Characterizing a novel endonuclease in *Trypanosoma brucei*

Samuel E Shelley

BSc Biochemistry

This thesis is submitted for the degree of MSc (by Research)

Biomedical Science

March 2020

Faculty of Health and Medicine

Declaration

This thesis has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated.

CONTENTS

TABLES AND FIGURES

INDEX OF TABLES

INDEX OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Dr Fiona Benson for their guidance, support and encouragement throughout this project , and for affording me the opportunity to pursue this research project under her supervision. I would also like to thank Dr Paul McKean for their help with this project as my secondary supervisor.

Additional thanks to Rebecca Overthrow, Dr Elisabeth Shaw, Dr Jayde Whittingham-Dowd.

ABSTRACT

Characterising a Novel Endonuclease in *Trypanosoma brucei* – Samuel E Shelley, BSc Biochemistry, September 2019

This Thesis is submitted for the degree of Master of Biomedicine

In 2012 the World Health Organisation published their Neglected Tropical Diseases Roadmap outlining 17 diseases which cause a significant burden across the world, namely in those areas stricken by poverty. The evidence outlined therein explained how with greater surveillance and management of these diseases, as well as increasing access to healthcare in the affect areas, should see these diseases brought under control, if not eradicated completely. One such disease is Human African Trypanosomiasis - or Sleeping Sickness – a parasitic disease caused by *Trypanosoma brucei*. Causing fevers joint pains and rashes at first, the disease progresses to cause a variety of neurological symptoms, before leading to a coma and eventually death. This parasite is able to evade the immune system by expressing an interchangeable coat of Variable Surface Glycoproteins coded by a vast library of different genes, allowing members of the parasite burden to escape immune detection long enough to be passed on to the next patient. To switch between these different genes, *T. brucei* relies on a series of different mechanisms, several of them underpinned by the process of homologous recombination. This report aims to explore the burden caused by this parasitic disease, the biology of the parasite itself and the mechanisms that underpin its pathogenicity. Furthermore, with previous studies having identified an enzyme with homology to the human FEN1 protein (an endonuclease responsible for cleaving the 5' flaps generated during DNA replication and repair), this report aims to examine the structure and function of this putative *Tb*GEN1 in homologous recombination and VSG-switching, and explore how this may help combat this disease. The research carried out for this report found that the putative *Tb*GEN1 showed minimal Holliday junction cleavage activity but was far more effective at cleaving flap junctions. Through site-directed mutagenesis, a total of 9 substitution mutants were identified that lost all 5'-flap cleavage activity, with 7 specific residues (D34, D90, E164, D183, D185, G235, and D237) being implicated as key to *Tb*FEN1's activity. Furthermore, localisation studies carried out *in vivo* in procyclic trypanosomes would support this hypothesis, as the fluorescent tagged protein was not observed outside the nucleus, although this also raised further questions about the exact role of the target protein. The findings of this report demonstrate that there is still much we do not know about the fine workings of this parasite's biology, but that with further research the possibility of finding a druggable target in *T. brucei* is very real.

1. LITERATURE REVIEW:

1.1. INTRODUCTION

As healthcare and medical research has improved in western More Economically Developed Countries (MEDCs) to the point where many life-threatening infectious diseases are controlled if not eradicated, organisations such as the World Health Organisation (WHO) and the Centres for Disease Control & Prevention (CDC) have directed researchers' attention to combating health concerns in the wider world, particularly Neglected Tropical Diseases. One specific example, and the focus of this study, is Human African trypanosomiasis (HAT), or African Sleeping Sickness, a tropical disease caused by the obligate parasitic protist *Trypanosoma brucei*. This review will cover the biology of trypanosomatids, the symptoms and epidemiology of African Sleeping Sickness, and how these parasites evade immune destruction, specifically via antigenic variation and the mechanisms that underpin this.

1.2. AFRICAN SLEEPING SICKNESS

1.2.1. CAUSES

Human African Trypanosomiasis (HAT), is a parasitic disease prevalent in Sub-Saharan Africa, where it is transmitted by the tsetse fly vector. The disease is caused by a blood-borne parasite, *Trypanosoma brucei*, which is found in two forms, typically localised to East and West Africa, roughly separated by the Great Rift Valley. The first, the slow progressing and chronic form, is caused by the subspecies *Trypanosoma brucei gambiense* and is localised to Central and West Africa. It is endemic in a total of 24 countries and accounts for around 98% of reported cases. This form exhibits anthroponotic transmission and can last up to 3 years. The second form, which exhibits a much faster – acute – progression, is caused by *Trypanosoma brucei rhodesiense* and is localised instead to East and Southern Africa. Responsible for a far smaller proportion of the disease prevalence (only 2% of reported cases), it is found in only 13 countries and is transmitted zoonotically. By comparison, patients with this form normally only live for a few weeks or months after contracting the parasite (Checchi, et al., 2008).

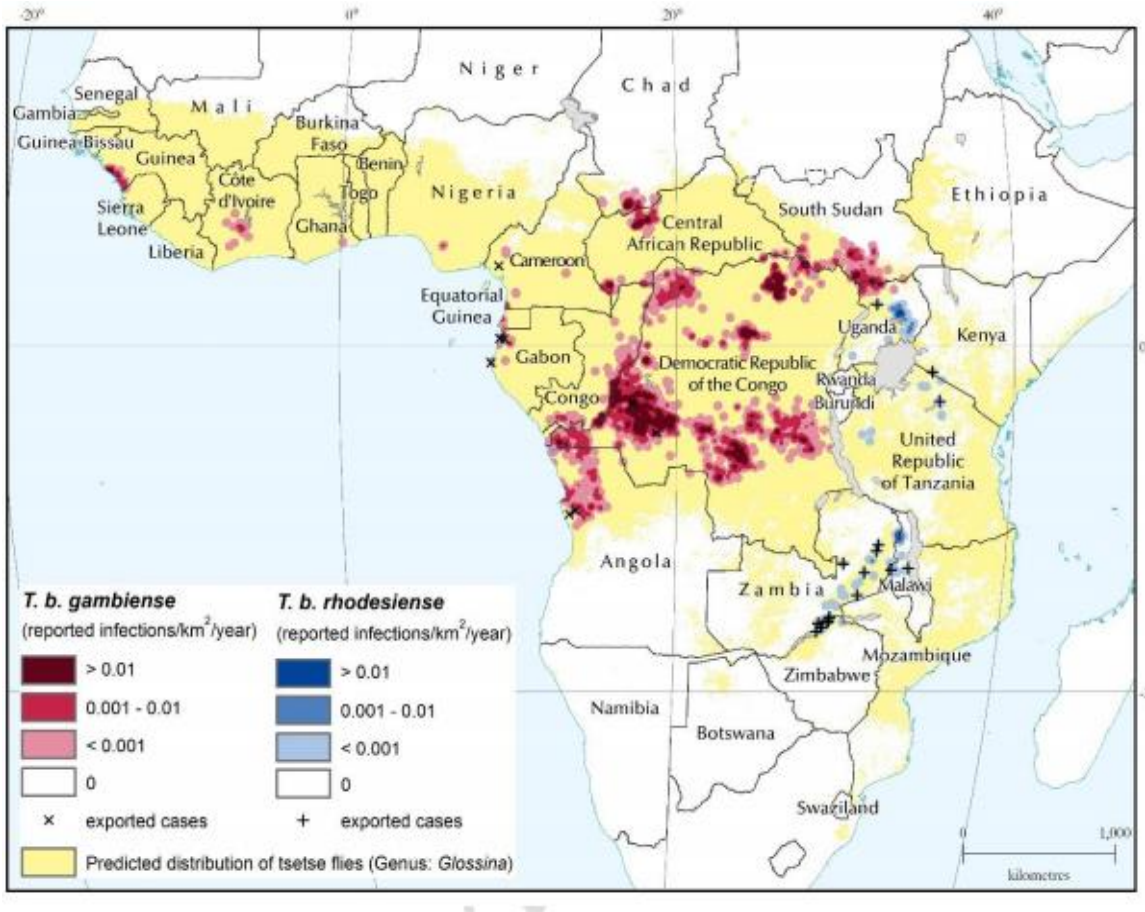


FIGURE 1.2-1 – GEOGRAPHIC DISTRIBUTION OF HUMAN AFRICAN TRYPANOSOMIASIS INFECTIONS REPORTED BETWEEN 2010-2014

T. B. GAMBIENSE **IS LOCALISED TO CENTRAL AND WEST AFRICA AND IS THE CAUSATIVE AGENT OF THE CHRONIC FORM OF HAT.** ***T. B. RHODESIENSE*** **IS LOCALISED TO EAST AND SOUTHERN AFRICA AND IS THE CAUSATIVE AGENT OF THE ACUTE FORM OF HAT. FIGURES FOR REPORTED INFECTIONS ARE FROM THE WHO ATLAS OF HAT. 'EXPORTED CASES' ARE PLOTTED AT THEIR ESTIMATED PLACE OF INFECTION. THE PREDICTED DISTRIBUTION OF TSETSE FLIES IS FROM THE PROGRAMME AGAINST AFRICAN TRYPANOSOMIASIS.**

ADAPTED FROM (BÜSHER, ET AL., 2017)

1.2.2. SYMPTOMS

Both forms of HAT progress through two distinct stages as the parasites move throughout the body. The early stage, referred to as the haemolymphatic stage, is characterised by *T. brucei* being present in the bloodstream and lymphatic system. The first symptom at this stage is a chancre, or painless ulcer, at the site of the tsetse fly bite. While this symptom can be seen in both forms, it is most observed in patients infected with *T. brucei rhodesiense*, while it is comparatively rare in *T. brucei gambiense* infections. Far more common in this form is the swelling of lymph nodes as the parasite migrates from the blood stream into the lymphatic system, especially along the back of the neck – posterior cervical lymphadenopathy. Other symptoms of this early stage are high fevers, swelling of the spleen, anaemia and inflammation of the heart tissue or even cardiac failure (Kennedy, 2004).

As the disease progresses into the late stage, or meningoencephalitic stage, and the parasites begin to break down the blood-brain barrier (BBB) and cross into the central nervous system, the pattern of symptoms changes. Firstly, patients begin to develop anorexia, lassitude and a number of minor neurological symptoms, e.g. tremors, partial limb paralysis, speech disorders, and fasciculations. As the disease progresses further a number of more severe symptoms may become apparent including hallucinations & delirium, hemiparesis, and akinesia. Most notably, and where the disease gets its name, is the dysregulation of the circadian rhythm leading to fragmentation of the sleep cycle. Finally patients fall into a coma, may contract concurrent infections due to a weakened immune system, and eventually die (Lundkvist, et al., 2004). While these symptoms vary a little between the two forms, as stated it is mainly the rate of progression that differentiates the two (MD, 2013).

1.2.3. EPIDEMIOLOGY

As previously stated, *T. brucei* parasites are transmitted by the bite of the tsetse fly, a blood sucking insect found throughout tropical Africa. Before the 19th century tsetse flies were comparatively rare in Sub-Saharan Africa, as they require extensive vegetation to reproduce, and the cattle that the myriad African tribes farmed would prevent grass and seedlings growing too high due to their grazing – effectively limiting tsetse population growth. However, in 1887 Italian explorers in Eritrea accidentally introduced the cattle virus rinderpest to Africa, which swept through the continent's ox and cattle at an extreme rate, and by 1898 an estimated 5.5 million cattle had died from the disease. Not only did this devastate the continent's infrastructure so that it was easy for colonists to take control in the subsequent years, but with the cattle gone the pasture was soon overgrown into the 'wild bush' we know today – a breeding ground for the tsetse flies. Furthermore, after the rinderpest epidemic had ended the wild animal populations grew much faster than that of the pastoral animals, not only providing an animal host for the tsetse fly, but the increased incidence of HAT, humans and the cattle they kept were prevented from re-grazing the bush (Pearce, 2000).

Since then, sleeping sickness epidemics have been classically associated with breakdowns in civil infrastructure, such as those during the second world war and during the civil unrest from the 1970s to the late 1990s (Figure 1.2.3-1). This is likely due to a combination of factors including relaxation of control measures, reduced awareness, and breakdowns in infrastructure (hygiene, medical care, refugee camps). Indeed, in the early 20th century, colonists concerned with the problems posed by HAT set up a number of successful control measures, reducing the transmission to only 4,435 new cases by the mid-1960s. However, this led to the false belief that HAT was on the decline and slowly surveillance and control efforts were reduced, leading to the disease making a resurgence by the end of the 20th century (Franco, et al., 2014). This process of control, surveillance and relaxation has been repeated time and again, and as previously stated is often linked with political and civil strife (Ford, 2007).

This century, 53 million people are expected to be at risk of infection from gambiense HAT (period 2012-2016) – with between 7000-10,000 cases reported annually, and 4 million at risk from rhodesiense HAT (period 2012-2016) – with a few hundred new cases reported annually (Franco & Priotto, 2018) (Centers for Disease Control and Prevention, 2012). In recent years however the number of new cases has been falling, and between 2000 and 2012 the incidence dropped by 73%. This reduction in the number of new cases has continued since then, with only 2804 cases reported in 2015, 84% of which can be attributed to the Democratic Republic of the

Congo. The WHO tropical diseases roadmap has targeted HAT for elimination as a public health problem by 2020 (WHO, 2018).

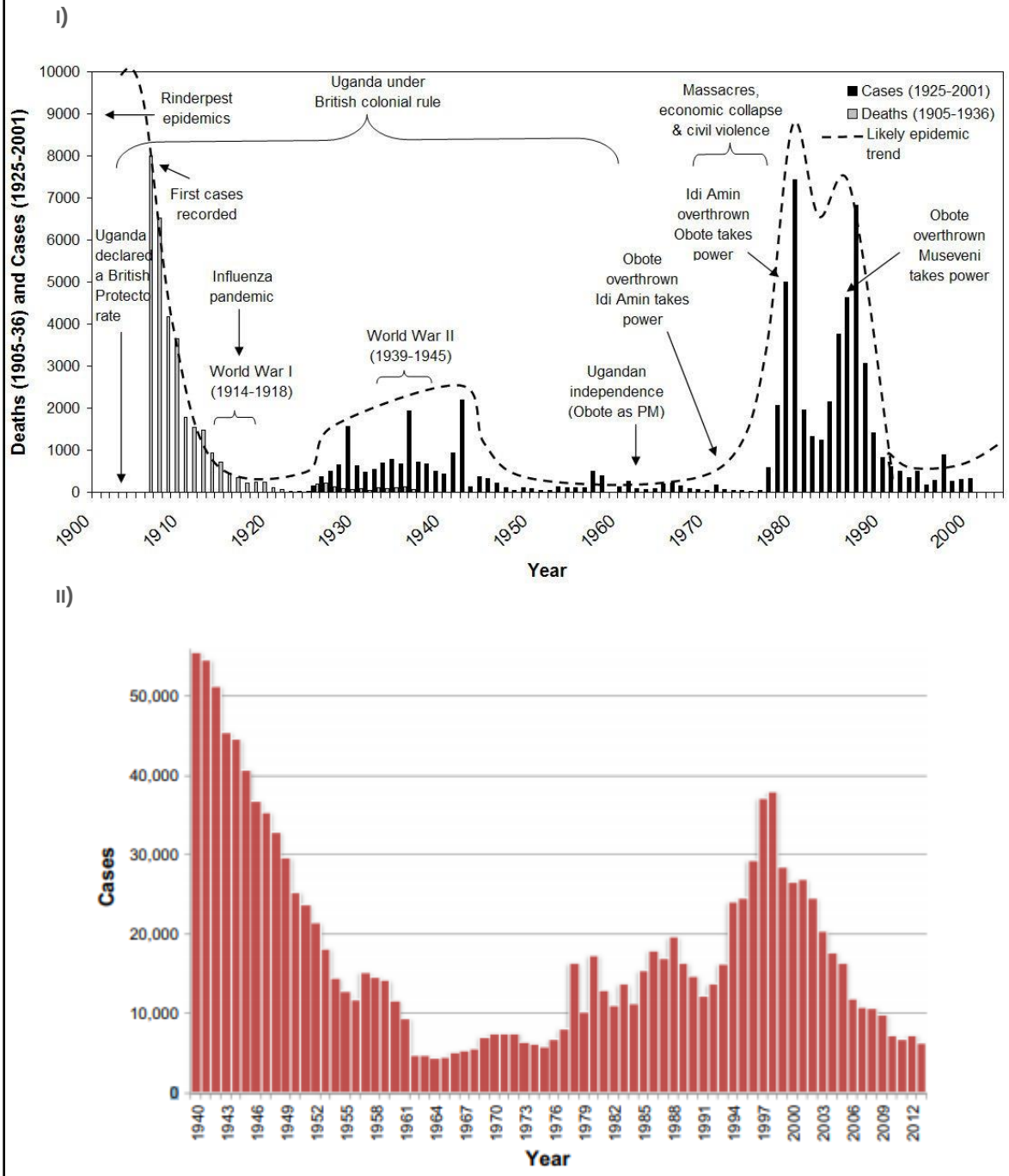


FIGURE 1.2-2 (I) SLEEPING SICKNESS EPIDEMICS AND MAJOR POLITICAL EVENTS IN UGANDA, 1905-2000

CASES FROM 1936 ONWARDS INCLUDE SOUTH-EASTERN UGANDA ONLY. ADAPTED FROM (FORD, 2007).

(II) TOTAL NUMBER OF NEW CASES OF HUMAN AFRICAN TRYPANOSOMIASIS REPORTED TO THE WORLD HEALTH ORGANISATION, 1940-2013

ADAPTED FROM (FRANCO, ET AL., 2014).

1.2.4. Control Measures

With *T. brucei* contributing to three major diseases (two human and one veterinary) in the developing world, major efforts are undertaken to control its spread and impact. In the past, a lack of coordination of such efforts and scientific understanding has led to repeated outbreaks and epidemics – such that these efforts were branded as a "failure of both science and public health" (Molyneux, et al., 2010). In more recent years, greater communication between scientific bodies and major funding from philanthropic bodies such as the Bill and Melinda Gates Foundation has finally made the possibility of eliminating neglected tropical diseases like HAT possible (Parker & Kingori, 2016). The zoonotic *T. B. rhodesiense* HAT has been the main focus of preventative measures, as its reliance upon animal reservoirs for tsetse transmission means that controlling either or both of the animal hosts exposure to the parasite or teste vector population has proved an effective, if complex, means to combat the disease's spread. The anthroponotic *T. b. gambiense* HAT prevention measures have instead largely relied upon diagnosis and treatment of the disease, and following plans of the WHO to eliminate gambiense HAT as a public health problem by 2020, the combined efforts of WHO, Sanofi-Aventis, Bayer and multiple nongovernmental organisations, have led to a reduction in the number of cases to only 1442 in 2017, and the goal to have sustainable elimination (zero cases) by 2030 (Aksoy, et al., 2017).

Due to the variant nature of *T. brucei*'s surface coat (covered in section 1.3.4.) the development of vaccines and other prophylactic treatments have been especially difficult, and thus, control has instead relied upon controlling the tsetse fly vector and the effective diagnosis and treatment of the disease within the population. Bayer and Aventis have however committed to manufacturing and freely providing the drugs necessary to treat HAT to the WHO, however with the high toxicity of current drugs, a correct and detailed diagnosis is paramount to effective treatment (Simarro, et al., 2012). Even diagnosis of HAT is costly and difficult however, as it relies upon the results of a lumbar puncture to determine the trypanosome subspecies and disease progression. Other molecular diagnostic tests do exist, but they have largely remained within research setting rather than being readily available to patients or governments for control of the disease. Serological diagnostic test are in development and could provide an affordable and effective method to test large proportions of the population, but further research is still needed in this field to increase the specificity and reliability of these devices (Matovu, et al., 2017) (Jamonneau, et al., 2015). On limiting transmission of the disease, the CDC advise steps to reduce the likelihood of tsetse fly bites. These include wearing medium-weight clothing that covers as much as the body as possible, and in neutral colours – as tsetse flies are attracted to high contrasting colours; avoiding bushes and other brush where the tsetse flies may reside; and using insect repellent and netting (CDC, 2012).

The lack of readily available means to test if patients have contracted HAT, combined with the logistics and expense (in training medical staff and in actual drug delivery) required to provide even free chemotherapy in the form of les toxic eflornithine, has meant that melarsoprol (which has a 5% drug induced death rate, as it causes reactive encephalopathy) has remained in widespread use (Burri, 2010). Even with the WHO providing kits containing the necessary materials for eflornithine administration, the slow-acting nature of the drug, complicated delivery and short-half life *in vivo* presents a problem of incomplete compliance to the full course of treatment and thus growing trypanosome resistance to the drug. Between 2003 and 2008, a huge clinical trial was carried out to determine if a combination therapy of eflornithine and nifurtimox may help combat some of these issues, and it concluded that the combination treatment was of a similar safety and efficacy to treatment with eflornithine alone, but with a reduced dose and treatment time. WHO once more developed medical kits to allow easy access to this combined treatment, with the two drugs provided again by Bayer and Aventis, and this

has proven to be an effective treatment against second stage gambiense HAT (Simarro, et al., 2012).

The drugs against the first stage of trypanosomiasis (Suramin for both forms of the disease, and Pentamidine for gambiense HAT) have been in use for the last 70 years and thankfully have raised little to no safety or efficacy concerns. For second stage rhodesiense HAT however, melarsoprol continues to be the only effective drug treatment. Thus, with growing resistance concerns and safety issues surrounding anti-trypanosomal drugs, there is still a great need to develop treatments for this collection of diseases.

1.3. *TRYPANOSOMA BRUCEI*

1.3.1. WHAT ARE THEY?

Trypanosomatids are a family of single-celled obligate parasites belonging to the Kinetoplastea taxonomic class. They are distinct from other protists, in their possession of an organelle referred to as a "kinetoplast", which contains the mitochondrial genome of these organisms. There are three different trypanosomatids that cause significant disease in humans: various species of *Leishmania*, which are transmitted by sand flies and cause leishmaniasis; *Trypanosoma cruzi* which is transmitted by triatomine bugs in the Americas and causes Chagas disease; and *Trypanosoma brucei* which is transmitted by tsetse flies in Africa and causes HAT. As the causative agent of HAT, this review focuses on *Trypanosoma brucei*.

Trypanosoma brucei is a heteroxenous blood-borne parasite, spending part of its life cycle in the tsetse fly vector and the other in the mammalian host. This single celled parasite exhibits a highly polarised cell architecture, with the vast majority of single organelles concentrated towards the posterior end of the cell. The cytoskeleton is all arranged uniformly, with the majority of microtubules in the subpellicular corset running from the anterior to the posterior of the cell (- to + ends). At the very rear of the cell is the flagellar pocket, an invagination at which the parasite's flagellum exits the cell while also acting as the sole site for endo and exocytosis, and particularly VSG recycling and coat clearance as a result. This single motile axonemal flagellum is responsible for cell movement and adhesion, and is associated with a dense bundle of filament proteins known as the paraflagellar rod, which has been implicated in flagellar beat frequency and strength (Portman & Gull, 2010).

1.3.2. LIFE CYCLE

Infection commences when a tsetse fly carrying the parasite in its salivary gland bites the patient to take a blood meal, injecting some of its saliva into the patient to prevent coagulation. At this point *T. brucei* is in the metacyclic trypomastigote stage of its life cycle, such that the parasites have already begun to express a surface coat of variable surface glycoprotein (VSG), and are unable to reproduce. Upon entering the blood stream, cells begin to transform into their bloodstream trypomastigote form, growing in length and gaining the ability to proliferate by binary fission, producing short stumpy daughter cells (Matthews, et al., 2004). While in long slender forms, the parasite continues to evade the host immune system by a combination of surface coat clearance and VSG switching – detailed below. This form is also able to burrow through the thin endothelium of the hosts blood vessels, gaining access to the lymphatic and central nervous systems (Langousis & Hill, 2014). The cells reproduce by binary fission, producing a new cell that is a clone of the original.

Eventually, a tsetse fly bites the host organism and takes a blood meal, taking up the short stumpy bloodstream trypomastigotes into its midgut. Here *T. brucei* again transform, this time into procyclic trypomastigotes. Their VSG coat is replaced by a dense coat of procyclins, a number of proteins with high numbers of glutamic acid and proline repeat units (EP repeats) and pentapeptide glycine, proline, glutamic acid, threonine repeats (GPEET repeats) at the C-terminus. While proteolytic enzymes within the fly's midgut can breakdown the N-terminus, these repeats at the protein's C-terminus, along with the heavily glycosylated GPI-anchor, are resistant to protease attack, and likely act as a protective barrier for other *T. brucei* proteins.

