat switching for the evasion of host immune systems ((Proudfoot and McCulloch, 2005; Glover, McCulloch and Horn, 2008) whilst key components of NHEJ are missing (Genois et al., 2014).

I analysed the presence of AAG in 37 different species which highlighted *T. vaginalis* and *T. foetus* to be the only excavates with a copy. The rest of the species with AAG homologues belonged to the animalia and fungi kingdoms. These results led me to question the origin of AAG in both *T. vaginalis* and *T. foetus*. An unrestricted blast search using *T. vaginalis* TvAAG1 and TvAAG2 to identify the five most highly similar AAG shows prokaryotic AAGs. Both TvAAG1 and TvAAG2 have a highly similar copy in *Parabacteroides chartae*. This bacterium is a member of the bacteroidetes

phylum. Bacteroidetes are highly abundant in the human gut (Eckburg, 2005) and their presence in the female urogeneital tract is associated with bacterial vaginosis (Ling et al., 2010). Bacterial vaginosis is also associated with an increased risk of *T. vaginalis* infection (Rathod et al., 2011; Brotman et al., 2012, Balkus et al., 2014). Horizontal gene transfer has been documented in *T. vaginalis*, as described in Section 1.2.6. Phylogenomic studies by Alsmark et al. (2013) demonstrate horizontal gene transfer between *T. vaginalis* and some bacteriodetes, it is inferred that 89% of HGT cases are from the *bacteriodes* species. Species of *bacteriodes* are prevalent in the human gut but not the female urogenital tract. An unrestricted BLAST search using the *T. foetus* AAG homologue as identified in Table 3.1 also reveals *Parabacteriodes chartae* as a top hit. It is possible that a gut dwelling ancestor of both *T. vaginalis* and *T. foetus* participated in horizontal gene transfer with an early ancestor of *Parabacteriodes chartae*. There is phylogenetic evidence that the AAG homologue in the land plant *Arabidopsis thaliana* is a product of an ancestral HGT event with bacteria (Fang et al., 2017). Therefore, it is likely that the original source of the genes encoding the two AAG homologues in *T. vaginalis* and AAG in *T. foetus* is also horizontal gene transfer.

3.5. Conclusion

The bioinformatical analysis outlined in this chapter provides a comprehensive report of the proteins and enzymes involved in four DNA repair pathways within two biologically important parasites, *Trichomonas vaginalis* and *Tritrichomonas foetus*. The analysis shows that both parasites have similar reduced libraries of DNA repair proteins and enzymes when compared to humans and other members of the excavata. *T. vaginalis* and *T. foetus* possess the proteins and enzymes that would

allow them to repair a variety of DNA lesions through base excision repair, mismatch repair and homologous recombination but not non-homologous end joining, of which only one protein was identified. I have also reported the presence or absence of AAG homologues in 37 different species across the kingdoms of life and described HGT as a possible source of the two *T. vaginalis* homologues characterised in Chapter 4.

4

Expression and Characterisation of Recombinant TvAAG1 and TvAAG2 from *Trichomonas vaginalis*

4.1 Introduction

In the second half of my project I was to continue work on the previously cloned recombinant protein *T. vaginalis* DNA-3-methyladenine glycosylase known as TvAAG1 (Daria Kania, manuscript in preparation) and to clone a possible second *T. vaginalis* DNA-3-methyladenine glycosylase homologue, referred to as TvAAG2 in this thesis. Both proteins were to be expressed and characterised by biochemical methods to identify their substrate specificities, determine their functionality and to identify their catalytic domains via site directed mutagenesis. This section of my project will try to understand why *T. vaginalis* has two AAG homologues and establish whether TvAAG2 is truly a DNA-3-methyladenine glycosylase.

4.2 Purification of TvAAG1 from *T. vaginalis*

Recombinant TvAAG1 was previously expressed in the laboratory using an expression vector derived from pET28a (see appendix 1 for plasmid map) containing the coding sequence insert for TvAAG1 between NdeI and XhoI restriction enzyme sites. The expression vector was transformed into the *E. coli* protein expression strain Rosetta 2 (DE3) pLysS and the protein was expressed using the T7 expression system. The T7 expression system works with a T7 RNA polymerase, which the Rosetta strain has a chromosomal copy of,8 that is under the control of the lac operator, and a T7 promoter region contained within the pET28a expression vector to allow transcription of the inserted gene upon induction on addition of the allolactose analogue, IPTG. The IPTG binds to the lac repressor allowing transcription of the gene by derepression of the T7 RNA polymerase. The protein was expressed from a 750 mL culture (LB medium) of cells.

Recombinant TvAAG1 could be purified using Ni-NTA affinity column chromatography. The pET28a expression vector used to express the recombinant protein contains an N-terminal 6xHis tag, which allows high affinity binding to the Ni-NTA resin on the column. The protein was purified through several fractionation steps using an increasing concentration of imidazole, which competes with the 6xHis tag to bind with Ni-NTA releasing the bound protein. To prepare the harvested cells for purification the bacterial cell pellet was resuspended in ice cold Ni-NTA containing 10 mM imidazole and sonicated to lyse the cells. The suspension was centrifuged and the supernatant ("Lysate", Figure 4.1) was incubated with the Ni-NTA resin before being poured through the column. The first collected fraction is composed of unbound

proteins that washed through the column (“Flowthrough”, Figure 4.1). The second fraction (‘’Wash 1’’, Figure 4.1) was eluted with 10 mM imidazole, the third fraction (‘’Wash 2’’, Figure 4.1) was eluted with 35 mM imidazole and the fourth fraction collected was the purified protein (‘’Elute’’, Figure 4.1) eluted using buffer containing 250 mM imidazole.

The fractions were collected and analysed, including the cell lysate from centrifugation, using an SDS-PAGE gel (Figure 4.1) and the eluted fraction was dialysed in protein dialysis buffer overnight. After dialysis the protein was centrifuged at 13,000 rpm for 10 minutes to remove any precipitate and the protein supernatant was collected. The protein was produced in good yield at the right size (23 kDa) and more than 90% pure as shown in Figure 4.1. The concentration was measured using UV-vis spectrometry and the overall yield was 1.46 mg, corresponding to 1.95 mg/ml of LB culture. The protein was stored in aliquots at -80°C.

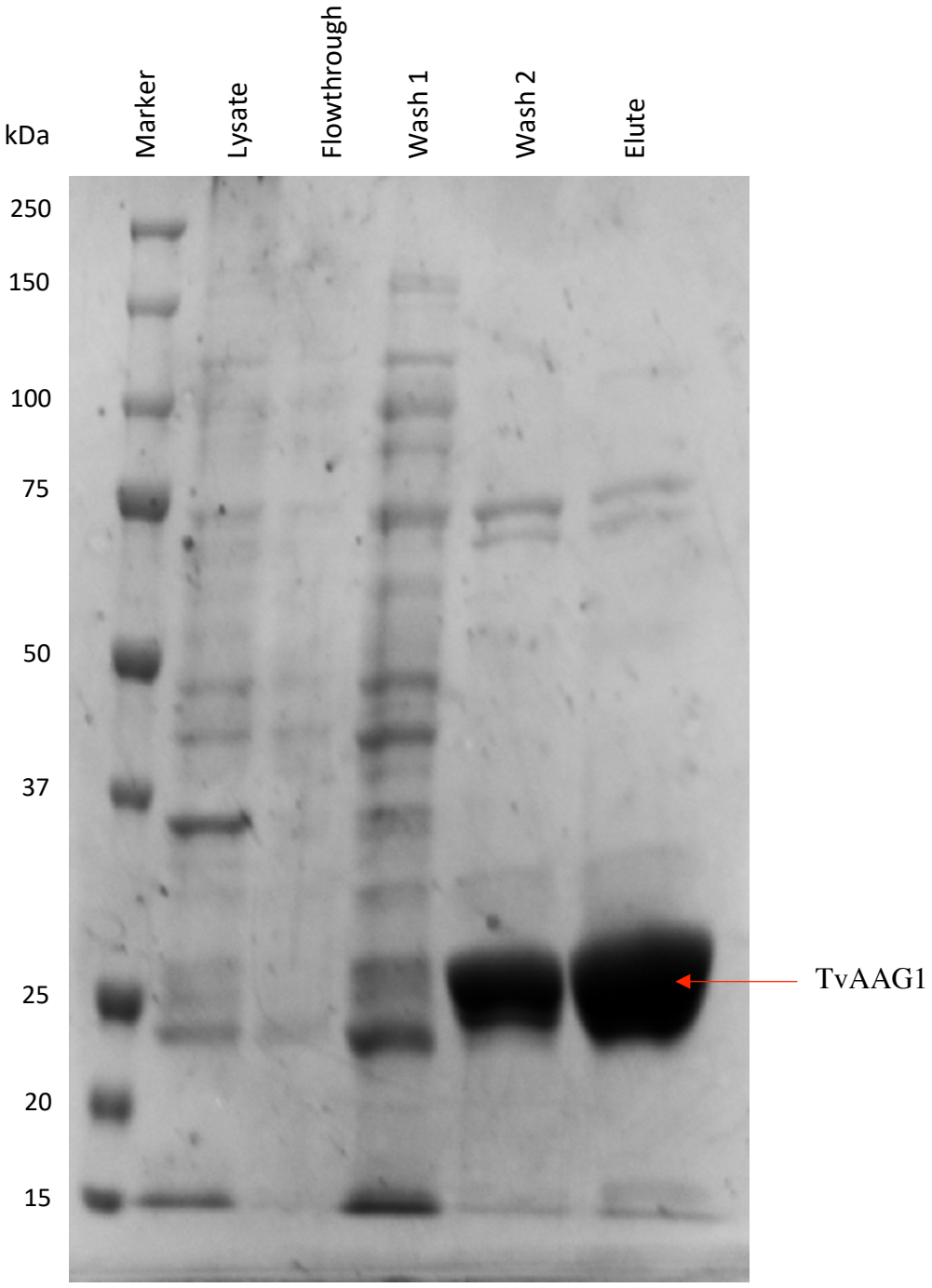


Figure 4.1 SDS-PAGE Analysis of TvAAG1 Ni-NTA Affinity Purification. TvAAG1 was expressed in Rosetta 2 BL21(DE3)pLysS cells and purified using Ni-NTA column chromatography. The protein marker is Precision Plus Protein™ All Blue from Bio-Rad. See accompanying text for details of samples.

4.3 TvAAG1 Glycosylase Activity Assays

Glycosylase activity assays were used to determine if the enzymes had activity with specific substrates and if so, to assess the amount of activity the enzyme had with each substrate. Activity was measured on the amount of substrate converted to product during the reaction. If active, the glycosylase will remove the modified base by cleaving the glycosidic bond. This leaves behind an AP site, which is labile to hydrolysis under alkaline conditions. Therefore, heating the product of the reaction with buffer containing NaOH will cut the oligomer (39mer) substrate to produce a shorter 18mer strand that will migrate further during denaturing urea PAGE (Figure 4.2).

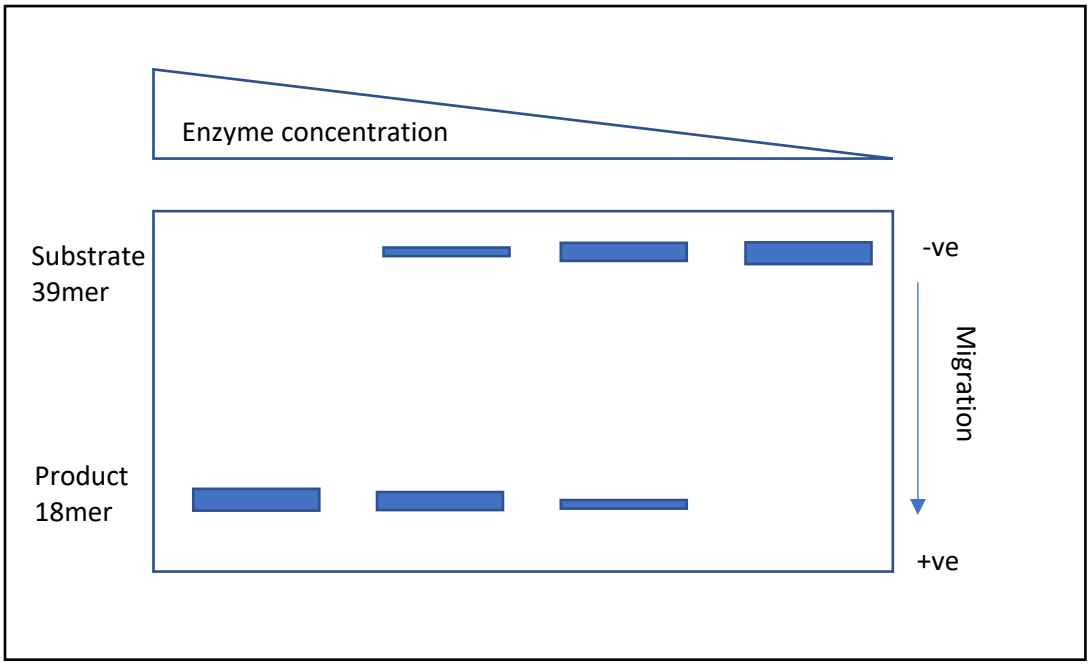


Figure 4.2 Glycosylase Activity Diagram. If the enzyme shows activity the substrate will be converted to product which can be observed in the migration of bands after cleavage with NaOH. The shorter 18mer product fragments migrate further than the 39mer substrate on the denaturing urea polyacrylamide gel.

The purified TvAAG1 protein was analysed using glycosylase activity assays with various double stranded or single stranded substrates (Figure 4.3) to try to characterise its substrate specificity. The protein was diluted in PDB to the appropriate concentrations for the assay. Substrate oligomers were labelled with Cy5 and/or TAMRA fluorescent labels for imaging. The substrates were made to a final concentration of 50 nM in a master mix. The master mix was added to TvAAG1 at varying concentrations to start the reaction. The reaction was left to incubate at 37°C for 15 minutes and the reaction was stopped by adding an equal volume of formamide buffer containing 10 mM NaOH (FA/NaOH). The samples were incubated at 95°C before loading onto an 8 M denaturing urea polyacrylamide gel. The gel was imaged using the Cy5 channel on a BioRad GelDoc imaging system.

The first substrate I investigated was deoxyinsoine (dI) opposite deoxythymidine (dT) (Substrate 1, Figure 4.3). Deoxyinosine arises through hydrolytic deamination of deoxyadenosine (dA) and causes transition mutations if not repaired (Yasui et al., 2008). dI across from thymine is a well-established substrate for AAG type enzymes (O'Brien and Ellenberger, 2003). Figure 4.4 shows the activity assay for TvAAG1 and substrate 1. A doubling of substrate bands can be seen on the gel due to the presence of both double-stranded and single-stranded forms. This is a consequence of using small gels that do not run at as high a temperature as larger gels, allowing the two strands of the substrate to re-anneal. However, the product can be clearly resolved from both the double-stranded and single-stranded forms of the substrate, so I continued to use the small gels throughout my project. TvAAG1 shows clear concentration dependant activity with substrate 1, as conversion to product decreases as TvAAG1

concentration decreases, however the concentration dependence is not linear. At 1 μM TvAAG1 all of the substrate was converted to product.

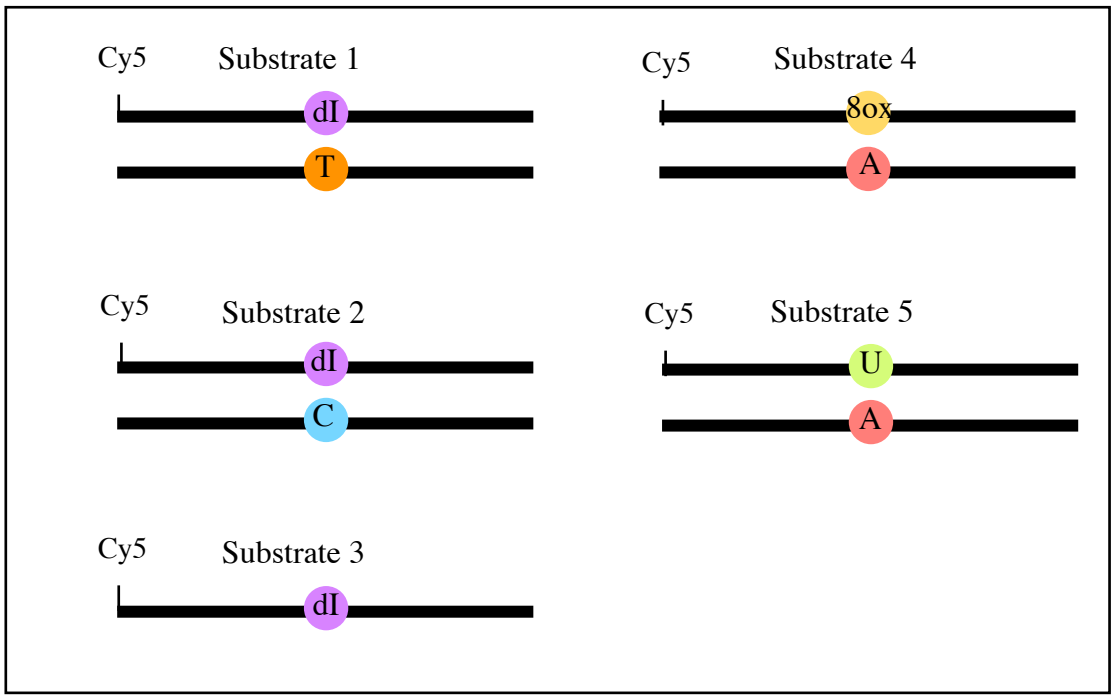


Figure 4.3 Diagram of Substrates used in Glycosylase Activity Assays. All substrates were 39 nucleotides long and made up to a final concentration of 50 nM for use in glycosylase assay. Substrate 1: Double stranded oligomer containing the modified base deoxyinosine across from deoxythymidine. Substrate 2: Double stranded oligomer containing deoxyinosine across from deoxycytidine. Substrate 3: Single stranded oligomer containing deoxyinosine. Substrate 4: Double stranded oligomer containing 8oxoG across from deoxyadenosine. Substrate 5: Double stranded oligomer containing uracil across from deoxyadenosine. A = deoxyadenosine C = deoxycytidine dI = deoxyinosine T = deoxythymidine U = deoxyuridine 8ox = 8-oxoguanine

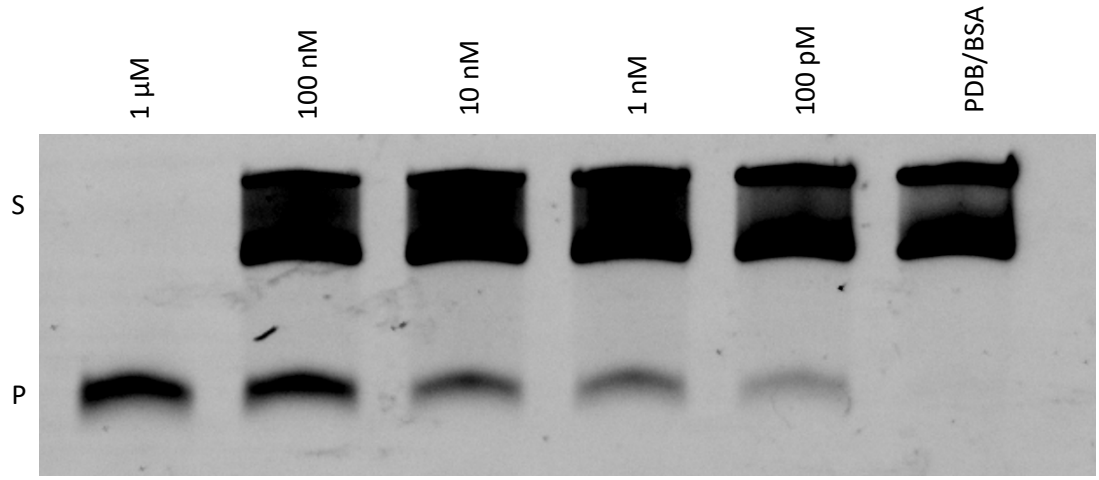


Figure 4.4 TvAAG1 Glycosylase Activity Assay with Substrate 1.

TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 1 at a final concentration of 50 nM. The final concentration of TvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated three times and a representative gel is shown.

The next substrate I investigated was dI across from cytosine (Substrate 2, Figure 4.3) TvAAG1 shows some activity at higher concentrations of the enzyme, 100 nM and 10 nM (Figure 4.5). At 100 nM all of the substrate appears to be cleaved however, there is only a weak signal for the product band. In substrate 2 the dI:dC pair forms a Watson-Crick base pair (Martin et al., 1985), where dI fits perfectly within the double helix, potentially making it harder for TvAAG1 to recognise the lesion for excision.