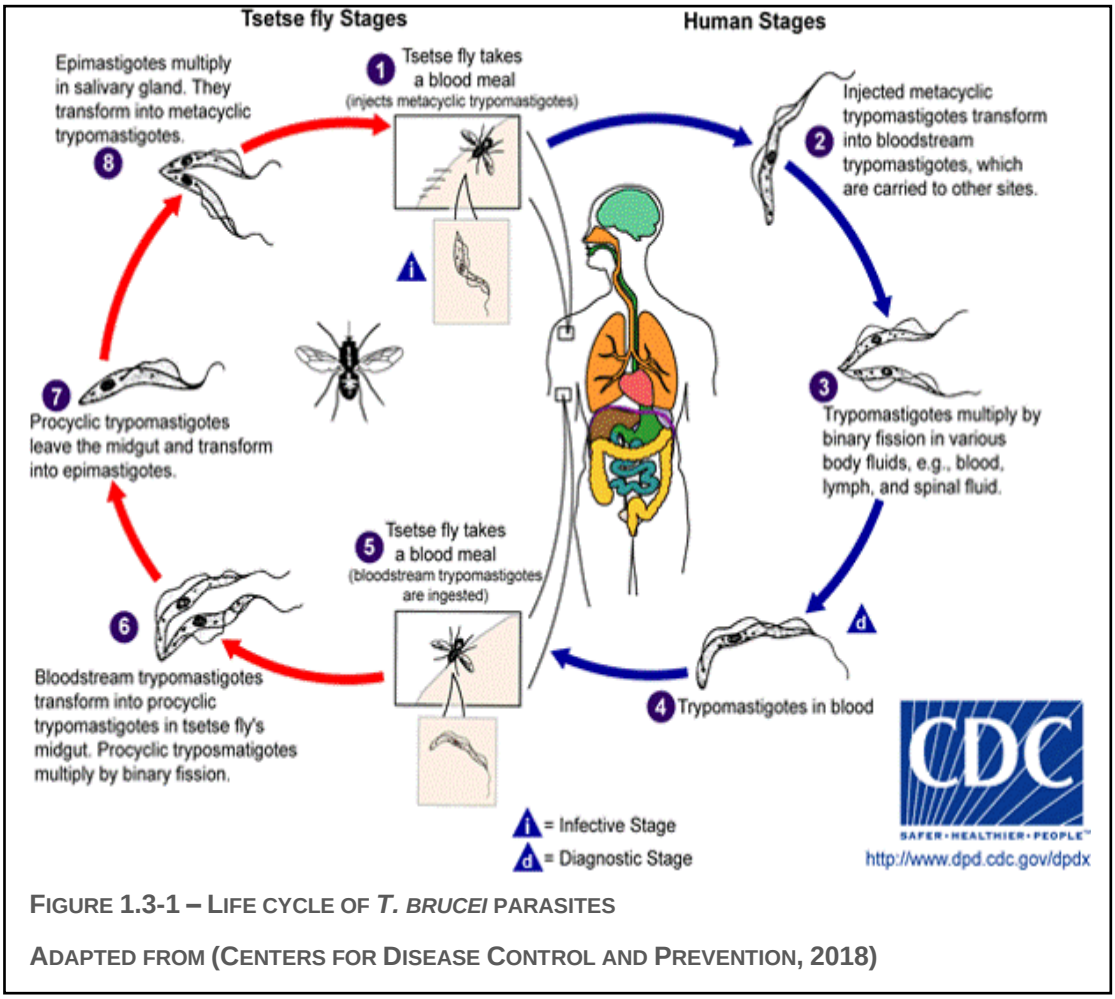


FIGURE 1.3-1 – LIFE CYCLE OF *T. BRUCEI* PARASITES

ADAPTED FROM (CENTERS FOR DISEASE CONTROL AND PREVENTION, 2018)

While in the procyclic form in the tsetse midgut, *T. brucei* can continue to multiply by binary fission (Vickerman, 1985).

As the *T. brucei* cells migrate towards the salivary gland they transform from trypomastigotes to epimastigotes, such that the kinetoplast and basal body have migrated past the nucleus towards the anterior of the cell, and the flagellum begins in the middle of the cell, projecting out in front of it. While in the salivary gland, the trypanosome's flagellum fulfils a secondary function, by allowing the parasite to attach to the fly's epithelium while they continue to proliferate. Finally, the attached epimastigote transform into metacyclic trypomastigotes which live freely within the salivary gland, ready to infect another animal when the tsetse fly next takes a blood meal (Sharm, et al., 2008) (Langousis & Hill, 2014).

1.3.3. Cell Cycle

In both the human and tsetse fly host, *T. brucei* parasites must multiply in number to reach a sufficient parasitaemia to guarantee a continued cycle of infection. As mentioned above, trypanosomes rely on a form of cell division known as binary fission (figure 1.3.3-1). Firstly, the basal body duplicates and a new flagellum begins to form, before the mitochondrial DNA within the kinetoplast is copied and the kinetoplast also divides, allowing the two basal bodies to separate. The nuclear DNA then also undergoes duplication prior to nuclear division, while the new flagellum grows alongside the existing one. Finally the nucleus divides in two, before the new daughter cell splits from the older cell by cytokinesis, starting at the anterior of the cell (Hammarton, 2007).

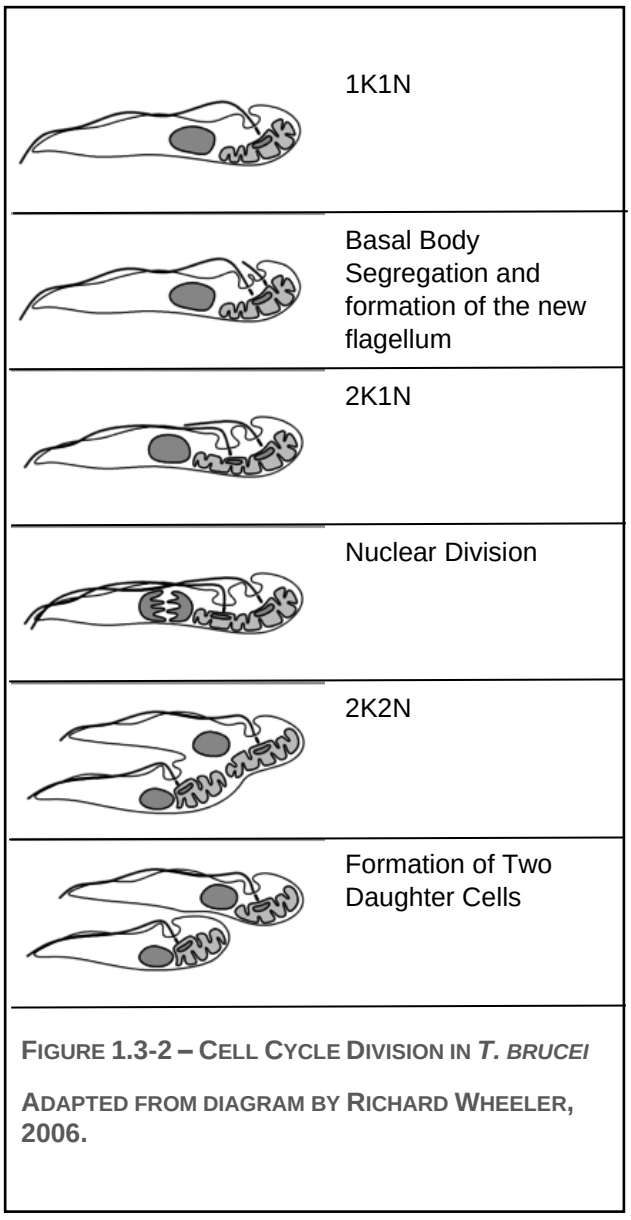


Figure 1.3-2 – Cell Cycle Division in *T. brucei*

Adapted from diagram by Richard Wheeler, 2006.

1.3.4. IMMUNE EVASION

In order to survive within their animal hosts, parasites like *T. brucei* must be able to evade their host's immune system. Many blood borne parasites evade immune detection and destruction by hiding within their host cells or in immune privileged tissues, such as the central nervous system (CNS) or the eye. Others actively dysregulate the host's immune response, secreting molecules that may mimic normal host molecules to misdirect or hinder the immune response (Schmid-Hempel, 2009). *T. brucei* cells however, are free-living within the host's bloodstream and tissues, and so must evade destruction by the immune system another way. To do so, *T. brucei* is covered in a dense coat of Variant Surface glycoproteins (VSGs) that is central to parasite immune evasion. Trypanosomes possess two core mechanisms that allow the parasite population to persist long enough within the host to be re-ingested by a new tsetse fly and continue the cycle of infection: surface coat clearance; and VSG switching. Both these mechanisms rely heavily upon the dense VSG coat to protect the cell from immune destruction, but they utilise it in different ways.

1.4. VARIABLE SURFACE GLYCOPROTEINS

1.4.1. STRUCTURE

Variable surface glycoproteins cover the entirety of the bloodstream form *T. brucei* cell, with approximately 10^7 proteins on each cell resulting in an electron dense shield across the surface membrane. Each VSG projects around 12-15 nm out from the cell's phospholipid bilayer where they are attached by a glycosylphosphatidylinositol (GPI) anchor on their C-terminus. Each VSG is composed of two identical dimerised subunits, each around 60 kDa in mass. They can be divided into two clear domains, the larger N-terminal domain (350-400 amino acid (aa) residues) which is highly variable and includes a 20 aa signal sequence; and the smaller C-terminal domains (20-40 residues), of which there are often more than one and are responsible for GPI-anchor interaction. The two domains are joined by flexible regions that act as linkers (Bartossek, et al., 2017) (Hutchinson, et al., 2003).

Although the sequence of amino acids can vary hugely between different VSGs, with the variable N-terminal domains sharing approximately 13-30% sequence homology, there is a conserved tertiary structure (Figure 1.4.1-1) (Schwede, et al., 2015). There are consistently 2 antiparallel α-helices present – linked by a short turn – which are then surrounded by a number of other secondary structures including 7 shorter α-helices and a 3-stranded β-pleated sheet, joined by a number of short loops and a longer loop towards the top of the molecule. So conserved is this structure that despite the vast differences in different VSGs sequences, molecules have been shown to be approximately 60% superimposable (Blum, et al., 1993).

Within both the N and C-terminal domains there are a number of conserved cysteine residues – which are involved in forming disulphide bridges between the monomers. These residues have been shown to occur in distinct patterns across different VSGS, 3 alternative patterns in the N-terminal domain and 6 in the C-terminal domain, which are referred to as Types A-C and 1-4 respectively. In the C-terminal domain, types 2, 4 & 5 all contain only 4 cysteine residues, while 1, 3 & 6 contain 8 cysteines, presumably arranged into 2 subdomains of 4 residues each. Meanwhile in the N-terminal domain, Type A domains have 4 residues, Type B have 8, and Type C have 6 (Carrington, et al., 1991).

Associated with the base of the N-terminal domain is an N-linked oligosaccharide, which is likely involved in folding efficiency – acting as a substrate for the unfolded glycoprotein glucosyltransferase enzyme, which adds glucose to a terminal mannose residue on unfolded proteins, targeting them for correct folding prior to release from the endoplasmic reticulum (ER). This ensures that only correctly folded VSGs are expressed on the cell surface, presumably to ensure homogeneity across the entire surface. Some larger oligosaccharides may too be involved in extra shielding for the invariant proteins beneath the VSG canopy (Schwede, et al., 2015).

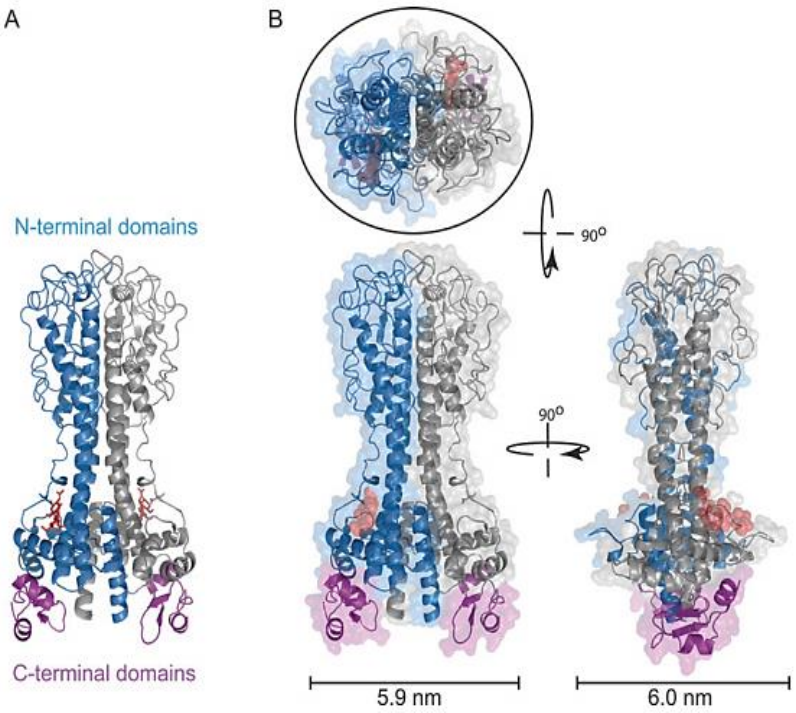


FIGURE 1.4-1 (A) 3-DIMENSIONAL STRUCTURE OF A VSG221 ILLUSTRATING THE CONSERVED STRUCTURE OF A VSG PROTEIN.

A RIBBON DIAGRAM SHOWING THE VARIOUS SECONDARY STRUCTURES AND OVERALL TERTIARY STRUCTURE OF VSG221. ONE N-TERMINAL DOMAIN MONOMER IS SHOWN IN BLUE, AND THE OTHER IN GREY. C TERMINAL DOMAINS ARE SHOWN IN PURPLE, AND N-LINKED OLIGOSACCHARIDES IN RED.

(B) SPACE-FILLING MODEL OF VSG221 IN 3 PLANES: X, Y, AND Z.

THE PREVIOUS RIBBON DIAGRAM IS SUPERIMPOSED OVER A SPACE-FILLING MODEL TO ILLUSTRATE THE SIZE AND SHAPE OF THE VSG221 PROTEIN, AND THE MODEL ROTATED 90° TO BE VIEWED IN ALL 3 PLANES. WIDTH MEASUREMENTS ARE DISPLAYED BELOW, AND AN ELLIPSE REPRESENTING APPROXIMATELY 28 Å^2 IS SHOWN OVER THE TOP-DOWN VIEW.

ADAPTED FROM (SCHWEDE, ET AL., 2015).

1.4.2. FUNCTION

Projecting approximately 14 nm from the cell surface, VSG proteins form a canopy across the entire membrane, hiding the majority of invariant cellular membrane proteins beneath from immune detection, such that the only valid target is the VSG. Furthermore, this dense coat of proteins also prevents the formation of the complement membrane attack complex, a collection of proteins that are assembled to form a pore in pathogen's cell membranes and thus disrupt the osmotic balance of the cell, causing cell lysis.

Some membrane proteins, including those translated from expression site-associated genes (ESAGs) and invariant surface glycoproteins (ISGs) may occupy a greater 3D space than the VSGs within the membrane topology and thus extend outside the VSG coat, meaning they could still be detected by the immune system. Equally, IgG has been shown to have some ability to penetrate the VSG coat, although it is mostly only as far as the base of the N-terminal domain (the widest part of the VSG) due to the density of VSG packing upon the cell surface. In fact VSGs each take up around 28 nm^2, and due to their high copy number only have between 28 nm^2 and 35 nm^2 available to them, such that at most there can only be approximately 7 nm^2 unoccupied space between them (Schwede, et al., 2015). Thus, due to the sheer ubiquity of VSGs, these invariant proteins are less readily detected by the immune system. Even when some are recognised, the antigen-antibody complex would then be subject to the same hydrodynamic flow as recognised VSGs, and thus also removed by surface coat clearance, which is explained below (Engstler, et al., 2007) (Mugnier, et al., 2016).

As the host immune system recognises the foreign VSG protein, the antigen is presented to the body's lymphocytes by various antigen presenting cells (such as dendritic cells and macrophages). These lymphocytes in turn seek out the targeted antigen for immune destruction. B lymphocytes in particular raise antibodies (Ab) against the target foreign epitope allowing for complement formation or destruction by macrophage cells. In order to escape this fate, the VSGs and other surface membrane proteins are continuously recycled through the trypanosomes' flagellar pocket, along with any bound Ab complexes (Rudenko, 2011). As the cell swims through the mammalian bloodstream, the hydrodynamic forces acting upon VSG-Ab complex causes it to move backwards along the cell surface, towards the flagellar pocket where it can be internalised by endocytosis and the antibody destroyed. It has been proposed that the rapid rate of antibody clearance is due to the complexes acting as molecular sails, as they project out from the sheer VSG coat, experiencing a higher drag force as the parasites pass through narrow blood vessels, with these increased hydrodynamic forces sweeping the complex backwards along the plasma membrane and recycling the entire membrane roughly every 12 minutes (Engstler, et al., 2007)

Not only does the VSG coat shield cell surface proteins from immune recognition, but VSG fragments released by dying trypanosomes may act to downregulate or misdirect the immune system as previously hidden epitopes are exposed upon release and denaturing of VSG dimers. These new epitopes are recognised by antigen presenting cells, directing lymphocytes to search for an epitope they will not find on live trypanosomes. Chemical messengers released as a result of the early infection and detection of these released VSG fragments can activate M2 suppression macrophages and actually downregulate the T-lymphocyte response against the VSG proteins (Mansfield & Paulnock, 2005).

While the concentration of anti-VSG antibodies in the hosts bloodstream and tissues remains low surface coat clearance is sufficient to prevent the *T. brucei* population from being destroyed by the immune system. However, as the infection persists and the immune response grows stronger, the concentration of antibodies eventually grows too great for coat clearance alone to be sufficient. This is where the second mode of immune evasion comes in: VSG Switching. While each individual *T. brucei* cell's VSGs are all identical, within their genome the parasites have a vast repertoire of VSG genes and pseudogenes, which can be randomly swapped out for the currently expressed gene, changing the phenotype and thus the detectable epitopes of that trypanosome. Thus, the immune system must once again go through the process of raising a new immune response of sufficient titre against this new VSG (Rudenko, 2011). The exact mechanisms of VSG switching are detailed in the next section.

1.5. ANTIGENIC VARIATION

1.5.1. VSG GENE EXPRESSION

Within *T. brucei*'s genome there are approximately 2000 VSG genes and pseudogenes, allowing for a huge cellular repertoire of VSG proteins. However, despite this tremendous number of distinct VSG-coding genes, individual parasites express only a single VSG at a time. VSGs can only be expressed from specific subtelomeric transcription units, which are highly polymorphic and found across several megabase chromosomes. These polycistronic sites are referred to as bloodstream expression sites (BESs), and alongside different VSG genes, they also contain different expression site-associated genes (ESAGs) which encode various proteins such as transferrin receptors that are transcribed alongside the VSG (Sima, et al., 2019). Transcription of these BESs by RNA Polymerase I (RNA Pol-1) produces a single polycistronic RNA molecule, before the individual genes have the capped splice leader trans-spliced ligated onto each gene to produce the final transcripts (Pays, 2005). Since the ESAG repertoire differs between different BESs, this allows *T. brucei* to better adapt to its environment by altering the proteins it expresses (Hertz-Fowler, et al., 2008). The number of BESs varies between different *T. brucei* subspecies and strains, but regardless of the exact number, parasites must express only a single VSG at a time for the VSG coat to prove effective. Allelic exclusion is assured by maintaining a discrete privileged Expression Site Body (ESB), outside the nucleolus, where RNA Pol-1 normally transcribes ribosomal RNA (Horn, 2014) (Navarro & Gull, 2001).

With such a huge collection of alternate VSG genes and pseudogenes, subsets of the *T. brucei* population within a patient are able to avoid immune clearance simply by expressing a different VSG protein to the one targeted by the immune system. When the host has generated a sufficiently robust immune response, characterised by a high antibody titre against the VSG molecule, the majority of the parasites in the population are destroyed. However, a small number of *T. brucei* cells within the population will have switched VSG expression and are unaffected by the host immune response, persisting within the host. This process is constantly occurring at random within the population and so at any one time a number of different VSGs are active within the population (Figure 1.5.1. – 1) (Mugnier, et al., 2016). The mechanisms by which *T. brucei* switches VSG expression are critical to understanding the immune evasion process and are explained in the next sections.

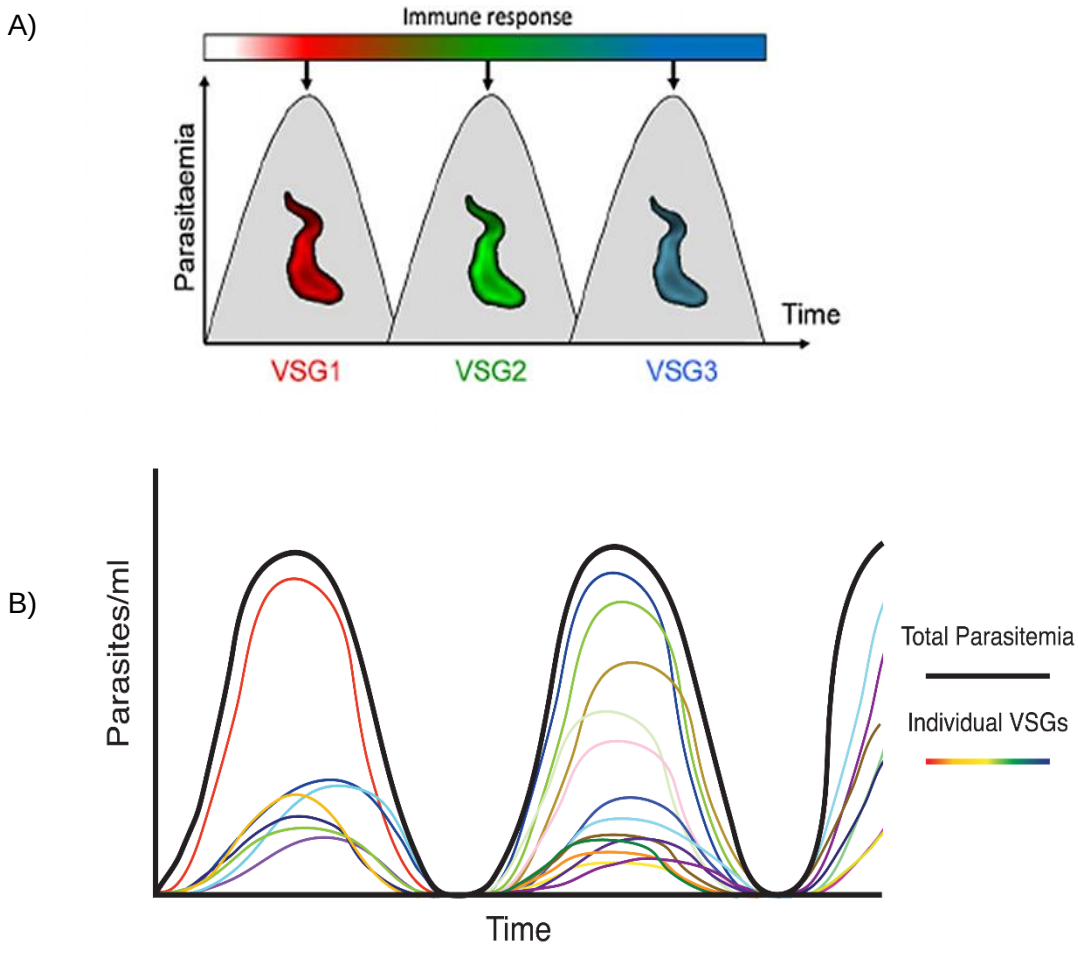


FIGURE 1.5-1 CHANGES IN VSG EXPRESSION ACROSS THE TRYPANOSOME POPULATION OVER THE COURSE OF INFECTION

A) GENERAL CHANGES IN PARASITAEMIA WITHIN THE HOST OVER THE COURSE OF INFECTION, AND EFFECT OF HOST IMMUNE RESPONSE. VSG PHENOTYPE SHOWN AS DIFFERENT COLOURS, WITH CORRESPONDING ANTI-VSG ANTIBODIES SHOWN ABOVE. AS THE ANTIBODY TITRE AGAINST THE RED VSG BECOMES MORE CONCENTRATED, THE PARASITAEMIA BEGINS TO PLATEAU AND THEN FALL AWAY. HOWEVER, WHEN VSG SWITCHING OCCURS AND THE NEW GREEN VSG PHENOTYPE ARISES, THE HOST ANTIBODY TITRE MUST REACH A SUFFICIENT TITRE ALL OVER AGAIN. (HORN, 2014).

B) INDIVIDUAL VSG EXPRESSION WITHIN THE TOTAL PARASITE POPULATION. WHILE THE FIRST GRAPH PROVIDES A GOOD SIMPLISTIC VIEW OF HOW VSG SWITCHING HELPS THE PARASITE OVERCOME THE CHALLENGES PRESENTED BY THE HOST IMMUNE SYSTEM, IT PRESENTS A FALSE IDEA THAT ONLY ONE NEW VSG PHENOTYPE ARISES AT A TIME. IN REALITY, EACH WAVE OF PARASITAEMIA IS COMPOSED OF HUNDREDS, IF NOT THOUSANDS DIFFERENT VSG PHENOTYPES. (MUGNIER, ET AL., 2016)

1.5.2. TRANSCRIPTIONAL SWITCHING

The first method of switching between the expression of different VSGs is that of transcriptional switching (or *in situ* switching), and it is perhaps the more straightforward method of VSG switching. Put simply, it is the termination of transcription of one BES and associated VSG and the activation of transcription at another site, and thus a new VSG. Since transcriptional switching not only allows a new VSG protein to be expressed but also a new set of ESAGs, this

allows *T. brucei* parasites greater adaptability to their environmental conditions. However, little to no multiallelic phenotypes are observed *in vitro*, so the expression of different BESs must be more tightly controlled than simply switching on and off expression at different sites. When these double-expression cells do arise, they appear to be highly unstable *in vitro*, occurring extremely rarely in culture ($\sim 10^{-7}$ per trypanosome). Furthermore, in the absence of selection pressures this double-expression phenotype is gradually lost (~12-15 cell doublings). Thus, mechanisms must act upon the transcriptional switching between different BESs to couple the activation of one site to the deactivation of another to allow long-term allelic exclusion (Chaves, et al., 1999).

While the availability of a single privileged expression domain might be sufficient to ensure allelic exclusion, the ability to temporarily house two expression sites simultaneously would undermine this hypothesis. Instead, evidence has shown that ES activation is more than likely controlled on an epigenetic level, through a number of mechanisms regulating RNA Pol-1 activity at different BESs, chromatin remodelling or telomere positioning. Molecules such as CIFTA (class I basal transcription initiation factor A) and SUMO proteins (small ubiquitin-like modifier) have both been demonstrated to play a role in RNA Pol-1 recruitment to the ES, while TDP1 (Trypanosome DNA binding protein 1) may be involved in maintaining an accessible chromatin structure at the active ES by histone depletion. Similarly, studies have shown that modifications to the chromatin structure to deplete an ES of nucleosomes will result in its activation, and that silent ESs have enriched histones H1, H2A, H3 or H3V (Maree & Patterton, 2014). A 6th nitrogenous base, β-D-Glucopyranosyloxymethyluracil (Base J), has been shown to be enriched at the telomeres and silent ESs, but was absent at active ESs, and has been shown to at least inhibit RNA Pol-2 transcription. Finally, many molecules have been identified that play a role in telomere location and integrity, and whose dysregulation (either through knockout or hyper-expression) can also affect either VSG switching or expression, e.g. Telomeric repeat binding factor (TRF), mini-chromosome maintenance-binding protein (MCM-BP), and nuclear peripheral protein-1 & 2 (NUP-1/-2) (Cestari & Stuart, 2018). Other molecules and mechanisms may be involved, but this is still being researched and is outside the scope of this report.

1.5.3. SWITCHING VIA RECOMBINATION PROCESSES

The remaining mechanisms for VSG switching all involve exchanging DNA from within the active BES with DNA from elsewhere in the trypanosome's genome, utilizing a process called homologous recombination. There are in fact three different mechanisms by which *T. brucei* can change VSG expression that involve homologous recombination: duplicative gene conversion, telomere exchange, and segmental gene conversion. This recombination of DNA from elsewhere in the genome is made possible by homology between the varied number of 70-bp repeats upstream of the donor and BESs, and the conserved 3' ends of VSG sequences (Taylor & Rudenko, 2006). Both telomere exchange and duplicative gene conversion introduce a different VSG gene into the active BES from elsewhere in the genome, while segmental gene conversion actually generates entirely novel VSG genes by combining a mosaic of DNA from pseudogenes throughout the genome.

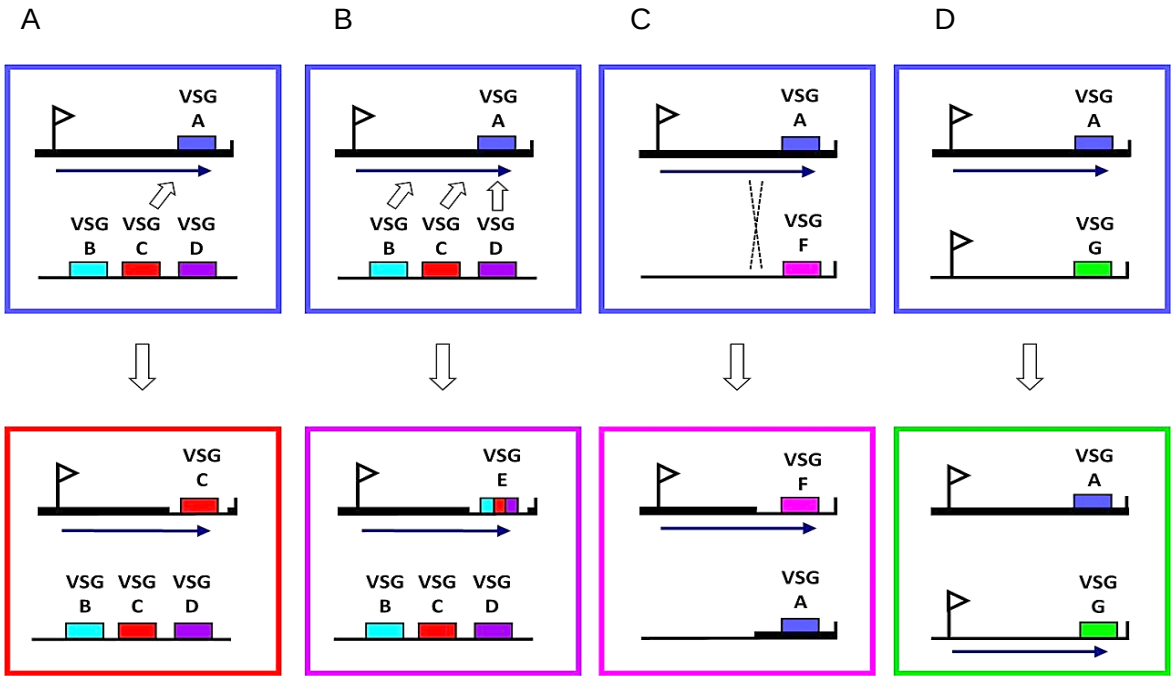


FIGURE 1.5-2 VSG SWITCHING MECHANISMS

MECHANISMS FOR VSG SWITCHING. EACH VSG GENE IS COLOUR CODED AND THE EXPRESSED VSG PHENOTYPE IS DEMONSTRATED BY THE COLOUR OF THE BOX SURROUNDING EACH SCENARIO.

A. DUPLICATIVE GENE CONVERSION
B. SEGMENTAL GENE CONVERSION
C. TELOMERE EXCHANGE
D. TRANSCRIPTIONAL SWITCHING

(VINK, ET AL., 2012)

The first means of introducing new VSG genes to the BES is telomere exchange. This involves exchanging the telomeric VSG gene from an inactive BES with the VSG gene in the active ES, via a cross-over event between two telomeres. This leaves the previous VSG gene intact and the new VSG gene transcribed along with any previous ESAGs. Duplicative gene conversion - by comparison to telomere exchange - is a destructive process, eliminating the previously active VSG gene from the genome as the new VSG gene is copied into the active BES from the vast libraries of VSG genes within chromosomal internal locations. While it may overwrite the previous VSG gene, this method of switching allows any potential VSG to be expressed from within the genome, rather than only those already present within ESs – vastly increasing the repertoire available to the *T. brucei* cell. Finally, segmental gene conversion uses the same mechanism as duplicative gene conversion, except that segments from a number of different genes and pseudogenes are recombined into a novel chimeric VSG gene. While this has the potential to create an entirely non-functional VSG, such cells would be quickly selected against by the immune system, meanwhile another novel VSG has the greatest chance of being entirely unrecognised by the host immune system (Rudenko, 2011). As discussed above, a key part of the puzzle for combatting *T. brucei's* ability to cause persistent infection lies in preventing its

continued evasion of the immune system, especially its ability to switch VSG coats via the mechanisms detailed here – and to do so, it is necessary to understand the process of homologous recombination.

1.6. HOMOLOGOUS RECOMBINATION

Although homologous recombination (HR) plays an important role in VSG switching, HR's role extends far beyond this specialised function and is an essential process operating in all organisms that acts to maintain genome stability and promote genetic diversity. The first model proposed to describe the mechanism of homologous recombination became known as the Holliday model as it was proposed by Robin Holliday in 1964, to account for discrepancies in the expected products of meiosis in fungi. A key intermediate in the Holliday model is the Holliday junction, a structure in which the DNA molecules participating in recombination are linked by a pair of exchanged DNA strands. Several models for mechanisms of homologous recombination have been developed since 1964, a key feature of these models is that recombination is initiated by the programmed introduction of a double-strand break (DSB). Many of them, most notably the double-strand break and repair model for homologous recombination, propose a key role for Holliday junction containing intermediates (Figure 1.6.1-1) (Haber, et al., 2004) (Szostak, et al., 1983).

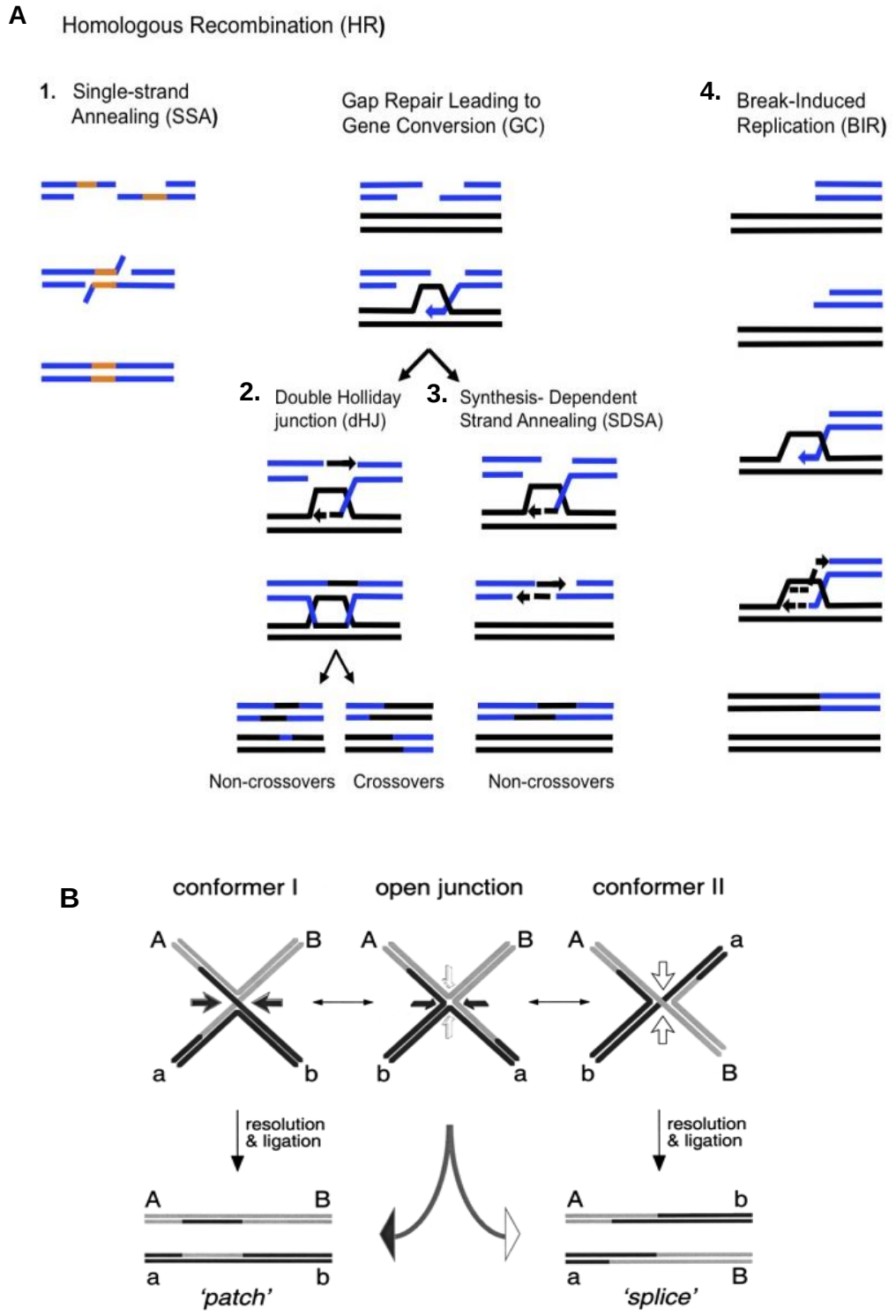


FIGURE 1.6.1-1 HOMOLOGOUS RECOMBINATION MECHANISMS AND HOLLIDAY JUNCTION RESOLUTION

A) DIAGRAM DEMONSTRATING THE PROCESS BY WHICH A DOUBLE STRAND BREAK CAN BE REPAIRED BY HOMOLOGOUS RECOMBINATION (SAKOFSKY, ET AL., 2012)

B) RESOLUTION OF DOUBLE HOLIDAY JUNCTIONS TO PRODUCE EITHER CROSSOVER OR NON-CROSSOVER PRODUCTS (VAN GOOL, ET AL., 1999)

1.6.1. VSG GENE STRUCTURE

To understand the role homologous recombination plays in VSG switching, it is key to first look at the structure of the VSG genes and the arrays they are located in. Of the thousands of VSG genes and pseudogenes encoded within the trypanosome's genome, most are located in subtelomeric gene arrays of *T. brucei's* 11 megabase chromosome pairs, while individual genes are located within about 30% of the one hundred or so minichromosome subtelomeres. In both the megabase chromosomes and the minichromosomes at least half of all VSG genes and pseudogenes are preceded by a 70 base pair repeat (which can be even longer in the expression sites). Furthermore, each VSG gene has an invariant motif of 14-bp (GATATATTTTAACA) in the untranslated region at the 3' end of the sequence, and both minichromosome and ES-linked genes have the downstream telomere repeats (Li, 2015). It is therefore presumed that these repeat sequences serve as areas of homology for recombination into the expression sites from elsewhere in the genome (Boothroyd, et al., 2009).

The first step in initiating this homologous recombination is the introduction of the DSBs that underpin this form of DNA repair. Firstly, it has been observed that the 70-bp repeated sequences that sit upstream of many VSG genes contain a large number of TTA repeats, which are known to cause instability during plasmid transcription and could form DNA motifs that are digested by a *T. brucei* endonuclease (Li, 2015) (Ohshima, et al., 1996). Studies using ligation-mediated PCR have in fact shown that DSBs occur naturally within the 70-bp repeat sequences during DNA replication, and with a frequency about 100 times greater than the rate of VSG-switching (Glover, et al., 2013). This would suggest that while there are plenty of DSB occurring that could result in homologous recombination induced VSG switching, not all DSBs cause VSG switching to occur. Indeed research by Glover, et al has shown that the precise location of DSB within the subtelomere can influence not only the probability of VSG switching but the mechanism by which it occurs (Glover, et al., 2013). Interestingly, work by Jehi, et al. into a novel *T. brucei* telomere protein has shed light on the regulation of VSG switching, as depletion the TRF-Interacting Factor 2 (*Tb*TIF2) protein they studied was shown to increase DSBs and VSG switching, and that together with *Tb*RAD51 may play a role in VSG-switching regulation (Jehi, et al., 2014). This would suggest therefore that the appearance of DSBs around the active ES VSG gene is a passive and random process, and it is only in the active repair of these double strand breaks by proteins such as *Tb*RAD51 and the potential *Tb*FEN1 discussed in this report, that the process of VSG switching is initiated. How this repair process leads to the introduction of a new VSG gene into the active ES is discussed below.

1.6.2. RECOMBINATION IN DNA REPAIR

Today, HR is recognised as referring to a number of different pathways that use homology between two different DNA molecules to repair damaged DNA. Recombination may be initiated at a DSB occurring as a result of damage, or at a DSB introduced enzymatically as part of a programmed recombination event. For example DSBs are introduced by Spo11 during meiosis to deliberately initiate recombination that ensures correct segregation of chromosomes and generates diversity in meiosis (Keeney, 2011). DNA damage occurs in all cells as a result of replication errors and exposure do both endogenous and exogeneous damaging agents. Endogenous DNA damage has been estimated to be as high as 70,000 lesions a day (Lindahl & Barnes, 2000). The repair of DNA damage is crucial to continued cell function, as the damaged DNA can block transcription; replication of the damaged DNA can lead to errors, which may be 'fixed' into harmful mutations upon a further round of replication; or replication of damaged DNA can convert damage affecting only one DNA strand into a DSB. It is estimated that DNA

damage affecting one strand is converted to a DSB at a rate of approximately 50 DSBs/cell/cell cycle (Vilenchik & Knudson, 2003). DSBs can be especially harmful if not repaired faithfully, as inaccurate repair may lead to major genome rearrangements.

1.6.3. Double Strand Break Repair and the Double Holliday Junction Pathway

Regardless of the exact mechanism, HR acts to repair the DSB at the point before the cell undergoes mitosis, immediately after DNA replication, such that the sister chromatids are readily available, as these are identical copies of each other. Before the DSB can be repaired, the molecular machinery involved must gain access to the damaged DNA, and thus unpack the target strand from the dense chromatin by remodelling its structure and modifying the histone proteins. In the case of *T. brucei* this is likely where a number of the molecules noted in Section 1.5.2. come into play, unpacking the chromatin at the site of both the current and replacement VSG, to allow HR to take place, however the exact mechanics of this remodelling are outside the scope of this review.

With the chromatin unpacked, the damaged DNA is exposed to the machinery necessary to repair it. In all recombination-based models for DSB repair the first step is to resect the DNA DSB, removing a small section of each 5' strand to form two 3' extended single stranded DNA tails. Research into human HR has shown that the MRN complex plays a pivotal role in end-resection, by binding DSBs, signalling for the cell cycle to pause while repair is carried out, identifying and selecting the required pathway and aiding the subsequent repair, e.g. tethering the broken sugar-phosphate backbones in place (Yuan & Chen, 2010). Work by Genois et al. has since identified MRN complex homologs in both trypanosomes and *Leishmania* that suggest this mechanism is conserved in trypanosomatids (Genois, et al., 2014). Following resection, the 3' single-stranded tail is bound by a recombinase (Rad51 in eukaryotes), which initiates homologous pairing and strand invasion by the 3' ss ends into an intact homologous DNA molecule. This creates a short region of heteroduplex DNA and displaces the complementary strand of the intact duplex into a D-loop (Bärtsch, et al., 2000) (Gupta, et al., 1997).

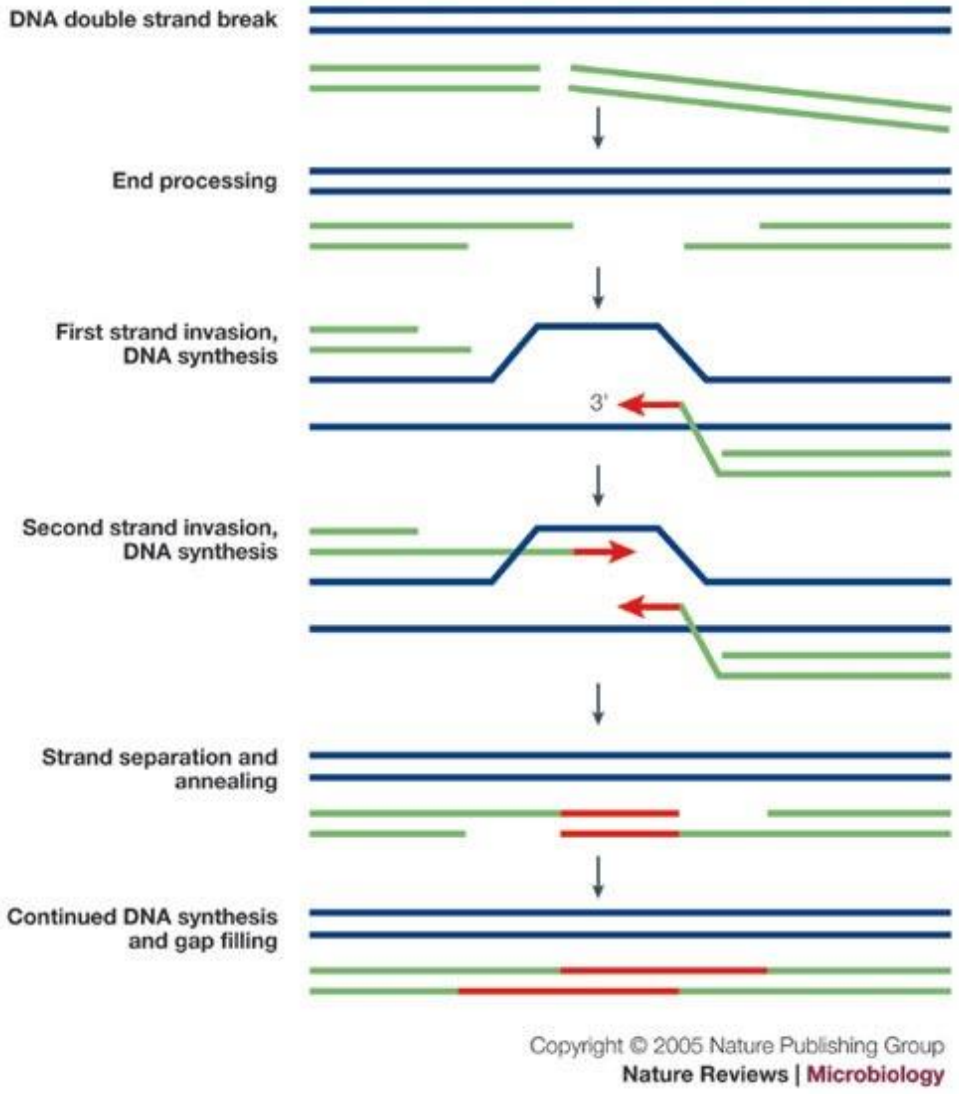


TABLE 1.6.3-1 MECHANISM OF SYNTHESIS DEPENDENT STRAND ANNEALING (COX & BATTISTA, 2005)

Following formation of heteroduplex, the invading DNA strand can prime new DNA synthesis, from its 3' end with new DNA being synthesised according to the sequence on the intact DNA molecule (Fig 1.6.3-1). The D-loop generated in the undamaged DNA molecule can then return to its original conformation of helical ds-DNA with the newly synthesised DNA annealing to the 3' end of the ssDNA tail originating from processing the damaged DNA on the other side of the initial DSB. This mechanism is termed synthesis-dependent strand annealing (SDSA) (As seen in mechanism 3 of Figure 1.6.1-1) (Cox & Battista, 2005) (McMahill, et al., 2007). Alternatively, continued DNA synthesis from the invading strand can move the D-loop along the intact DNA molecule such that the displaced DNA is able to anneal to the 3' ssDNA tail generated from the other side of the DSB gap. This generates an intermediate containing two Holliday junctions – referred to as a double Holliday Junction. Holliday junctions, as shown in figure 1.6.2-2, are X-shaped DNA confirmations that arise from 4 incoming dsDNA strands that share sequence homology, with a nearly square planar confirmation. From here, the remaining DNA is synthesised on the 3' ends to fill the gaps on both DNA molecules. Recombination intermediates are then resolved into products by introduction of a pair of nicks across each of

the Holliday junctions. Recombination is completed by ligating the remaining nicks to generate intact DNA molecules (Haber, et al., 2004) (Punatar, et al., 2017). In summary it is clear that Holliday Junctions are a central intermediate in homologous recombination, and potentially to the recombination mechanisms that ultimately allow *T. brucei* to substitute new VSG genes into the BES.

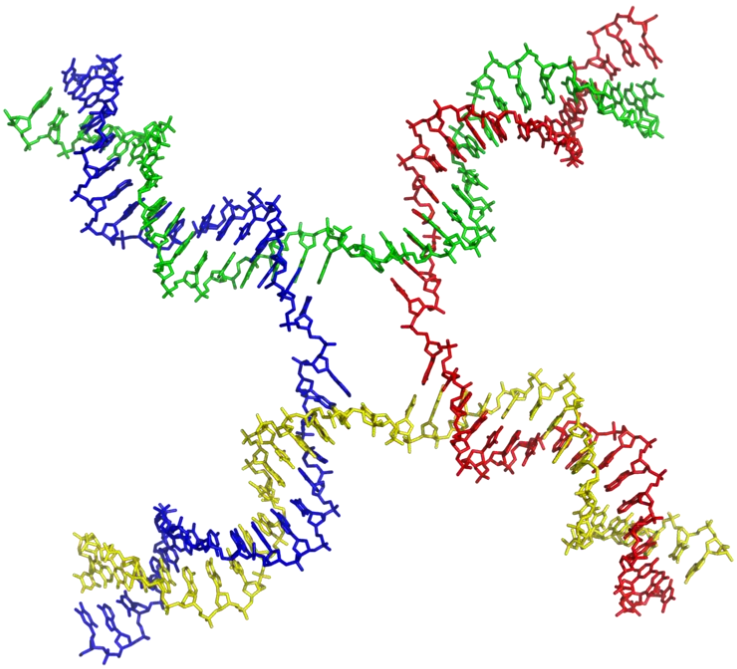


FIGURE 1.6.3-2 MOLECULAR STRUCTURE OF A HOLLIDAY JUNCTION ((ZEPHYRIS), 2007)

1.6.4. HOLLIDAY JUNCTIONS

As stated above, Holliday junctions (HJ) were first proposed by Robin Holliday in 1964, as part of his model for homologous recombination, to explain how apparently linked genes can end up segregated independently (Holliday, 1964). When two Holliday junctions arise together – as with the dHJ pathway – their resolution can either lead to non-crossover products (NCOs) or full crossover products (COs) (Figure 1.6.4-1). In the case of NCOs, the final DNA products remain ultimately unchanged from their original, undamaged state, with only a small 'patch' of DNA exchanged. COs, by comparison, can lead to the exchange of entire chromosome regions and while this is potentially harmful in somatic cells, it is key to the introduction of genetic variation during meiosis, as it creates new combinations of genes on the same chromosome (Heyer, 2004). If recombination occurs between two identical replicated sister chromatids (i.e. after replication) then regardless of whether NCOs or COs would be formed, there is no new genetic material exchanged. However, when COs occur between non-sister chromatids in mitotic cells (i.e. chromatids from replicated homologous chromosomes – one from each parent), then this can result in loss of heterozygosity depending upon how the new genetic material is segregated into daughter cells (Colavito, et al., 2010) (Swuec & Costa, 2014). This is crucial to eukaryotic cell survival, as loss of heterozygosity can decrease the individual's genetic variation and ability to survive future mutation events. But in *T. brucei*, the resolution

of dHJs to produce COs can provide a useful mechanism to introduce new VSG genes into the active BES, and thus are key to VSG switching. So how do they occur?