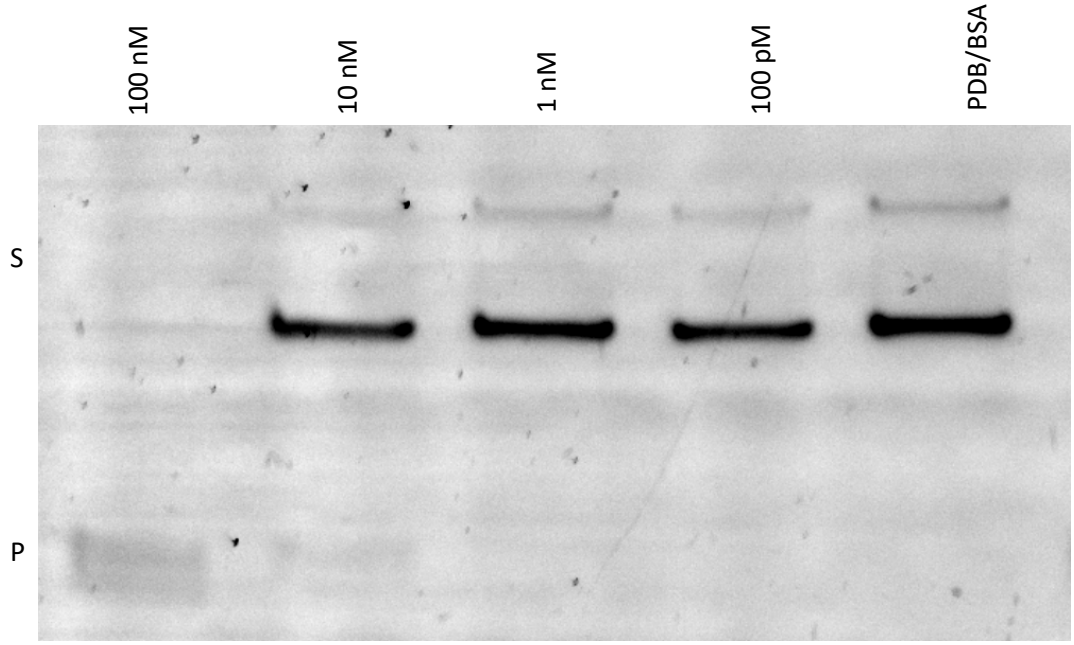


Figure 4.5 TvAAG1 Glycosylase Activity Assay with Substrate 2. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 2 at a final concentration of 50 nM, the final concentration of TvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated two times and a representative gel is shown.

Following the results of TvAAG1 with substrate 2, I wanted to investigate the activity with a single stranded dI substrate (substrate 3, Figure 4.3). TvAAG1 showed minimal activity with substrate 3, less activity than with substrate 1 and 2 as the only activity seen is at 100 nM TvAAG1 where not all of the substrate was converted to product (see Figure 4.6).

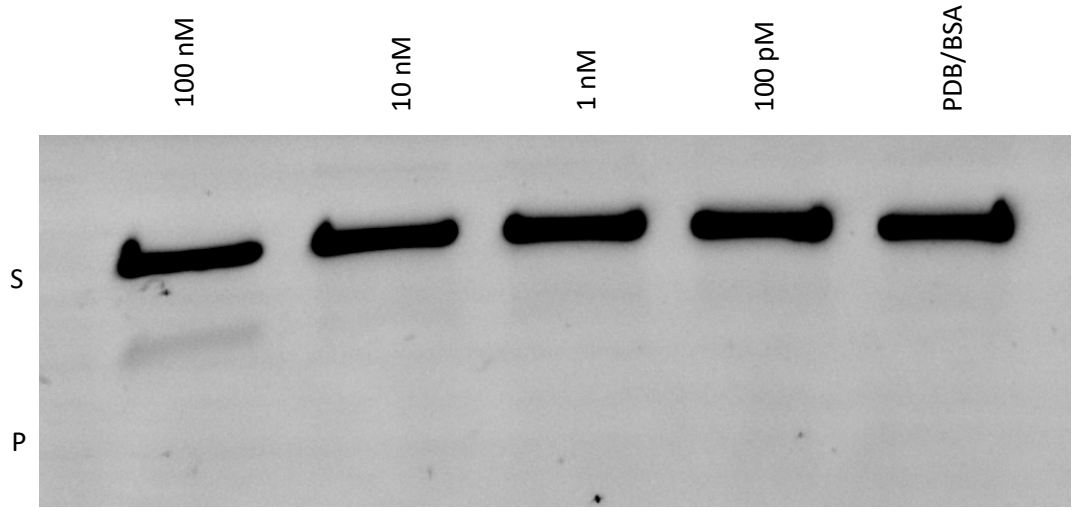


Figure 4.6 TvAAG1 Glycosylase Activity Assay with Substrate 3. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 3 at a final concentration of 50 nM, the final concentration of TvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated two times and a representative gel is shown.

Bioinformatic analysis (Chapter 3) indicates that *T. vaginalis* lacks a glycosylase for removal of 8-oxoguanine. This is a surprising observation because 8oxoG is a frequently occurring modified base (Cheng et al., 1992). I therefore decided to investigate whether TvAAG1 acts to remove 8oxoG using a double stranded oligomer containing 8oxoG across from (Substrate 4, Figure 4.3). TvAAG1 did not show any activity with substrate 4 at the concentrations tested (Figure 4.7). Human AAG has been reported to have some uracil DNA glycosylase activity (Lee et al., 2009) therefore I tested TvAAG1 activity with a substrate containing uracil (Substrate 5, Figure 4.3) TvAAG1 shows no activity with substrate 5 at the concentrations tested (Figure 4.8).

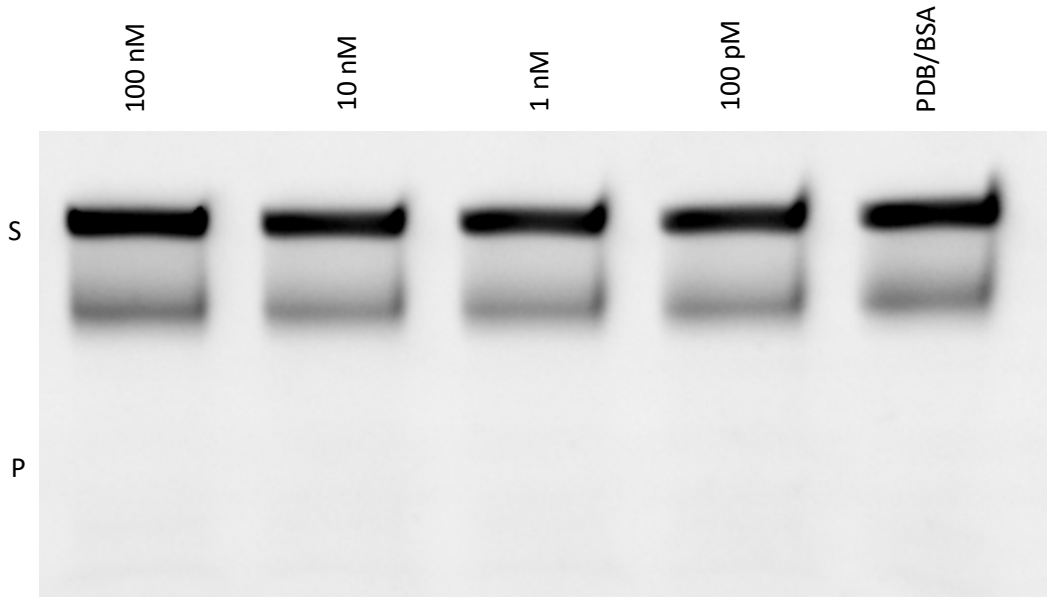


Figure 4.7 TvAAG1 Glycosylase Activity Assay with Substrate 4. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 4 at a final concentration of 50 nM, the final concentration of TvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated two times and a representative gel is shown.

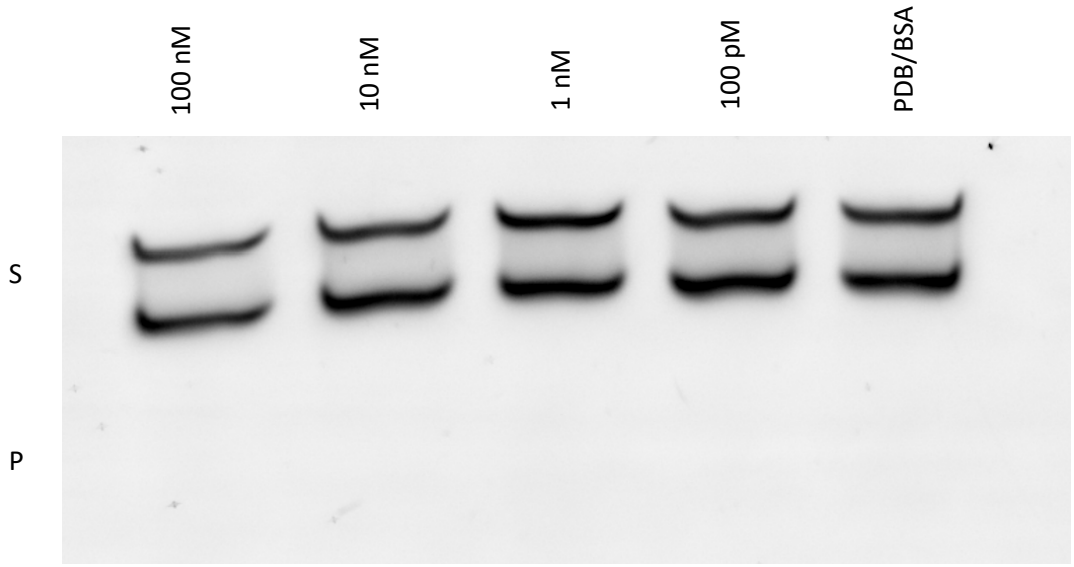


Figure 4.8 TvAAG1 Glycosylase Activity Assay with Substrate 5. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 5 at a final concentration of 50 nM, the final concentration of tvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated two times and a representative gel is shown.

4.4 Schiff Base Assay with TvAAG1

Human AAG is a monofunctional glycosylase and uses an activated water molecule as the attacking nucleophile (Lau et al., 1998). However, preliminary data in our laboratory (Daria Kania, personal communication) suggested that TvAAG1 might have an alternative mechanism and uses a lysine as the attacking nucleophile, with associated AP lyase activity (i.e. it may be a bifunctional enzyme). To determine whether TvAAG1 is indeed bifunctional, I carried out a Schiff base assay.

The Schiff base assay works on the principle of trapping a covalent intermediate that can be identified using SDS-PAGE. The intermediate is only formed if the glycosylase is bifunctional and has β-lyase activity. Borohydride reduction of the Schiff base results in formation of a covalent DNA-glycosylase intermediate which can be trapped and visualised on the gel.

TvAAG1 was incubated with substrate 1 (figure 4.3) in the presence or absence of 2.5 mM cyanoborohydride (CB) for 15 minutes at 37°C and loaded onto the SDS PAGE gel after heating at 95°C for 5 minutes (Figure 4.9). Human NTH1 (hNTH1), a known bifunctional glycosylase, was used as the control lane without CB (lane 1, Figure 4.9) and with CB (lane 2, Figure 4.9). The enzyme hNTH1 was previously purified in the lab. As expected, there are no bands in the TvAAG1 without CB lane (lane 3, Figure 4.9) or the no protein control lane (lane 5, Figure 4.9). There are two bands in the TvAAG1 with CB lane (lane 4, Figure 4.9) which suggests TvAAG1 forms an intermediary TvAAG1-DNA adduct characteristic of bifunctional glycosylases, when excising dI.

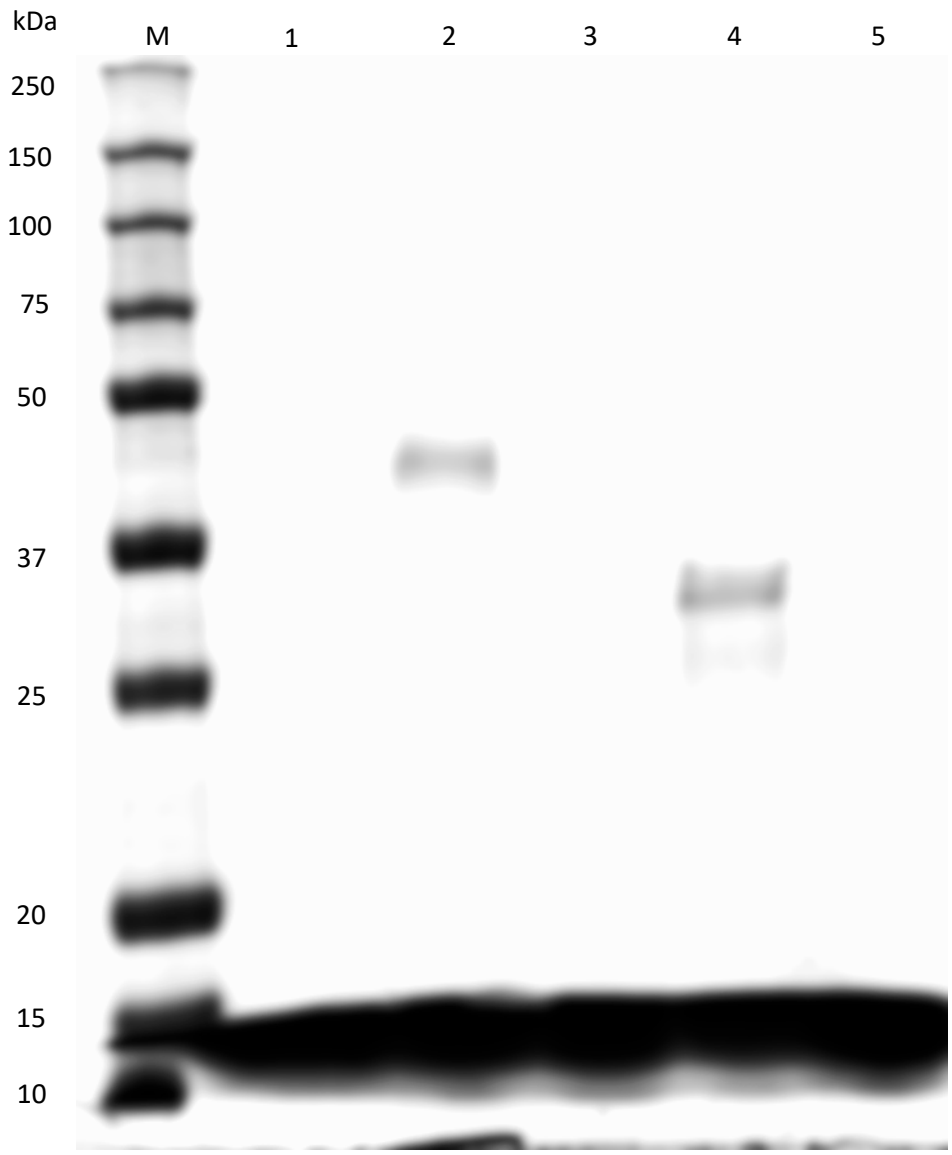


Figure 4.9 Schiff Base Assay with TvAAG1 and Cyanoborohydride. TvAAG1 and the control enzyme hNTH1 were incubated with or without cyanoborohydride (CB) at a final concentration of 2.5 mM in addition to master mix for 15 minutes at 37°C. Samples were stained with SDS loading buffer preheated for 5 minutes at 95°C and loaded alongside Precision Plus Protein™ All Blue (Bio-Rad). The gel was ran for 45 minutes at 220 V with 1 X TGS buffer and visualised with Chemidoc XRS+ Fluro orange setting. M) Marker. 1) NTH1. 2) NTH1 + CB. 3) TvAAG1. 4) TvAAG1 + CB. 5) No protein

A second Schiff base assay was conducted to test the bifunctional activity of TvAAG1 on an abasic site. An abasic site was generated in-situ by the monofunctional enzyme uracil-DNA glycosylase which acts on uracil and leaves an AP site. In theory the AP site could be acted upon by a bifunctional glycosylase. When TvAAG1 is added to the in-situ generated AP site (lane 4, Figure 4.10) there are three clear bands produced suggesting formation of the covalent intermediate formed through bifunctional glycosylase activity and borohydride trapping. As expected, there are no bands in the uracil only or TvAAG1 and uracil lanes (lanes 1 and 3, Figure 4.10). Two bands can be seen in the uracil and UDG lane (lane 2 Figure 4.10). UDG is a monofunctional glycosylase so the formation of a band is unexpected. Human UDG used in the Schiff base assay was sourced from Sigma Aldrich. The production of the covalent intermediates is possibly due to contamination of the commercial UDG used by other bacterial enzymes.

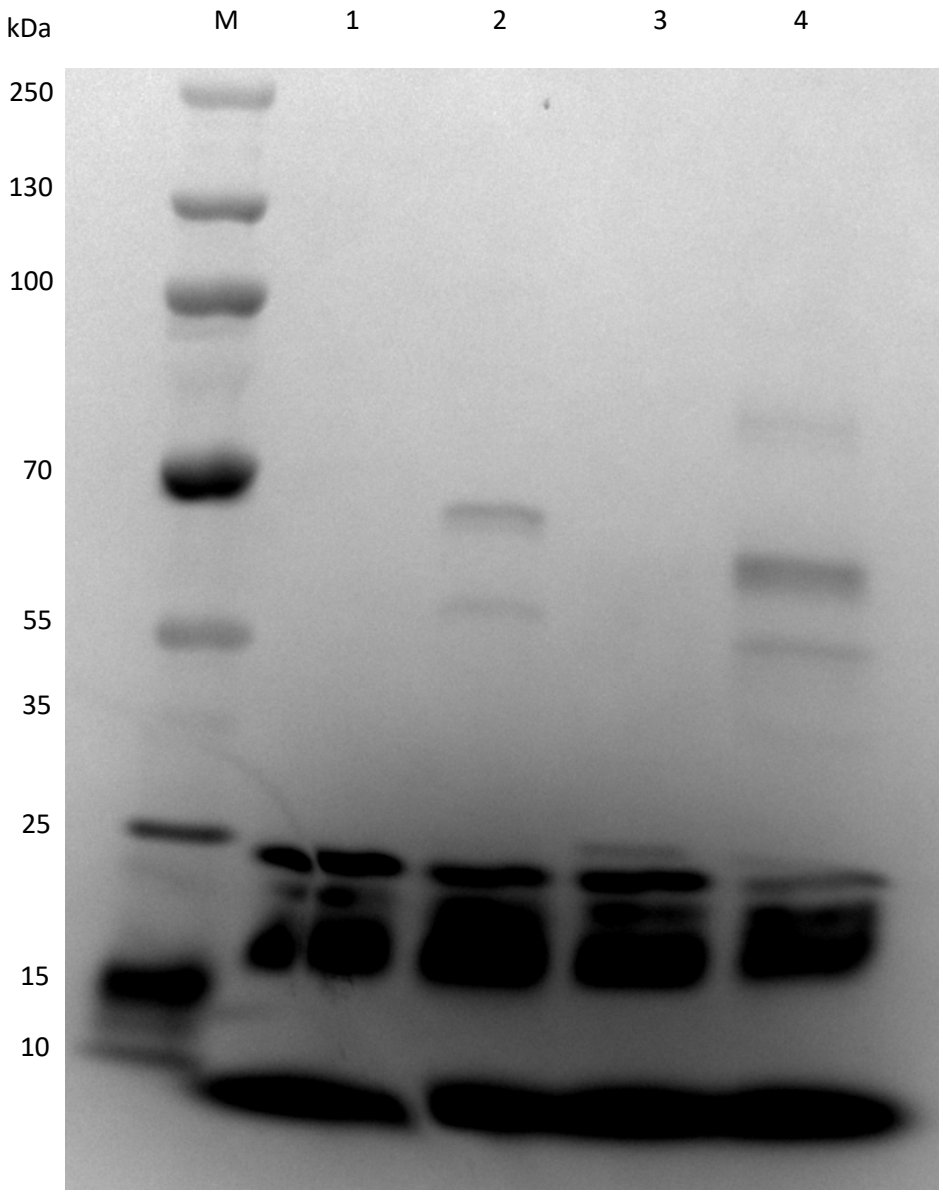


Figure 4.10 Schiff Base Assay with TvAAG1 and an *in situ* Generated Abasic Site. TvAAG1 and the control enzyme UDG were incubated with or without cyanoborohydride (CB) at a final concentration of 2.5 mM in addition to master mix for 15 minutes at 37°C. Samples were stained with SDS loading buffer preheated for 5 minutes at 95°C and loaded alongside Precision Plus Protein™ All Blue (Bio-Rad). The gel was ran for 45 minutes at 220 V with 1 X TGS buffer and visualised with Chemidoc XRS+ Fluro orange setting. M) Marker 1) Uracil. 2) Uracil + UDG. 3) TvAAG1 + uracil. 4) TvAAG1 + uracil + UDG.

4.5 Molecular Cloning and Expression of TvAAG2

The second enzyme I investigated in my project is a possible second *T. vaginalis* AAG homologue identified during bioinformatic analysis (Chapter 3) and known throughout this thesis as TvAAG2. My aim was to successfully clone and try to determine the substrate specificity for this enzyme.

A cloning vector containing the DNA sequence coding for TvAAG2 flanked by NdeI and XhoI restriction sites was obtained from the Invitrogen gene synthesis service. The expression vector pET-28a was obtained from Novagen. The TvAAG2 insert was cloned into the pET28-a vector (Figure 4.11) using NdeI and XhoI restriction enzymes. Success of the restriction reaction to produce the insert was identified via agarose gel electrophoresis. The band was cut out of the agarose gel and purified with a gel extraction kit. The insert was then ligated to pET28a cut with the same restriction enzymes. The ligation products were used to transform NEB Turbo *E. coli* competent cells, which were plated on kanamycin selection agar. Several colonies were obtained after overnight incubation. LB broth was inoculated with single colonies and these cultures were grown overnight. The cells were harvested and the plasmids from each clone isolated using Gene JET Plasmid Miniprep kit (Fisher). Plasmids were screened for the presence of insert by digestion with NdeI/XhoI (see Figure 4.12). Two of the clones contained insert of approximately the right size (lanes 2 and 3, Figure 4.12). The plasmid was then sent off for sequencing to confirm successful cloning of TvAAG2.

Protein was expressed from this vector in Rosetta^{TM} BL21(DE3)pLyyS cells using the T7 expression system. The Rosetta strain has a chromosomal copy of T7 RNA polymerase, and the pET28-a expression vector used to construct the plasmid contains a T7 promoter region and *lacI* coding sequence upstream of the insert allowing the transcription of the gene by the T7 RNA polymerase upon IPTG induction. Rosetta cells were transformed with the plasmids and plated on LB agar plates with kanamycin/chloramphenicol antibiotics. Single colonies were used to inoculate 5 ml LB broth cultures, containing kanamycin and chloramphenicol at the appropriate concentration, which was incubated overnight and used to inoculate 750 ml LB broth cultures containing kanamycin and chloramphenicol. The cells were left to grow to mid-log phase (0.6 OD). Up until this point cells were replicating but TvAAG2 was not transcribed, as the T7 promoter region on the pET28a plasmid was repressed by the lac repressor, coded for by the *lacI* gene, which binds to the lac operon upstream of the T7 RNA Polymerase gene (see Section 4.2 for more detailed explanation). After passing mid-log phase isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cultures to a final concentration of 1 mM to induce protein expression. IPTG is a molecular mimic of allolactose which can bind to the lac repressor causing a conformational change and dissociation from the operon allowing T7 RNA polymerase to transcribe the TvAAG2 sequence. The cultures were left to incubate for a further three hours after IPTG induction to allow for accumulation of TvAAG2. The cells were harvested using a centrifuge.