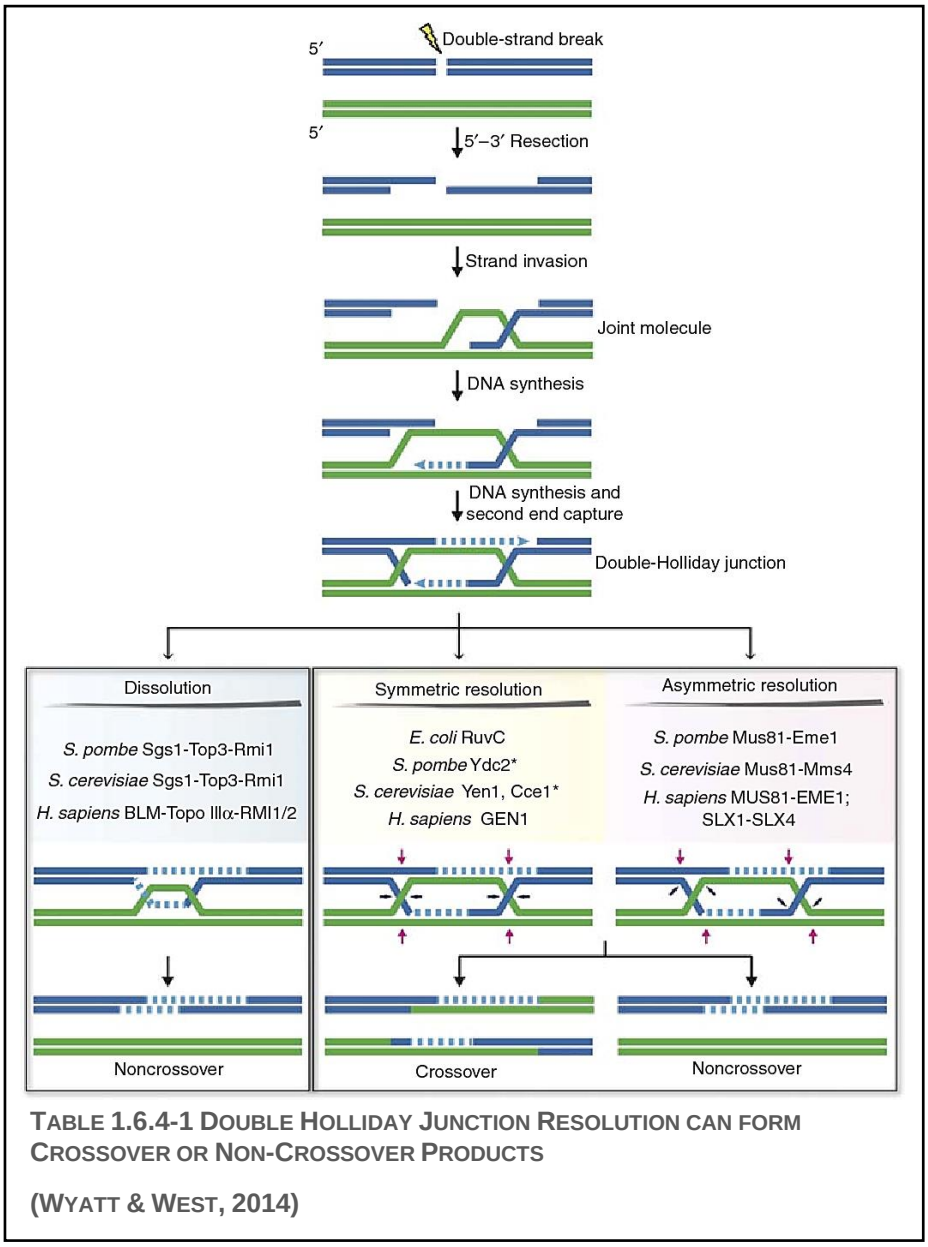


TABLE 1.6.4-1 DOUBLE HOLLIDAY JUNCTION RESOLUTION CAN FORM CROSSOVER OR NON-CROSSOVER PRODUCTS

(WYATT & WEST, 2014)

Double HJs can be processed in a further two ways after their formation during DSBR, either by dissolution or resolution. Dissolution will always lead to the critical NCOs and thus maintain heterozygosity between sister chromatids. It relies upon the BLM-TopoIIIα-RMI1-RMI2 (BTR) complex to catalyse the migration of the 2 HJs together to form a hemicatenane, before separating the linked DNA. The other method of processing dHJs is by resolution, and it is only through this path that cross-over events can occur – depending upon the conformation of the nicks made by the Holliday junction resolvase enzyme (Fig 1.6.4-1) (Wyatt & West, 2014) (Chan & West, 2015). These enzymes act to cleave 2 of the 4 involved strands with symmetrical nicks in either a horizontal or vertical axis, where cleavage in the same plane produces the 'patched' NCO events as it leaves the parent DNA strands covalently linked to one another. Cleavage in different planes (one cut in the horizontal plane and one vertical) meanwhile results in CO events, as the DNA from originally different strands is left covalently bonded together, and thus is ligated to form recombinant products (Heyer, 2004) (Oxford Academic, 2014). As

stated, it is the role of a number of HJ resolvase enzymes found throughout all domains of life, including viruses, to tightly control this process to ensure genetic material, and thus information, is safely preserved.

1.7. HOLLIDAY JUNCTION RESOLVASES

The family of enzymes responsible for cleaving Holliday Junctions to form 2 separate duplex DNA strands are collectively known as Holliday Junction Resolvases. As discussed above, they introduce symmetrical nicks in one of either 2 conformations to produce recombinant or non-recombinant products upon resolution. In the 1990s, work by a number of scientists first identified RuvC as a Holliday junction resolvase (Iwasaki, et al., 1989) (Connolly, et al., 1991) (Dunderdale, et al., 1991). RuvC is composed of 2 identical 19-kDa subunits which together form two DNA binding clefts, approximately 30 Å apart. Binding HJs then allows the 2 diametrically opposed nicks to be introduced into the bound DNA strands, which occurs preferentially (although not conditionally) at 5'-$^{A}/_{T}TT^{G}/_{C}$-3' sequences (Punatar, et al., 2017).

Subsequent to characterisation of prokaryotic HJ resolvases, eukaryotic HJ resolvase activity was simultaneously identified in yeast and in humans, associated with the Yen1 and GEN1 proteins respectively (Ip, et al., 2008). While these two enzymes were functionally similar to RuvC, they belong to a separate family of enzymes known as the XPG/Rad2 family, while RuvC is related to the retroviral integrase superfamily (Tsutakawa, et al., 2011) (Gorecka, et al., 2013). Since then, a two further HJ resolvases have been identified in human cells – the non-canonical SLX-MUS complex (SLX1-SLX4-MUS81-EME1), and the non-canonical BTR complex (BLM-topoisomerase IIIα-RMI1-RMI2) (Sarbajna & West, 2014). The BTR complex however acts to resolve dHJs by dissolution and so is outside the scope of this review.

Unlike the BTR complex, GEN1 and the SLX-MUS complex are both structure-selective endonucleases and both follow the resolution pathway of HJ cleavage. Importantly the SLX-MUS complex cleaves HJs asymmetrically – such that short sections of single stranded (ss) DNA left forming gaps and flaps, requiring further downstream processing - while GEN1 performs symmetrical cleavage. Patients with Bloom's Syndrome who are deficient in the BTR complex thus show an increase incidence of sister chromatid exchanges due to SLX-MUS and GEN1 being the only available enzymes for HJ resolution (Sarbajna & West, 2014). Furthermore, while scientists are still working to determine the precise differences between the cellular roles of these two enzymes, it is clear they are of paramount importance to proper DNA maintenance since their depletion is lethal to cells – while they must occupy different niches since GEN1 can only moderately compensate for loss of SLX4 *in vivo* (Wechsler, et al., 2011) (Wyatt, et al., 2013) (Garner, et al., 2013).

1.7.1. XPG/RAD2 FAMILY, GEN1

As stated, GEN1 belongs to the XPG/Rad2 enzyme family, a group of 5'-flap endonucleases found throughout a number of eukaryotic kingdoms, which contains the two eponymous enzymes: xeroderma pigmentosum complementation group G protein (XPG) from humans and Rad2 from *Saccharomyces cerevisiae* – however these enzymes are primarily involved in nucleotide excision repair in response to DNA damage, rather than HJ resolution. In fact, the family consists of 4 different subclasses of enzymes that fulfil different roles: I) Nucleotide excision repair endonucleases, II) Replication flap endonucleases, III) Recombination/repair exonucleases, and IV) Holliday junction resolvases. With such a range of different functionalities, the group is largely characterised by three domains conserved across all its members. Firstly, the N-region, located at the extreme N-terminus of the protein, consisting of

the first 95-105 amino acids. Secondly, an internal domain (I-region) of approximately 140 residues. And finally a helix-hairpin-helix domain located just along from the I-region, which is implicated in catalysis and DNA binding (Figure 1.7.1-1) (Ip, et al., 2008) (Rass, et al., 2010). The importance of this family of enzymes has been demonstrated in that a lack of XPG in humans can lead to xeroderma pigmentosum, or Cockayne syndrome, whereby patients exhibit microcephaly, impaired development of the nervous system, premature aging and a failure to grow and gain weight (Zafeiriou, et al., 2001).

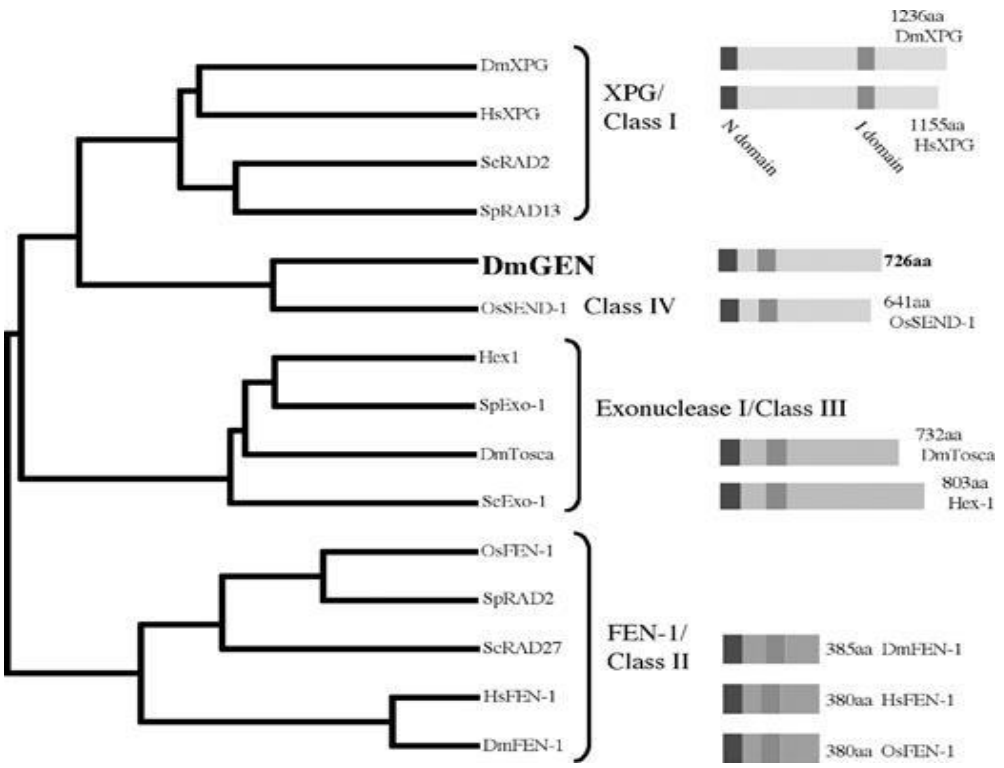


FIGURE 1.7-1 DENDROGRAM DEPICTING RELATIONSHIPS AMONG FEN-1/RAD2 FAMILY MEMBERS. ON THE RIGHT ARE DEPICTED PROTEIN STRUCTURES WITH HIGHLY CONSERVED DOMAINS (XPG-N AND XPG-1 DOMAINS) HIGHLIGHTED

(ISHIKAWA, ET AL., 2004)

As stated, of particular relevance to the focus of this review is the class IV XPG/Rad2 family member, GEN1. The enzyme was first isolated in 2008 by Ip et al. and has orthologues in a number of different eukaryotic kingdoms, e.g. *D. melanogaster* (Ishikawa, et al., 2004) *and S. cerevisiae* (Ip, et al., 2008). While being allocated to the HJ resolvase class, *in vitro* characterisation of GEN1 has demonstrated it retains its ability to also recognise 5' flaps, replication fork intermediates and ss gaps, and in fact GEN1's flap cleavage activity is far more efficient than its HJ cleavage activity. However, its primary function *in vivo* would appear to be a final measure to remove dHJs and thus conjoined chromatids, prior to cytokinesis. To fulfil such a variety of functions, GEN1 has a sophisticated structure consisting of a homodimer between two identical subunits, which forms around the HJ to introduce symmetrical nicks, but operates as a monomer to cleave flaps (Chan & West, 2015).

A single GEN1 monomer resembles an open right hand, with the catalytic domain forming the palm and the DNA-binding domain located within the ball of the thumb. This catalytic core contains both the XPG-N and XPG-I regions, forming a 7-stranded β-sheet surrounded by 9 helices which harbours the active site. Contained therein are a number of negatively charged residues – D30, E75, E134, E136, D155, D157, D208 – which are responsible for catalysis. Inserted within the helix-hairpin-helix (α10-α11) upstream of the I-region that forms the exonuclease domain (along with hairpins α12- α 13 and α14-α15), is a 78 amino acid insertion that forms an additional helix (α12b), which is packed loosely on the end of the 'fingers' of the protein. The palm and fingers then form 2 characteristic DNA binding surfaces, separated by a hydrophobic wedge within the ball of the thumb. Finally a small globular chalice-shaped chromodomain is located at the wrist, forming an aromatic cage which is thought to secure the DNA-GEN1 complex (Lee, et al., 2015).

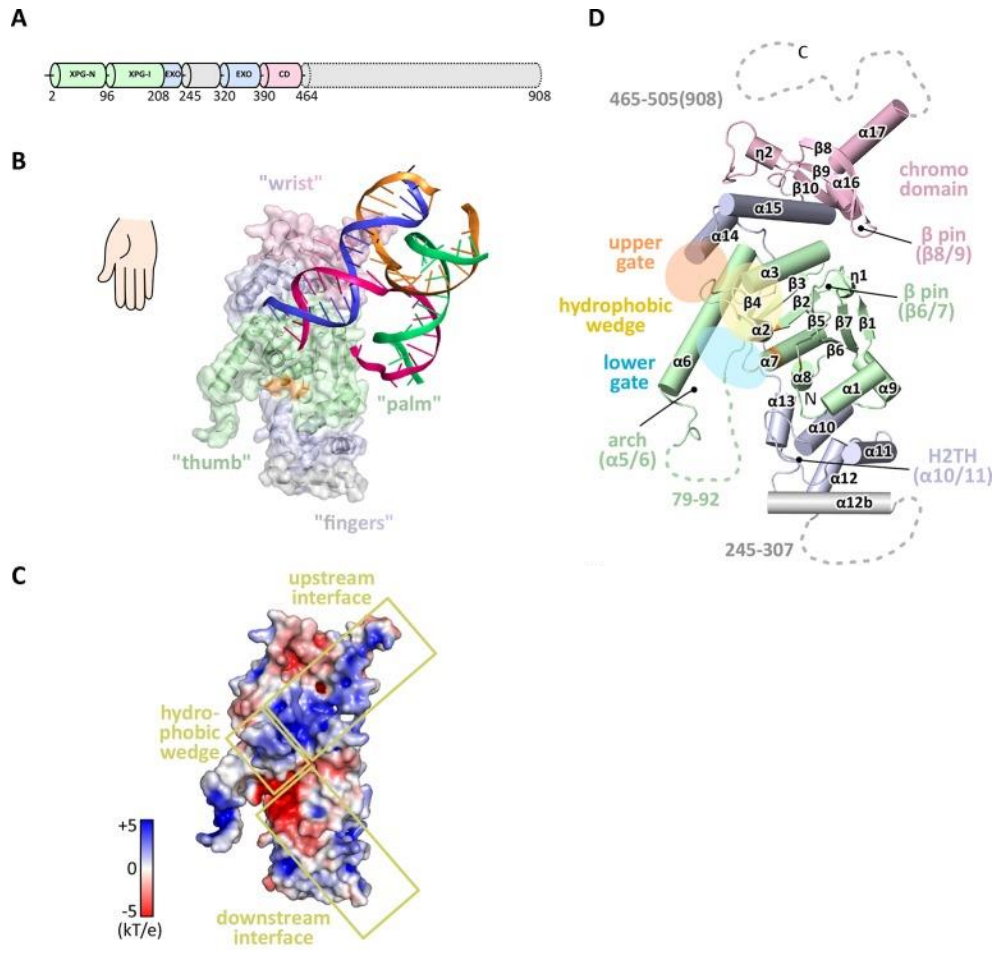


FIGURE 1.7-2 – ARCHITECTURE OF HUMAN GEN1

(A) SIMPLE AMINO ACID STRUCTURE AND DOMAINS OF HUMAN GEN1. (B) CATALYTIC CORE OF GEN1 PROTEIN BOUND TO HJ DNA. DOMAINS ARE COLOUR CODED TO MATCH THOSE IN PART A: INTERTWINING XPG-N AND XPG-I IN GREEN, 5'->3' EXONUCLEASE C-TERMINAL DOMAIN IN BLUE, CHROMODOMAIN IN PINK, ACTIVE SITE RESIDUES (E134, E136, D115, D157) IN ORANGE, AND UNASSIGNED REGIONS IN GREY. THE ARCHITECTURE OF THE FOLDED PROTEIN RESEMBLES A DOWNWARD-ORIENTATED RIGHT HAND, AS ILLUSTRATED. THE SURFACE MODEL OF THE PROTEIN IS SHOWN IN A SEMI-TRANSPARENT REPRESENTATION, WITH SECONDARY STRUCTURE UNDERNEATH. THE HJ IS OVERLAID WITH DIFFERENT STRANDS SHOWN IN DIFFERING COLOURS. (C) ELECTROSTATIC SURFACE POTENTIAL OF GEN1. COLOUR-CODED TO SHOW POTENTIAL FROM -5 KT/E IN RED, TO +5 KT/E IN BLUE. THE DNA-BINDING REGIONS AND HYDROPHOBIC WEDGE ARE HIGHLIGHTED IN YELLOW. (D) SECONDARY STRUCTURE ELEMENTS OF GEN1'S CATALYTIC CORE, COLOUR CODED AS BEFORE. DOTTED LINES REPRESENT UNRESOLVED REGIONS IN THE CRYSTAL STRUCTURE. NUMBERING FOLLOWS A UNIFIED SCHEME FOR RAD2/XPG FAMILY FOR A-HELICES, B-SHEETS AND 310-HELICES (H).

ADAPTED FROM (LEE, ET AL., 2015).

To cleave HJs, the GEN1 homodimer introduces two symmetrical incisions across the intersection of the characteristic X-shape, through the coordinated action of both subunits. The first of the 2 nicks is rate-limiting, but the second is almost simultaneous, and the correct positioning of the arms of the HJ substrate within the subunits is critical to successful cleavage, hence the importance of the novel chromodomain, as this provides another anchorage site for

the DNA-binding interface upstream. Of particular note is that analysis of GEN1 structure-function relationships have suggested that its ability to cleave HJs is far slower than 5' flaps, despite being recognised as a HJ resolvase (Lee, et al., 2015).

Due to the apparent functional degeneracy with other endonucleases and particularly other the HJ resolvases discussed above, determining the cellular role of GEN1 has so far been difficult. Thus it has been suggested that GEN1 may simply function as a secondary backup pathway for HJ resolution compared to the SLX-MUS complex (Garner, et al., 2013). However other studies have suggested GEN1 may occupy a separate niche of specifically resolving nicked HJs following their conversion to single intact HJs after the SLX-MUS complex has failed to resolve them, as well as being implicated in centrosome maintenance (García-Luis & Machín, 2014) (Gao, et al., 2012). In most cells, the preference for HJ dissolution and thus prevention of CO products is enforced by sequestering GEN1 outside the nucleus until M phase, as well as by phosphorylation (Matos, et al., 2011).

1.7.2. Flap Endonuclease 1

Another member of the XPG/Rad2 family of note, is the class II flap endonuclease 1, or FEN1. This smaller protein has been shown to play a key role in DNA replication during both mitosis and DNA repair. As DNA polymerase δ extends primer DNA upon the lagging strand during both DNA replication, downstream DNA can be displaced, creating 5'-end flap structures in the DNA. FEN1 is responsible for cleaving these 5'-flaps in a sequence-independent manner, allowing DNA Ligase to then seal the sugar phosphate backbone (Henneke, et al., 2003). Furthermore, FEN-1 also cleaves bifurcated 5'-flap junctions and the single base 3'-flaps created during the extension of Okazaki fragments for long-patch base excision repair and homologous recombination, as well as playing a role in telomere maintenance in human and *S. cerevisiae* cells (Gary, et al., 1999) (Saharia, et al., 2008) (Parenteau & Wellinger, 2002) (Stodola & Burgers, 2016). As such, this critical enzyme (or orthologs thereof) can be found in all domains of life, and depletion of FEN1 has been shown to be lethal in both mice and yeast (Larsen, et al., 2003) (Parenteau & Wellinger, 1999) (Reagan, et al., 1995).

A number of studies have revealed that several key regions of FEN1 have been conserved in homologues from distantly related organisms, including *Methanococcus jannaschii,* and *Pyrococcus furiosus* (Hwang, et al., 1998) (Hosfield, et al., 1998). Comparison of these structures has demonstrated a strong similarity in the topology of the core structure of this enzyme, with a pocket formed of a β-sheet and two α-helices housing the enzymes active site, and the seven conserved acidic amino acids which surround two distinct Mg^{2+} ions – one which regulates substrate binding and the other which influences conformational changes in the enzyme's 3D structure – which are key to nuclease function. In addition to this active site, FEN1 also has a flexible loop region (residues 87-134) containing a helical motif and which is lined with positively charged residues. This is thought to form a 8 x 25 Å hole which would accommodate single-stranded DNA for flap junction cleavage (Henneke, et al., 2003). Much like GEN1, the FEN1 enzyme would seem to envelop the target single stranded DNA within this loop domain, and then a 'thumb' structure would contain the flap itself, such that the cleavage site sits within the active site. Meanwhile interaction with the 3' nucleotide of the upstream strand by a hydrophobic pocket on FEN1 prevents further extension of the DNA by Pol δ and thus, prevents any further strand displacement. This interaction may also help to stabilise the interaction of FEN1 around the flap junction, allowing for precise and efficient cleavage that DNA Ligase can repair (Henneke, et al., 2003) (Storici, et al., 2002).

Of particular relevance to this report is the N-terminal 50% sequence homology and 5'-flap endonuclease activity shared by FEN1 with the previously discussed GEN1 enzyme (Rass, et al., 2010).

1.8. *TB*FEN1, A NOVEL ENDONUCLEASE IN *TRYPANOSOMA BRUCEI*

As covered earlier in the review, *T. brucei* utilises homologous recombination to carry out VSG switching - introducing new VSG genes from elsewhere in their genome to the active BES, as well as creating novel VSG genes via segmental gene conversion. To do so, they rely on the DSBR pathway and the resolution of the arising dHJs to copy VSG genes across into the active ES from a number of VSG gene libraries throughout their genome. With homologous recombination playing such a key role in the mechanism that underpins VSG switching, it presents a major focal point for research into how to potentially interrupt the parasites immune evasion and to ideally treat the disease. By studying the proteins involved in DNA repair, the DSBR pathway and homologous recombination, and what it is about their structure and molecular make-up that allows them to fulfil their roles, it may be possible to design drugs or other treatments that inhibit those roles.

1.8.1. A TRYPANOSOMAL ORTHOLOG OF GEN1

Previous studies in this laboratory have identified the need for a Trypanosomal ortholog of the GEN1 enzyme and attempted to locate such an enzyme within *T. brucei*'s proteome. Following a protein BLAST search using the human GEN1 amino acid sequence, two *T. brucei* proteins with sequence homology were identified: putatively labelled as FEN1 (Tb927.3.930) and Rad2 (Tb927.9.11760) proteins. However, initial fluorescence studies of Rad2 demonstrated it to have poor localisation to the trypanosome nucleus, and so the 45kDa FEN1 (Tb927.3.930) (referred to hereafter as *Tb*FEN1) protein was identified as the more likely HJ resolvase enzyme for further study (McAllister & Benson, 2015). If *Tb*FEN1could be shown to possess either HJ cleavage activity or flap cleavage activity this would prove to be a major steppingstone towards designing a drug that could inhibit homologous recombination in *T. brucei*.

To this end, work within this laboratory previous to this project has endeavoured to express and functionally characterise the *Tb*FEN1 protein. The target *Tb*FEN1 protein was expressed in *E. coli* Tuner cells, and then purified by a two-step purification process with HisTrap and anion exchange liquid chromatography (McAllister & Benson, 2015). The purified *Tb*FEN1 was then later used in nuclease assays with a number of different novel synthetic DNA substrates to study its endonuclease activity compared to RuvC, and T7endonuclease I. The results of these studies showed that while the *Tb*FEN1 protein does possess 5'-flap endonuclease activity, little to no HJ cleavage was seen under the conditions used in that study (Noblett & Benson, 2017). With this new information about *Tb*FEN1's cleavage activity, the next step in understanding its role was to further characterise the mechanism by which it cleaves these DNA structures, so that it might be possible to identify possible residues that could be targeted in future drug development.

To continue looking into the role of the *Tb*FEN1 protein in *T. brucei*, this project aimed to:

- Investigate whether further *Tb*FEN1 has HJ resolution activity.
- Compare human FEN1 with *Tb*FEN1 to identify and predict critical catalytic residues.
- Use site directed mutagenesis to generate *E. coli* plasmids to express mutant *Tb*FEN1 proteins containing key residue substitutions.
- Express mutant proteins in *E. coli* and purify these for use in enzymatic assays to determine which residues were critical for *Tb*FEN1 function

- Examine *Tb*FEN1 localisation in *T. brucei* cells

2. MATERIALS AND METHODS

2.1. BUFFERS, SOLUTIONS, ANTIBIOTICS AND *E. COLI* STRAINS

2.1.1. BUFFERS AND SOLUTIONS

Buffers and media were prepared using Milli-Q water.

Buffer/Solution	**Ingredients**
Agarose Gel	Agarose low EEO, used at 0.8% w/v with 1x TAE
Dithiothreitol (DTT)	Stock solution made at 1M
Isopropyl β-D-1 thiogalactopyranoside (IPTG)	Stock solution made at 1M
Luria Broth (LB) Low Salt Agar	20 g/L LB and 10 g/L Agar
LB Broth	20 g/L LB
10x Phosphate Buffered Saline (PBS)	1.37 M Sodium chloride, 27 mM Potassium chloride, 101 mM Disodium hydrogen phosphate, 18 mM Potassium dihydrogen phosphate, pH 7.4
Protein Purification Native – Immobilised-metal Affinity Chromatography (IMAC)	**Native Lysis Buffer** 0.01% lysozyme in Native Lysis Buffer **Native Binding Buffer** 0.02% Triton-X, 20 mM Tris-HCl (pH 8), 500 mM Sodium chloride, 20 mM Imidazole, 10% Glycerol **Native Elution Buffer** 0.02% Triton-X, 20 mM Tris-HCl (pH 8), 500 mM Sodium chloride, 500 mM Imidazole, 10% Glycerol
Protein Purification – Ion-exchange Chromatography (IEX)	**Low Salt Buffer** 20 mM Tris-HCl (pH 8.5), 1mM EDTA, 10% Glycerol **High Salt Buffer** 20 mM Tris-HCl (pH 8.5), 1mM EDTA, 10% Glycerol, 500 mM Sodium chloride

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) – 12.5% Gel	**12.5% Resolving Gel** 375 mM Tris-HCl pH8.8 , 2.5% Acrylamide (37.5:1 acrylamide:bis-acrylamide), 0.1 % SDS, 0.1% ammonium persulphate, 0.1% TEMED **5% Stacking Gel** 125 mM Tris-HCl pH 6.8. 2.5% acrylamide (37.5:1 acrylamide:bis-acrylamide), 0.1 % SDS, 0.1% ammonium persulphate, 0.1% TEMED **SDS Loading Buffer** 125 mM Tris-HCl (pH 6.8), 4% SDS, 0.004% Bromophenol blue, 20% glycerol, 100 mM DTT **SDS-PAGE Running Buffer (10x)** 250 mM Tris-base, 1.92 M Glycine, 35 mM SDS
Tris-Acetic acid EDTA (TAE) Buffer (50x)	1 X 40 mM Tris, 20 mM acetate, and 1 mM EDTA, pH ~8.3
Denaturing Urea-PAGE	**Urea Buffer (5x)** 25% 10x Tris Borate EDTA (TBE), 24% w/v Urea **20% Resolving Gel** 40% Acrylamide/Bis, 5x Urea Buffer, 0.1% TEMED, 10% APS **Formamide Loading Buffer** 10 mg/ml Blue Dextran, 98% deionised formamide, 10 mM EDTA
Native-PAGE	**Native Loading Buffer** 0.5 M EDTA, 50% Glycerol, 1mg/ml bromophenol blue **10% Resolving Gel** 10x TBE, 30% Acrylamide, 10% APS, 0.1% TEMED
Nuclease Assays	**Cleavage Buffer (1x)** 50 mM Tris HCl (pH 7.5), 1 mM Magnesium chloride, 1 mM DTT **Stop Buffer (2x)**

	100 mM Tris HCl (pH 7.5), 50 mM EDTA, 2.5% SDS
SOC Media	0.5% Yeast Extract, 2% Tryptone, 10 mM sodium chloride, 2.5 mM Potassium chloride, 10 mM Magnesium chloride, 10 mM Magnesium sulphate, 20mM Glucose
Mutagenesis Buffer	5.5% 10x reaction buffer, 11.1% 2 ng μL^{-1} of plasmid pET24a, 1.1% dNTP mix, and 3.3% QuikSolution
Zimmerman's postfusion medium	132 mM NaCl, 8 mM KCl, 8 mM $Na_{2H}PO_4$, 1.5 mM KH_2PO_4, 0.775 mM $Mg(CH_3COO)_2$, 0.063 mM $Ca(CH_3COO)_2$
Localisation studies	**Blocking Buffer** 1x PBS, 0.05% Tween-20, 1% Bovine serum albumin **Wash Buffer** 1x PBS with 0.05% Tween-20

2.1.2. GENERATED PLASMIDS

Plasmid Name	**Specific Mutagenesis of *Tb*FEN1**
pJG04	*Tb*FEN1 with C-terminal 6 His tag, generated by J. Owen within the Benson Lab (Owen, 2015) Hereafter referred to as 'pJG04 *Tb*FEN1', used as the base sequence for all mutagenesis products generated as part of this project
pSS1	pJG04 *Tb*FEN1 D34A
pSS2	pJG04 *Tb*FEN1 K47A
pSS3	pJG04 *Tb*FEN1 R74A
pSS4	pJG04 *Tb*FEN1 D90A
pSS5	pJG04 *Tb*FEN1 E162A
pSS6	pJG04 *Tb*FEN1 E164A
pSS7	pJG04 *Tb*FEN1 D183A
pSS8	pJG04 *Tb*FEN1 D185A
pSS9	pJG04 *Tb*FEN1 G235A
pSS10	pJG04 *Tb*FEN1 D237A
pSS11	pJG04 *Tb*FEN1 D34K

pSS12	pJG04 *Tb*FEN1 D90K
pSS13	pJG04 *Tb*FEN1 E162K
pSS14	pJG04 *Tb*FEN1 E164K
pSS15	pJG04 *Tb*FEN1 D183K
pSS16	pJG04 *Tb*FEN1 D185K
pSS17	pJG04 *Tb*FEN1 G235K
pSS18	pJG04 *Tb*FEN1 D237K
pSS19	pJG04 *Tb*FEN1 K47D
pSS20	pJG04 *Tb*FEN1 K74D
pSS21	pJG04 *Tb*FEN1 G235D
pRO1	pJG04 *Tb*FEN1 Q341A
pRO2	pJG04 *Tb*FEN1 Q341A G342A

2.1.3. ANTIBIOTICS AND DRUG STOCKS

Antibiotic/Drug	**Concentration**
Kanamycin	Stock solution made at 30 mg/ml diluted in Milli-Q. Used at 30 µg/ml in LB agar and broth
Chloramphenicol	Stock solution made at 25 mg/ml diluted in Milli-Q. Used at 25 µg/ml in LB Agar and 12.5 µg/ml in LB Broth
Blasticidin	Stock solution made at 10 mg/ml diluted in Milli-Q Used at 10 µg/ml

2.1.4. *E. COLI* STRAINS

Strain Name	**Full Genotype**
DH5 α	*fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
Tuner™ (DE3) pLacI	F-*ompT hsdS*$_B$ (r$_B$-m$_B$) *gal dcm lacY1*(*DE3*) pLysS (Cam^R)
XL-10 Gold Competent	Tet^{r}Δ(*mcr*A)183 Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac*

	Hte [F´ *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^{r}) Amy Cam^{r}]

2.2. BIOINFORMATICS

2.2.1. OBTAINING SEQUENCE DATA AND ALIGNMENT

The amino acid sequence and information about key residues of both the Human FEN1 and GEN1 proteins and the *Trypanosoma brucei brucei* FEN1 protein were acquired from their UniProt entries ('P39748 FEN1_HUMAN', 'Q17RS7 GEN_HUMAN' & 'Q57WW6 FEN1_TRYB2' respectively). The amino acid sequence of these 3 proteins were compared both to each other and to the FEN1 proteins a number of other organisms (Fruit Fly (*Drosophila melanogaster*), African Clawed Frog (*Xenopus laevis*), Chicken (*Gallus gallus*), Sheep (*Ovis aries*), Gorilla (*Gorilla gorilla*), Chimpanzee (*Pan troglodytes*), Dog (*Canis lupus familiaris*), Cattle (*Bos taurus*), House Mouse (*Mus musculus*), and Brown Rat (*Rattus norvegicus*)), using the MEGA™ alignment software (Table 2.2.1-1). A second alignment was also performed between only the Human and *T. brucei* FEN1 amino acid sequences. Both alignments were performed by ClustalW.

Organism	**Protein**	**UniProt Accession Number**
T. brucei	GEN1/FEN1	Q57WW6
H. Sapiens	GEN1	Q17RS7
	FEN1	P39748
D. melanogaster	FEN1	Q7K7A9
X. laevis	FEN1	P70054
G. gallus	FEN1	Q5ZLN4
O. aries	FEN1	C8BKD0
G. gorilla	FEN1	G3RA03
P. troglodytes	FEN1	H2Q3U7
C. lupus familiaris	FEN1	J9PB88
B. taurus	FEN1	Q58DH8
M. musculus	FEN1	Q8C5X6
R. norvegicus	FEN1	Q5XIP6

TABLE 2.2.1-1 ORTHOLOGS OF TBFEN1 USED TO ANALYSE CONSERVED SEQUENCE HOMOLOGY

The aligned sequences were then used to identify conserved residues across the different species and study possible evolutionary relationships between the different endonucleases. The alignment between human and *T. brucei* FEN1 was further studied to highlight the differences between the two sequences, and to allow direct comparison of key residues. Using the UniProt entry for the key amino acids in the human protein, 12 positions were identified that are implicated in the function of the human protein (from literature and conserved residues). By comparison of these positions in the *T. brucei*/*H. sapiens* FEN1 alignment it was possible to identify these corresponding residues in the *T. brucei* protein, and as such select them as targets for site-directed mutagenesis.

2.3. EXPRESSION IN *E. COLI*

2.3.1. ROUTINE CULTURE OF *E. COLI*

Bacteria were routinely cultured in liquid LB or upon LB agar plates containing antibiotic as appropriate, at 37°C (unless stated otherwise) and 200 rpm for shaking liquid cultures.

2.3.2. BACTERIAL TRANSFORMATION

Plasmid DNA was incubated with competent cells on ice for 30 mins, before the mixture was heat shocked at 42°C for 30 seconds and returned to the ice bath. The mixture was then incubated in nutrient media for 1 hour at 37°C. Transformants were plated onto LB agar plates containing appropriate antibiotics and incubated at 37°C for ~16 hours. Transformant colonies were restreaked on selective plates and incubated for a further ~16 hours. The resulting single colonies were used to inoculate liquid cultures containing appropriate antibiotics for expression studies/plasmid purification.

Throughout the project three *E. coli* strains were used. DH5α and XL10-Gold strains were used for plasmid manipulation, while Tuner (DE3) pLysS was used for expression. Expression system works by a gene encoding the T7 RNA polymerase sitting on a defective lambda phage (known as DE3) under the control of a *lac* promoter. The pLysS plasmid carries the chloramphenicol resistance gene, so inclusion of chloramphenicol in media selects for the presence of pLysS plasmid. It also encodes bacteriophage T7 lysozyme which both inhibits any low-level expression of T7 RNA polymerase prior to induction and also helps cells lysis. The recombinant plasmid (pJG04) used to express the *Tb*FEN1 protein is derived from the kanamycin resistant pET24a vector and expressed the *Tb*FEN1 protein with a C-terminal His tag.

2.3.3. SITE DIRECTED MUTAGENESIS

Using the previously identified key amino acids positions in the *T. brucei* protein, the Agilent primer design program was then used to design oligonucleotide primers to introduce specific mutations into the DNA, to change codons within the open reading frame that would result in amino acid substitutions upon translation. The key wild type residues were substituted for alanine (as a small non-polar amino acid) or an opposing amino acid (basic targets were switched for Aspartic Acid, and acidic targets for Lysine). The primers were dissolved in sterile water to a concentration of 25ng μL^{-1}, and aliquoted for use in site-directed mutagenesis.

Amino Acid Substitution	Primers	
	Sequence (5'-3')	Sequence (5'-3')
D34A	GGCCATGGAGGCAGCGATTGCGATGCG	CGCATCGCAATCGCTGCCTCCATGGCC
K47A	TGACCCTCTTGAAAGCCTGCCATGGCGATGACGAACTG	CAGTTCGTCATCGCCATGGCAGGCTTTCAAGAGGGTCA
R74A	CCTTCATCTATCATTCTTAATGTAGCAAAAAAAATTCCACTTAAATGAGACGTAAC	GTTACGTCTCATTTAAGTGGAATTTTTTTTGCTACATTAAGAATGATAGATGAAGG
D90A	GGTGGGCGGCTTACCAGCAAACACGTATATAGGACG	CGTCCTATATACGTGTTTGCTGGTAAGCCGCCCACC
E162A	CTGGGCTTCTGCCGCAGAGGGTGCTTG	CAAGCACCCTCTGCGGCAGAAGCCCAG
E164A	GCGCACTGGGCTGCTGCCTCAGAGG	CCTCTGAGGCAGCAGCCCAGTGCGC
D183A	GCCAGGGCATCCATAGCCTCTGTTCCTACCG	CGGTAGGAACAGAGGCTATGGATGCCCTGGC
D185A	AAACGCCAGGGCAGCCATATCCTCTGTTCCTAC	GTAGGAACAGAGGATATGGCTGCCCTGGCGTTT
G235A	AAATTCTGGGGACGTAATCACAAGCAAGAAGAATACATAAATCTATG	CATAGATTTATGTATTCTTCTTGCTTGTGATTACGTCCCCAGAATTT
D237A	ATGTATTCTTCTTGGTTGTGCTTACGTCCCCAGAATTTCAG	CTGAAATTCTGGGGACGTAAGCACAACCAAGAAGAATACAT
Q341A	CGCACTTACGAAGAAAACCGCAGGTCGCTTGGACCAATTT	GCGTGAATGCTTCTTTTGGGGTCCAGCGAACCTGGTTAAA
Q341A G342A	TACGAAGAAAACCGCAGCTCGCTTGGACCAATTTT	ATGCTTCTTTTGGCGTCGAGCGAACCTGGTTAAAA
D34K	TACGGCCATGGAGGCCTTGATTGCGATGCGTCG	CGACGCATCGCAATCAAGGCCTCCATGGCCGTA
D90K	TGGGCGGCTTACCCTTAAACACGTATATAGGACGAAGTCCT	AGGACTTCGTCCTATATACGTGTTTAAGGGTAAGCCGCCCA
E162K	GGCTTCTGCCTTAGAGGGTGCTTGAACAACAGG	CCTGTTGTTCAAGCACCCTCTAAGGCAGAAGCC
E164K	GCGCACTGGGCTTTTGCCTCAGAGGGT	ACCCTCTGAGGCAAAAGCCCAGTGCGC
D183K	CGCCAGGGCATCCATCTTCTCTGTTCCTACCGC	GCGGTAGGAACAGAGAAGATGGATGCCCTGGCG
D185K	CAAACGCCAGGGCCTTCATATCCTCTGTTCCTACCG	CGGTAGGAACAGAGGATATGAAGGCCCTGGCGTTTG
G235K	AAATTCTGGGGACGTAATCACACTTAAGAAGAATACATAAATCTATGAACTGATGCA	TGCAGCAGTTCATAGATTTATGTATTCTTCTTAAGTGTGATTACGTCCCCAGAATTT
D237K	AATTCCTGAAATTCTGGGGACGTACTTACAACCAAGAAGAATACATAAATC	GATTTATGTATTCTTCTTGGTTGTAAGTACGTCCCCAGAATTTCAGGAATT
K47D	CTGACCCTCTTGAAAGCCATCCATGGCGATGACGAACTG	CAGTTCGTCATCGCCATGGATGGCTTTCAAGAGGGTCAG
R74D	GTCCTTCATCTATCATTCTTAATGTATCAAAAAAAATTCCACTTAAATGAGACGTAACAT	ATGTTACGTCTCATTTAAGTGGAATTTTTTTTGATACATTAAGAATGATAGATGAAGGAC
G235D	AAATTCTGGGGACGTAATCACAATCAAGAAGAATACATAAATCTATG	CATAGATTTATGTATTCTTCTTGATTGTGATTACGTCCCCAGAATTT

TABLE 2.3.3-2.3.3-1 – TABLE OF PRIMERS

TABLE SHOWING THE FULL LIST OF SITE-DIRECTED MUTAGENESIS TARGETS AND THE FORWARD AND REVERSE PRIMERS USED TO INTRODUCE THEM

The site-specific mutations were introduced into plasmid DNA using the Agilent QuikChange mutagenesis kit, and the included instructions were followed in full, with the exception that the reactions were scaled down by 50%. For site-directed mutagenesis 2.5 µL of forward and reverse primers (Table 2.3.3.-1) were mixed with 5 µL 2ng μL^{-1} of the plasmid pET24a, 2.5 µL

of 10x reaction buffer, 0.5 µL od dNTP mix and 1.5 µL of QuikSolution; before being made up to 50 µL with ddH_2O. The solution was then cycled according to conditions in table 2.3.3-2, along with 0.5 µL of 2.5 U μL^{-1} PfuUltra HF DNA polymerase. Following thermal cycling, the parental dsDNA was digested with Dpn 1 restriction enzyme according to the manufacturer's instructions.

In order to recover transformants with the desired mutations, 1µL of β-mercaptoethanol mix was added to 22.5 µL of ultracompetent cells and incubated on ice for 10 minutes with gentle agitation, before addition of 1 µL of the prepared plasmid DNA. The mixture was incubated for a further 30 minutes on ice before heat-pulsing the solution at 42°C for 30 seconds and returning to ice for at least 2 minutes. The cells were then cultured as detailed in section 2.3.1.

Segment	Cycles	Temperature	Time
1	1	95	1 minute
2	18	95	50 seconds
		60	50 seconds
		68	1 minute/kb plasmid length
3	1	68	7 minutes

TABLE 2.3.3-2– PCR CYCLING PARAMETERS

2.3.4. PLASMID PURIFICATION

Single transformant colonies from the site-directed mutagenesis were inoculated into 5ml of LB media with a final concentration of 30 µg ml^{-1} Kanamycin and incubated overnight as detailed in section 2.3.1. Bacteria from these cultures were then pelleted by centrifugation at 3000rpm and 4°C for 5 minutes and the supernatant decanted. The plasmids were then purified from pelleted bacteria using the Qiagen QIAprep Miniprep kit and plasmid concentration determined by measuring absorbance at 260 nm. A sample (typically 10 µL of a 100 ng uL^{-1} solution) of the purified plasmids was sent for sequencing using either the T7 reverse or T7 forward sequencing primers as appropriate to the site of the introduced mutation.

2.3.5. DETERMINATION OF SUCCESSFUL TRANSFORMANTS

The sequencing data from the purified plasmids was compared to the original *Tb*FEN1 sequence by first analysing the new sequencing data on Chromas™ to check for the desired point mutation to the DNA. The forward and reverse sequences were then assembled into a single contiguous sequence using the online CAP3 Sequence Assembly Program (http://doua.prabi.fr/software/cap3), and subsequently translated into an amino acid sequence using the online ExPASy Translate tool (https://web.expasy.org/translate/). This could then also be checked for the desired amino acids substitution before running a Blast search against the new sequence to check it was still identical to the original sequence in every other regard.

2.3.6. SMALL SCALE INDUCTION OF PROTEIN EXPRESSION

An aliquot (25 ml) of fresh LB medium containing selective antibiotics was inoculated with 0.5 ml of overnight culture and incubated at 37°C with shaking. Samples (0.5 ml) were taken at 1 hour, and every subsequent 30 minutes, and the optical density at 600nm measured until it reached approximately an absorbance of 0.4. IPTG was subsequently added to the remaining

culture to a final concentration of 1mM, and the incubation continued. Further 0.5 ml samples were taken every hour for the next 3 hours.

Samples were processed by centrifuging at 4000 rpm for 1 minute to pellet the bacteria, before discarding the supernatant. The pellet was then resuspended in 50 µL of SDS-Gel Sample Buffer (SDS-GSB) and heated to 98°C for 3 minutes to solubilise and denature all the bacterial proteins and pulsing the contents to the bottom of the tube. The samples were then stored at 4°C for later analysis. After 3 hours the remaining liquid culture was pelleted by centrifuging at 4000 rpm for 10 minutes, the supernatant discarded, and the pellet frozen at -80°C to allow subsequent purification of recombinant protein.

The time-point samples from the induction of expression were run on a 12% SDS-PAGE gel for 1 hour at 180 V with a Mark-12 MW marker, before staining overnight in 20 ml InstantBlue. The gel was then analysed to determine the efficiency of expression, so as to optimise the conditions prior to scaling up.

2.3.7. Large Scale Induction

Single colonies from agar plates were picked and used to inoculate 5 ml of LB medium, which was incubated at 37°for approximately 6-7 hours before making the culture up to 25 ml and incubating for a further ~16 hours. An aliquot (16 ml) of the overnight culture was then used to inoculate 800ml of LB medium across two 400 ml cultures in 1 litre flasks. The cultures were then grown until their OD_{600} reached 0.4. As with the small-scale induction, a 1 ml aliquot removed to be pelleted and resuspended in SDS-GSB, before IPTG was added to a final concentration of 1 mM. The cultures were then incubated at 30°C with shaking for 3 hours and another 1 ml aliquot removed and frozen. The remaining cultures were recombined in a single 1 litre centrifuge tube and spun at 4000 rpm for 20 minutes, at 4°C. The supernatant was then discarded, and the pellet frozen at -20°C to be lysed and the recombinant protein purified at a later date.

2.4. PROTEIN PURIFICATION AND ANALYSIS

2.4.1. Small Scale Lysis & Purification using Ni-NTA Spin Columns

Small scale lysis and purification was performed the frozen pellet from the earlier induction was resuspended in 1 ml of Qiagen Qproteome bacterial lysis buffer by pipetting up and down before being incubated on ice for 20 minutes. A 50 µL sample was removed and labelled as the Whole Cell Extract, while the remaining sample was transferred to a fresh microcentrifuge tube and spun at 13000 rpm for 20 minutes. The supernatant was then transferred to a fresh tube, and 0.5 ml of supernatant was added to 0.5 ml of 2x Binding Buffer (100 mM Na phosphate pH 8, 600 mM NaCl, 40 mM Imidazole), before removing a 50 µL sample to a fresh labelled tube as the soluble fraction.

A Ni-NTA spin column was prepared by equilibrating with 600 µL of wash buffer (50 mM Na phosphate pH 8, 300 mM NaCl, 20 mM Imidazole) and spinning at 2000 rpm for 2 minutes, before discarding the flow-through. 600 µL of the soluble extract/binding buffer mixture was

then spun through the column, again at 2000 rpm for 2 minutes. The flow through was transferred to a fresh microcentrifuge tube labelled as the non-binding *E. coli* proteins. The column was washed 3 times with 600 µL of wash buffer, as above, with each flow through sample transferred to clean labelled tubes. The His-tagged *Tb*FEN1 protein was then eluted by applying 200 µL of elution buffer to the column and spinning for a further 2 minutes at 2000 rpm, before transferring the flow through to a clean labelled tube. This final elution step was repeated a further time. The samples were then frozen at -20°C for later analysis by SDS-PAGE, by addition of 50 µL from each fraction to an equal volume of SDS-GSB.

2.4.2. Preparation of Large Scale Lysate

Pellets from large scale induction were thawed on ice for at least 15 minutes before resuspending in 16 ml of lysis buffer. The suspension was then sonicated for 3 cycles of 20 seconds in an ice bath, before being transferred to ultracentrifuge tubes. The sample was then spun at 30000 rpm for 1 hour at 4°C in a 45 Ti rotor, and the supernatant transferred to a fresh 50 ml tube and held on ice.

2.4.3. His-Tag Chromatography

The large scale cell lysate was loaded into a 50 ml superloop and attached to an ÄKTA-prime liquid chromatography system, before purifying His-tagged protein from the sample using a HisTrap HP column (GE Healthcare). Protein was loaded, the column washed with a 20 mM imidazole wash buffer (20 mM Tris pH 8, 500 mM NaCl, 20 mM Imidazole, 0.02% Triton X, 10% glycerol), and His-tagged protein eluted with 20 mM-500 mM linear imidazole gradient (final elution buffer 20 mM Tris pH 8, 500 mM NaCl, 500 mM Imidazole, 0.02% Triton X,10% glycerol). Fractions (1 ml) were collected across the imidazole gradient. Samples of the peak fractions were analysed by SDS-PAGE, as below, and assuming minimal contaminants were present, the 3 most concentrated fractions pooled for purification by anion exchange.

2.4.4. Anion-Exchange Chromatography

The pooled fraction from the His-tag purification were diluted at a ratio of 1:10 with the low sodium wash buffer (20 mM Tris pH 8.5, 10% glycerol, 1mM EDTA) and loaded onto the column. The solution was then run over a MonoQ 5/50 GL column using an ÄKTA FPLC system with a 0-100% concentration gradient between the low sodium wash buffer and a 0.5 M NaCl elution buffer (20 mM Tris pH 8.5, 10% glycerol, 1 mM EDTA, 0.5 M NaCl), over which 1 ml fractions were collected. The peak fractions were again analysed by SDS-PAGE.

2.4.5. SDS-PAGE analysis of Purification Products

Purified proteins from both the small and large scale purifications were all analysed by 12% SDS-PAGE gel. In each case, 10 µL was withdrawn from the sample and mixed with 40 µL of SDS-GSB. The mixture was then heated to 98°C for 3 minutes before pulsing to the bottom of the tube. 10 µL of each sample was then loaded into the respective well and the gel was run for 1 hour at 180 V, along with a Mark-12 MW marker. The gels were then stained overnight in 20 ml InstantBlue before being photographed under white light in a GelDoc. Band intensity and location was used to qualitatively determine those fractions with the highest concentration of the desired protein.

2.5. CHARACTERISATION OF PROTEIN CLEAVAGE ACTIVITY ON DNA SUBSTRATES

The absorbance readings at 280nm of the 3 most concentrated fractions of each mutant from the anion exchange chromatography were used to determine the concentration of each fraction and thus to produce 10 µM stocks of each protein by dilution in the low sodium buffer, prior to carrying out assays.

All substrates used in this project to assay for endonuclease cleavage were a gift from Dr F. Benson and were derived from those previously used to characterise prokaryotic and eukaryotic Holliday junction resolvases (Constantinou & West, 2004). In brief, they were prepared by annealing a Cy5 5' labelled 60mer oligonucleotide (Cy5 X26.1) with an excess of partially complementary oligonucleotides (Table 2.5-1). The substrates generated were then isolated from excess unlabelled oligonucleotides by polyacrylamide electrophoresis, eluted from a polyacrylamide gel slice, ethanol precipitated and re-dissolved in 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 100 mM NaCl buffer. Holliday junction substrates with a mobile core (indicated by nucleotides in bold in Table 2.5-1) were prepared by annealing Cy5-X26.1 with an excess of oligonucleotides X26.2, X26.3 and X26.4; a double-strand DNA marker was prepared by annealing Cy5-X26.1 with an excess of oligonucleotide S26.5; and the 5' FLAP substrate (with a Cy5 label positioned at the end of the single-stranded DNA) prepared by annealing Cy5-X26.1 with an excess of F26.6 and F26.7. A 5' Cy5 labelled 30-mer oligonucleotide (M30) was used as a marker to compare with the products of cleavage of the FLAP substrate.