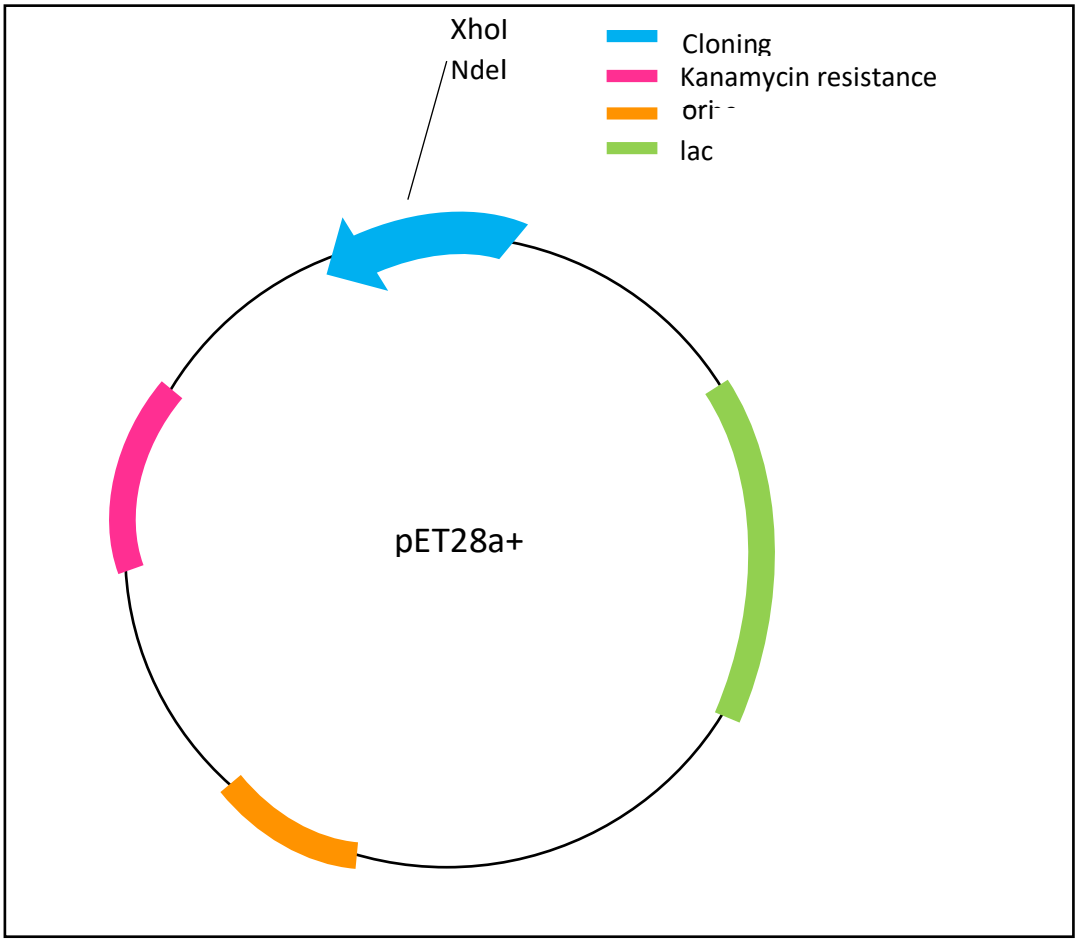


Figure 4.11 Simplified Map of pET28a+ Expression Vector

The pET28a vector was obtained from Novagen and used with NdeI and XhoI restriction enzymes to express TvAAG2.

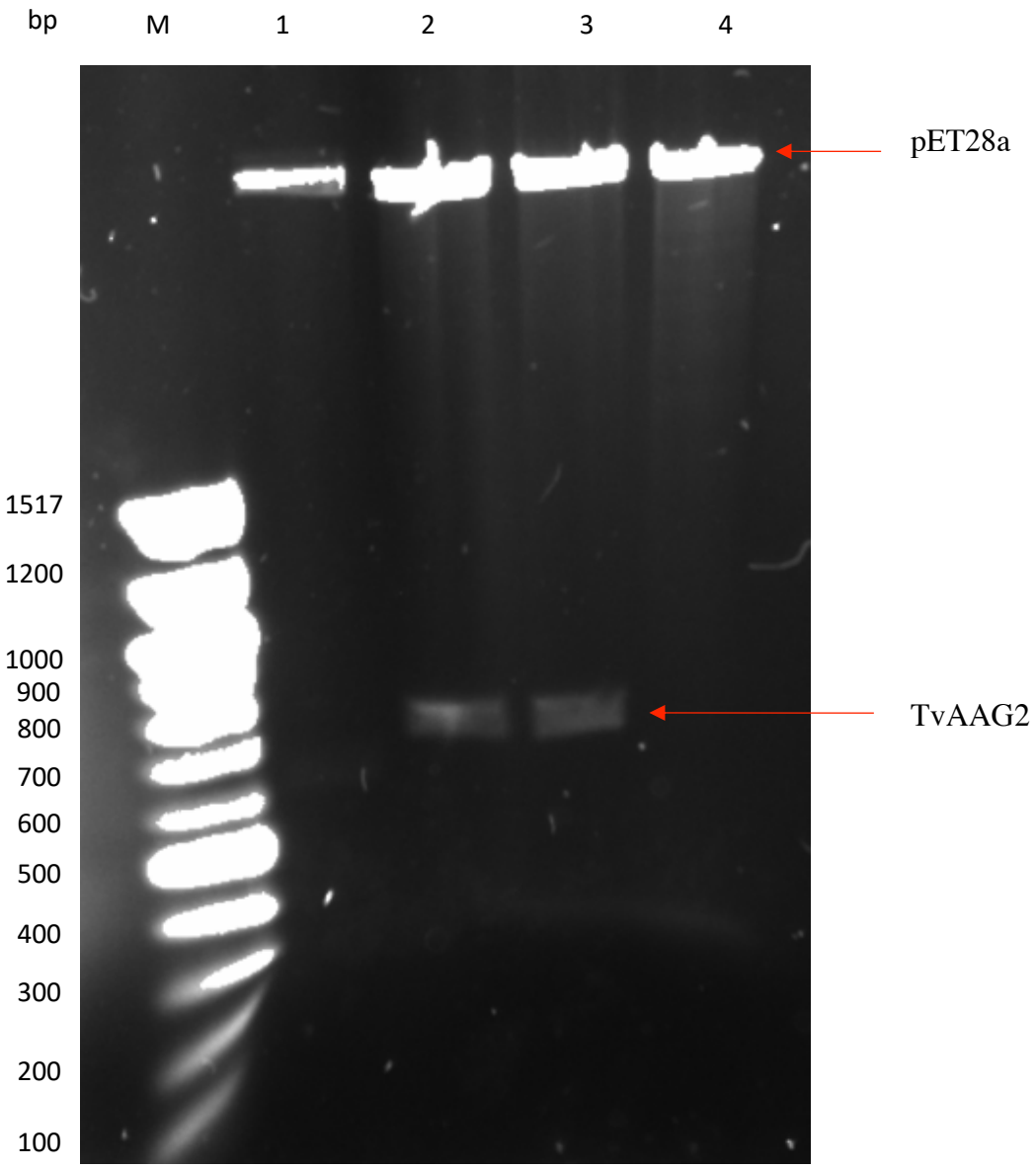


Figure 4.12 Restriction Digests of Plasmids Isolated from Clones Obtained from Ligation of TvAAG2 Coding Sequence into pET28a Vector. Agarose gel electrophoresis loaded with 100 bp DNA ladder (NEB) (lane 1) and restriction digests (lane 1-4) ran for 1 hour at 90 V. The band in lane 2 was cut for further purification.

4.6 Purification of TvAAG2

TvAAG2 was expressed in Rosetta™ BL21(DE3)pLysS and successfully purified using the same Ni-NTA purification method as described in Section 4.2. Figure 4.13 shows the purification of TvAAG2 and shows the protein was successfully produced at the right size (22 kDa) in good yield (0.42 mg) and more than 90% pure.

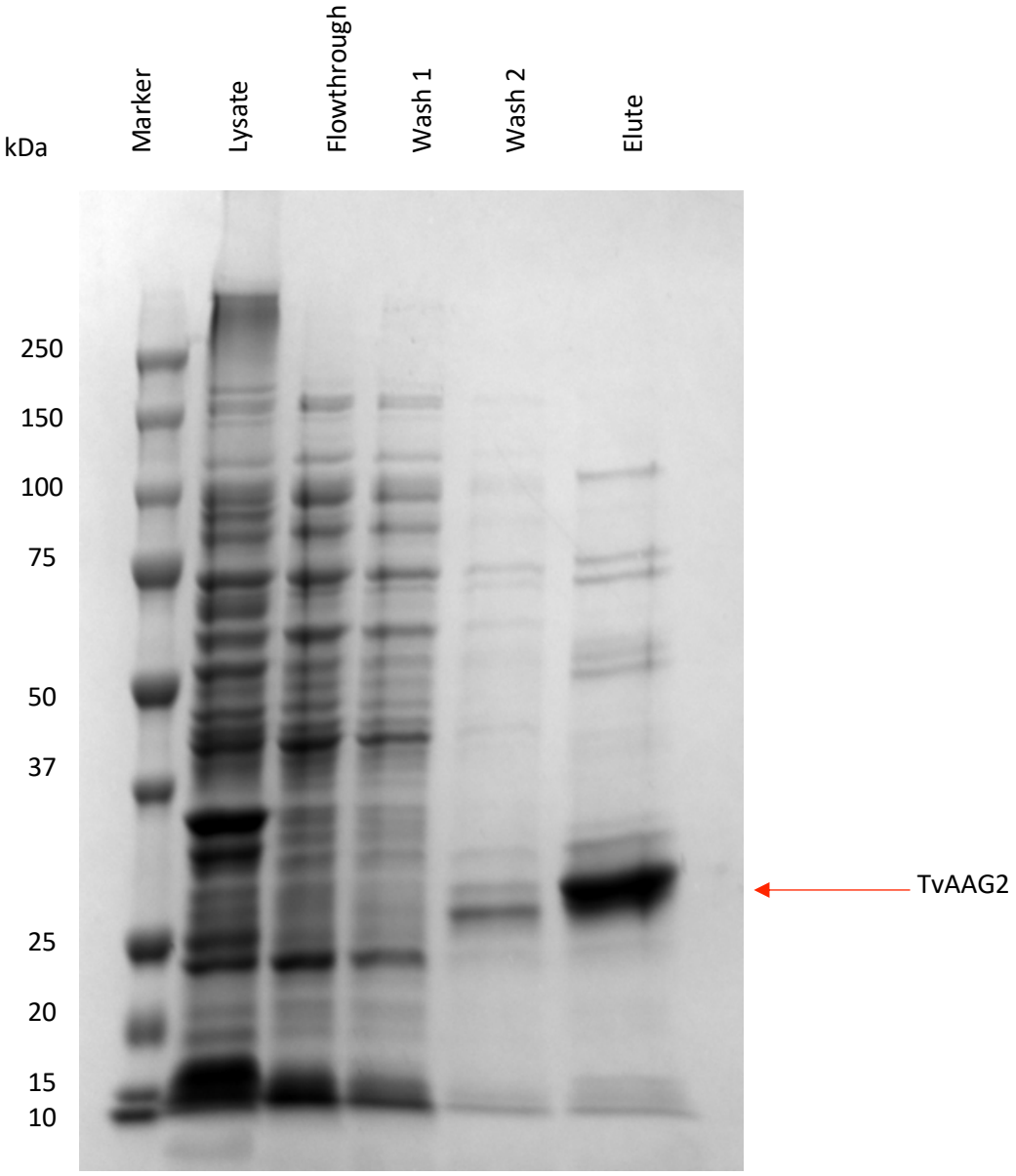


Figure 4.13 SDS-PAGE of TvAAG2 Ni-NTA Affinity Purification. TvAAG2 was expressed in Rosetta BL21(DE3)pLysS cells and purified using Ni-NTA column chromatography. The protein marker is Precision Plus Protein™ All Blue from Bio-Rad. The elute fraction was frozen at -80°C to be used in further experiments.

4.7 TvAAG2 Glycosylase Activity Assays

TvAAG2 was subjected to the same glycosylase activity assays with the same conditions described in Section 4.3 with the substrates described in Figure 4.3. The first substrates tested were the ones containing deoxyinosine; substrate 1, substrate 2 and substrate 3. TvAAG2 shows some activity with substrate 1 (Figure 4.14, A) and minimal activity with substrate 2 and 3 (Figure 4.14, B and C respectively). In all three cases TvAAG2 had far less activity compared with TvAAG1 activity on the same substrates suggesting that, although TvAAG2 does recognise dI as a lesion, dI may not be its primary substrate due to the limited activity seen with the activity assays. The second set of substrates tested with TvAAG2 were uracil and 8oxoG. TvAAG2, like TVAAG1, had no activity with substrate 4 or substrate 5 at the concentrations tested (Figure 4.15, A and B respectively).

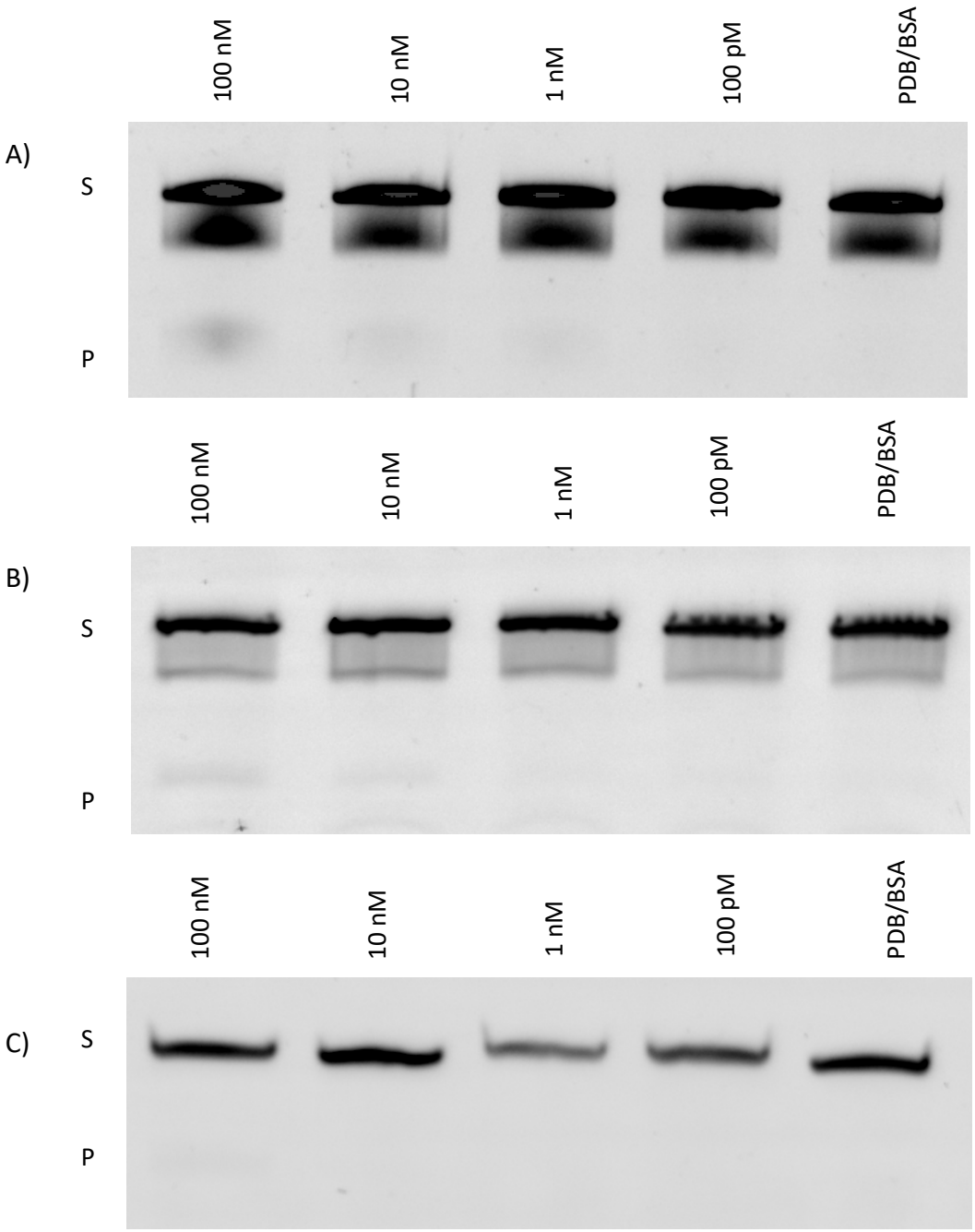


Figure 4.14 Glycosylase Activity Assays for TvAAG2 with Substrates 1-3. TvAAG2 was diluted in PDB to the appropriate concentrations and added to substrate 3 at a final concentration of 50 nM, the final concentration of TvAAG2 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S – substrate P – product. A) substrate 1. B) substrate 2. C) substrate 3. The experiments were repeated three times and a representative gel is shown.

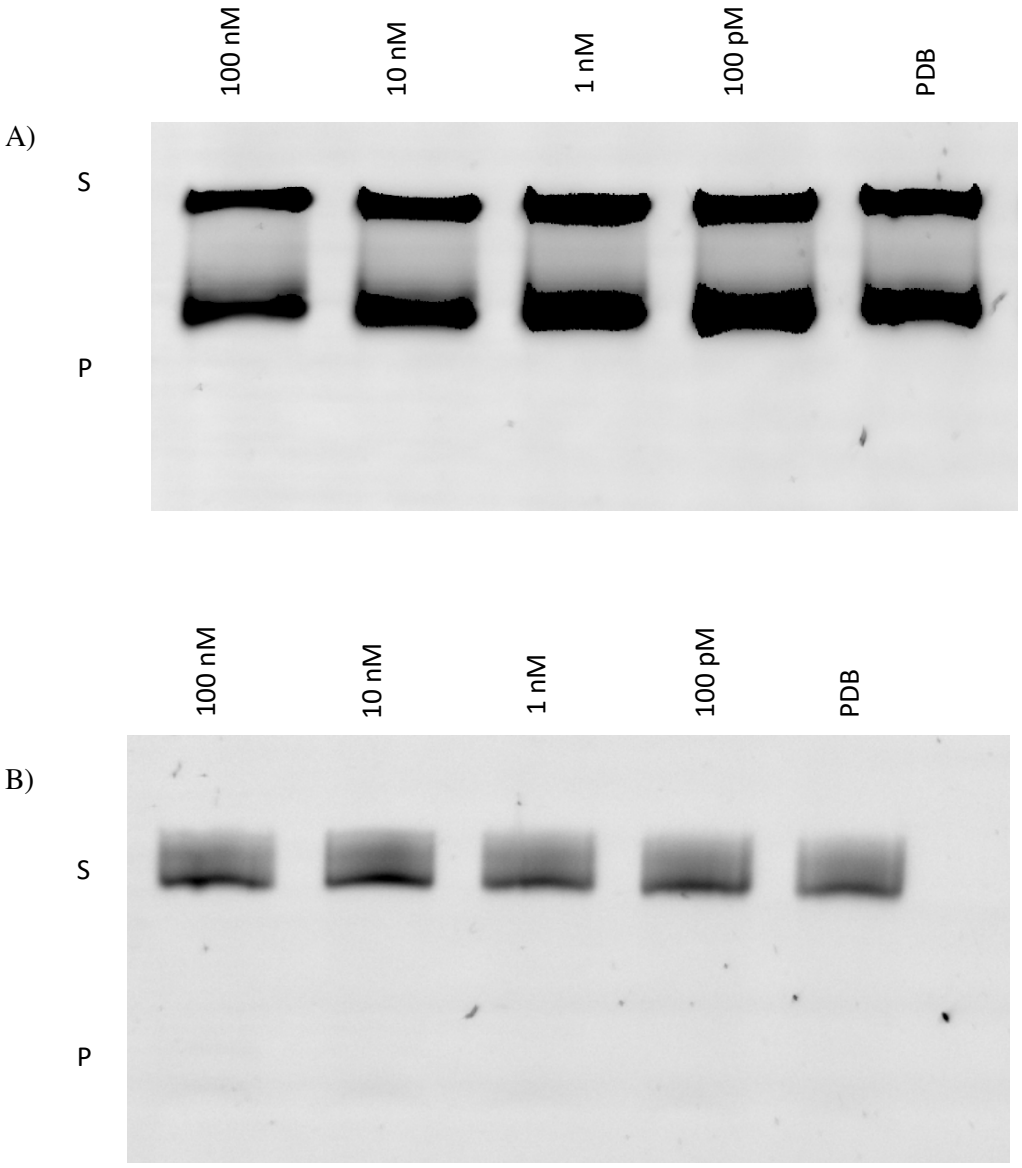


Figure 4.15 Glycosylase Activity Assays for TvAAG2 Substrate 4 and 5. TvAAG2 was diluted in PDB to the appropriate concentrations and added to substrate 5 at a final concentration of 50 nM, the final concentration of TvAAG2 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S – substrate P – product. A) substrate 4. B) substrate 5. The experiment was repeated three times and a representative gel is shown.

4.8 Schiff Base Assay with TvAAG2

TvAAG2 was used in the same Schiff base assay as described in Section 4.4. Figure 4.16 shows the Schiff base assay with NTH1 and TvAAG2. NTH1 without and with CB are the control lanes (lanes 1 and 2 respectively, Figure 4.16). As expected, there is no band in lane 1 and one band in lane 2. There is one band in the TvAAG2 without CB (lane 3, Figure 4.16) and two bands in the TvAAG2 plus CB (lane 4, Figure 4.16). These bands are possibly due to some intrinsic fluorescent component in the TvAAG2, as they are seen in the TvAAG2 sample without CB, and not bifunctional activity.

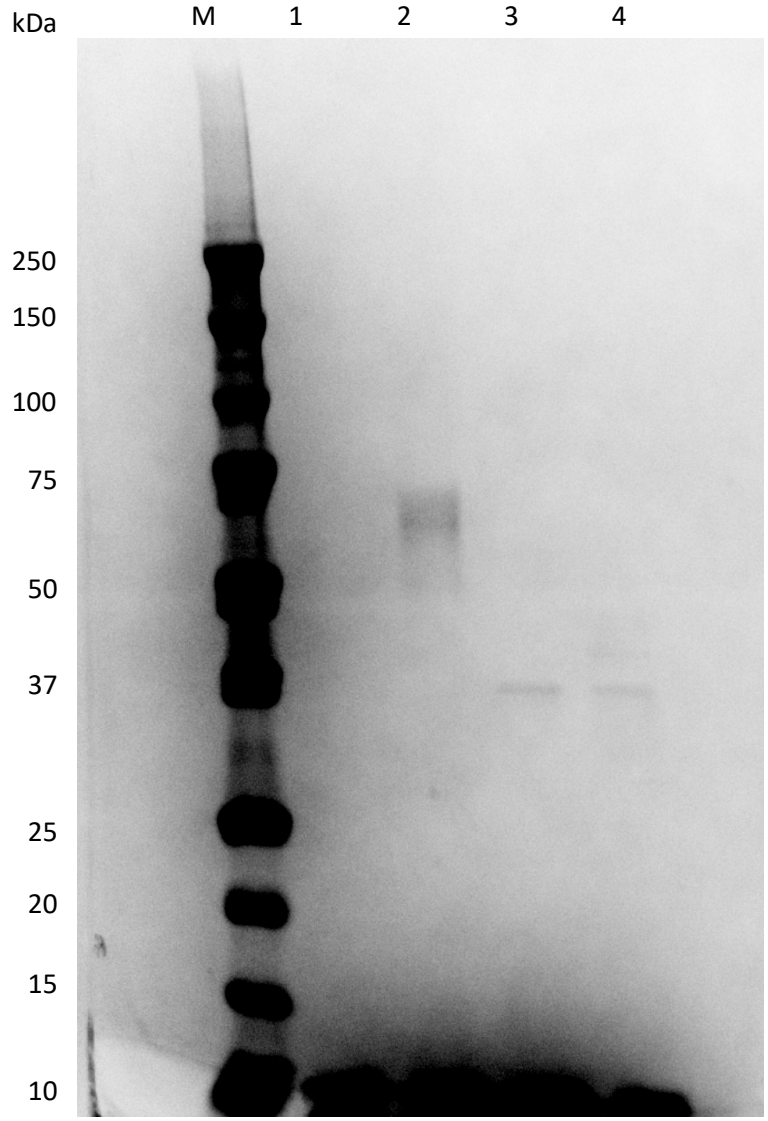


Figure 4.16 Schiff Base Assay with TvAAG2. TvAAG2 and the control enzyme NTH1 were incubated for 15 minutes at 37°C with 1 µl of 50 nM cyanoborohydride (CB) in addition to master mix. Samples were stained with SDS loading buffer preheated for 5 minutes at 95°C and loaded alongside Precision Plus Protein™ All Blue (Bio-Rad). The gel was ran for 45 minutes at 220 V with 1 X TGS buffer and visualised with Chemidoc XRS+ Fluro orange setting. 1) NTH1. 2) NTH1 + CB. 3) TvAAG2. 4) TvAAG2 + CB. 5) No protein.

4.9 Identification of Conserved Residues for Site Directed Mutagenesis Targets

To identify possible catalytic domains of TvAAG1 and TvAAG2, the NCBI accession numbers for each gene (XP_001311901.1 and XP_001315594.1 respectively) were used in NCBI protein BLAST search to find closely related species with homologues of the gene. Species close to the top of the list for TvAAG1and TvAAG2 were chosen for an alignment with CLUSTAL Omega (Table 4.1). This alignment was used to identify conserved lysine residues, as bifunctional glycosylases require an active site lysine (Jacobs and Schär, 2011). Only one uniquely conserved lysine was found and is highlighted in the alignment in Figure 4.17. This lysine was the target for site directed mutagenesis for both genes.

Species	**NCBI accession number**
Draconibacterium sediminis	WP_045025985.1
Draconibacterium orientale	SES79290.1
Parabacteroides chartae	WP_079683634.1
Tritrichomonas foetus	OHT07844.1

Table 4.1 Species Chosen for Clustal Omega Alignment. Species names and their NCBI accession numbers for the species chosen to be used in an alignment with *T. vaginalis* TvAAG1and TvAAG2 sequences

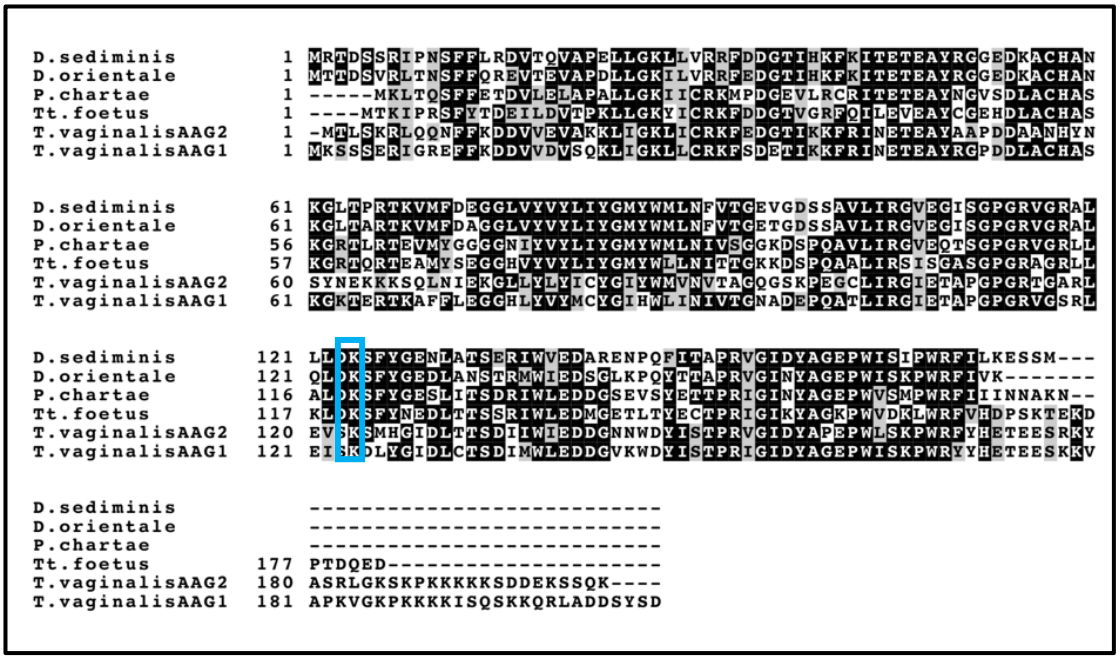


Figure 4.17 Multiple Sequence Alignment of *Trichomonas vaginalis* TvAAG1 and TvAAG2 with Closely Related Species with 3-methyladenine DNA Glycosylases. Species were found with a BLAST search of *T. vaginalis* TvAAG1 and TvAAG2. The FASTA sequences of the genes were collected from NCBI and pasted into Clustal Omega. The output from Clustal Omega was copied and used with BoxShade to colour conserved (black) and similar (grey) residues. This alignment was used to identify conserved lysines to be targeted in site directed mutagenesis. The only lysine conserved between all five sequences is highlighted in blue.

4.10 Site Directed Mutagenesis and Purification of TvAAG1 and TvAAG2 Mutants

Primers for site directed mutagenesis (SDM) were designed using the Quikchange Primer Design Program at Agilent genomics website. Primers were designed for two-point mutations at lysine 124 of TvAAG1: lysine to alanine (known as K124A mutant), which should abolish activity and lysine to arginine (K124R mutant) which should partially reduce the activity (Liu and Roy, 2001). Primers were designed for the same point mutations at lysine 123 of TvAAG2: lysine to alanine (K123A mutant) and lysine to arginine (K123R mutant). SDM was carried out using Phusion Flash DNA polymerase within a thermal cycler (Figure 4.18). After incubation with DpnI to digest the wild-type template plasmid, the SDM mixture was used to transform Turbo *E. coli* competent cells. Plasmids were isolated from individual clones and sent for sequencing as previously described.

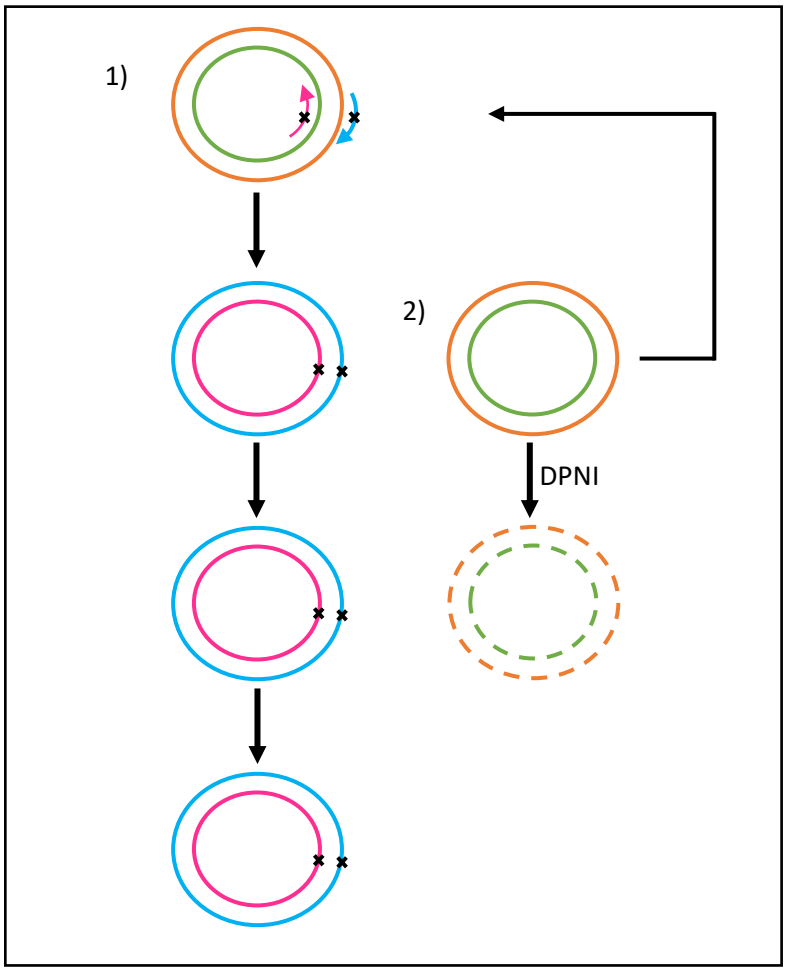


Figure 4.18: Quik Change Site Directed Mutagenesis. 1) Thermal cycling denatures the DNA template, anneals the mutagenic primers and primers are extended with Phusion Flash DNA polymerase. 2) Parental DNA is digested with DpnI. After digestion the SDM mixture is transformed into *E. coli* competent cells.

After successful identification of mutated plasmids, Rosetta™ 2 DE3 pLysS cells were transformed with these plasmids and the mutant proteins purified as previously described (Section 4.2) Samples for SDS PAGE gels were taken during purification for each mutant protein; the supernatant ''Lysate'', the unbound proteins washed through the column ''Flowthrough'', the first elution ''Wash 1'', the second elution ''Wash 2'' and the purified protein ''Elute''. First attempts at purifying these

mutants were unsuccessful, with very low yields. Eventually, successful purifications of each mutant were carried out. Figure 4.19 – 4.22 show the SDS PAGE gels of the mutant protein purifications. The overall yield for the mutant proteins were; 0.11 mg (K124A, Figure 4.19), 0.4 mg (K124R Figure 4.20), 0.16 mg (K123A Figure 4.21), 0.26 mg (K123R, Figure 4.22).

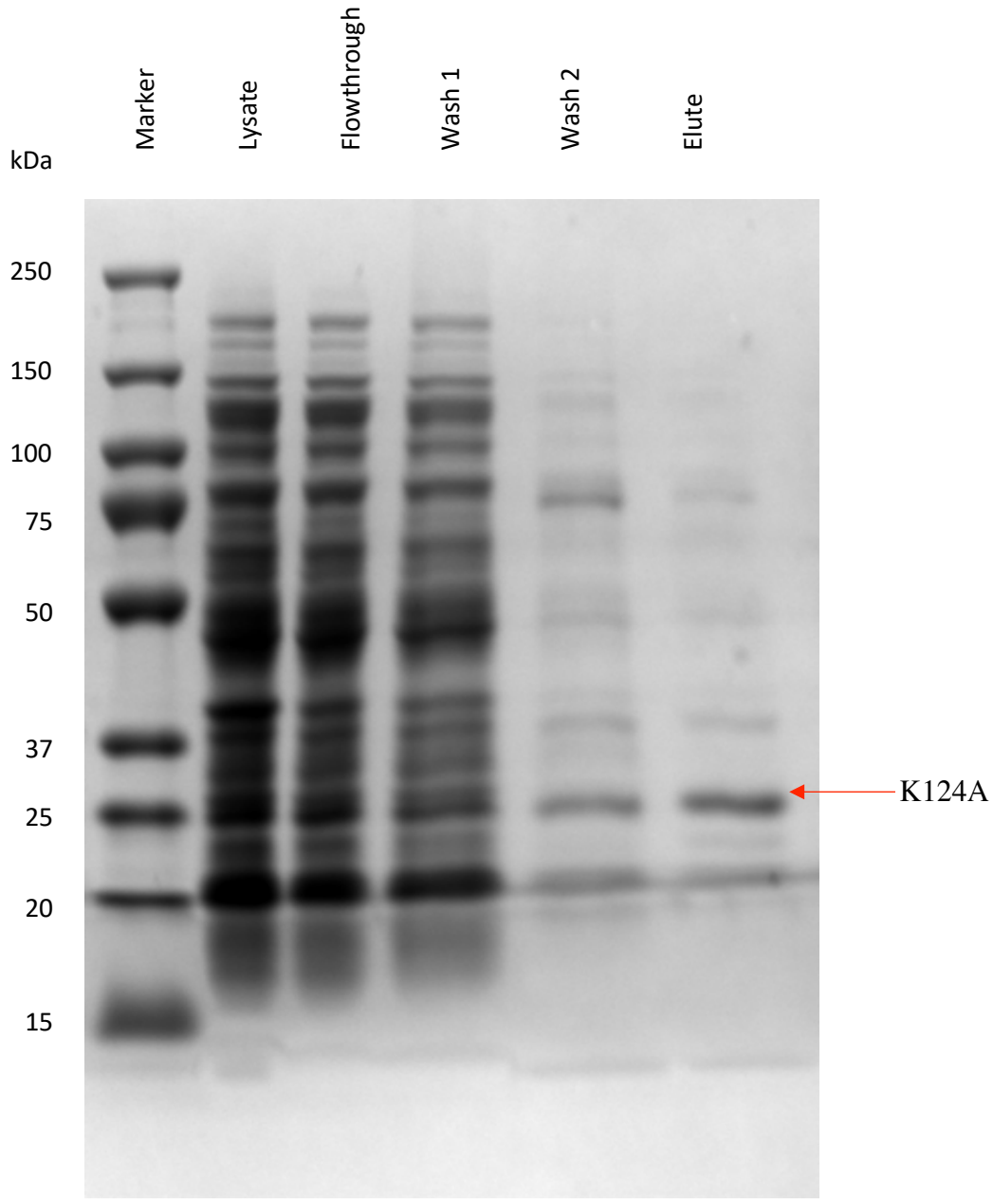


Figure 4.19: SDS-PAGE Gel of Purification Fractions Collected from Ni-NTA Affinity Column from TvAAG1 K124A Mutant Proteins. TvAAG1 K124A was expressed in Rosetta 2 BL21(DE3)pLysS cells and purified using Ni-NTA column chromatography. The protein marker is Precision Plus Protein™ All Blue from Bio-Rad. See accompanying text for details of samples.

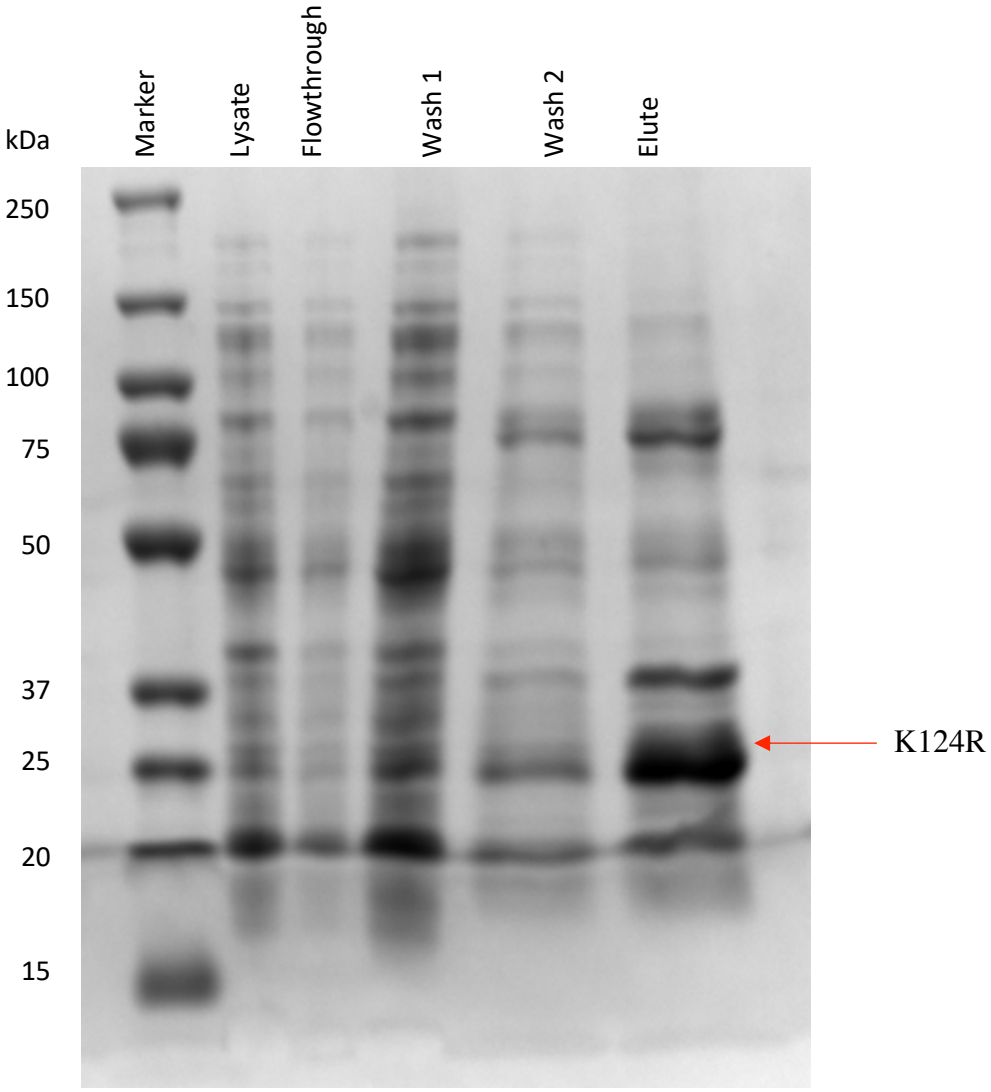


Figure 4.20: SDS-PAGE Gel of Purification Fractions Collected from Ni-NTA Affinity Column from TvAAG1 K124R Mutant Proteins. TvAAG1 K124R was expressed in Rosetta 2 BL21(DE3)pLysS cells and purified using Ni-NTA column chromatography. The protein marker is Precision Plus Protein™ All Blue from Bio-Rad. See accompanying text for details of samples.