Name	Base sequence (5'-3')
Fluorescently labelled oligonucleotides	
CY5-X26.1	CY5-CCGCTACCAGTGATCAC**CAATGGATTGCTAGGACATCTTTGCC**CACCTGCAGGTTCACCC (60)
Unmodified oligonucleotides	
X26.2	TGGGTGAACCTGCAGGTG**GGCAAAGATGTCCTAGCAATCCATTG**TCTATGACGTCAAGCT (60)
X26.3	GAGCTTGACGTCATAGA**CAATGGATTGCTAGGACATCTTTGCC**GTCTTGTCAATATCGGC (60)
X26.4	TGCCGATATTGACAAGAC**GGCAAAGATGTCCTAGCAATCCATTG**GTGATCACTGGTAGCGG (61)
S26.5 (ds)	GGGTGAACCTGCAGGTG**GGCAAAGATGTCCTAGCAATCCATTG**GTGATCACTGGTAGCGG (60)
F26.6 (Flap)	TGGGTGAACCTGCAGGTGGGCAAAGATGTCTCTATGACGTCAAGCTGTCTTGTCAATATC (60)
F26.7 (Short Flap)	GATATTGACAAGACAGCTTGACGTCATAGA (30)
M30.	CCGCTACCAGTGATCACCAATGGATTGCT

TABLE 2.5-1 BASE SEQUENCE (5' – 3') OF SYNTHETIC DNA CONSTRUCTS

2.5.1. ENDONUCLEASE CLEAVAGE ASSAY

For each assay a 20 µL reaction mixture containing 1 µM of the protein of interest and 200 nM Cy5 5'lablled Flap or HJ substrate in 1× Cleavage Buffer (50 mM Tris pH 7.5, 1 mM $MgCl_2$, 1 mM DTT) was assembled. The solutions were then incubated at 37°C in a water bath for 20 minutes, before addition of 2 µL of Stop Buffer (100 mM Tris pH 7.5, 50 mM EDTA, 2.5% SDS, 10 mg mL^{-1} proteinase K), and then returned to the water bath for a further 15 minutes. The resulting solutions were then split for analysis by either native-PAGE (non-denaturing) or Urea-PAGE (denaturing) assays (visualising double and single stranded DNA respectively)

DNA markers were also prepared by mixing 2 µL of 2 µM DNA substrates (Cy5 labelled 5'-Flap junction, Double Stranded DNA, or a Cy5 labelled 30mer oligonucleotide) with 2 µL 10x Cleavage buffer, and 16 µL of water to give a finally concentration of 200 nM. The double strand marker appeared much dimmer than other lanes and so was eventually raised to 4 µL in 14 µL water and 2 µL 10x Cleavage Buffer, to a final concentration of 400 nM.

2.5.2. NATIVE-PAGE ANALYSIS OF DNA CLEAVAGE

All native PAGE was performed on neutral 10% acrylamide (37.5:1) gels, at 60 V for 105 minutes, using 1xTBE as a running buffer. 10 µL of each cleavage reaction and substrate control was added to 5µL of native gel loading buffer, before running 10 µL on the gel. The gel was

then visualised on the Bio-Rad ChemiDoc™ MP imaging system, looking at the Cy5-labelled substrates.

2.5.3. DENATURING UREA-PAGE

Denaturing Urea-PAGE was performed on denaturing 12% acrylamide (19:1) urea gels, at 42 W for 2.5 hours, using 1xTBE as a running buffer. The cleavage assay samples were mixed 1:1 with 10 µL of formamide loading buffer, before being heated to 95°C for 5 minutes. The gel was pre-run for 50 minutes prior to loading 10 µL of sample into the appropriate wells. Denaturing gels were then visualised using the Typhoon™ FLA biomolecular imager.

2.6. LOCALISATION STUDIES

Throughout this section of work, procyclic *T. brucei* (927Smox and S427) were cultured in SDM-79 medium supplemented with 10% w/v foetal bovine serum and haemin.

2.6.1. PREPARATION OF TRANSFORMED *T. BRUCEI* FLUORESCENT-TAGGED GEN1

To localise *Tb*FEN1 protein in *T. brucei* cells throughout the cell cycle, a PCR only tagging (pPOT) tagging strategy was used to generate DNA constructs enabling expression of *TbF*EN1 fusion proteins in which the open reading frame for mNeon Green was fused in frame at the N-terminus.

Component	**Volume**	**Concentration**
Forward Primer	5 µL	5 µM
Reverse Primer	5 µL	5 µM
pPOTv6 mNG template DNA	3 µL	25 µg/ml
H_2O	12 µL	
2 X Pfu High Fidelity PCR Mix	25 µL	1X

TABLE 2.6.1-1 PCR MIX

amplified by high fidelity PCR using gene-specific primers. The reaction mixture was then cycled according to Table 2.6.1-2.

Number of Cycles	**Stages**	**Temperature (°C)**	**Time**
1	Initial Denaturation	94	5 mins
29	Denaturation	94	30 secs
	Annealing	55	30 secs
	Elongation	72	1 min/1Kb
1	Extension	72	2 mins
	Hold	4	

TABLE 2.6.1-2 TABLE OF PCR CYCLING PARAMETERS

The amplified pPOT DNA was purified using a GeneJET PCR purification kit and transfected directly into the *T. brucei* cells. Each transfection required $3x10^7$ cells ml^{-1}, and so an appropriate volume of cell culture (see above) was pelleted by centrifugation at 400 x g for 10 minutes before being resuspended in 0.5 mL of Zimmerman's postfusion medium (ZPFM) before transferring the sample to 0.4 mm electroporation cuvettes (Bio-Rad) prior to addition of the pPOT DNA (5-10 ng). The cell samples were electroporated using a BTX Electro Square Porator ECM830, by applying 3 sets of 100 µs pulses at 1700 V with 200 ms intervals. The cells were then transferred into 10 ml of SMD-79 media (supplemented with 10% foetal calf serum and 0.008 mM Hemin) and left for >16 hours to recover. The cells were then diluted to $5x10^5$ cells ml^{-1} in SMD-79 and selected using blasticidin.

2.6.2. Slide Preparation

Procyclic trypanosomes expressing the mNG::*Tb*FEN1 fusion protein were centrifuged at 1811 × *g* for 3 minutes in a bench top micro-centrifuge. The supernatant was then discarded and approximately 500 µL of 1x PBS used to wash the pellet before re-centrifugation at 1811 × *g* for a further 3 minutes. The cell pellet was then resuspended in a further 500 µL of 1x PBS and pipetted onto slides before leaving the cells to settle and removing the excess solution. Whole cells were then fixed in 100% methanol precooled to -20°C prior to use.

Slides were removed from the 100% methanol and rehydrated in 1x PBS for ~10 mins, prior to 1-hour incubation in blocking buffer in a humidified chamber. Excess buffer was removed, and the slides incubated in ~50 µL of appropriate primary antibody to tag the cell flagellum for one hour, prior to washing 3 times in wash buffer. The washed slides were subsequently incubated with the appropriate secondary antibody in the humidified chamber for 1 hour. Slides were then washed a further 3 times in wash buffer, the excess buffer removed, and the slide mounted using a single drop of Vectashield Mounting Medium containing DAPI (Vectorlabs). A coverslip was placed over the sample and the edges sealed with nail varnish before storing the slides wrapped in foil at 4°C prior to use.

Antibody	**Specificity**	**Dilution**
Primary – L8C4 mouse monoclonal IgG	Anti-PFR2	1:25 with 1% BSA
Secondary – IgG TRITC-conjugated	Anti-mouse IgG	1:10000 in 1% BSA

Table 2.6.2-1 - Table of Antibodies

2.6.3. Deconvolution Microscopy

Cells were imaged using an Applied Precision DeltaVision Deconvolution microscope and images processed using SoftWoRx software. Images were captured in a 30 image z-stack (0.15 µm interval). The resulting files were deconvolved through 10 cycles and finally the sections and fluorescent channels merged to produce a final focused image. The final selection of image panes was corrected simultaneously, so as to produce the clearest possible image while maintaining the same setting across all panes.

3. IDENTIFICATION OF KEY AMINO ACIDS AND GENERATION OF TARGETED AMINO ACID SUBSTITUTIONS

3.1. BIOINFORMATICS

3.1.1. *Tb*FEN1 SEQUENCE ANALYSIS AND PHYLOGENY STUDIES

To study how highly conserved *Tb*FEN1 is between species, a BLAST search was carried out upon the sequence and a number of sequences from a wide variety of species were selected. In particular, the decision was made to focus on species that have been extensively studied by the scientific community for various reasons, e.g. *D. melanogaster* which has often been used in biological research into genetics, and life history evolution; and *X. laevis* which has been used in a large amount of molecular and developmental research due to its close evolutionary relationship with humans and its experimental tractability. The *P. troglodyte* and *G. gorilla* sequences were chosen to provide a good comparison for hose conserved the human sequence has been compared to other closely related species. Meanwhile, organisms such as *O. aries, B. taurus* and *C. l. familiaris* were specifically chosen for how distinct they were likely to be from *T. brucei*.

The amino acid sequences of human GEN1, human FEN1 and *T. brucei* FEN1 were analysed alongside 10 other chosen organisms to study the possible phylogeny of the different endonucleases. The sequences were analysed using MEGA7, first by aligning them all by ClustalW and then identifying the number of conserved residues – 358/937 variable sites (38.2%) and 229/937 singleton sites (24.4%) – although the majority of the alignment length was due to the far larger human GEN1 protein, and repeated alignments without HsGEN1 produced a higher proportion of variable sites (249/423 or 58.9%) and singleton sites (170/423 or 40.2%). MEGA was then used to produce a maximum-likelihood phylogenetic tree of the different endonucleases, shown in Figure 3.1.1-1.

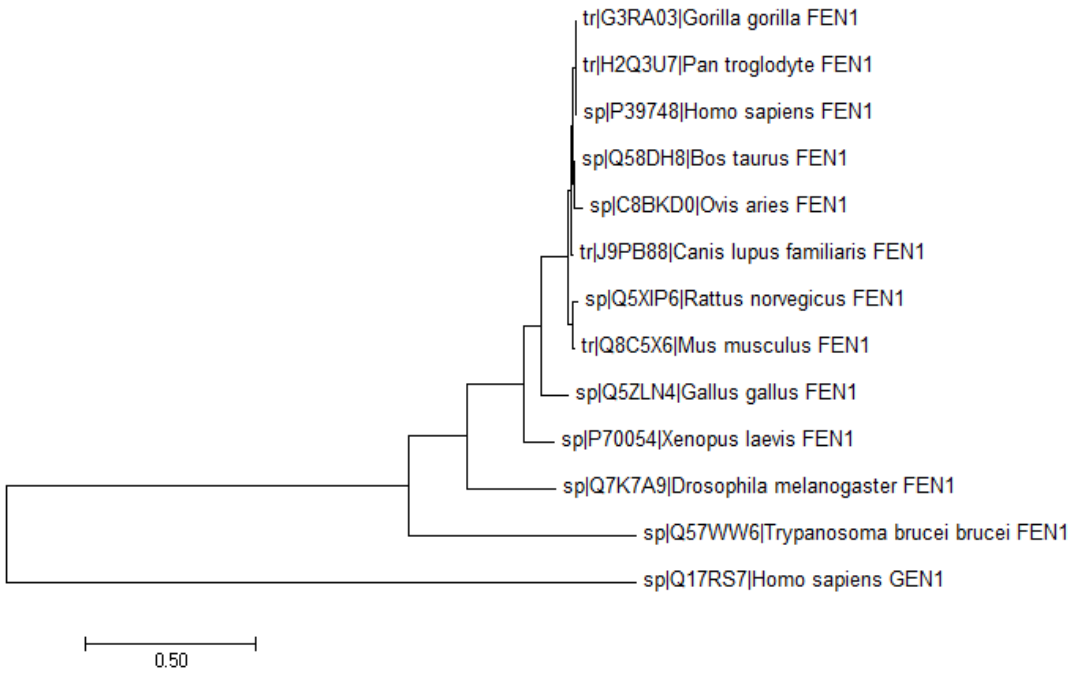


FIGURE 3.1-1 MOLECULAR PHYLOGENETIC ANALYSIS BY MAXIMUM LIKELIHOOD METHOD

THE EVOLUTIONARY HISTORY WAS INFERRED BY USING THE MAXIMUM LIKELIHOOD METHOD BASED ON THE LE_GASCUEL_2008 MODEL. THE TREE WITH THE HIGHEST LOG LIKELIHOOD (-3697.7480) IS SHOWN. INITIAL TREE(S) FOR THE HEURISTIC SEARCH WERE OBTAINED AUTOMATICALLY BY APPLYING NEIGHBOUR-JOIN AND BIONJ ALGORITHMS TO A MATRIX OF PAIRWISE DISTANCES ESTIMATED USING A JTT MODEL, AND THEN SELECTING THE TOPOLOGY WITH SUPERIOR LOG LIKELIHOOD VALUE. A DISCRETE GAMMA DISTRIBUTION WAS USED TO MODEL EVOLUTIONARY RATE DIFFERENCES AMONG SITES (5 CATEGORIES (+G, PARAMETER = 1.5321)). THE RATE VARIATION MODEL ALLOWED FOR SOME SITES TO BE EVOLUTIONARILY INVARIABLE ([+I], 0.2817% SITES). THE TREE IS DRAWN TO SCALE, WITH BRANCH LENGTHS MEASURED IN THE NUMBER OF SUBSTITUTIONS PER SITE. THE ANALYSIS INVOLVED 13 PROTEIN SEQUENCES. ALL POSITIONS CONTAINING GAPS AND MISSING DATA WERE ELIMINATED. THERE WERE A TOTAL OF 355 POSITIONS IN THE FINAL DATASET. EVOLUTIONARY ANALYSES WERE CONDUCTED IN MEGA7.

As well as assembling a maximum-likelihood phylogenetic tree, the human and *D. melanogaster* GEN1 sequences (UniProt entries Q17RS7 and Q9VRJ0, respectively) were aligned with the *Tb*FEN1 sequence in order to take a closer look at how much the different domains of the well-studied *Hs*GEN1 protein translated onto the *Tb*FEN1 protein and to study the levels of conservation with any areas of homology between the different proteins (figure. 3.1.1-2). There were a number of conserved and semiconserved residues, especially those that were identified as key to human GEN1 from previous research, with areas of homology generally clustered around these regions. This would support the hypothesis that these proteins occupy similar biological niches, as the conserved domains would likely fold to create areas of similar topology, and ultimately create similar 3D structures. There were, however, also large sections of unconserved regions, and indeed both GEN1 sequences were approximately twice the length of the *Tb*FEN1 sequence. Of the *Hs*GEN1 domains, *Tb*FEN1 had a large amount of homology with *Hs*GEN1 in the areas around the XPG-n domain (resides 2-96 of *Hs*GEN1) and the XPG-I domain (residues 122-208), but comparatively little homology was present in the areas around the 5'-3' exonuclease domain (residues 208-384) and the chromodomain (residues

390-464) of *Hs*GEN1. All data for this alignment and subsequent analysis was obtained from UniProt

```
CLUSTAL O(1.2.4) multiple sequence alignment


SP|Q57WW6|FEN1_TRYB2 MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE 60
SP|Q17RS7|GEN_HUMAN  MGVNDLWQILEP----VKQHIPLRNLGGKTIAVDLSLWVCEAQTVKK-----------MM 45
SP|Q9VRJ0|GEN_DROME  MGVKELWGVLTP----HCERKPINELRGKKVAIDLAGWVCESLNVVD-----------YF 45
                     ***  *  :*        :.  :.   *: :*:* :  * :   . .

SP|Q57WW6|FEN1_TRYB2 AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE 120
SP|Q17RS7|GEN_HUMAN  GSVMKPHLRNLFFRISYLTQMDVKLVFVMEGEPPKLKADVISKRNQSRYGSSG------- 98
SP|Q9VRJ0|GEN_DROME  V-HPRHHLKNLFFRTCYLIWEQVTPVFVLEGVAPKLKSQVIAKRNELQFRGVK------P 98
                           ** .:***   :    :  ::*::*  *.**:. : .*.:    .

SP|Q57WW6|FEN1_TRYB2 EGDDEAMEKMSKRMVRVGR----DQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKA 176
SP|Q17RS7|GEN_HUMAN  ----------KSWSQKTGRSHFKSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGGHV 148
SP|Q9VRJ0|GEN_DROME  KNSPECTQ-SQPSKGDKGRSRFNHVLKQCETLLLSMGIQCVQGPGEAEAYCAFLNKHGLV 157
                              .      **      :.:   :*  :**  **. .**** ** *   . .

SP|Q57WW6|FEN1_TRYB2 WAVGTEDMDALAFGSRVMLRHLTYGEAKK-----RPIAEYHLDEILEASGFSMQQFIDLC 231
SP|Q17RS7|GEN_HUMAN  DGCLTNDGDTFLYGAQTVYRNFTMNTKD------PHVDCYTMSSIKSKLGLDRDALVGLA 202
SP|Q9VRJ0|GEN_DROME  DGVISQDSDCFAYGAVRVYRNFSVSTQGAQAAAGGAVDIYDMREITSRMDFGQQKIIVMA 217
                      .  ::* * : :*:  : *::: .           :  * : .* .  .:. : :: :.

SP|Q57WW6|FEN1_TRYB2 ILLGCDYVPR-ISGIGPHKAWEGIKKYGSLEAFI--ESLDG-------------------- 269
SP|Q17RS7|GEN_HUMAN  ILLGCDYLPKGVPGVGKEQALKLIQILKGQSLLQRFNRWNETSCNSSPQLLVTKKLAHCS 262
SP|Q9VRJ0|GEN_DROME  LLCGCDYCPDGIGGIGKDGVLKLFNKYKETEILDRMRSWRGETDKYNALEIRVDDKSICS 277
                     :* **** *  : *:* . . : ::     . :   .

SP|Q57WW6|FEN1_TRYB2 ----------------------TRYVVPEEFNYK--------DA------------------ 283
SP|Q17RS7|GEN_HUMAN  VCSHPGSPKDHERNGCRLCKSDKYCEPHDYEYCCPCEWHRTEHDRQLSEVENNIKKKACC 322
SP|Q9VRJ0|GEN_DROME  NCGHIGKTQSHTKSGCSVCRTHKGCDESL--------WKEQ---RLSIKSELTLRRKALL 326
                                    :

SP|Q57WW6|FEN1_TRYB2 ----------RNFFLEPEVTPGEEIDIQFREPDEEGLIKFLVDEKLFSKERVLKGIQRLR 333
SP|Q17RS7|GEN_HUMAN  CEGFPFHEVIQEFLLNKDKLV--K-VIRYQRPDLLLFQRFTLEKMEWPNHY--------- 370
SP|Q9VRJ0|GEN_DROME  SPDFPNEEIIAEFLSEPDTIP--NLNLNWRQPNLVKFIKQIGHLLQWPEIY--------- 375
                               :*: : :     :  :.::.*:   : :   .   : :

SP|Q57WW6|FEN1_TRYB2 DALTKKTQGRLDQFFTITKPQKQVNSEAS----------TAGTKRNRGAV-AL------- 375
SP|Q17RS7|GEN_HUMAN  ---------ACEKLLVLLTHYDMIERKLGSRNSNQLQPIRIVKTRIRNGVHCFEIEWEKP 421
SP|Q9VRJ0|GEN_DROME  ---------CFQKFFPILTRWQVQQSK---QEKILIQPHEIIKKRTVKGVPSLELRWHDP 423
                               :::: : .  .  : :              ..*   .* .:

SP|Q57WW6|FEN1_TRYB2 ----PGVLQRKSS----SGHKKAVKK---------------------------------- 393
SP|Q17RS7|GEN_HUMAN  EHYAM-------------EDKQHGEFALLTIEEESLFEAAYPEIVAVYQKQKLEIKGKKQ 468
SP|Q9VRJ0|GEN_DROME  SGIFKGLIPDKQIAEYEAEHPKGIEELYYTIEPLDMLETAYPDLVAAFLKSKEKPAKKTT 483
                                   . :  :

SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  KRIK---PKENNLPEPDEVMS-FQSHMTLKP--TCEIFHKQNSKLNSGISPDPTLPQESI 522
SP|Q9VRJ0|GEN_DROME  RKKKTASEEENKENEPNSKPKRVVRKIKAQPEENQPLLHQFLGRKKEGTPVKAPAPQ--- 540


SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  SASLNSLLLPKNTPCLNAQEQFMSSLRPLA----IQQIKAVSKSLISES-SQPNTSSH-- 575
SP|Q9VRJ0|GEN_DROME  -----------RQQCSTPITKFLPSDLESDCDAEEFDMSDIVKGIISNPNAKPALTNHDG 589


SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  -NI---SVIADLHLSTIDWEGTSFSNSPAIQRNTFSHDLKSEVESELSAIPDGFENIPEQ 631
SP|Q9VRJ0|GEN_DROME  HQLHYEPMAEDLSLRLAQMSLGNVNESPKVETKRDLSQVDQLPQSKRFSLEDSFDLLV-- 647
```

```
SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  LSCESERYTANIKKVLDEDSDGISPEEHLLSGITDLCLQDLPLKERIFTKLSYPQDNLQP 691
SP|Q9VRJ0|GEN_DROME  ----------------KGDL----------QKLARTPVERFKMQHRISEKIPTPVKPLD- 680


SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  DVNLKTLSILSVKESCIANSGSDCTSHLSKDLPG---IPLQNESRDSKILKGDQLLQEDY 748
SP|Q9VRJ0|GEN_DROME  -----NISYF-------FNQSSDNADVF-EELMNSSLVPQDQEDNAEDEEEDDLVVISD- 726


SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  KVNTSVPYSVSNTVVKTCNVRPPNTALDHSRKVDMQTTRKILMKKSVCLDRHSSDEQSAP 808
SP|Q9VRJ0|GEN_DROME  ------------------------------------------------------------


SP|Q57WW6|FEN1_TRYB2 ------------------------------------------------------------
SP|Q17RS7|GEN_HUMAN  VFGKAKYTTQRMKHSSQKHNSSHFKESGHNKLSSPKIHIKETEQCVRSYETAENEESCFP 868
SP|Q9VRJ0|GEN_DROME  ------------------------------------------------------------


SP|Q57WW6|FEN1_TRYB2 ----------------------------------------
SP|Q17RS7|GEN_HUMAN  DSTKSSLSSLQCHKKENNSGTCLDSPLPLRQRLKLRFQST 908
SP|Q9VRJ0|GEN_DROME  ----------------------------------------
```

FIGURE 3.1-2 ALIGNMENT OF *HS*GEN1, *DM*GEN1 AND *TB*FEN1

***T.* BRUCEI PUTATIVE FEN1 AMINO ACID SEQUENCE (UNIPROT ID: Q57WW6) ALIGNED BY CLUSTAL OMEGA WITH *H. SAPIENS* GEN1 SEQUENCE (UNIPROT ID: Q17RS7) AND *D. MELANOGASTER (UNIPROT ID: Q9VRJ0).* GAPS IN THE ALIGNMENT ARE DENOTED BY HYPHENS (-). BELOW THE SEQUENCES IS SYMBOL BASED KEY DENOTING AREAS OF SIMILARITY, WITH CONSERVED SEQUENCES MARKED BY AN ASTERIX (*), CONSERVATIVE MUTATIONS MARKED BY A COLON (:)SEMI CONSERVATIVE MUTATIONS MARKED BY A PERIOD (.) AND BLANK SPACE MARKING UNCONSERVED SEQUENCES ( )KEY RESIDUES ARE HIGHLIGHTED IN GREEN.**

The human FEN1 and *T. brucei brucei* FEN1 sequences were also compared directly, to highlight the variance between them, with 196/394 (49.7%) variable sites, and a pairwise distance of approximately 0.5 amino acid differences per site.

```
TbFEN1 1    MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE  60
            MG+ GL+KL+ D  P AI+E ++K YFGR++AIDASM++YQF+IA++   +G  V L NE
HsFEN1 1    MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVR---QGGDV-LQNE  56

TbFEN1 61   AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE  120
             G+ TSHL G+F+RT+RM++ G++P+YVFDGKPP LK+ EL  R +R  +A+ + ++A+
HsFEN1 57   EGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQA  116

TbFEN1 121  EGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKAWAVG  180
             G ++ +EK +KR+V+V +   +E K LL LMGIP + APSEAEA CA LVK  K +A
HsFEN1 117  AGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKAGKVYAAA  176

TbFEN1 181  TEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP  240
            TEDMD L FGS V++RHLT  EAKK PI E+HL  IL+  G + +QF+DLCILLG DY
HsFEN1 177  TEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVDLCILLGSDYCE  236

TbFEN1 241  RISGIGPHKAWEGIKKYGSLEAFIESLDGTRYVVPEEFNYKDARNFFLEPEVTPGEEIDI  300
             I GIGP +A + I+K+ S+E  +  LD  +Y VPE + +K+A   FLEPEV   E +++
HsFEN1 237  SIRGIGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEVLDPESVEL  296

TbFEN1 301  QFREPDEEGLIKFLVDEKLFSKERVLKGIQRLRDALTKKTQGRLDQFFTIT  351
            ++ EP+EE LIKF+  EK FS+ER+  G++RL  +    TQGRLD FF +T
HsFEN1 297  KWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDFFKVT  347
```

FIGURE 3.1-3 ALIGNMENT OF TBFEN1 WITH HSFEN1

T. BRUCEI PUTATIVE FEN1 AMINO ACID SEQUENCE (ENTREZ GENE ID: 3656016) ALIGNED WITH H. SAPIENS FEN1 AMINO ACID SEQUENCE (ENTREZ GENE ID: 2237). KEY RESIDUES ARE HIGHLIGHTED IN GREEN.

3.1.2. Identification of Key Amino Acids in *Tb*FEN1

Following alignment, the UniProt entries of the human FEN1 and *T. brucei brucei* FEN1 protein were used, in conjunction with the previously generated alignment to identify the residues in *T. brucei* equivalent to those within the human FEN1 protein with characterised functional roles (Table 3.1.2-1). By identifying those residues that are key to the function of the *Hs*FEN1 protein, it would be possible to infer that those respective residues were also key to the function of the *Tb*FEN1 protein. Therefore, those residues should serve as appropriate targets for mutagenesis experiments to determine if any drastic changes to their characteristics had an effect of *Tb*FEN1's cleavage activity.