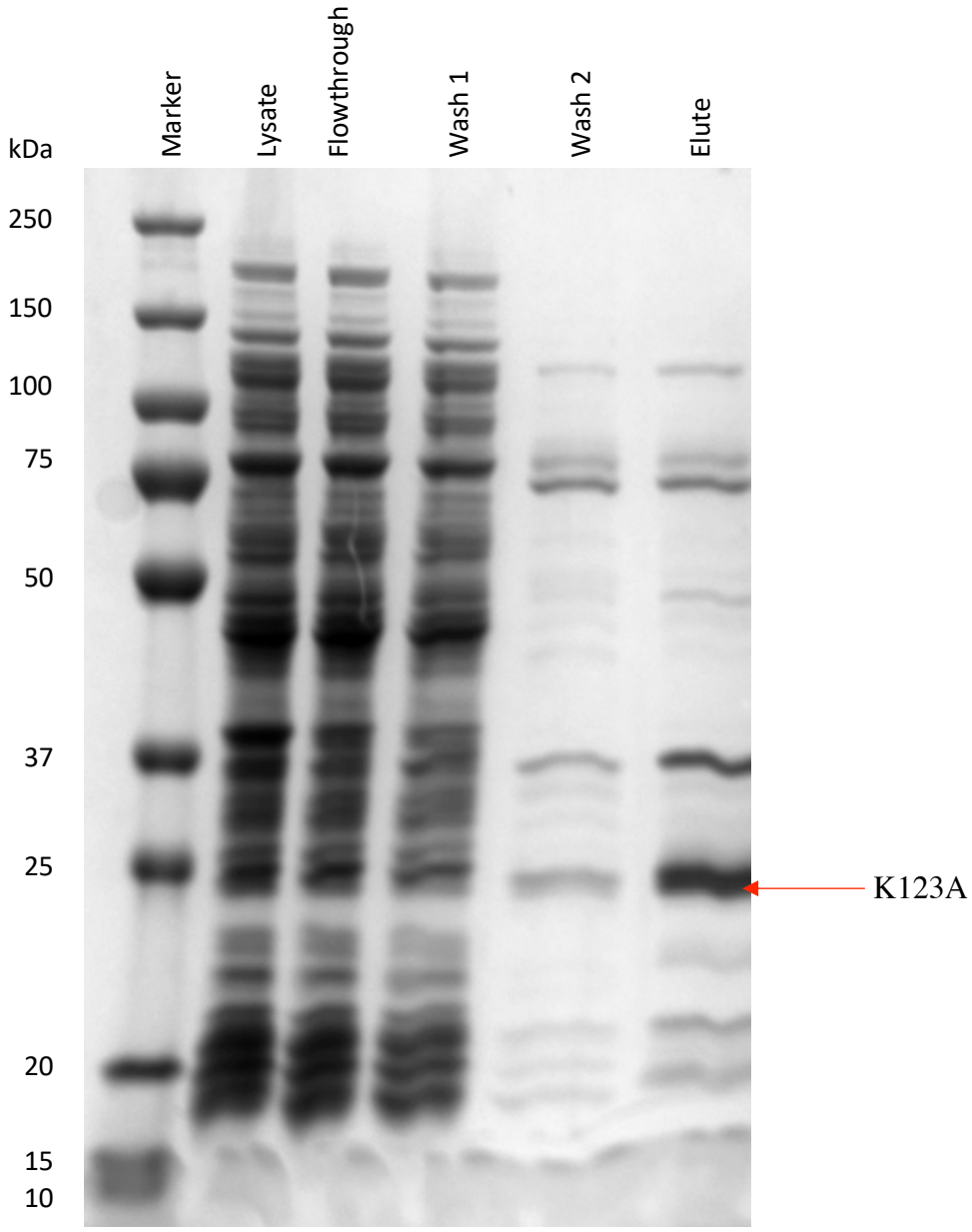


Figure 4.21: SDS-PAGE Gel of Purification Fractions Collected from Ni-NTA Affinity Column from TvAAG2 K123A Mutant Proteins. TvAAG2 K123A was expressed in Rosetta 2 BL21(DE3)pLysS cells and purified using Ni-NTA column chromatography. The protein marker is Precision Plus Protein™ All Blue from Bio-Rad. See accompanying text for details of samples.

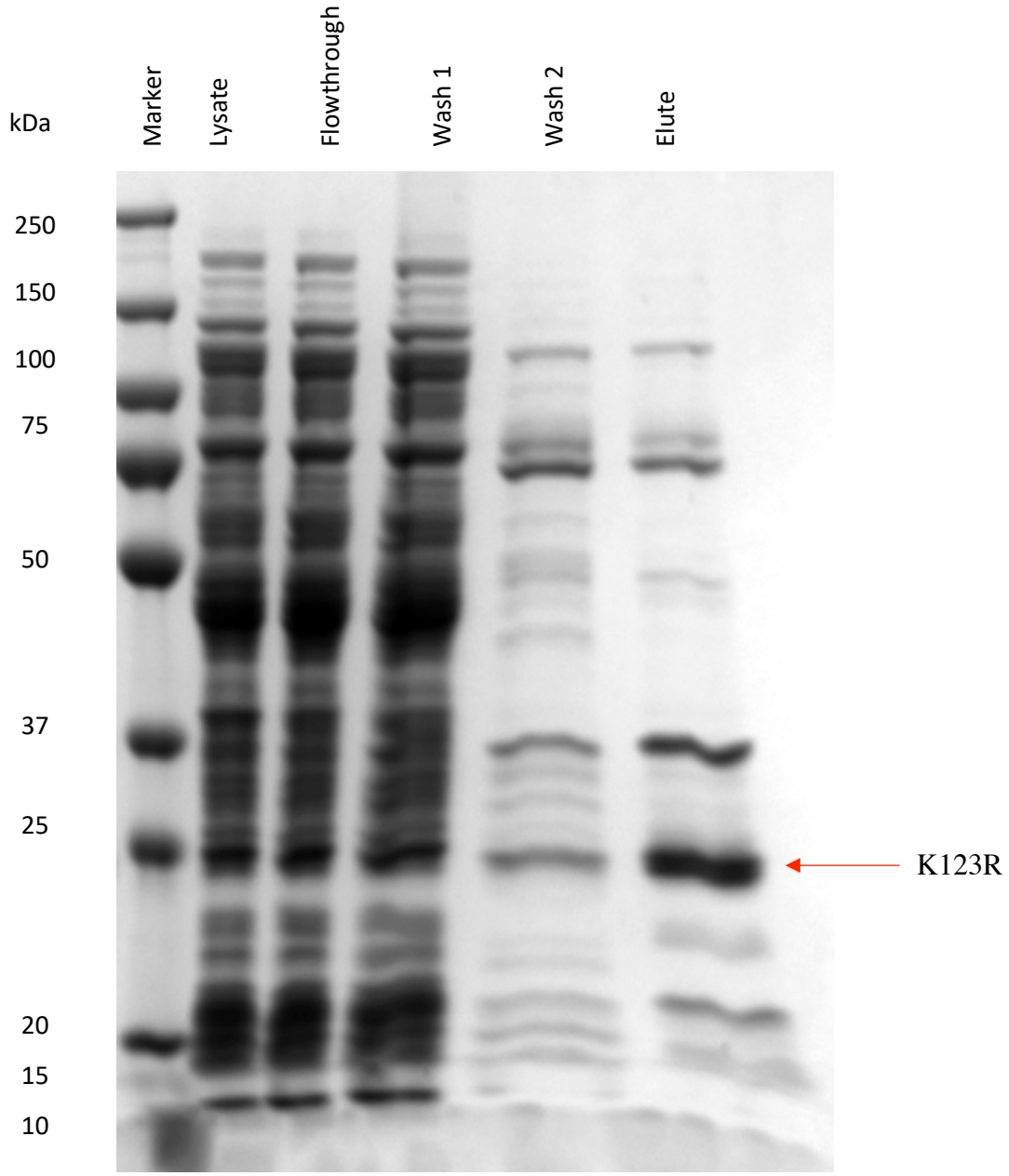


Figure 4.22: SDS-PAGE Gel of Purification Fractions Collected from Ni-NTA Affinity Column from TvAAG2 K123R Mutant Proteins. TvAAG2 K123R was expressed in Rosetta 2 BL21(DE3)pLysS cells and purified using Ni-NTA column chromatography. The protein marker is Precision Plus Protein™ All Blue from Bio-Rad. See accompanying text for details of samples.

4.11 TvAAG1 and TvAAG2 Mutant Glycosylase Activity Assays

Glycosylase activity assays were carried out with TvAAG1 K124A, K124R and TvAAG2 K124A and K24R mutants as described in Section 4.3 with the substrates described in Figure 4.3. Wildtype TvAAG1 or TvAAG2 at 100 nM and 1 pM were used as control lanes.

The K124A mutation abolished TvAAG1 activity as seen in Figure 4.23 the K124A mutant had no activity with substrate 1 or substrate 2 at the concentrations tested (A and B respectively). The K124R mutation reduced activity with substrate 1, at 100 nM K124R not all of the substrate was converted to product compared to the100 nM wild type control lane in which all the substrate was converted to product (Figure 4.24, A). K124R also reduced activity with substrate 2 (Figure 4.24, B) which again can be seen in the reduction of activity in the 100 nM K124R lane compared to 100 nM TvAAG1 control. The next experiment I carried out was to assess the extent of the reduction in activity by K124R by using a wider range of enzyme concentrations with substrate 1. The wildtype TvAAG1 control shows clear activity on substrate 1 from 100 nM to 6.25 nM, whereas activity at 50 nM K124R is far less. Activity on substrate 1 then decreases to minimal levels at concentrations 25 nM to 6.25 nM K124R (Figure 4.25).

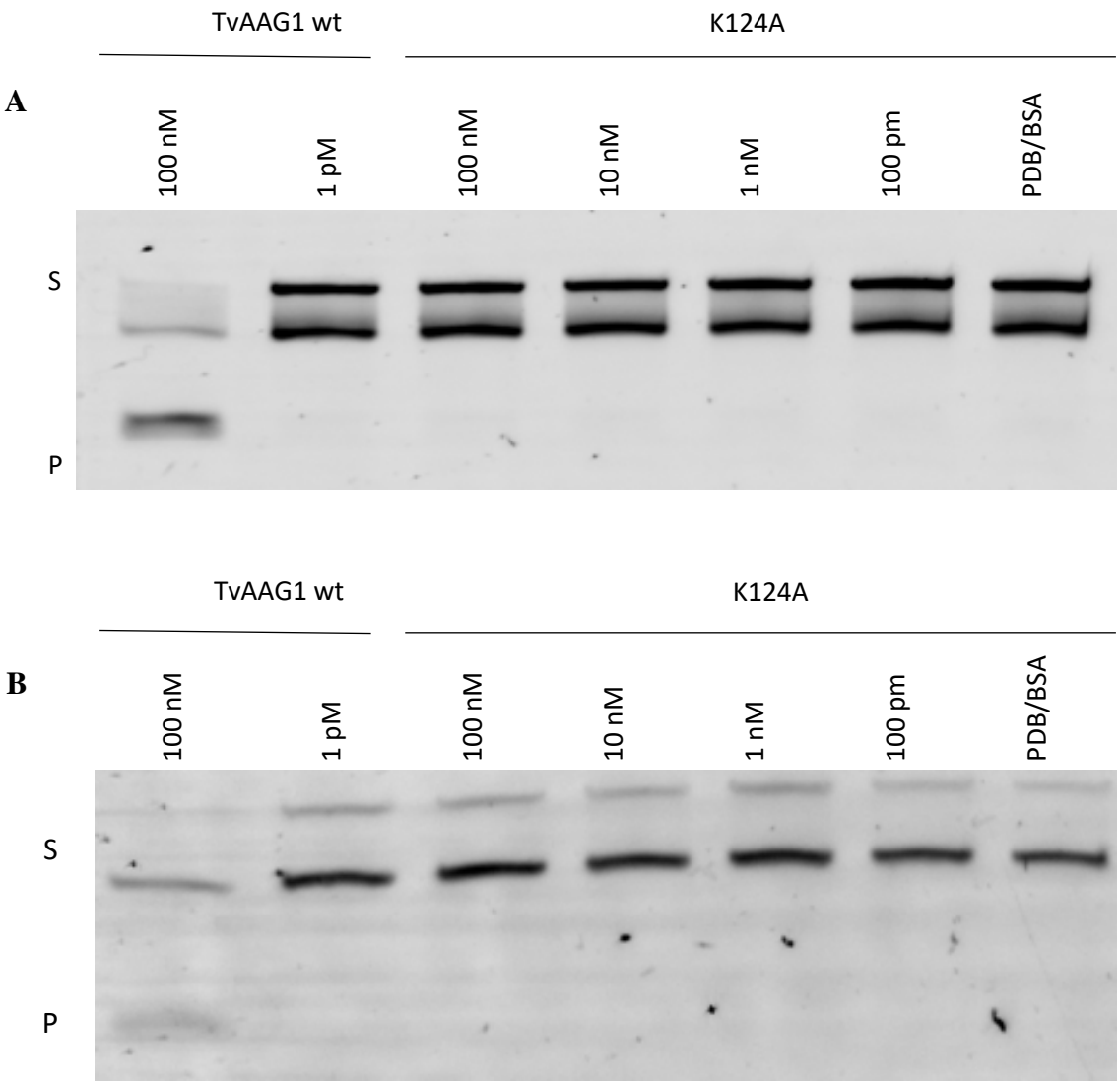


Figure 4.23: Glycosylase Activity Assay TvAAG1 K124A Substrate 1 and 2. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 2 at a final concentration of 50 nM, the final concentration of tvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. A) Substrate 1. B) Substrate 2. The experiment was repeated three times and a representative gel is shown.

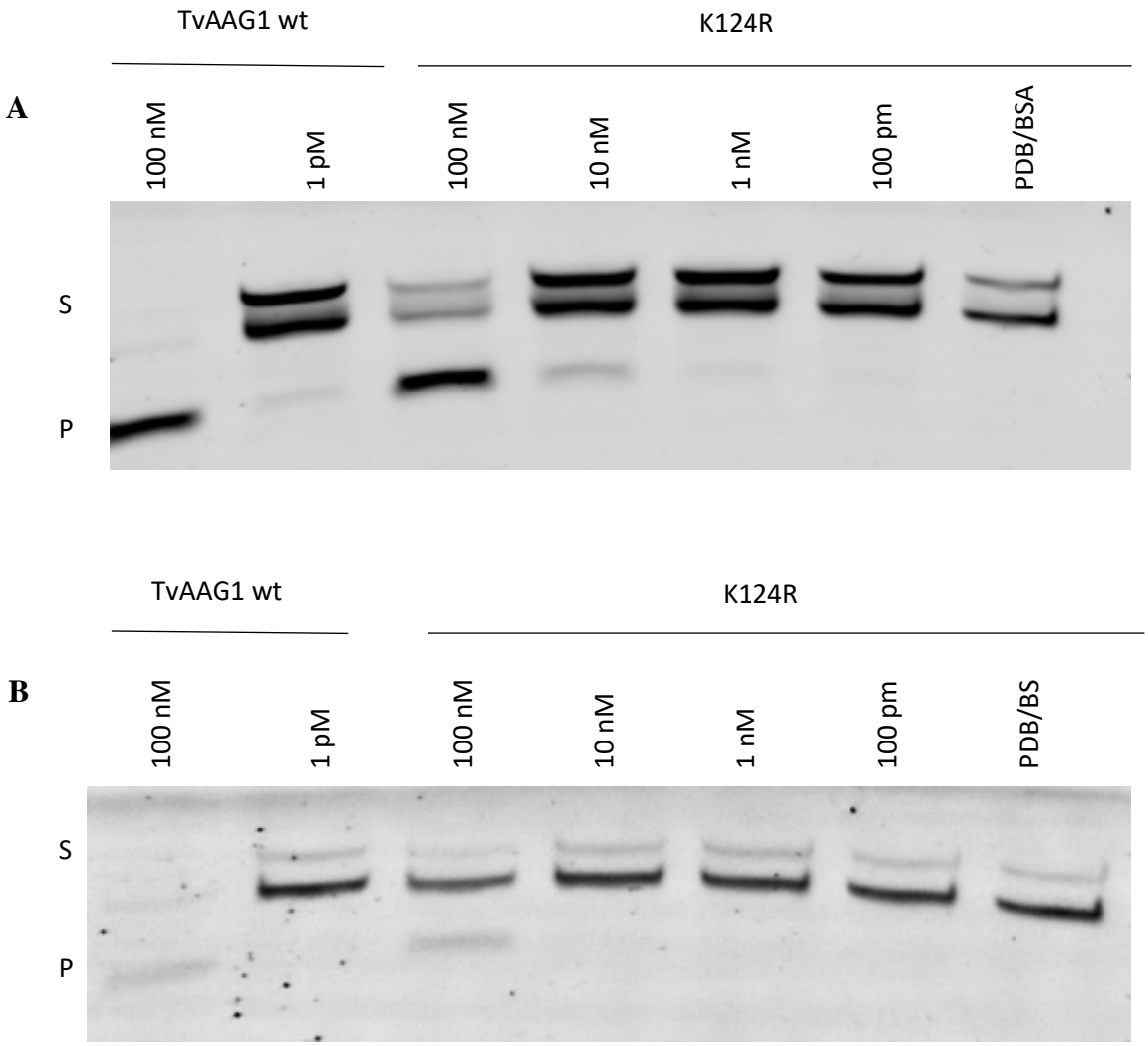


Figure 4.24: Glycosylase Activity Assays for TvAAG1 K124R Substrate 1 and 2. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 2 at a final concentration of 50 nM, the final concentration of tvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. A) Substrate 1. B) Substrate 2. The experiment was repeated three times and a representative gel is shown.

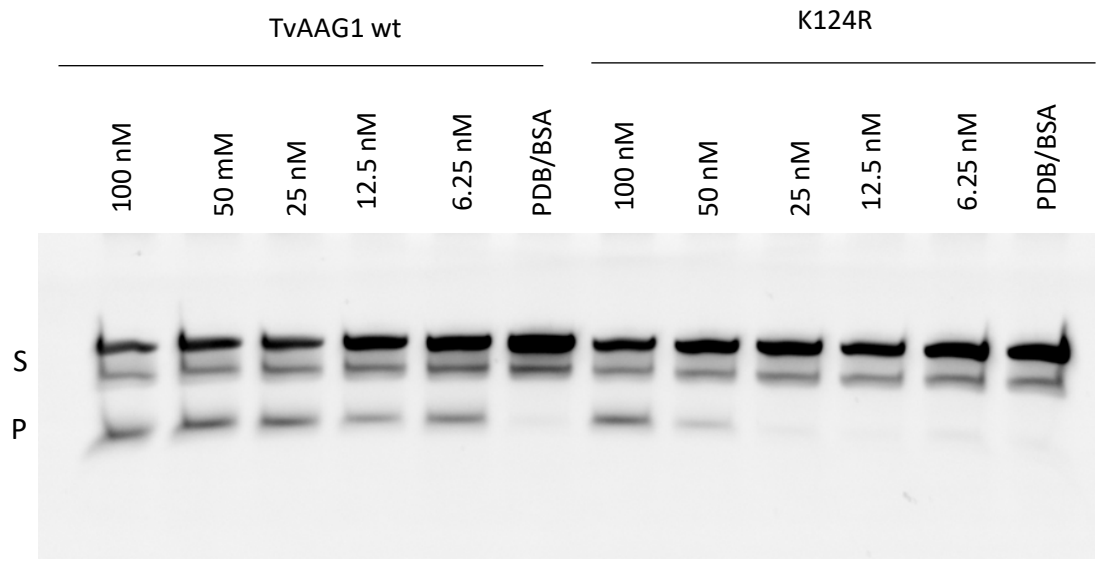


Figure 4.25: Glycosylase Activity Assay TvAAG1 and TvAAG1 K124R with dIdT Substrate at Concentrations between 100 nM and 10 nM. TvAAG1 was diluted in PDB to the appropriate concentrations and added to substrate 1 at a final concentration of 50 nM, the final concentration of TvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated two times and a representative gel is shown.

Glycosylase activity assays for TvAAG2 mutants were carried out using substrate 1 and 2, despite TvAAG2 only showing minimal activity with these substrates, as a suitable substrate which was not identified during Section 4.7. K123A abolished the minimal activity seen at 100 nM TvAAG2 with substrate 1 (Figure 4.26, A) as expected. K123R slightly reduced the minimal activity seen at 100 nM TvAAG2 in Figure 4.27, A. The control lanes with TvAAG2 and substrate 2 (Figure 4.26 B, Figure 4.27 B) showed no activity therefore the effects of K123A and K123R activity on substrate 2 could not be determined.