Position - Human	Human Amino Acid	Description	T. brucei Amino Acid	Position – T. brucei	Supporting Experimental Data
34	Aspartic Acid (D)	Metal Binding – Magnesium 1	Aspartic Acid (D)	34	Mutagenesis (Shen, et al., 1996)
47	Arginine (R)	Binding Site – DNA Substrate	Lysine (K)	47	Mutagenesis (Qui, et al., 2002)
70	Arginine (R)	Binding Site – DNA Substrate	Arginine (R)	74	Mutagenesis (Qui, et al., 2002)
86	Aspartic Acid (D)	Metal Binding – Magnesium 1	Aspartic Acid (D)	90	Mutagenesis (Shen, et al., 1996)
158	Glutamic Acid (E)	Metal Binding – Magnesium 1 Binding Site – DNA Substrate	Glutamic Acid (E)	162	Mutagenesis (Shen, et al., 1996)
160	Glutamic Acid (E)	Metal Binding – Magnesium 1	Glutamic Acid (E)	164	Mutagenesis (Frank, et al., 1998)
179	Aspartic Acid (D)	Metal Binding – Magnesium 2	Aspartic Acid (D)	183	Mutagenesis (Shen, et al., 1996)
181	Aspartic Acid (D)	Metal Binding – Magnesium 2	Aspartic Acid (D)	185	Mutagenesis (Shen, et al., 1996)
231	Glycine (G)	Binding Site – DNA Substrate	Glycine (G)	235	Mutagenesis (Shen, et al., 1996)
233	Aspartic Acid (D)	Metal Binding – Magnesium 2 Binding Site – DNA Substrate	Aspartic Acid (D)	237	Mutagenesis (Shen, et al., 1996)

Table 3.1.2-1 Key Amino Acids in Human FEN1 and Their Equivalents in T. brucei FEN1

Negatively charged residues are highlighted in purple, positively charged residues in green, and non-polar in orange.

3.2. SELECTION OF MUTANTS

3.2.1. Identify Appropriate Substitutions at each Key Site

With the key residues identified for mutagenesis, it was decided that each would be substituted for a neutral residue – Alanine – and an oppositely charged residue – Lysine in place of any

acidic residues, Aspartic Acid for any basic residues and in the case of the 235 Glycine residue, substitutions were then made for all of the above cases (Table 3.2.1-1). Using neutral residues and oppositely charged residues would hopefully highlight how crucial each target was both on its own – since the neutral substitution should simply prevent it from interacting with either the DNA or Mg^{2+} ions – and in the larger picture of the structure-function relationships of the protein – since oppositely charged residues should actively disrupt the proteins structure, or at the very least work antagonistically to the other key residues.

Position	**Original Residue**	**Substituted Residues**
34	Aspartic Acid (D)	Alanine (A)
		Lysine (K)
47	Lysine (K)	Alanine (A)
		Aspartic Acid (D)
74	Arginine (R)	Alanine (A)
		Aspartic Acid (D)
90	Aspartic Acid (D)	Alanine (A)
		Lysine (K)
162	Glutamic Acid (E)	Alanine (A)
		Lysine (K)
164	Glutamic Acid (E)	Alanine (A)
		Lysine (K)
183	Aspartic Acid (D)	Alanine (A)
		Lysine (K)
185	Aspartic Acid (D)	Alanine (A)
		Lysine (K)
235	Glycine (G)	Alanine (A)
		Aspartic Acid (D)
		Lysine (K)
237	Aspartic Acid (D)	Alanine (A)
		Lysine (K)

TABLE 3.2.1-1 RESIDUES TARGETED FOR SITE DIRECTED MUTAGENESIS AND CHOSEN SUBSTITUTIONS

3.2.2. PRIMER DESIGN & MUTAGENESIS

With the list of mutants prepared, forward and reverse primers were designed using the Agilent primer design software for each of the 21 mutants in table 3.2.1-1. The primers were then used to introduce the desired mutations via site-directed mutagenesis into the pET24a plasmid with the *Tb*FEN1 gene already present from previous studies. The new mutant plasmids were transformed into *E. coli* for amplification and then the purified plasmids sent off for sequencing.

3.2.3. Confirming Successfully Modified Plasmids

When the sequencing data was received it was used to determine if the mutagenesis had been completely successful, rather than waste time and resources introducing a redundant change to the target sequence. The introduced mutation was first located on both the forward and reverse sequences by searching for the complementary section to each primer on the Chromas™ software (Figure 3.2.3-1). The nucleotide sequence was then converted to the amino acid sequence using the online ExPASy translation tool and confirming the presence of the new amino acid. Finally, a BLASTP search was carried out on the new amino acid sequence to compare it to the wild type sequence and ensure no other mutations had been introduced in the process. With the successful mutagenesis confirmed, the plasmids were transformed into Tuner cells for expression. In total, 23 mutant *Tb*FEN1 proteins were designed and plasmids created for each, with 13 successfully expressed in *E. coli*.

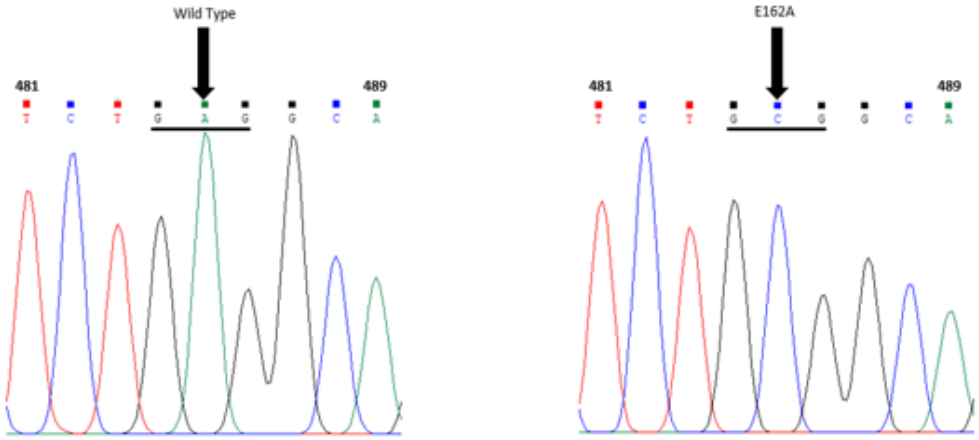
Figure 3.2-1 Sequence Analysis for the TbFEN1 E162A amino acid substitution

*i) A Chromas trace illustrating the single nucleotide substitution introduced from the Wild type base triplet **GAG** coding for glutamic acid, to the base triplet **GCG** coding for alanine.*

*ii) A protein **BLAST** showing that alanine is encoded instead of glutamic acid at position **162**.*

*Nucleotides are numbered according to nucleotide **1** representing the adenine base of the start codon ATG, at the beginning of the **TbFEN1** gene*

i)



ii) TbGEN1 protein [*Trypanosoma brucei brucei*]
Identities = 392/393 (99%)

```
Query  1    MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE  60
Sbjct  1    MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE  60

Query  61   AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE  120
Sbjct  61   AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE  120

Query  121  EGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSAAEAQCAELVKKNKAWAVG  180
Sbjct  121  EGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKAWAVG  180

Query  181  TEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP  240
Sbjct  181  TEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP  240

Query  241  RISGIGPHKAWEGIKKYGSLEAFIESLDGTRYVVPEEFNYKDARNFFLEPEVTPGEEIDI  300
Sbjct  241  RISGIGPHKAWEGIKKYGSLEAFIESLDGTRYVVPEEFNYKDARNFFLEPEVTPGEEIDI  300

Query  301  QFREPDEEGLIKFLVDEKLFSKERVLKGIQRLRDALTKKTQGRLDQFFTITKPQKQVNSE  360
Sbjct  301  QFREPDEEGLIKFLVDEKLFSKERVLKGIQRLRDALTKKTQGRLDQFFTITKPQKQVNSE  360

Query  361  ASTAGTKRNRGAVALPGVLQRKSSSGHKKAVKK  393
Sbjct  361  ASTAGTKRNRGAVALPGVLQRKSSSGHKKAVKK  393
```

4. EXPRESSION OF MUTANT AND WILD TYPE *TB*FEN1 IN E. COLI

4.1. EXPRESSION OF *TB*FEN1 PROTEINS IN TUNER CELLS

4.1.1. SMALL-SCALE INDUCTION AND PURIFICATION

In order to express recombinant *Tb*FEN1 proteins, expression plasmids with the desired mutations were transformed into the *E. coli* strain Tuner (DE3) pLysS. The pLysS expression system works by a gene encoding the T7 RNA polymerase sitting on a defective lambda phage (known as DE3) under the control of a *lac* promoter. The pLysS plasmid carries the chloramphenicol resistance gene, so inclusion of chloramphenicol in media selects for the presence of pLysS plasmid. It also encodes bacteriophage T7 lysozyme which both inhibits any low-level expression of T7 RNA polymerase prior to induction and also helps cells lysis. The recombinant plasmid (pJGO4) used to express the *Tb*FEN1 protein is derived from the kanamycin resistant pET24a vector and expressed the *TbF*EN1 protein with a C-terminal His tag.

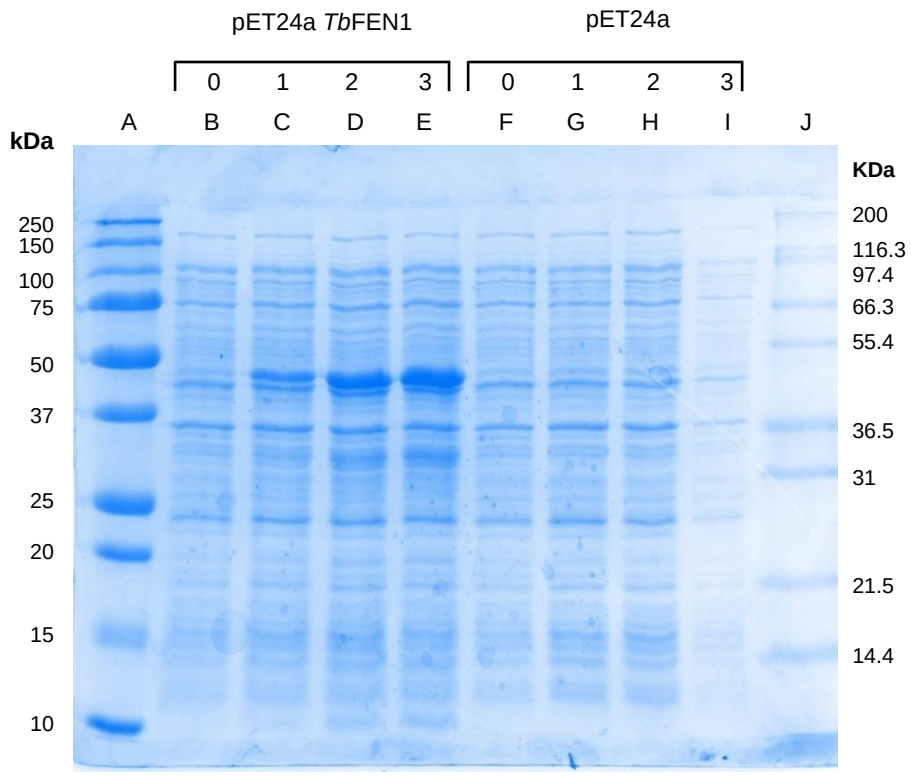


FIGURE 4.1-1 TIME COURSE OF TBFEN1 EXPRESSION IN TUNER CELLS CONTAINING PET24A GEN1 OVER 3 HOURS, COMPARED TO TUNER CELLS CARRYING THE PET24A VECTOR, FOLLOWING ADDITION OF IPTG

LANES A BIORAD PRECISION PLUS PROTEIN STANDARD MARKER (PROTEIN SIZES INDICATED IN KDA, LEFT). LANES B-E, TIME COURSE CELL EXTRACTS PREPARED FROM TUNER CARRYING RECOMBINANT PET24A-TBFEN1 PLASMID. LANES F-I, CELL EXTRACTS PREPARED FROM TUNER CARRYING PET24A VECTOR. LANE J, THERMOFISCHER MARK12 UNSTAINED STANDARD MARKER

All mutant proteins were first purified on a small scale to ensure the desired expression occurred, as well as providing an opportunity to optimise the induction protocol for the highest quantity of soluble protein extract. Samples were taken throughout the induction and purification process and visualised on SDS-PAGE gels. Figure 4.1.1. - 1 shows proteins visualised from bacterial cell extracts prepared from Tuner cells carrying either vector pET24a or recombinant plasmid pET24a FEN1 over a 3 hour period of expression. The presence of the *Tb*FEN1 protein can be clearly seen from the increasingly dense bands around 45 kDa, in lanes C, D and E.

Following successful demonstration of *TbF*EN1 expression, the bacterial pellet was collected and frozen at the end of the time course and frozen at -80°C. Prior to purification the desired pellet was thawed, lysed and pushed over a 45 micron filter, before applying the resulting soluble fraction to a QIAGEN Ni-NTA spin column. Following sequential washes with a low imidazole buffer to remove proteins binding non-specifically to the Ni-NTA resin, bound proteins were eluted using a buffer containing 500 mM imidazole. Samples collected from the induced bacterial culture, protein loaded onto the column, and from each wash and elution step were analysed by SDS-PAGE. An example of a small-scale purification on a Ni-NTA spin column is presented in Figure 4.1.1-2. Analysis of the whole cell extract (lane B) shows the presence of an abundant induced protein band with a predicted mass of 45 kDa, confirming high level expression of *Tb*FEN1. However, a comparison of the abundance of *Tb*FEN1 in the soluble fraction (lane c) with that in the whole cell extract (lane B) indicated that the majority of *Tb*FEN1 expressed under these conditions is insoluble. The majority of proteins loaded on the column do not bind to the column and appear in the flow through fraction (lane D) or are removed with low imidazole (lane E). Some of the His-tagged *TbF*EN1 is however retained on the column and eluted with 500 mM imidazole (lanes H and I).

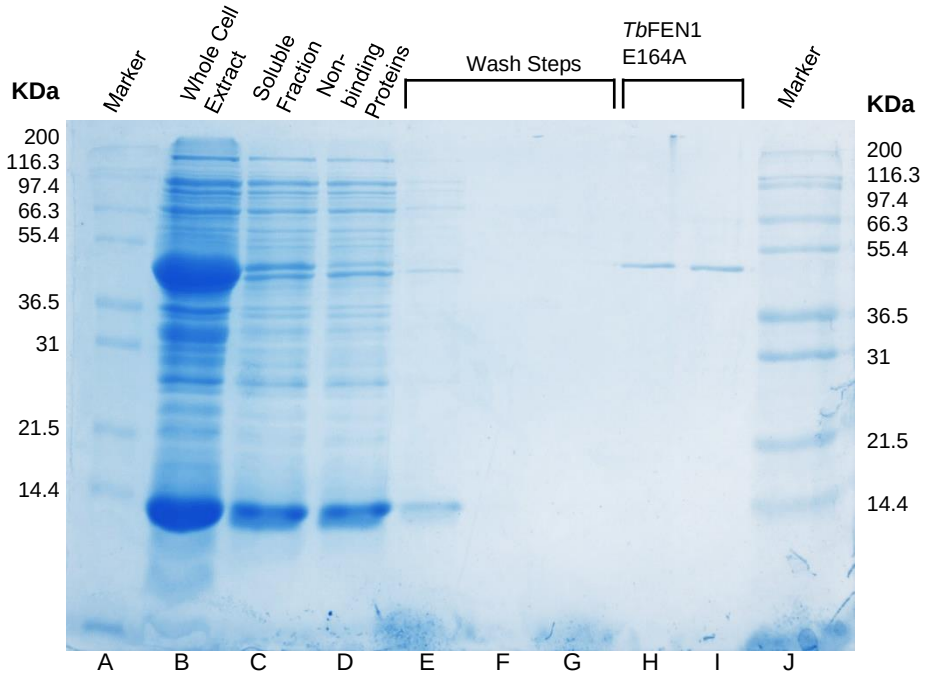


FIGURE 4.1-2 NI2+NTA SPIN COLUMN PURIFICATION OF TBFEN1 E164A MUTANT AT 37°C

HIS-TAGGED TBFEN1 WAS PURIFIED FROM BACTERIA PELLETED FROM APPROXIMATELY 20 ML OF INDUCED CULTURE AS DESCRIBED IN METHOD 1.3.6.

LANES A & J, MARK 12 MW MARKERS (PROTEIN SIZES INDICATED IN KDA, ON LEFT). LANE B, WHOLE CELL EXTRACT FROM TUNER CARRYING RECOMBINANT PET24A-TBFEN1 E164A PLASMID AFTER INDUCTION WITH IPTG. LANE C, SOLUBLE FRACTION. LANE D, NON-BINDING E.COLI PROTEINS. LANES E-G, WASH STEPS. LANES H & I, ELUTED HIS-TAGGED TBFEN1 E164A PROTEIN.

To try and optimise the yield, the effect of changing the temperature of expression conditions upon solubility was investigated. Therefore, mutant E164A was expressed again but this time at 30°C rather than 37°C as previously (Figure 4.1.1 – 3). Analysis of the whole cell extract (lane B) shows that the total expression of *TbF*EN1 has fallen slightly after lowering the temperature of the expression step, but has still remained high enough for this method to be effective. However, comparison of the abundance of *TbF*EN1 in the soluble fraction (lane C) with the whole cell extract (lane B), demonstrates that although there is still a significant difference, there is a smaller disparity than at 37°C suggesting a greater abundance of soluble *Tb*FEN1 at 30°C. Furthermore, the amount of His-tagged *TbF*EN1 finally eluted from the column with a 500 mM imidazole wash appears far greater at this lower temperature.

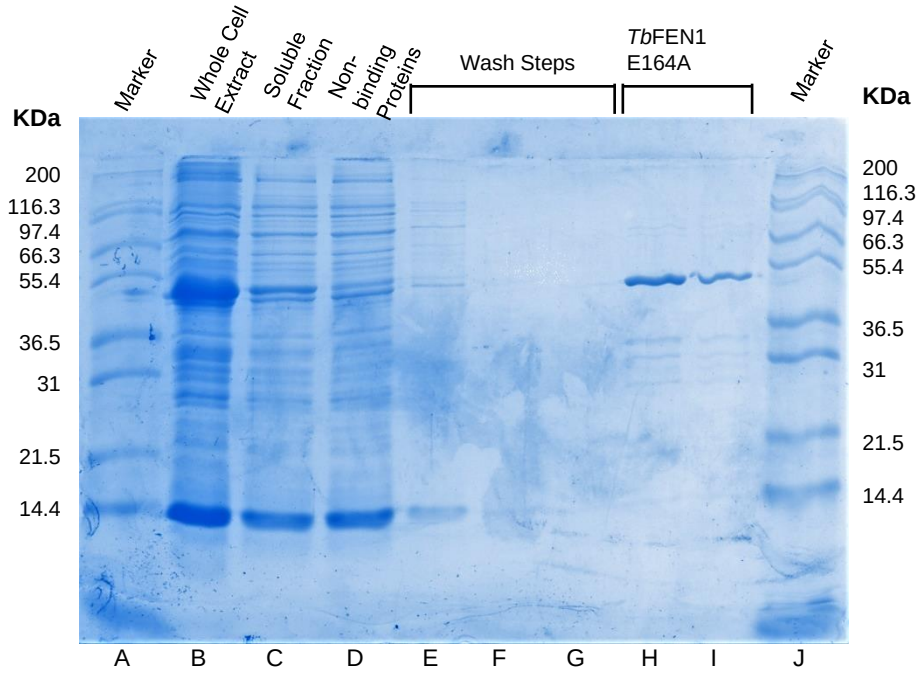


FIGURE 4.1.1-3 NI2+-NTA SPIN COLUMN PURIFICATION OF TBFEN1 E164A MUTANT AT 30°C

LANES A & J, MARK 12 MW MARKERS (PROTEIN SIZES INDICATED IN KDA, ON LEFT). LANE B, WHOLE CELL EXTRACT FROM TUNER CARRYING RECOMBINANT PET24A-TBFEN1 E164A PLASMID AFTER INDUCTION WITH IPTG. LANE C, SOLUBLE FRACTION. LANE D, NON-BINDING E.COLI PROTEINS. LANES E-G, WASH STEPS. LANES H & I, ELUTED HIS-TAGGED TBFEN1 E164A PROTEIN.

4.1.2. LARGE-SCALE INDUCTION & LYSIS

Having tested His-Tag purification as an effective method of initial purification and determined that induction at 30°C produced a higher concentration of soluble protein, the process was scaled up to produce sufficient quantities of each mutant for carrying out cleavage assays. This was mostly a simple case of scaling up the volume of liquid cultures from 25 ml to 800 ml and implementing the optimisation of 30°C induction with IPTG. However, after purifying a first batch of wild type *TbF*EN1, the point was raised that the lysis buffer that the pellets were resuspended in contained the endonuclease Benzonase, which may have an adverse effect on the perceived activity of the *TbF*EN1 we were trying to analyse – itself being an endonuclease. Although the Benzonase should be purified from the suspension by the subsequent His-tag and Anion-exchange liquid chromatography, the decision was therefore made to carry out subsequent lysis without adding Benzonase to the lysis buffer to be sure that *TbF*EN1 was responsible for any DNA cleavage in later assays.

4.1.3. His-Tag Purification and Pooling

Wild type and mutant *TbF*EN1 proteins were harvested from pelleted expression cultures by lysis, centrifugation to remove insoluble material, and subsequent purification by liquid chromatography over an immobilised Ni^{2+} affinity column using the AKTA-prime machine as detailed in method 1.4.3. Samples were taken from the whole cell extract, flow through, and from those fractions indicated to have the greatest protein abundance by a 280 nm UV trace. The collected samples were then analysed by SDS-PAGE according to method 1.4.5.

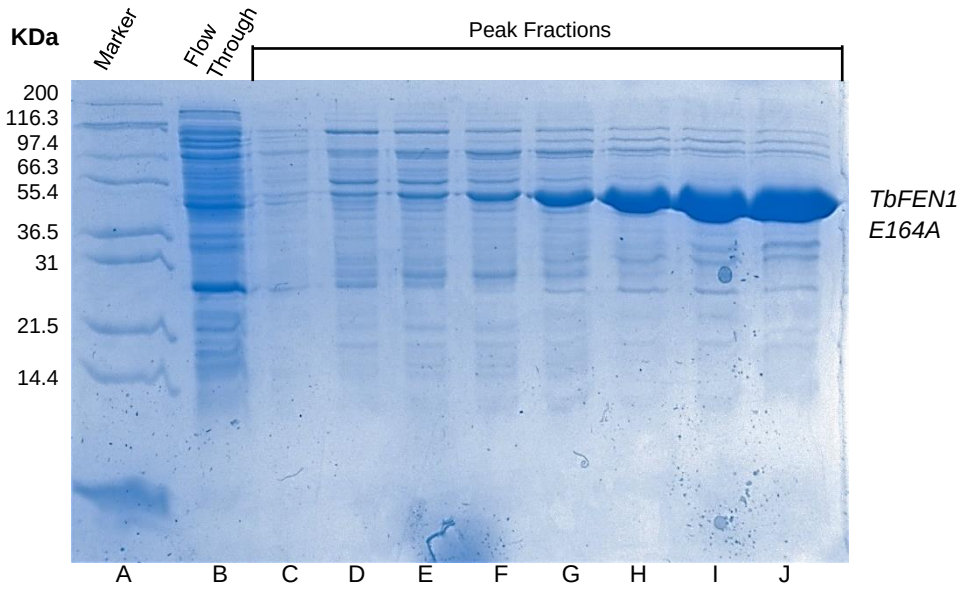


Figure 4.1.3-1 Large scale purification of TbFEN1 E164A by Ni2+ affinity liquid chromatography

Lane A, Mark-12 MW marker (Protein sizes indicated in kDa, on left). Lane B, Flow through. Lanes C-J, Fractions 6-13 collected across the UV peak. Fractions 11-13 (Lanes H-J) were identified as having the highest abundance of TbFEN1 E164A and thus pooled for further purification by Anion Exchange Liquid Chromatography.

As shown in Figure 4.1.3-1, purifying *Tb*FEN1 protein using a Ni^{2+} column and a linear gradient of 20-500 mM imidazole buffer allows for the gradual elution of all the proteins in the cell extract, according to how strongly they are bound to the column. The majority of bacterial proteins do not bind and are washed off immediately (lane B) or at low imidazole concentrations. As the concentration of imidazole in the elution buffer rises the more strongly bound proteins begin to elute, which corresponds to a gradual incline on the UV trace. If the amount of target protein expressed is sufficient then this will give rise to a peak on the UV trace as the protein is finally displaced by the imidazole. However, the UV trace only gives a rough idea of which fractions the eluted protein is in. therefore it is necessary to analyse those fractions across the peak of the trace using SDS-PAGE (Figure 4.1.3-1). Analysis of the different fractions (lanes C-J) shows a gradually increasing band around 45 kDa, corresponding to the purified *Tb*FEN1 protein, and specifically that the greatest abundance of *Tb*FEN1 is in those fractions corresponding to lanes H, I and J. It is also clear from this figure that while the His-tag purification was effective in concentrating the protein of interest from 16 ml of cell lysate to simply three 1 ml fractions, and in purifying a significant quantity of unwanted *E. coli* proteins from the suspension, there was still a large amount of contaminant protein in these 3

fractions – seen as fainter bands above and below the *Tb*FEN1 bands. It is for this reason that a second purification step using anion exchange chromatography was also performed.