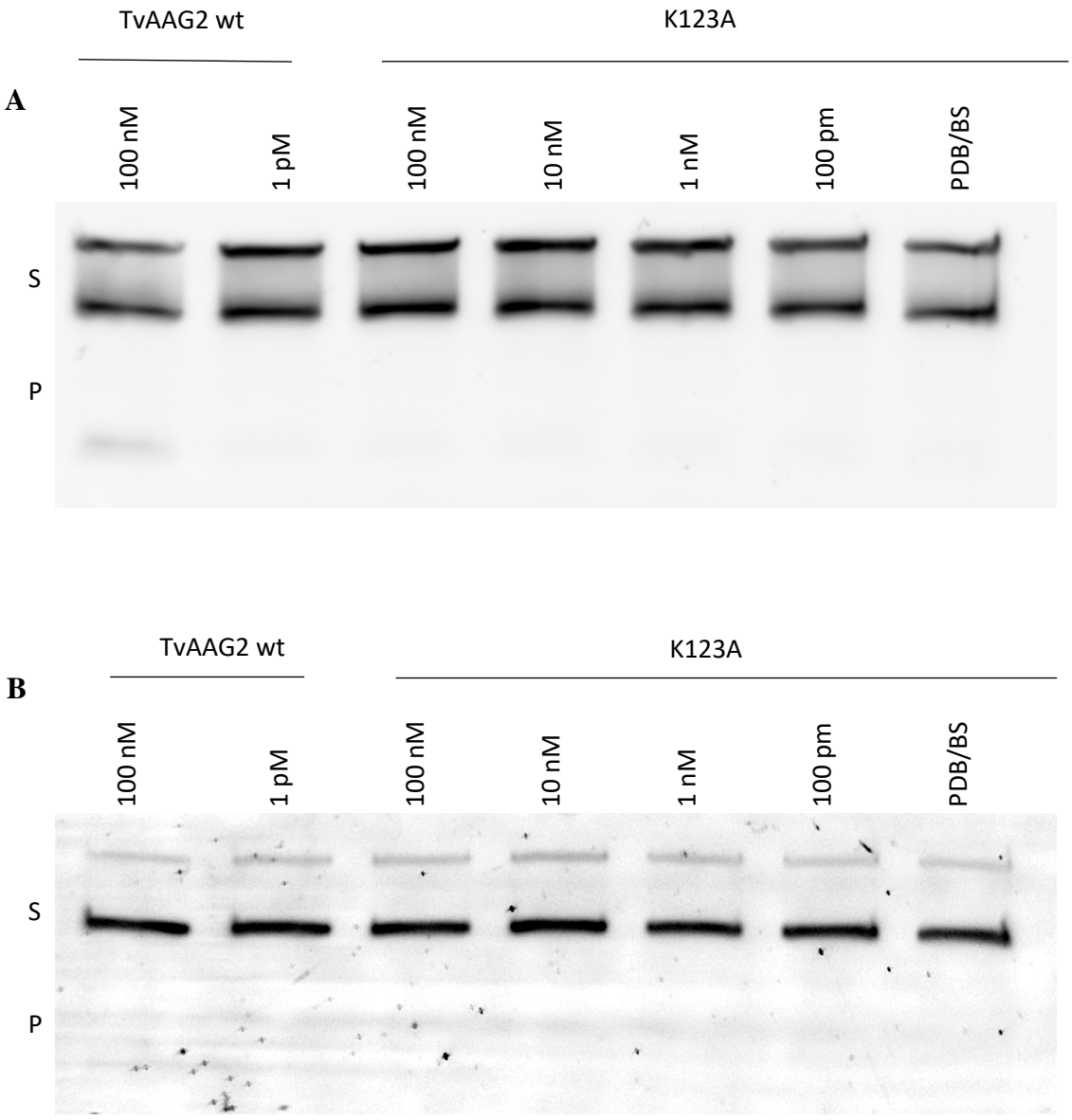


Figure 4.26: Glycosylase Activity Assays for TvAAG2 K123A with Substrate 1 and 2. TvAAG2 was diluted in PDB to the appropriate concentrations and added to substrate 2 at a final concentration of 50 nM, the final concentration of TvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. A) Substrate 1. B) Substrate 2. The experiment was repeated two times and a representative gel is shown

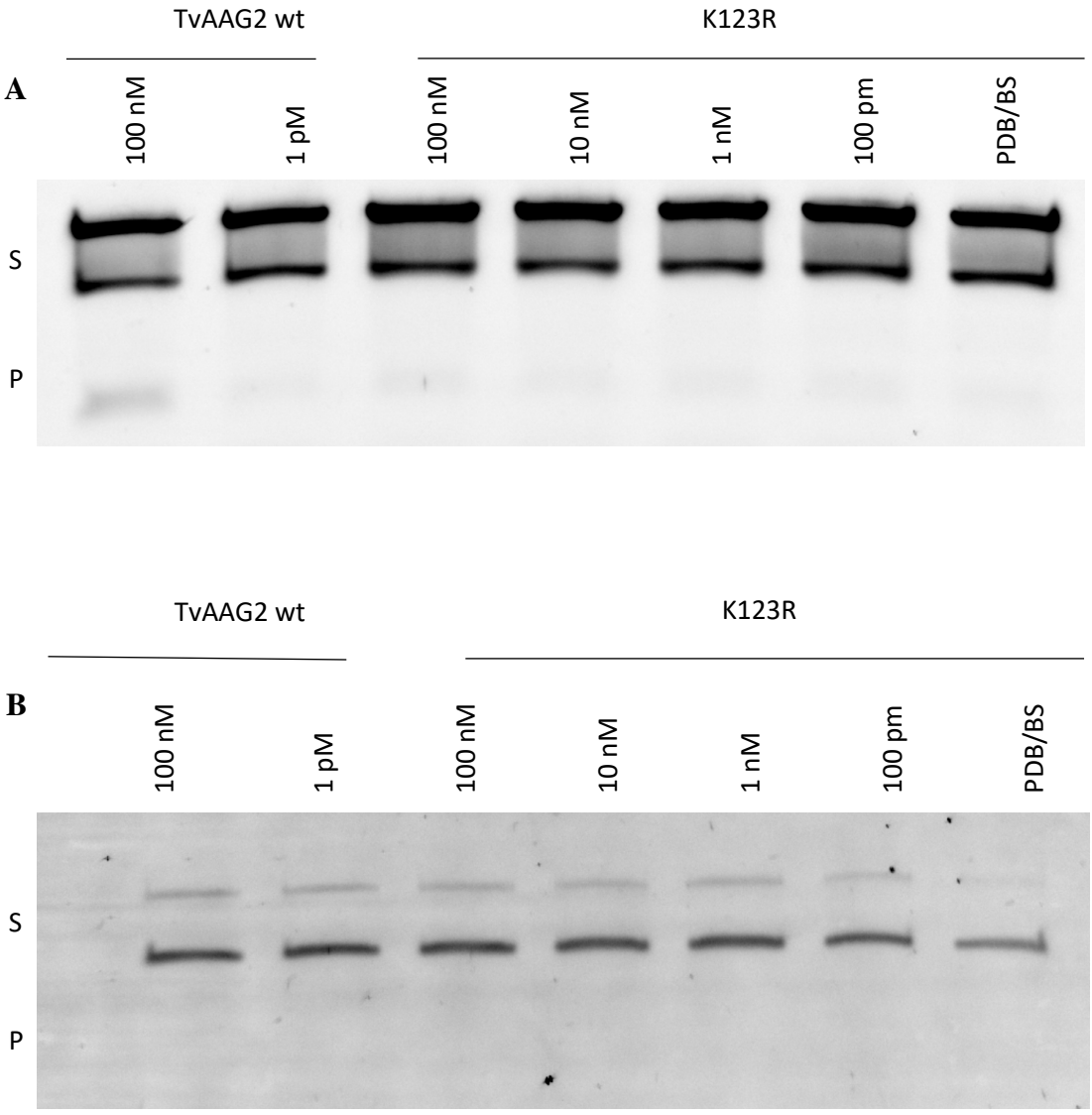


Figure 4.27: Glycosylase Activity Assays for TvAAG2 K123R Substrate 2. TvAAG2 was diluted in PDB to the appropriate concentrations and added to substrate 2 at a final concentration of 50 nM, the final concentration of tvAAG1 in each reaction is indicated above the relevant lanes. The reaction was incubated for 15 minutes at 37°C. FA/NaOH was added to stop the reaction and samples were heated for 5 minutes at 95°C before loading onto an 8 M Urea PAGE gel and run for 1 hour at 200 V. S - substrate P – product. The experiment was repeated three times and a representative gel is shown.

4.12 Discussion

In my project I have successfully demonstrated that TvAAG1 is a bifunctional glycosylase by the use of Schiff base assays. TvAAG1 is the first bifunctional AAG type enzyme to be identified, as there are currently non-described within the literature. The human AAG enzyme MPG is monofunctional as evidenced by the activated water molecule in its active site (Lau et al., 1998). *E. coli* has two AAG type enzymes TAG and AlkA which are both monofunctional (Thomas, Yang and Goldthwait, 1982; Nakabeppu et al 1984). As are the AAG type enzymes in; *Bacillus subtilis* (Aamodt et al., 2004) *Saccharomyces cerevisiae* (Chen, Derfler and Samson, 1990) *Helicobacter pylori* (O'Rourke et al., 2000) and *Thermotoga maritima* (Begley et al., 1999).

I have also characterised TvAAG1 substrate specificity with the use of glycosylase activity assays. TvAAG1 has activity with deoxyinosine in a double stranded oligomer across from deoxythymidine and across from deoxycytidine. The glycosylase has significantly more activity with dI:dT than dI:dC, this preference is consistent with other studies within the literature (Wyatt and Samson, 2000; Connor and Wyatt, 2002). Deoxyinosine is structurally similar to deoxyguanine and as such forms a Watson-Crick base pair with deoxycytidine making it harder for TvAAG1 to recognise dI within the DNA backbone (Figure 4.28). O'Brien and Ellenberger (2003) show that a normal G:C Watson-Crick base pair is 50-fold more resistant to human AAG compared to a G:T mismatch. They suggest a way of ensuring unmodified bases are not excised by AAG is to discriminate against Watson-Crick base pairs making them less likely to be flipped into the active site. Even though TvAAG1 can recognise dI as a substrate, its structure when in a base pair with dC reduces TvAAG1 ability to flip it into the active site for catalysis and therefore reduces rate of excision. TvAAG1 has minimal activity with

single stranded deoxyinosine substrate and, unlike human AAG, does not have activity with uracil containing substrates.

In the bioinformatics section of this project (Chapter 3) I discovered that *T. vaginalis* lacks an 8oxoG specific glycosylase, some AAGs have been shown to act on 8oxoG lesions lesions (Bessho et al., 1993 and Wyatt and Samson 2000). However, no activity was seen with the 8oxoG substrate. It can be assumed that TvAAG1 and TvAAG2 do not compensate for the lack of OGG1 or similar enzyme.

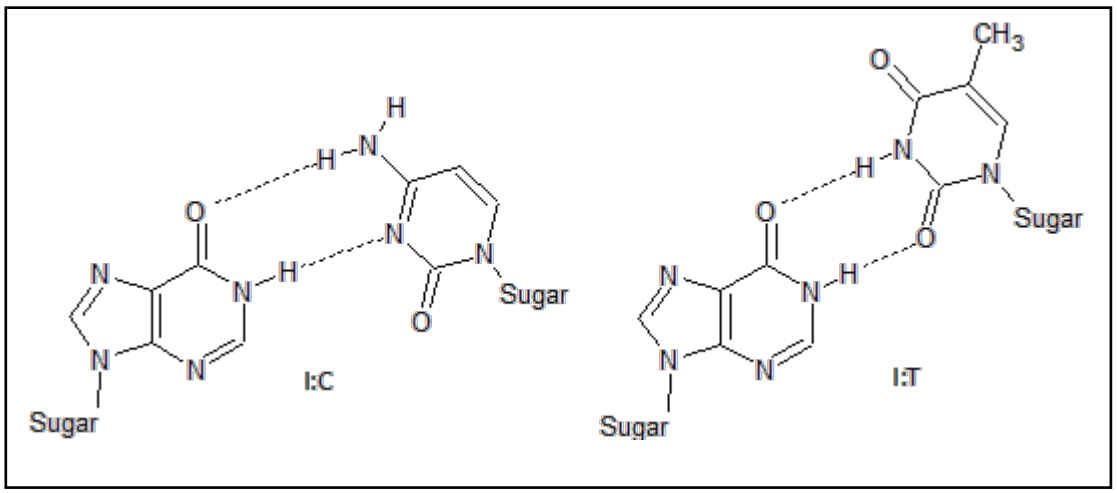


Figure 4.28: Deoxyinosine Paired with Deoxycytidine and Deoxythymidine. Deoxyinosine (I) paired to deoxycytidine (C) forms a Watson-Crick base pair whilst deoxyinosine paired with deoxythymidine (T) forms a less stable pair.

In regard to TvAAG2 I have not been able to identify a substrate that the enzyme is significantly active with. TvAAG2 has very small, minimal activity on both dI:dT and dI:dC and the single stranded dI substrate. TvAAG2, like TvAAG1, has no activity with 8oxoG or uracil substrates. Other AAG type enzymes are known to be active with highly mutagenic etheno adducts produced from lipid peroxidation and environmental pollutants. Human AAG, *S. cerevisiae* MAG and *E. coli* AlkA enzymes can recognise and cleave the etheno adduct 1,N6-ethenoadenine (Saparbaev, Kleibl and Laval, 1995;

Hang et al., 1997; Lingaraju et al., 2008). TvAAG2 may be specific for this type of damage within *T. vaginalis*.

The site-directed mutagenesis I carried out successfully identified the critical lysine residue present in both TvAAG1 and TvAAG2 active sites, strongly supporting our hypothesis that the enzyme is bifunctional, in contrast to previously characterised AAGs. The lysine – arginine mutations reduced TvAAG1 activity whilst lysine – alanine mutations completely abolished activity. TvAAG2 mutations did not show much of a reduction of activity as the optimal substrate for TvAAG2 was not identified for use in the mutant assays.

4.13 Conclusion

The work detailed in this chapter identifies that dI:dT is the optimal substrate for an alkyladenine glycosylase homologue from *Trichomonas vaginalis*, and also provides strong evidence that it has a bifunctional reaction mechanism. To our knowledge, this is the first such bifunctional AAG homologue to be characterised. I have also expressed and carried out the first preliminary characterisation of the activity of a second AAG homologue present in *T. vaginalis*. Further studies should be carried out to characterise TvAAG2 substrate specificity, in particular studies using εA substrates. As a substrate was not found it cannot be said if TvAAG2 is truly a TvAAG1 homologue, but the minimal activity with deoxyinosine substrates suggests that it has glycosylase activity.

5
Conclusion

DNA repair pathways have evolved to protect and ensure the faithful replication of DNA against damage arising from endogenous and exogenous sources. The significance of DNA repair is evidenced by the fact that it is a mechanism that is conserved, in some form, throughout all life. *Trichomonas vaginalis* is the causative agent of the most prevalent, non-viral, sexually transmitted infection worldwide (WHO, 2012). The published draft genome of *T. vaginalis* has enabled further research to be undertaken on this once elusive parasite (Carlton et al., 2007). As a member of the excavata (Adl et al., 2012), an early diverging group of eukaryotes, insights from *T. vaginalis* DNA repair pathways may begin to explain the evolution of DNA repair in other eukaryotes.

The two enzymes described in this research are DNA 3-methyladenine homologues. Bioinformatical analysis of a wide range of eukaryotes reveals almost no other eukaryotes, except from *T. vaginalis, T. foetus* and some animal species, to possess an AAG homologue. This raises the question of how and why these two parasites have a base excision repair glycosylase when it is not conserved in any of their evolutionary relatives, *Leishmania, Trypanosoma* or indeed the majority of other eukaryotes. Horizontal gene transfer is documented in other parasites as a source of novel genes which can allow the evolution of pathogenic traits (Alsmark et al., 2013;

Trasviña-Arenas et al., 2019). As discussed in 3.4 the likely source of TvAAG1 and TvAAG2 is through horizontal gene transfer between an ancient ancestor of both *T. vaginalis* and *T. foetus* and a prokaryote related to the bacteroidetes lineage. The AAG type homologue in *Arabidopsis thaliana* is also product of horizontal gene transfer (Fang et al., 2017). It is unknown if the parasites utilize the AAG glycosylases for survival or if, as knockout studies in mice show the enzyme does not contribute to lethality (Engelward et al., 1997), the genes are redundant. Some eukaryotes have a wide range of DNA glycosylases with overlapping substrate specificities which may explain how other eukaryotes can repair deamination and oxidation damage without AAG.

Another major question identified by bioinformatic analysis of the BER is the repair of the highly mutagenic lesion 8oxoG. Both parasites lack an 8oxoG specific enzyme (i.e OGG1) and neither TvAAG1 nor TvAAG2 compensates for the repair of this lesion. The lack of an 8oxoG specific enzyme is an unusual finding as 8oxoG is a frequently occurring lesion, the parasites may have an alternate mechanism to repair this type of DNA damage. This finding is consistent with research on the parasitic protist *E. histolytica* who also lacks OGG1 (López-Camarillo et al., 2009).

Both *T. vaginalis* and *T. foetus* lack nearly all of the genes used in the Non-homologous end joining pathway. The only gene present, MRE11, is also used in homologous recombination. This suggests double strand breaks can only be repaired via homologous recombination during S and G2 phases of the cell cycle. However, as the cell cycle in Trichomonads is not well studied, homologous recombination may not be limited as it is in mammalian cells. The lack of genes involved in NHEJ is also

consistent with findings in other excavates including *Plasmodium falciparum* (Gardner et al., 2002), *T. brucei, T. cruzi, L. infantum* and *L. major* (Genois et al., 2014).

In this project I have successfully demonstrated TvAAG1 as the first bifunctional AAG glycosylase, as non are currently described within the literature. Additionally, I have defined its optimal substrate, deoxyinosine across from deoxythymine, and carried out the first characterisation of TvAAG2 throughglycosylase activity assays. I have also created a comprehensive atlas of DNA repair proteins and enzymes for both *T. vaginalis* and *Tritrichomonas foetus* by bioinformatical means. The first thorough documentation of these pathways with regard to these particular parasites.

This research establishes the foundation for new research on both *T. vaginalis* and *T. foetus* and their DNA repair pathways. The first report of CRISP/Cas9 gene knockout studies in *T. vaginalis* cells was published last year (Janssen et al., 2018). In light of this, future research could use CRISPR/Cas9 knockouts of TvAAG1 and TvAAG2 to establish any changes in phenotype caused by the loss of one, or both genes. Any lethality or hypersensitivity to DNA damaging agents seen in *T. vaginalis* cells after gene knockouts would allow further characterisation of both TvAAG1 and TvAAG2 and the establishment of whether the genes are an evolutionary advantage or indeed, redundant. In light of increased reports of antibiotic resistant *T. vaginalis* strains (Schwebke and Barrientes, 2006; Upcroft et al., 2009) research into the origins of its DNA repair proteins may be advantageous as homologues of bacterial origin, that differ enough from the corresponding human homologues, could be used as novel drug targets or for vaccine design.

Bibliography

Aamodt, R., Falnes, P., Johansen, R., Seeberg, E. and Bjørås, M. (2004). The *Bacillus subtilis* Counterpart of the Mammalian 3-Methyladenine DNA Glycosylase Has Hypoxanthine and 1,N6-Ethenoadenine as Preferred Substrates. *Journal of Biological Chemistry*, 279(14), pp.13601-13606.

Adl, S., Simpson, A., Lane, C., Lukeš, J., Bass, D., Bowser, S., Brown, M., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., le Gall, L., Lynn, D., McManus, H., Mitchell, E., Mozley-Stanridge, S., Parfrey, L., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C., Smirnov, A. and Spiegel, F. (2012). The Revised Classification of Eukaryotes. *Journal of Eukaryotic Microbiology*, 59(5), pp.429-514.

Al-Tassan, N., Chmiel, N., Maynard, J., Fleming, N., Livingston, A., Williams, G., Hodges, A., Davies, D., David, S., Sampson, J. and Cheadle, J. (2002). Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nature Genetics*, 30(2), pp.227-232.

Alderete, J., Benchimol, M., Lehker, M. and Crouch, M. (2002). The complex fibronectin–*Trichomonas vaginalis* interactions and Trichomonosis. Parasitology International, 51(3), pp.285-292.

Alderete, J., Provenzano, D. and Lehker, M. (1995). Iron mediates *Trichomonas vaginalis* resistance to complement lysis. *Microbial Pathogenesis*, 19(2), pp.93-103.

Allsworth, J., Ratner, J. and Peipert, J. (2009). Trichomoniasis and Other Sexually Transmitted Infections: Results From the 2001–2004 National Health and Nutrition Examination Surveys. Sexually Transmitted Diseases, 36(12), pp.738-744.

Alsmark, C., Foster, P., Sicheritz-Ponten, T., Nakjang, S., Martin Embley, T. and Hirt, R. (2013). Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biology*, 14(2), p.R19.

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3), pp.403-410.

Antoccia, A., Kobayashi, J., Tauchi, H., Matsuura, S. and Komatsu, K. (2006). Nijmegen Breakage Syndrome and Functions of the Responsible Protein,NBS1. *Genome and Disease*, pp.191-205.

Arroyo, R. and Alderete, J. (1989). *Trichomonas vaginalis* surface proteinase activity is necessary for parasite adherence to epithelial cells. *Infection and Immunity*, 57(10), pp.2991-2997.

Arroyo, R., González-Robles, A., Martínez-Palomo, A. and Alderete, J. (1993). Signalling of *Trichomonas vaginalis* for amoeboid transformation and adhesin synthesis follows cytoadherence. Molecular Microbiology, 7(2), pp.299-309.

Azuara-Liceaga, E., Betanzos, A., Cardona-Felix, C., Castañeda-Ortiz, E., Cárdenas, H., Cárdenas-Guerra, R., Pastor-Palacios, G., García-Rivera, G., Hernández-Álvarez, D.,

Trasviña-Arenas, C., Diaz-Quezada, C., Orozco, E. and Brieba, L. (2018). The Sole DNA Ligase in *Entamoeba histolytica* Is a High-Fidelity DNA Ligase Involved in DNA Damage Repair. *Frontiers in Cellular and Infection Microbiology*, 8.

Badugu, S., Nabi, S., Vaidyam, P., Laskar, S., Bhattacharyya, S. and Bhattacharyya, M. (2015). Identification of *Plasmodium falciparum* DNA Repair Protein Mre11 with an Evolutionarily Conserved Nuclease Function. *PLOS ONE*, 10(5), p.e0125358.

Balkus, J., Richardson, B., Rabe, L., Taha, T., Mgodi, N., Kasaro, M., Ramjee, G., Hoffman, I. and Abdool Karim, S. (2014). Bacterial Vaginosis and the Risk of *Trichomonas vaginalis* Acquisition Among HIV-1–Negative Women. *Sexually Transmitted Diseases*, 41(2), pp.123-128.

Barbarella, G., Tugnoli,V. and Zambianchi,M. (1991) Imidazole ring opening of 7-methylguanosine at physiologic pH. Nucl. Nucl. , 10, 1759–1769.

Barratt, J., Gough, R., Stark, D. and Ellis, J. (2016). Bulky Trichomonad Genomes: Encoding a Swiss Army Knife. *Trends in Parasitology*, 32(10), pp.783-797.

Begley, T., Haas, B., Noel, J., Shekhtman, A., Williams, W. and Cunningham, R. (1999). A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA. *Current Biology*, 9(12), pp.653-656.

Benchimol, M. (2004). Trichomonads under Microscopy. *Microscopy and Microanalysis*, 10(05), pp.528-550.

Beranek, D. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 231(1), pp.11-30.

Bessho, T., Roy, R., Yamamoto, K., Kasai, H., Nishimura, S., Tano, K. and Mitra, S. (1993). Repair of 8-hydroxyguanine in DNA by mammalian N-methylpurine-DNA glycosylase. *Proceedings of the National Academy of Sciences*, 90(19), pp.8901-8904.

Bjornson, K., Blackwell, L., Sage, H., Baitinger, C., Allen, D. and Modrich, P. (2003). Assembly and Molecular Activities of the MutS Tetramer. *Journal of Biological Chemistry*, 278(36), pp.34667-34673.

Breen AP, Murphy JA. (1995). Reactions of oxyl radicals with DNA. - *Free Radic Biol Med* 18:1033–1077.

Brotman, R., Bradford, L., Conrad, M., Gajer, P., Ault, K., Peralta, L., Forney, L., Carlton, J., Abdo, Z. and Ravel, J. (2012). Association Between *Trichomonas vaginalis* and Vaginal Bacterial Community Composition Among Reproductive-Age Women. *Sexually Transmitted Diseases*, 39(10), pp.807-812.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. and Modrich, P. (2001). In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences*, 98(12), pp.6765-6770.

Cappelli, E., Taylor, R., Cevasco, M., Abbondandolo, A., Caldecott, K. and Frosina, G. (1997). Involvement of XRCC1 and DNA Ligase III Gene Products in DNA Base Excision Repair. *Journal of Biological Chemistry*, 272(38), pp.23970-23975.

Carlton, J., Hirt, R., Silva, J., Delcher, A., Schatz, M., Zhao, Q., Wortman, J., Bidwell, S., Alsmark, U., Besteiro, S., Sicheritz-Ponten, T., Noel, C., Dacks, J., Foster, P., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G., Westrop, G., Muller, S., Dessi, D., Fiori, P., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F., Simoes-Barbosa, A., Brown, M., Hayes, R., Mukherjee, M., Okumura, C., Schneider, R., Smith, A., Vanacova, S., Villalvazo, M., Haas, B., Pertea, M., Feldblyum, T., Utterback, T., Shu, C., Osoegawa, K., de Jong, P., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S., Logsdon, J., Henze, K., Gupta, A., Wang, C., Dunne, R., Upcroft, J., Upcroft, P., White, O., Salzberg, S., Tang, P., Chiu, C., Lee, Y., Embley, T., Coombs, G., Mottram, J., Tachezy, J., Fraser-Liggett, C. and Johnson, P. (2007). Draft Genome Sequence of the Sexually Transmitted Pathogen Trichomonas vaginalis. *Science*, 315(5809), pp.207-212.

Carney, J., Maser, R., Olivares, H., Davis, E., Le Beau, M., Yates, J., Hays, L., Morgan, W. and Petrini, J. (1998). The hMre11/hRad50 Protein Complex and Nijmegen Breakage Syndrome: Linkage of Double-Strand Break Repair to the Cellular DNA Damage Response. *Cell*, 93(3), pp.477-486.

Carter, J. and Whithaus, K. (2008). Neonatal Respiratory Tract Involvement by *Trichomonas vaginalis*: A Case Report and Review of the Literature. *The American Journal of Tropical Medicine and Hygiene*, 78(1), pp.17 - 19.

Cdc.gov. (2017). *CDC - Trichomoniasis Treatment*. [online] Available at: https://www.cdc.gov/std/trichomonas/treatment.htm [Accessed 4 Jun. 2018].

Cdc.gov. (2018). *CDC - Parasites - Neglected Parasitic Infections (NPIs) in the United States*. [online] Available at: https://www.cdc.gov/parasites/npi/index.html [Accessed 15 Aug. 2018].

Chen, J., Derfler, B. and Samson, L. (1990). *Saccharomyces cerevisiae* 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of *E. coli* and is induced in response to DNA alkylation damage. *The EMBO Journal*, 9(13), pp.4569-4575.

Cheng K C, Cahill D S, Kasai H, Nishimura S, Loeb LA. 1992. 8-Hydroxy- guanine, an abundant form of oxidative DNA damage, causes G—- T and A—. C substitutions. *Journal of Biological Chemistry* 267:166–172.

Connor, E. and Wyatt, M. (2002). Active-Site Clashes Prevent the Human 3-Methyladenine DNA Glycosylase from Improperly Removing Bases. *Chemistry & Biology*, 9(9), pp.1033-1041.

Cornelius, D., Robinson, D., Muzny, C., Mena, L., Aanensen, D., Lushbaugh, W. and Meade, J. (2012). Genetic Characterization of *Trichomonas vaginalis* Isolates by Use of Multilocus Sequence Typing. *Journal of Clinical Microbiology*, 50(10), pp.3293-3300.

Cosar, C. and Julou, L. (1959). The activity of 1-(2-hydroxyethyl)-2-methyl-5-nitroimidazole (R. P. 8823) against experimental *Trichomonas vaginalis* infection. *Ann. Inst. Pasteur.*, 96,

pp.238-241.

Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P. and Vindigni, A. (2007). Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair*, 6(6), pp.712-722.

Costes, S., Chiolo, I., Pluth, J., Barcellos-Hoff, M. and Jakob, B. (2010). Spatiotemporal characterization of ionizing radiation induced DNA damage foci and their relation to chromatin organization. *Mutation Research/Reviews in Mutation Research*, 704(1-3), pp.78-87.

Cotch, M., Pastorek, J., Nugent, R., Hillier, S., Gibbs, R., Martin, D., Eschenbach, D., Edelman, R., Carey, C., Regan, J., Krohn, M., Klebanoff, M., Rao, V. and Rhoads, G. (1997). *Trichomonas vaginalis* Associated With Low Birth Weight and Preterm Delivery. *Sexually Transmitted Diseases*, 24(6), pp.353-360.

Crucitti, T., Jespers, V., Mulenga, C., Khondowe, S., Vandepitte, J. and Buvé, A. (2011). Non-Sexual Transmission of *Trichomonas vaginalis* in Adolescent Girls Attending School in Ndola, Zambia. PLoS ONE, 6(1), p.e16310.

Davis, A. and Chen, D. (2013). DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res*, 2(3), pp.130-143.

De Bont, R. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, 19(3), pp.169-185.

De Bont, R.; van Larebeke, N. Endogenous DNA damage in humans: A review of quantitative data. Mutagenesis 2004, 19, 169–185.)

Demirezen, S., Safi, Z. and Beksaç, S. (2000). The interaction of *Trichomonas vaginalis* with epithelial cells, polymorphonuclear leucocytes and erythrocytes on vaginal smears: light microscopic observation. Cytopathology, 11(5), pp.326-332.

Dianov, G., Price, A. and Lindahl, T. (1992). Generation of single-nucleotide repair patches following excision of uracil residues from DNA. *Molecular and Cellular Biology*, 12(4), pp.1605-1612.

Doolittle, W. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics*, 14(8), pp.307-311.

Duboucher C, Noel C, Durand-Joly I, Gerbod D, Delgado-Viscogliosi P, Jouveshomme S, Leclerc C, Cartolano GL, Dei-Cas E, Capron M, Viscogliosi E, 2003. Pulmonary coinfection by *Trichomonas vaginalis* and Pneumocystis sp. as a novel manifestation of AIDS. *Hum Pathol* 34 : 508–511α

Eckburg, P. (2005). Diversity of the Human Intestinal Microbial Flora. *Science*, 308(5728), pp.1635-1638.

Engelward, B., Weeda, G., Wyatt, M., Broekhof, J., de Wit, J., Donker, I., Allan, J., Gold, B., Hoeijmakers, J. and Samson, L. (1997). Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proceedings of the National Academy of Sciences*, 94(24), pp.13087-13092.

Fair, W., Couch, J. and Wehner, N. (1976). Prostatic antibacterial factor identity and significance. *Urology*, 7(2), pp.169-177.

Fang, H., Huangfu, L., Chen, R., Li, P., Xu, S., Zhang, E., Cao, W., Liu, L., Yao, Y., Liang, G., Xu, C., Zhou, Y. and Yang, Z. (2017). Ancestor of land plants acquired the DNA-3-methyladenine glycosylase (MAG) gene from bacteria through horizontal gene transfer. *Scientific Reports*, 7(1).

Fenoy, I., Bogado, S., Contreras, S., Gottifredi, V. and Angel, S. (2016). The Knowns Unknowns: Exploring the Homologous Recombination Repair Pathway in *Toxoplasma gondii*. *Frontiers in Microbiology*, 7(627)

Fettweis, J., Serrano, M., Huang, B., Brooks, J., Glascock, A., Sheth, N., Strauss, J., Jefferson, K. and Buck, G. (2014). An Emerging Mycoplasma Associated with Trichomoniasis, Vaginal Infection and Disease. *PLoS ONE*, 9(10), p.e110943.

Figueroa-Angulo, E., Rendón-Gandarilla, F., Puente-Rivera, J., Calla-Choque, J., Cárdenas-Guerra, R., Ortega-López, J., Quintas-Granados, L., Alvarez-Sánchez, M. and Arroyo, R. (2012). The effects of environmental factors on the virulence of *Trichomonas vaginalis*.

Microbes and Infection, 14(15), pp.1411-1427.

Flegr, J., Čerkasov, J. and Štokrová, J. (1988). Multiple populations of double-stranded RNA in two virus-harbouring strains of *Trichomonas vaginalis*. *Folia Microbiologica*, 33(6), pp.462-465.

Fortini, P., Pascucci, B., Parlanti, E., Sobol, R., Wilson, S. and Dogliotti, E. (1998). Different DNA Polymerases Are Involved in the Short- and Long-Patch Base Excision Repair in Mammalian Cells. *Biochemistry*, 37(11), pp.3575-3580.

Gao, Y., Katyal, S., Lee, Y., Zhao, J., Rehg, J., Russell, H. and McKinnon, P. (2011). DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature*, 471(7337), pp.240-244.

Gardner, M., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R., Carlton, J., Pain, A., Nelson, K., Bowman, S., Paulsen, I., James, K., Eisen, J., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M., Vaidya, A., Martin, D., Fairlamb, A., Fraunholz, M., Roos, D., Ralph, S., McFadden, G., Cummings, L., Subramanian, G., Mungall, C., Venter, J., Carucci, D., Hoffman, S., Newbold, C., Davis, R., Fraser, C. and Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 419 (6906), pp.498–511

Gary, R., Kim, K., Cornelius, H., Park, M. and Matsumoto, Y. (1999). Proliferating Cell Nuclear Antigen Facilitates Excision in Long-patch Base Excision Repair. *Journal of Biological Chemistry*, 274(7), pp.4354-4363.

Genois, M., Paquet, E., Laffitte, M., Maity, R., Rodrigue, A., Ouellette, M. and Masson, J. (2014). DNA Repair Pathways in Trypanosomatids: from DNA Repair to Drug Resistance. *Microbiology and Molecular Biology Reviews*, 78(1), pp.40-73.

Gibbs, P., McGregor, W., Maher, V., Nisson, P. and Lawrence, C. (1998). A human homolog of the *Saccharomyces cerevisiae* REV3 gene, which encodes the catalytic subunit of DNA polymerase. *Proceedings of the National Academy of Sciences*, 95(12), pp.6876-6880.

Gill, E. and Fast, N. (2007). Stripped-down DNA repair in a highly reduced parasite. *BMC Molecular Biology*, 8(1), p.24.

Giloni L, Takeshita M, Johnson F, Iden C, Grollman AP. (1981). Bleomy- cin-induced strand-scission of DNA. Mechanism of deoxyribose cleavage. *J Biol Chem* 256:8608–8615

Glover, L., McCulloch, R. and Horn, D. (2008). Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Research*, 36(8), pp.2608-2618.

Goodarzi, A., Yu, Y., Riballo, E., Douglas, P., Walker, S., Ye, R., Härer, C., Marchetti, C., Morrice, N., Jeggo, P. and Lees-Miller, S. (2006). DNA-PK autophosphorylation facilitates

Artemis endonuclease activity. *The EMBO Journal*, 25(16), pp.3880-3889.

Goodman, R., Freret, T., Kula, T., Geller, A., Talkington, M., Tang-Fernandez, V., Suciu, O., Demidenko, A., Ghabrial, S., Beach, D., Singh, B., Fichorova, R. and Nibert, M. (2011). Clinical Isolates of *Trichomonas vaginalis* Concurrently Infected by Strains of Up to Four Trichomonasvirus Species (Family Totiviridae). *Journal of Virology*, 85(9), pp.4258-4270.

Goodman, R., Ghabrial, S., Fichorova, R. and Nibert, M. (2010). Trichomonasvirus: a new genus of protozoan viruses in the family Totiviridae. *Archives of Virology*, 156(1), pp.171-179.

Goodsell, D. (2004). The Molecular Perspective: Nicotine and Nitrosamines. *The Oncologist*, 9(3), pp.353-354.

Gottlieb, T. and Jackson, S. (1993). The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. *Cell*, 72(1), pp.131-142.

Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M. and Lieber, M. (2007). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *The EMBO Journal*, 26(4), pp.1010-1023.

Gunn-Moore, D., McCann, T., Reed, N., Simpson, K. and Tennant, B. (2007). Prevalence of Tritrichomonas foetus infection in cats with diarrhoea in the UK. *Journal of Feline Medicine & Surgery*, 9(3), pp.214-218.

Han, L., Masani, S., Hsieh, C. and Yu, K. (2014). DNA Ligase I Is Not Essential for Mammalian Cell Viability. *Cell Reports*, 7(2), pp.316-320.

Hanahan, D. and Weinberg, R. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), pp.646-674.

Hang, B., Singer, B., Margison, G. and Elder, R. (1997). Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1,N6-ethenoadenine and hypoxanthine but not of 3,N4-ethenocytosine or 8-oxoguanine. *Proceedings of the National Academy of Sciences*, 94(24), pp.12869-12874.

Harper, J. and Elledge, S. (2007). The DNA Damage Response: Ten Years After. *Molecular Cell*, 28(5), pp.739-745.

Hewitt, G., Jurk, D., Marques, F., Correia-Melo, C., Hardy, T., Gackowska, A., Anderson, R., Taschuk, M., Mann, J. and Passos, J. (2012). Telomeres are favoured targets of a persistent DNA damage response in ageing and stress-induced senescence. *Nature Communications*, 3(1).

Honigberg, B. and King, V. (1964). Structure of *Trichomonas vaginalis* Donne. *The Journal of Parasitology*, 50(3), p.345.

Huffman JL, Sundheim O, Tainer JA. 2005. DNA base damage recognition and removal: new twists and grooves. Mutat Res 577:55–76.

Ivens, A., Peacock, C., Worthey, E., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R., Cronin, A., Cruz, A., Davies, R., Gaudenzi, J., Dobson, D., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T., Price, C., Purnelle, B., Quail, M., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J., Rutter, S., Saundeers, D., Schafer, M., Schein, J., Schwartz, D., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D., Blackwell, J., Stuart, K., Barrell, B. and Myler, P. (2005). The Genome of the Kinetoplastid Parasite, Leishmania major. *Science*, 309(5733), pp.436-442.

Jacobs, A. and Schär, P. (2011). DNA glycosylases: in DNA repair and beyond. *Chromosoma*, 121(1), pp.1-20.

Jacobs, Aα and Schär, P. (2011). DNA glycosylases: in DNA repair and beyond. *Chromosoma*, 121(1), pp.1-20.

Janssen, B., Chen, Y., Molgora, B., Wang, S., Simoes-Barbosa, A. and Johnson, P. (2018). CRISPR/Cas9-mediated gene modification and gene knock out in the human-infective

parasite *Trichomonas vaginalis*. *Scientific Reports*, 8(1).

Johnson, R., Klassen, R., Prakash, L. and Prakash, S. (2015). A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. *Molecular Cell*, 59(2), pp.163-175.

Johnson, R., Washington, M., Haracska, L., Prakash, S. and Prakash, L. (2000). Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature*, 406(6799), pp.1015-1019.

Kadyrov, F., Dzantiev, L., Constantin, N. and Modrich, P. (2006). Endonucleolytic Function of MutLα in Human Mismatch Repair. *Cell*, 126(2), pp.297-308.

Kim, J., Mouw, K., Polak, P., Braunstein, L., Kamburov, A., Tiao, G., Kwiatkowski, D., Rosenberg, J., Van Allen, E., D'Andrea, A. and Getz, G. (2016). Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nature Genetics*, 48(6), pp.600-606.

Kirkcaldy, R., Augostini, P., Asbel, L., Bernstein, K., Kerani, R., Mettenbrink, C., Pathela, P., Schwebke, J., Secor, W., Workowski, K., Davis, D., Braxton, J. and Weinstock, H. (2012). *Trichomonas vaginalis* Antimicrobial Drug Resistance in 6 US Cities, STD Surveillance Network, 2009–2010. *Emerging Infectious Diseases*, 18(6), pp.939-943.

Kissinger, P. and Adamski, A. (2013). Trichomoniasis and HIV interactions: a review. *Sexually Transmitted Infections*, 89(6), pp.426-433.

Kondo, N., Takahashi, A., Ono, K. and Ohnishi, T. (2010). DNA Damage Induced by Alkylating Agents and Repair Pathways. *Journal of Nucleic Acids*, 2010, pp.1-7.

Krieger, J. and Rein, M. (1982). Zinc Sensitivity of *Trichomonas vaginalis*: In Vitro Studies and Clinical Implications. *Journal of Infectious Diseases*, 146(3), pp.341-345.

Laga, M., Alary, M., Behets, F., Goeman, J., Piot, P., Nzila, N., Manoka, A., Tuliza, M. and St Louis, M. (1994). Condom promotion, sexually transmitted diseases treatment, and declining incidence of HIV-1 infection in female Zairian sex workers. *The Lancet*, 344(8917), pp.246-248.

Lahue, R. and Modrich, P. (1988). Methyl-directed DNA mismatch repair in *Escherichia coli*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 198(1), pp.37-43.

Larsen, B. and Hwang, J. (2010). Mycoplasma,Ureaplasma, and Adverse Pregnancy Outcomes: A Fresh Look. *Infectious Diseases in Obstetrics and Gynecology*, 2010, pp.1-7.

Larson, K., Sahm, J., Shenkar, R. and Strauss, B. (1985). Methylation-induced blocks to in vitro DNA replication. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 150(1-2), pp.77-84.

Lau, A., Schärer, O., Samson, L., Verdine, G. and Ellenberger, T. (1998). Crystal Structure of a Human Alkylbase-DNA Repair Enzyme Complexed to DNA. *Cell*, 95(2), pp.249-258.

Lau, Y., Lee, W., Gudimella, R., Zhang, G., Ching, X., Razali, R., Aziz, F., Anwar, A. and Fong, M. (2016). Deciphering the Draft Genome of Toxoplasma gondii RH Strain. *PLOS ONE*, 11(6), p.e0157901.

Lee, C., Delaney, J., Kartalou, M., Lingaraju, G., Maor-Shoshani, A., Essigmann, J. and Samson, L. (2009). Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG)†. *Biochemistry*, 48(9), pp.1850-1861.

Lehker, M. (1990). Specific erythrocyte binding is an additional nutrient acquisition system for *Trichomonas vaginalis*. Journal of Experimental Medicine, 171(6), pp.2165-2170.

Lehker, M. and Alderete, J. (2000). Biology of trichomonosis. Current Opinion in Infectious Diseases, 13(1), pp.37-45.

Levin, D., McKenna, A., Motycka, T., Matsumoto, Y. and Tomkinson, A. (2000). Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Current Biology*, 10(15), pp.919-S2.

Levy, M., Gookin, J., Poore, M., Birkenheuer, A., Dykstra, M. and Litaker, R. (2003). *Tritrichomonas foetus* and not *Pentatrichomonas hominis* is the etiologic agent of feline trichomonal diarrhea. *Journal of Parasitology*, 89(1), pp.99-104.

Lim, S., Park, S., Ahn, K., Oh, D., Ryu, J. and Shin, S. (2010). First Report of Feline Intestinal Trichomoniasis Caused by *Tritrichomonas foetus* in Korea. *The Korean Journal of Parasitology*, 48(3), p.247.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362(6422), pp.709-715.

Lindahl, T. and Wood, R. (1999). Quality Control by DNA Repair. *Science*, 286(5446), pp.1897-1905.

Lindahl, T.; Barnes, D.E. (2000), Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* 65, 127–133

Ling, X., Kong, M., Liu, F., Zhu, B., Chen, Y., Wang, Z., Li, J., Nelson, K., Xia, X. and Xiang, C. (2010). Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis. *BMC Genomics*, 11(1), p.488.

Lingaraju, G., Kartalou, M., Meira, L. and Samson, L. (2008). Substrate specificity and sequence-dependent activity of the *Saccharomyces cerevisiae* 3-methyladenine DNA glycosylase (Mag). *DNA Repair*, 7(6), pp.970-982.

Liu, D., Keijzers, G. and Rasmussen, L. (2017). DNA mismatch repair and its many roles in eukaryotic cells. *Mutation Research/Reviews in Mutation Research*, 773, pp.174-187.

Liu, X. and Roy, R. (2001). Mutation at Active Site Lysine 212 to Arginine Uncouples the Glycosylase Activity from the Lyase Activity of Human Endonuclease III. *Biochemistry*, 40(45), pp.13617-13622.

Longley, M., Pierce, A. and Modrich, P. (1997). DNA Polymerase δ Is Required for Human Mismatch Repair in Vitro. *Journal of Biological Chemistry*, 272(16), pp.10917-10921.

López-Camarillo, C., Lopez-Casamichana, M., Weber, C., Guillen, N., Orozco, E. and Marchat, L. (2009). DNA repair mechanisms in eukaryotes: Special focus in *Entamoeba histolytica* and related protozoan parasites. *Infection, Genetics and Evolution*, 9(6), pp.1051-1056.

Lushbaugh, W., Turner, A., Klykken, P. and Gentry, G. (1989). Characterization of a Secreted Cytoactive Factor from *Trichomonas Vaginalis*. *The American Journal of Tropical Medicine and Hygiene*, 41(1), pp.18-28.

Malik, S.B., Pightling, A.W., Stefaniak, L.M., Schurko, A.M., Logsdon, J.M., 2008. An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. PLoS One 3.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C., Lu, S., Chitsaz, F., Derbyshire, M., Geer, R., Gonzales, N., Gwadz, M., Hurwitz, D., Lu, F., Marchler, G., Song, J., Thanki, N., Wang, Z., Yamashita, R., Zhang, D., Zheng, C., Geer, L. and Bryant, S. (2016). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(D1), D200-D203.

Mardones, F., Perez, A., Martínez, A. and Carpenter, T. (2008). Risk factors associated with *Tritrichomonas foetus* infection in beef herds in the Province of Buenos Aires, Argentina. *Veterinary Parasitology*, 153(3-4), pp.231-237.

Margarita, V., Rappelli, P., Dessì, D., Pintus, G., Hirt, R. and Fiori, P. (2016). Symbiotic Association with *Mycoplasma hominis* Can Influence Growth Rate, ATP Production, Cytolysis and Inflammatory Response of *Trichomonas vaginalis*. *Frontiers in Microbiology*, 7.

Marti, T., Kunz, C. and Fleck, O. (2002). DNA mismatch repair and mutation avoidance pathways. *Journal of Cellular Physiology*, 191(1), pp.28-41.

Martin, D., Zozaya, M., Lillis, R., Myers, L., Nsuami, M. and Ferris, M. (2013). Unique Vaginal Microbiota That Includes an Unknown Mycoplasma-Like Organism Is Associated With *Trichomonas vaginalis* Infection. *The Journal of Infectious Diseases*, 207(12), pp.1922-1931.

Martin, F., Castro, M., Aboul-ela, F. and Tinoco, I. (1985). Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Research*, 13(24), pp.8927-8938.

Matsumoto, Y. and Kim, K. (1995). Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science*, 269(5224), pp.699-702.

Mavaddat, N., Peock, S., Frost, D., Ellis, S., Platte, R., Fineberg, E., Evans, D., Izatt, L., Eeles, R., Adlard, J., Davidson, R., Eccles, D., Cole, T., Cook, J., Brewer, C., Tischkowitz,

M., Douglas, F., Hodgson, S., Walker, L., Porteous, M., Morrison, P., Side, L., Kennedy, M., Houghton, C., Donaldson, A., Rogers, M., Dorkins, H., Miedzybrodzka, Z., Gregory, H., Eason, J., Barwell, J., McCann, E., Murray, A., Antoniou, A. and Easton, D. (2013). Cancer Risks for BRCA1 and BRCA2 Mutation Carriers: Results From Prospective Analysis of EMBRACE. *JNCI: Journal of the National Cancer Institute*, 105(11), pp.812-822.

McClelland, R., Sangaré, L., Hassan, W., Lavreys, L., Mandaliya, K., Kiarie, J., Ndinya-Achola, J., Jaoko, W. and Baeten, J. (2007). Infection with *Trichomonas vaginalis* Increases the Risk of HIV-1 Acquisition. *The Journal of Infectious Diseases,* 195(5), pp.698-702.

McCulloch, S., Gu, L. and Li, G. (2002). Bi-directional Processing of DNA Loops by Mismatch Repair-dependent and -independent Pathways in Human Cells. *Journal of Biological Chemistry*, 278(6), pp.3891-3896.

McIlwraith, M., Vaisman, A., Liu, Y., Fanning, E., Woodgate, R. and West, S. (2005). Human DNA Polymerase η Promotes DNA Synthesis from Strand Invasion Intermediates of Homologous Recombination. *Molecular Cell*, 20(5), pp.783-792.

McLaren LC, Davis LE, Healy GR, James CG, 1983. Isolation of *Trichomonas vaginalis* from the respiratory tract of infants with respiratory disease. *Pediatrics* 71 : 888–890.

Mechanic, L., Frankel, B. and Matson, S. (2000). *Escherichia coli* MutL Loads DNA Helicase II onto DNA. *Journal of Biological Chemistry*, 275(49), pp.38337-38346.

Midlej, V. and Benchimol, M. (2009). *Trichomonas vaginalis* kills and eats – evidence for phagocytic activity as a cytopathic effect. *Parasitology*, 137(01), p.65.

Miyabe, I., Mizuno, K., Keszthelyi, A., Daigaku, Y., Skouteri, M., Mohebi, S., Kunkel, T., Murray, J. and Carr, A. (2015). Polymerase δ replicates both strands after homologous recombination–dependent fork restart. *Nature Structural & Molecular Biology*, 22(11), pp.932-938.

Moore, L., Le, T. and Fan, G. (2012). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1), pp.23-38.

Muller, M., Mentel, M., van Hellemond, J., Henze, K., Woehle, C., Gould, S., Yu, R., van der Giezen, M., Tielens, A. and Martin, W. (2012). Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. *Microbiology and Molecular Biology Reviews*, 76(2), pp.444-495.

Nakabeppu, Y., Kondo, H. and Sekiguchi, M. (1984). Cloning and characterization of the alkA gene of *Escherichia coli* that encodes 3-methyladenine DNA glycosylase II. *Journal of Biological Chemistry*, 259(22), pp.13723-13729.

Nielsen, F., Jäger, A., Lützen, A., Bundgaard, J. and Rasmussen, L. (2003). Characterization of human exonuclease 1 in complex with mismatch repair proteins, subcellular localization and association with PCNA. *Oncogene*, 23(7), pp.1457-1468.

Nimonkar, A., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J., Wyman, C., Modrich, P. and Kowalczykowski, S. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN

constitute two DNA end resection machineries for human DNA break repair. *Genes & Development*, 25(4), pp.350-362.

O'Brien, P. and Ellenberger, T. (2003). Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase. *Journal of Biological Chemistry*, 279(11), pp.9750-9757.

O'Rourke, E., Chevalier, C., Boiteux, S., Labigne, A., Ielpi, L. and Radicella, J. (2000). A Novel 3-Methyladenine DNA Glycosylase from *Helicobacter pylori* Defines a New Class within the Endonuclease III Family of Base Excision Repair Glycosylases. *Journal of Biological Chemistry*, 275(26), pp.20077-20083.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., Kainuma, R., Sugino, A. and Iwatsuki, N. (1968). In Vivo Mechanism of DNA Chain Growth. *Cold Spring Harbor Symposia on Quantitative Biology*, 33(0), pp.129-143.

Ordoñez-Quiroz, A., Ortega-Pierres, M., Bazán-Tejeda, M. and Bermúdez-Cruz, R. (2018). DNA damage induced by metronidazole in *Giardia duodenalis* triggers a DNA homologous recombination response. *Experimental Parasitology*, 194, pp.24-31.

Pandya, G. and Moriya, M. (1996). 1,N6-Ethenodeoxyadenosine, a DNA Adduct Highly Mutagenic in Mammalian Cells. Biochemistry, 35(35), pp.11487-11492.

Parikh, S. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *The EMBO Journal*, 17(17), pp.5214-5226.

Parsons, J. and Dianov, G. (2013). Co-ordination of base excision repair and genome stability. *DNA Repair*, 12(5), pp.326-333.

Paz-Elizur, T., Sevilya, Z., Leitner-Dagan, Y., Elinger, D., Roisman, L. and Livneh, Z. (2008). DNA repair of oxidative DNA damage in human carcinogenesis: Potential application for cancer risk assessment and prevention. *Cancer Letters*, 266(1), pp.60-72.

Pereira-Neves, A., Ribeiro, K. and Benchimol, M. (2003). Pseudocysts in Trichomonads – New Insights. Protist, 154(3-4), pp.313-329.

Petrini, J., Xiao, Y. and Weaver, D. (1995). DNA ligase I mediates essential functions in mammalian cells. *Molecular and Cellular Biology*, 15(8), pp.4303-4308.

Pilié, P., Johnson, A., Hanson, K., Dayno, M., Kapron, A., Stoffel, E. and Cooney, K. (2017). Germline genetic variants in men with prostate cancer and one or more additional cancers. *Cancer*, 123(20), pp.3925-3932.

Press, N., Chavez, V., Ticona, E., Calderon, M., Apolinario, I., Culotta, A., Arevalo, J. and Gilman, R. (2001). Screening for Sexually Transmitted Diseases in Human Immunodeficiency Virus-Positive Patients in Peru Reveals an Absence of *Chlamydia trachomatis* and Identifies *Trichomonas vaginalis* in Pharyngeal Specimens. *Clinical*

Infectious Diseases, 32(5), pp.808-814.

Proudfoot, C. and McCulloch, R. (2005). Distinct roles for two RAD51-related genes in *Trypanosoma brucei* antigenic variation. *Nucleic Acids Research*, 33(21), pp.6906-6919.

Provenzano, D. and Alderete, J. (1995). Analysis of human immunoglobulin-degrading cysteine proteinases of *Trichomonas vaginalis*. *Infection and Immunity*, 63(9), pp.3388–3395.

Pruski, P., Lewis, H., Lee, Y., Marchesi, J., Bennett, P., Takats, Z. and MacIntyre, D. (2018). Assessment of microbiota:host interactions at the vaginal mucosa interface. *Methods*, 149, pp.74-84.

Quinlivan, E., Patel, S., Grodensky, C., Golin, C., Tien, H. and Hobbs, M. (2012). Modeling the Impact of *Trichomonas vaginalis* Infection on HIV Transmission in HIV-Infected Individuals in Medical Care. *Sexually Transmitted Diseases*, 39(9), pp.671-677.

Rae, D., Crews, J., Greiner, E. and Donovan, G. (2004). Epidemiology of *Tritrichomonas foetus* in beef bull populations in Florida. *Theriogenology*, 61(4), pp.605-618.

Rappelli, P., Addis, M., Carta, F. and Fiori, P. (1998). *Mycoplasma hominis* parasitism of *Trichomonas vaginalis*. *The Lancet*, 352(9145), p.2023.

Rathod, S., Krupp, K., Klausner, J., Arun, A., Reingold, A. and Madhivanan, P. (2011). Bacterial Vaginosis and Risk for *Trichomonas vaginalis* Infection: A Longitudinal

Analysis. *Sexually Transmitted Diseases*, 38(9), pp.882-886.

Rendón-Maldonado, J., Espinosa-Cantellano, M., González-Robles, A. and Martı´nez-Palomo, A. (1998). *Trichomonas vaginalis*: In Vitro Phagocytosis of Lactobacilli, Vaginal Epithelial Cells, Leukocytes, and Erythrocytes. *Experimental Parasitology*, 89(2), pp.241-250.

Robson, C. and Hickson, I. (1991). Isolation of cDNA clones encoding a human apurini/apyrimidinic endonuclease that corects DNA repair and mutagenisis defects inE.coli xth(exonuclease III) mutants. *Nucleic Acids Research*, 19(20), pp.5519-5523.

Røttingen, J., Cameron, D. and Garnett, G. (2001). A Systematic Review of the Epidemiologic Interactions Between Classic Sexually Transmitted Diseases and HIV: How much is really known?. *Sexually Transmitted Diseases*, 28(10), pp.579-597.

Rydberg B, Lindahl T. 1982. Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO J 1:211–216.

Sandoval-Cabrera, A., Zarzosa-Álvarez, A., Martínez-Miguel, R. and Bermúdez-Cruz, R. (2015). MR (Mre11-Rad50) complex in *Giardia duodenalis*: In vitro characterization and its response upon DNA damage. Biochimie, 111, pp.45-57.

Saparbaev, M., Kleibl, K. and Laval, J. (1995). *Escherichia coil,Saccharomyces cerevisia*e, rat and human 3-methyladenine DNA glycosylases repair 1,N6-ethenoadenine when present

in DNA. *Nucleic Acids Research*, 23(18), pp.3750-3755.

Sartori, A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J. and Jackson, S. (2007). Human CtIP promotes DNA end resection. *Nature*, 450(7169), pp.509-514.

Schönknecht, G., Weber, A. and Lercher, M. (2013). Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays*, 36(1), pp.9-20.

Schumacher, B., Garinis, G. and Hoeijmakers, J. (2008). Age to survive: DNA damage and aging. *Trends in Genetics*, 24(2), pp.77-85.

Schwebke, J. and Barrientes, F. (2006). Prevalence of *Trichomonas vaginalis* Isolates with Resistance to Metronidazole and Tinidazole. Antimicrobial Agents and Chemotherapy, 50(12), pp.4209-4210.

Sekelsky, J. J., Brodsky, M. H., and Burtis, K. C. (2000). DNA repair in *Drosophila*: insights from the *Drosophila* genome sequence. *The Journal of cell biology*, 150(2), pp.31–36.

Shapiro, R. and Shiuey, S. (1969). Reaction of nitrous acid with alkylaminopurines. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 174(1), pp.403-405.

Shrivastav, N., Li, D. and Essigmann, J. (2009). Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*, 31(1), pp.59-70.

Sieber, K., Bromley, R. and Dunning Hotopp, J. (2017). Lateral gene transfer between prokaryotes and eukaryotes. *Experimental Cell Research*, 358(2), pp.421-426.

Sieber, O., Lipton, L., Crabtree, M., Heinimann, K., Fidalgo, P., Phillips, R., Bisgaard, M., Orntoft, T., Aaltonen, L., Hodgson, S., Thomas, H. and Tomlinson, I. (2003). Multiple Colorectal Adenomas, Classic Adenomatous Polyposis, and Germ-Line Mutations inMYH. *New England Journal of Medicine*, 348(9), pp.791-799.

Silver, B., Guy, R., Kaldor, J., Jamil, M. and Rumbold, A. (2014). *Trichomonas vaginalis* as a Cause of Perinatal Morbidity. *Sexually Transmitted Diseases*, 41(6), pp.369-376.

Simsek, D., Furda, A., Gao, Y., Artus, J., Brunet, E., Hadjantonakis, A., Van Houten, B., Shuman, S., McKinnon, P. and Jasin, M. (2011). Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. *Nature*, 471(7337), pp.245-248.

Sleeth, K., Robson, R. and Dianov, G. (2004). Exchangeability of Mammalian DNA Ligases between Base Excision Repair Pathways. *Biochemistry*, 43(40), pp.12924-12930.

Snipes, L., Gamard, P., Narcisi, E., Beard, B., Lehmann, T. and Secor, W. (2000). Molecular Epidemiology of Metronidazole Resistance in a Population of *Trichomonas vaginalis* Clinical Isolates. *Journal of Clinical Microbiology*, 38(8), pp.3004-3009.

Stodola, J. and Burgers, P. (2016). Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nature Structural & Molecular Biology*, 23(5), pp.402-408.

Sung, P., Krejci, L., Van Komen, S. and Sehorn, M. (2003). Rad51 Recombinase and Recombination Mediators. *Journal of Biological Chemistry*, 278(44), pp.42729-42732.

Swygard, H. (2004). Trichomoniasis: clinical manifestations, diagnosis and management. Sexually Transmitted Infections, 80(2), pp.91-95.

Thomas, L., Yang, C. and Goldthwait, D. (1982). Two DNA glycosylases in *E. coli* which release primarily 3-methyladenine. *Biochemistry*, 21(6), pp.1162-1169.

Tomkinson, A., Tappe, N. and Friedberg, E. (1992). DNA ligase I from Saccharomyces cerevisiae: physical and biochemical characterization of the CDC9 gene product. *Biochemistry*, 31(47), pp.11762-11771.

Trasviña-Arenas, C., David, S., Delaye, L., Azuara-Liceaga, E. and Brieba, L. (2019). Evolution of Base Excision Repair in *Entamoeba histolytica* is shaped by gene loss, gene duplication, and lateral gene transfer. *DNA Repair*, 76, pp.76-88.

Tropp BE. (2011). Molecular biology, from Gene to Protein, 4th ed. *Jones & Bartlett Learning*. pp 1100.

Trussell, R. and Plass, E. (1940). The Pathogenicity and Physiology of a Pure Culture of *Trichomonas Vaginalis*. *American Journal of Obstetrics and Gynecology*, 40(5), pp.883-890.

Tsai, C., Kim, S. and Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proceedings of the National Academy of Sciences*, 104(19), pp.7851-7856.

Uematsu, N., Weterings, E., Yano, K., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P., van Gent, D., Chen, B. and Chen, D. (2007). Autophosphorylation of DNA-PKCSregulates its dynamics at DNA double-strand breaks. *The Journal of Cell Biology*, 177(2), pp.219-229.

Upcroft, J., Dunn, L., Wal, T., Tabrizi, S., Delgadillo-Correa, M., Johnson, P., Garland, S., Siba, P. and Upcroft, P. (2009). Metronidazole resistance in *Trichomonas vaginalis* from highland women in Papua New Guinea. *Sexual Health*, 6(4), p.334.

Vancini, R. and Benchimol, M. (2007). Entry and intracellular location of *Mycoplasma hominis in Trichomonas vaginalis*. *Archives of Microbiology*, 189(1), pp.7-18.

Vancini, R., Pereira-Neves, A., Borojevic, R. and Benchimol, M. (2007). *Trichomonas vaginalis* harboring *Mycoplasma hominis* increases cytopathogenicity in vitro. *European Journal of Clinical Microbiology & Infectious Diseases*, 27(4), pp.259-267.

Warren, J., Forsberg, L. and Beese, L. (2006). The structural basis for the mutagenicity of O6-methyl-guanine lesions. *Proceedings of the National Academy of Sciences*, 103(52), pp.19701-19706.

Watson, J. and Crick, F. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356), pp.737-738.

Wei, Q., Cheng, L., Hong, W. and Spitz, M. (1996). Reduced DNA Repair Capacity in Lung Cancer Patients. *American Association for Cancer Research*, 56(18), pp.4103-4107.

Who.int. (2012). *WHO | Global incidence and prevalence of selected curable sexually transmitted infections – 2008*. [online] Available at: https://www.who.int/reproductivehealth/publications/rtis/stisestimates/en/ [Accessed 20 Aug. 2018].

Wildenberg, J. and Meselson, M. (1975). Mismatch repair in heteroduplex DNA. *Proceedings of the National Academy of Sciences*, 72(6), pp.2202-2206.

Wyatt, M. and Samson, L. (2000). Influence of DNA structure on hypoxanthine and 1,N6-ethenoadenine removal by murine 3-methyladenine DNA glycosylase. *Carcinogenesis*, 21(5), pp.901-908.

Yang, S., Zhao, W., Wang, H., Wang, Y., Li, J. and Wu, X. (2018). *Trichomonas vaginalis* infection-associated risk of cervical cancer: A meta-analysis. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 228, pp.166-173.

Yano, K., Morotomi-Yano, K., Wang, S., Uematsu, N., Lee, K., Asaithamby, A., Weterings, E. and Chen, D. (2007). Ku recruits XLF to DNA double-strand breaks. *EMBO reports*, 9(1), pp.91-96.

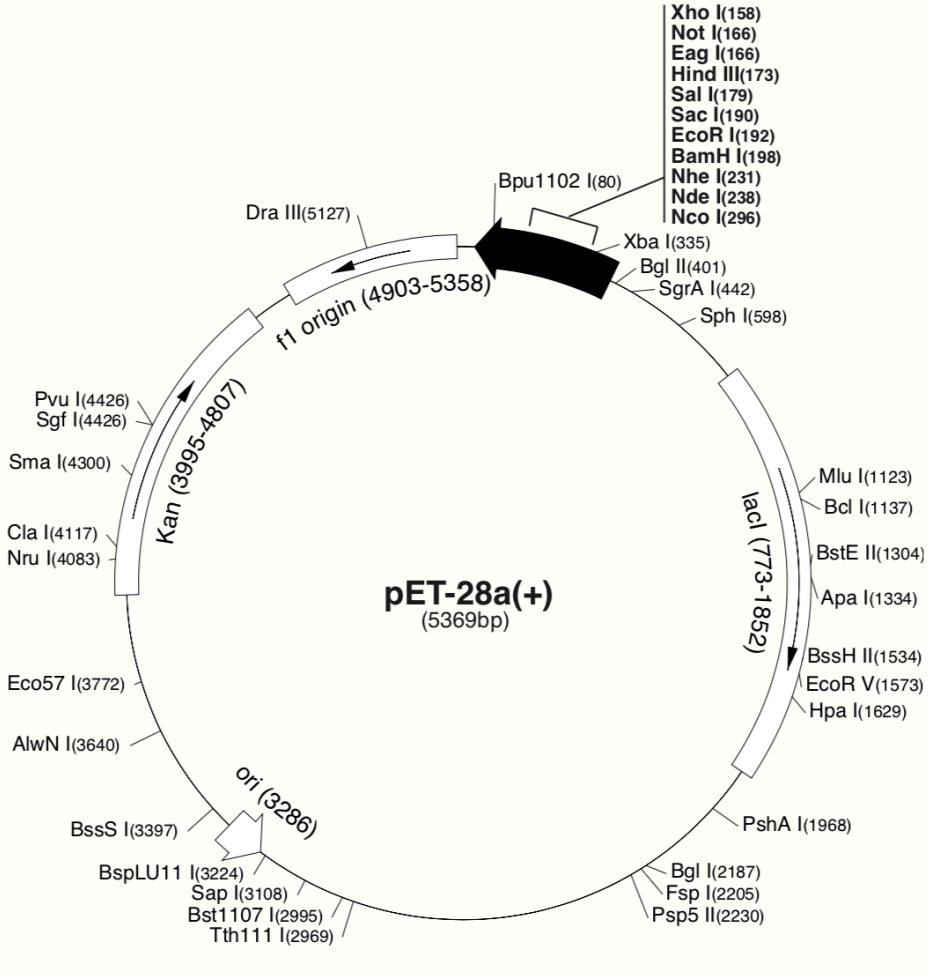
Yasui, M., Suenaga, E., Koyama, N., Masutani, C., Hanaoka, F., Gruz, P., Shibutani, S., Nohmi, T., Hayashi, M. and Honma, M. (2008). Miscoding Properties of 2′-Deoxyinosine, a Nitric Oxide-Derived DNA Adduct, during Translesion Synthesis Catalyzed by Human DNA Polymerases. *Journal of Molecular Biology*, 377(4), pp.1015-1023.

Yu, H., Harrison, F. and Xia, F. (2018). Altered DNA repair; an early pathogenic pathway in Alzheimer's disease and obesity. *Scientific Reports*, 8(1).

Yu, S. and Lee, S. (2017). Ultraviolet radiation: DNA damage, repair, and human disorders. *Molecular & Cellular Toxicology*, 13(1), pp.21-28.

Zhang, Z. and Begg, C. (1994). Is *Trichomonas vaginalis* a Cause of Cervical Neoplasia? Results from a Combined Analysis of 24 Studies. *International Journal of Epidemiology*, 23(4), pp.682-690.

Zhaxybayeva, O. and Doolittle, W. (2011). Lateral gene transfer. *Current Biology*, 21(7), pp.R242-R246.



Appendix 1: pET28a+ Map : Image taken from the Novagen pET28a+ vector map accessed at (merkmillipore.com)