4.1.4. Anion-Exchange Purification

Following initial purification of the target *Tb*FEN1 protein from the cell lysate it was still necessary to further isolate it from a number of contaminant co-purified *E. coli* proteins still present in the fraction. Therefore, pooled fractions were prepared according to method 1.4.4. and loaded on to a MonoQ 5/50 GL column, to be further purified by anion-exchange chromatography. Anion exchange chromatography was chosen following previous work within the lab by McAllister, et al. which had identified that this method gave a very good separation between *Tb*FEN1 and any remaining contaminants, although even this method was not 100% effective at completely isolating *Tb*FEN1 (McAllister & Benson, 2015). Bound proteins were then eluted using a NaCl concentration gradient. The 280 nm UV trace was again used to identify peak fractions and samples analysed by SDS-PAGE (Figure 4.1.4-1) to determine which fractions contained the highest concentration of purified target protein. Analysis of the eluted fractions (lanes C-I) shows the gradual increase in abundance of the distinctive *Tb*FEN1 band at 45 kDa. Comparing the lanes on this gel with the greatest abundance of *Tb*FEN1 (lanes D, E and F) to those on Figure 4.1.3-1 containing the greatest abundance of *Tb*FEN1 (lanes H-J), it is clear that the second purification step has been successful in removing many contaminants. Comparing lanes D-F to the flow through (lane B) it is apparent that most of those proteins greater than 45 kDa in mass have been isolated from the final sample. However, there are still multiple bands with smaller mass than *Tb*FEN1 in those samples – these are most likely degradation products however.

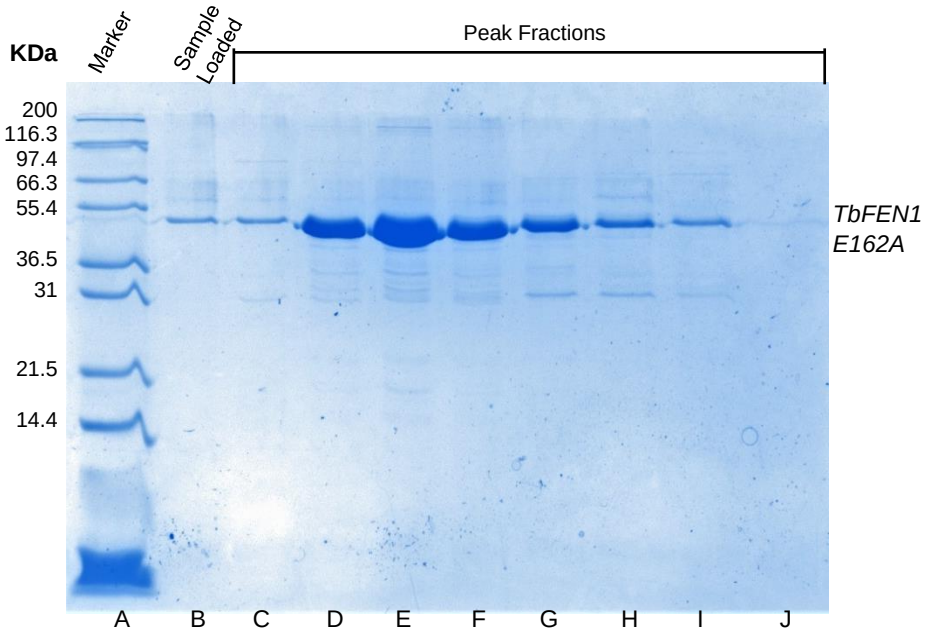


FIGURE 4.1-1 LARGE SCALE PURIFICATION OF TBFEN1 E162A BY ANION EXCHANGE LIQUID CHROMATOGRAPHY

POOLED FRACTIONS FROM HIS-TAG PURIFICATION WERE DILUTED 1 IN 10 WITH 0 M NACL AND RUN THROUGH A MONOQ COLUMN USING THE ÄKTA FPLC SYSTEM TO PURIFY THE PROTEIN OF INTEREST FROM ANY CONTAMINANT E. COLI PROTEINS.

LANE A, MARK-12 MW MARKER (PROTEIN SIZES INDICATED IN KDA, ON LEFT). LANE B, SAMPLE LOADED ONTO THE COLUMN. LANES C-I, FRACTIONS A14-B11 (SERPENTINE FRACTION COLLECTION WAS USED). FRACTIONS A15, B15 & B14 WERE SELECTED FOR USE IN CLEAVAGE ASSAYS

Purified protein fractions were then analysed to determine the concentration of protein in each fraction by measuring the absorbance of each fraction at 280 nm compared to a blank elution buffer. The absorbance was then used to calculate the protein concentration as detailed in section 2.5. The measured concentrations of each protein can be found in Appendix I.

5. ASSAYING MUTANT CLEAVAGE

5.1. NATIVE PAGE

5.1.1. *Tb*FEN1 AS A HJ RESOLVASE COMPARED TO T7 ENDONUCLEASE 1

In order to investigate whether purified *Tb*FEN1 was able to cleave Holliday junctions, a synthetic HJ was incubated with a range of concentrations of purified wild type protein, and following deproteinization, products and substrates were separated by native PAGE. Substrates and products with Cy5 5' end labels were visualised using the Cy5 channel on a Bio-Rad Gel doc system. In parallel with assays using *Tb*FEN1, as a positive control the synthetic Holliday junction was incubated with bacteriophage T7 endonuclease 1, which should validate the synthetic substrate and provide a comparison for *Tb*FEN1's cleavage activity (Figure 5.1.1.-1). Both enzymes were used at a number of different concentrations to ensure that this was not a limiting factor. If *Tb*FEN1 could cleave HJs, then the larger band produced by the HJ substrate at the top of the gel should be replaced by a band further down, in line with the dsDNA marker, as the HJ was cleaved into two dsDNA molecules, thus confirming *Tb*FEN1s activity which should mirror that of known HJ resolvase – T7endo1.

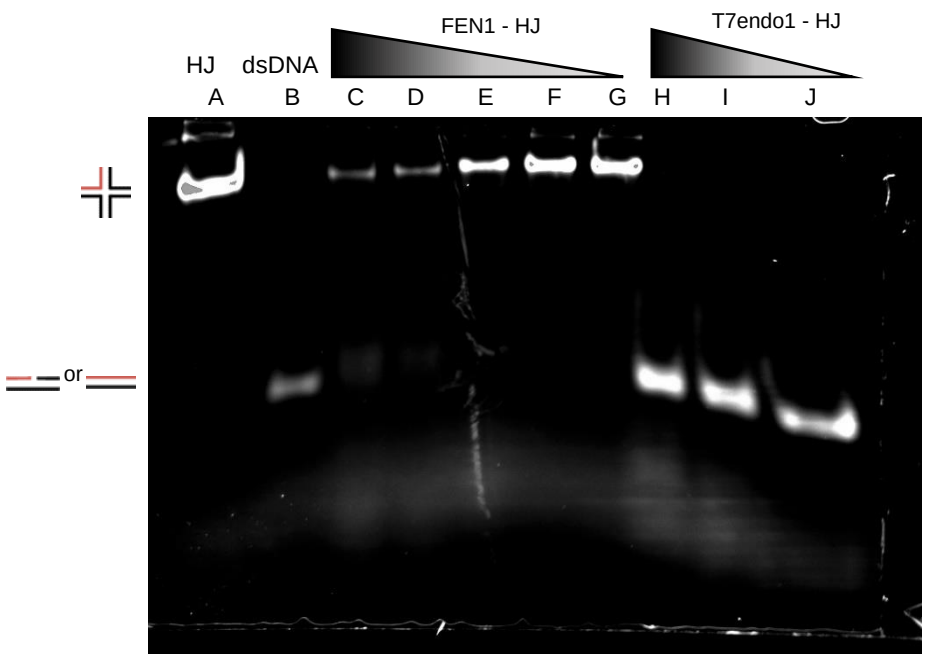


FIGURE 5.1.1-1 NATIVE PAGE ANALYSIS OF TBFEN1 AND T7ENDO1 HOLLIDAY JUNCTION CLEAVAGE

SYNTHETIC HOLLIDAY JUNCTION SUBSTRATES WERE USED TO ANALYSE THE CLEAVAGE ACTIVITY OF TBFEN1 COMPARED TO KNOWN HJ RESOLVASE T7ENDO1. TWO-FOLD DILUTIONS OF THE ENZYMES IN 10X CLEAVAGE BUFFER AND WATER WERE USED TO ENSURE THAT A WIDE RANGE OF ENZYME CONCENTRATIONS WERE TESTED IN ORDER TO REMOVE THIS AS A POSSIBLE LIMITING FACTOR.

LANE A, 200 NM HOLLIDAY JUNCTION SUBSTRATE. LANE B, 200 NM DSDNA SUBSTRATE. ALL OF LANES C-J, CONTAIN 200 NM HJ SUBSTRATE. LANE C, 4 µM TBFEN1. LANE D, 2 µM TBFEN1. LANE E, 1 µM TBFEN1. LANE F, 0.5 µM TBFEN1. LANE G, 0.25 µM TBFEN1. LANE H, 2 µM T7ENDO1. LANE I, 1 µM T7ENDO1. LANE J, 0.5 µM T7ENDO1.

In actuality, the results presented in figure 5.1.1-1 show that the HJ is cleaved efficiently by T7 endonuclease I to produce products with a mobility similar to that of a 60 bp ds DNA molecule (compare lanes H-J and lane B), indicative of symmetrical cleavage across the HJ by a resolvase activity. In contrast, although the substrate HJ band appears to decrease with increasing *Tb*FEN1 protein, this is not accompanied by an appearance of a band consistent with introduction of symmetrically placed nicks across the HJ. From this it can be concluded that under the conditions used in this assay *Tb*FEN1 is not cleaving HJs. Increased concentrations of *Tb*FEN1 lead to the increase in a smeared band with an increased mobility compared with the dsDNA marker, suggesting some non-specific nucleolytic activity of *Tb*FEN1 on the HJ substrate. As expected, the T7endo1 cleaved HJs normally, with greater cleavage seen at higher concentrations. However, *Tb*FEN1 showed almost no cleavage of HJs at any concentrations. While there appears to be a very minute amount of HJ cleavage into duplex products at the highest concentration of 4 µM, it is so minimal as to not be significant. This may suggest that ensuring HJ cleavage by *Tb*FEN1 is simply a case of optimising the conditions for cleavage to occur, but due to time limitations this was not explored. It is also possible that this faint band is simply minor contamination from the duplex marker in the lane next to it. Similarly, the buffer conditions here had not been optimised beforehand for the most efficient *Tb*FEN1 cleavage, so this may also be a factor.

5.1.2. *Tb*FEN1 AND MUTANTS AS A 5’-FLAP ENDONUCLEASE

In the absence of robust evidence that the *Tb*FEN1 protein was able to cleave HJs, experiments to investigate the cleavage activity on synthetic 5’- flap substrates were initiated, under the possibility *Tb*FEN1 may be a more effective Flap endonuclease. Along with the WT *Tb*FEN1, a number of the purified mutants were also assayed for 5’-flap nuclease activity. As with the HJ cleavage assay, by incubating the *Tb*FEN1 protein (and its mutagenesis products) with a 5’-flap junction and running the resulting solution on a gel, it should be possible to observe if the junction has been cleaved by the presence (or absence, if cleavage was unsuccessful) of a band towards the bottom of the gel, caused by the fluorescent tagged 30mer, compared to the much larger tagged flap junction at the top of the gel.

Firstly, mutants E162A and E164A (glutamic acid residues presumed to be involved in magnesium and DNA substrate binding in the wild type) were assayed alongside 2 µM *Tb*FEN1 on their ability to cleave 5’-flaps (Figure 5.1.2-1). Analysis of *Tb*FEN1’s 5’-flap cleavage activity (lane C) shows that 2 µM *Tb*FEN1 is capable of cleaving the substrate to a significant degree – with almost no substrate remaining and plenty of cleavage product visible at the base of the gel. Analysis of E162A’s activity (lanes D-F) also demonstrates successful 5’-flap cleavage across all 3 concentrations, however at 1 µM a significant amount of uncleaved substrate can still be seen, and less duplex product is visible than at lower concentrations of the enzyme. This unexpected phenomenon remains a mystery, as it was the only time such an event occurred, and an overabundance of enzyme should result in more product or at least a faster rate. E164A’s cleavage activity (lanes G-I) however seems totally absent, suggesting this residue is key to cleavage activity.

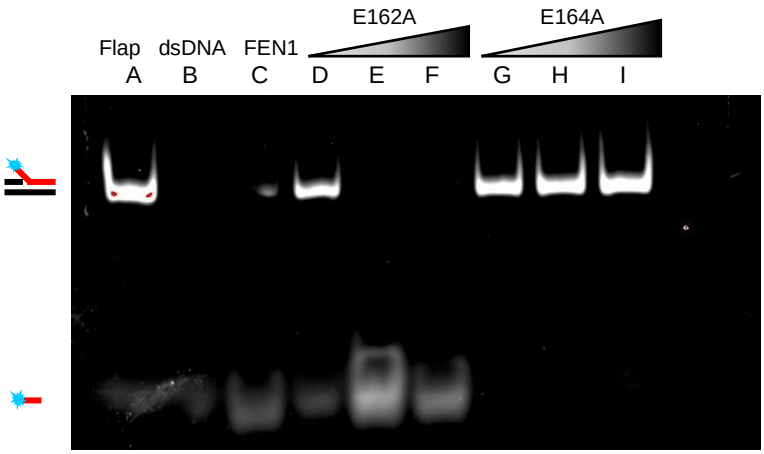


FIGURE 5.1.2-1 TBFEN1, E162A & E164A 5'-FLAP NUCLEASE ACTIVITY

SYNTHETIC 5'-FLAP JUNCTION SUBSTRATES WERE USED TO ANALYSE THE CLEAVAGE ACTIVITY OF TBFEN1 AND TBFEN1 MUTANTS: E162A AND E164A. TWO-FOLD DILUTIONS OF THE ENZYMES IN 10X CLEAVAGE BUFFER AND WATER WERE USED TO ENSURE THAT A WIDE RANGE OF ENZYME CONCENTRATIONS WERE TESTED IN ORDER TO REMOVE THIS AS A POSSIBLE LIMITING FACTOR.

LANE A, 200 NM 5'-FLAP JUNCTION SUBSTRATE. LANE B, 200 NM DSDNA SUBSTRATE. ALL OF LANES C-J, CONTAIN 200 NM 5'-FLAP JUNCTION SUBSTRATE. LANE C, 2 µM TBFEN1. LANE D, 250 NM E162A. LANE E, 500 NM E162A. LANE F, 1 µM E162A. LANE G, 250 NM E164A. LANE H, 500 NM E164A. LANE I, 1 µM E164A.

Upon studying the results from the E162A/E164A assay, the dsDNA marker was very hard to visualise and so the decision was made to use 400 µM dsDNA in all future assays. The next mutants to be assayed were D90K and D90A (magnesium coordination) (Fig. 5.1.2-2), and D185A and D185K (magnesium coordination) (Fig. 5.1.2-3). All 4 of these mutants showed no significant cleavage activity, suggesting that each of the mutated residues is key to *Tb*FEN1's cleavage activity. In both these 2 further assays, 2 µM *Tb*FEN1 cleaves 5'-flap junctions significantly, validating the earlier result. Having demonstrated that the synthetic 5'-flap substrates could be cleaved by both the wild-type *Tb*FEN1 and E162A *Tb*FEN1, further assays were visualised on denaturing Urea-page to provide a greater resolution of the proteins' nuclease activity.

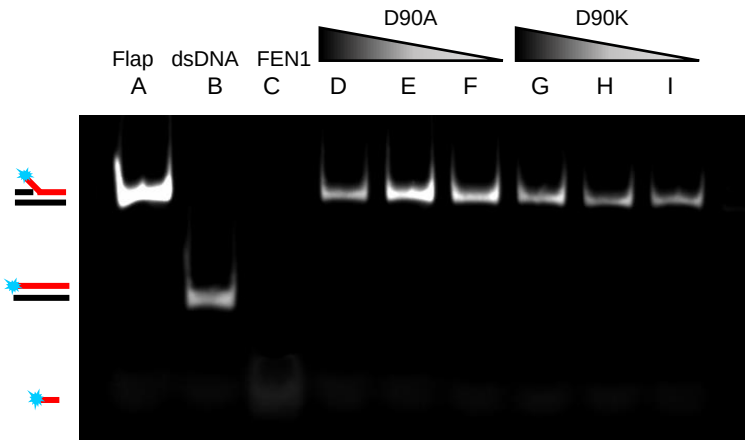


FIGURE 5.1.2-2 TBFEN1, D90A & D90K 5'-FLAP NUCLEASE ACTIVITY

SYNTHETIC 5'-FLAP JUNCTION SUBSTRATES WERE USED TO ANALYSE THE CLEAVAGE ACTIVITY OF TBFEN1 AND TBFEN1 MUTANTS: D90A AND D90K. TWO-FOLD DILUTIONS OF THE ENZYMES IN 10X CLEAVAGE BUFFER AND WATER WERE USED TO ENSURE THAT A WIDE RANGE OF ENZYME CONCENTRATIONS WERE TESTED IN ORDER TO REMOVE THIS AS A POSSIBLE LIMITING FACTOR.

LANE A, 200 NM 5'-FLAP JUNCTION SUBSTRATE. LANE B, 400 NM DSDNA SUBSTRATE. ALL OF LANES C-J, CONTAIN 200 NM 5'-FLAP JUNCTION SUBSTRATE. LANE C, 2 µM TBFEN1. LANE D, 1 µM D90K. LANE E, 500 NM D90K. LANE F, 250 NM D90K. LANE G, 1 µM D90A. LANE H, 500 NM D90A. LANE I, 250 NM D90A.

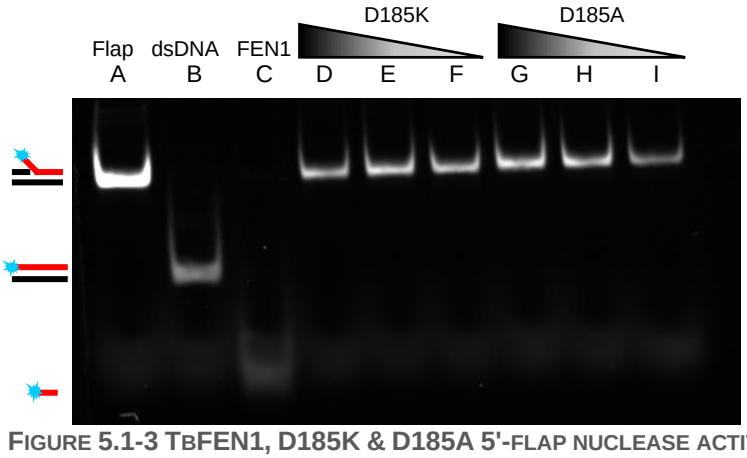


FIGURE 5.1-3 TBFEN1, D185K & D185A 5'-FLAP NUCLEASE ACTIVITY

SYNTHETIC 5'-FLAP JUNCTION SUBSTRATES WERE USED TO ANALYSE THE CLEAVAGE ACTIVITY OF TBFEN1 AND TBFEN1 MUTANTS: D185A AND D185K. TWO-FOLD DILUTIONS OF THE ENZYMES IN 10X CLEAVAGE BUFFER AND WATER WERE USED TO ENSURE THAT A WIDE RANGE OF ENZYME CONCENTRATIONS WERE TESTED IN ORDER TO REMOVE THIS AS A POSSIBLE LIMITING FACTOR.

LANE A, 200 NM 5'-FLAP JUNCTION SUBSTRATE. LANE B, 400 NM DSDNA SUBSTRATE. ALL OF LANES C-J, CONTAIN 200 NM 5'-FLAP JUNCTION SUBSTRATE. LANE C, 2 µM TBFEN1. LANE D, 1 µM D185K. LANE E, 500 NM D185K. LANE F, 250 NM D185K. LANE G, 1 µM D185A. LANE H, 500 NM D185A. LANE I, 250 NM D185A

5.2. DENATURING UREA-PAGE

5.2.1. ANALYSIS OF 5'-FLAP NUCLEASE ACTIVITY OF *TB*FEN1 AND ITS MUTANTS

The same 5'-flap cleavage assay was carried out for all the mutants purified by the end of the project (*Tb*FEN1. D34A, D90A, E162A, E164A, D185A, D237A, Q341A, Q341A+G342A, D90K, D183K, D185K, D237K, G235D), before running each sample on a 12% Urea-PAGE gel to denature any DNA products and thus provide a greater resolution, enabling visualisation of the cleavage products at a resolution of 1 nucleotide. This would therefore provide a much clearer picture of the exact cleavage reaction taking place – and thus rule out if *Tb*FEN1 simply cleaves single stranded DNA structures, or specifically cleaves around Flap junctions to form a neat nick.

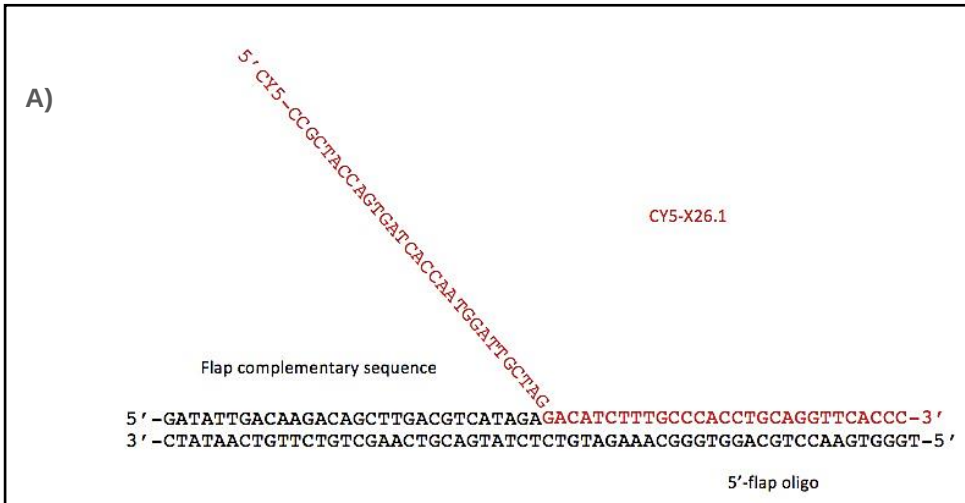


FIGURE 5.2-1 STRUCTURE OF THE SYNTHETIC 5'-CY5-LABELLED FLAP SUBSTRATE AND DOUBLE STRANDED DNA SUBSTRATE

ALL OLIGONUCLEOTIDES WERE SYNTHESISED BY EUROFINS BEFORE BEING PREPPED AND PURIFIED BY NOBLETT, ET AL. PRIOR TO THIS PROJECT (NOBLETT & BENSON, 2017).

- **A) THE 5'-FLAP STRUCTURE USED WAS BY COMBINING THEIR 60 BP CY5-X26.1 OLIGONUCLEOTIDE (RED) WITH A 60 BP 5'-FLAP OLIGO AND A 30 BP OLIGO WITH A COMPLIMENTARY SEQUENCE TO THE FLAP. THIS FORMS A 5'FLAP JUNCTION WITH A 31-BP FLAP.**
- **B) THE DS-DNA SUBSTRATE WAS PREPARED COMBINING THE 60-BP CY5-X26.1 (RED) AND 60-BP X26.5 OLIGOS**

The samples were diluted in formamide buffer to ensure complete denaturation, and the final gel was visualised on a Typhoon™ FLA biomolecular imager (Figure 5.2.1-2). A 5'-CY5- end labelled 5'-flap was used this time to allow visualisation of the cleavage products, and a 30-mer used as a marker (Figure 5.2.1-1).

Analysing the Urea-PAGE gels, it is clear that wt-*Tb*FEN1, E162A (magnesium coordination and DNA binding), and pip-box mutants Q341A and Q341A+G342A (lanes C and G on Gel 1, and lanes C, I and J on Gel 2) were all successful in cleaving the 5'-flap junction. This is demonstrated by the decrease or absence of a band at the position of a full length 60-mer at the top of the gel, and the appearance of three to five products bands at the base, in line with the 30-mer marker. It is certainly not surprising that the PIP box mutants' cleavage activity was unaffected as such mutations should only interfere with their ability to interact with PCNA in vivo, rather than affect catalytic activity. Furthermore, the multiple bands seen in these lanes suggest there is some 'breathing' at the junction, resulting in products estimated to be 31, 32, 33, 34 and 35 nucleotides in length. This may go some way to explaining the blurred product bands shown on the lower resolution native-page gels.

Meanwhile mutants D34A (Gel1 – lane D, implicated in magnesium coordination), D90A (Gel1 – lane E, implicated in), D90K (Gel1 – lane F, implicated in magnesium coordination), E164A (Gel1 – lane H, implicated in magnesium coordination), D183K (Gel1 – lane I, implicated in magnesium coordination), D185A (Gel2 – lane D, implicated in magnesium coordination), D185K (Gel2 – lane E, implicated in magnesium coordination), G235D (Gel2 – lane F, implicated in DNA binding), D237A (Gel2 – lane G, implicated in magnesium coordination and DNA binding), and D237K (Gel2 – lane H, implicated in magnesium coordination and DNA binding) all failed to cleave the 5'-flap junctions, suggesting these residues are key to the *Tb*FEN1 enzymes catalytic activity. Taking the successful cleavage results in conjunction with these failed cleavage results validates both the synthetic junctions as viable substrates as well as confirming the nuclease activity of those enzymes that appear to cleave 5'-flaps. The results from the cleavage assays for each mutagenesis target are summarised in Table 5.2-1.

Mutagenesis	**Observations**
D34A	Loss of 5'-flap cleavage activity
D90A	Loss of 5'-flap cleavage activity
D90K	Loss of 5'-flap cleavage activity
E162A	No effect on 5'-flap cleavage activity
E164A	Loss of 5'-flap cleavage activity
D183K	Loss of 5'-flap cleavage activity
D185A	Loss of 5'-flap cleavage activity
D185K	Loss of 5'-flap cleavage activity
G235D	Loss of 5'-flap cleavage activity
D237A	Partial loss of 5'-flap cleavage activity
D237K	Loss of 5'-flap cleavage activity
Q341A	No effect on 5'-flap cleavage activity
Q341A + G342A	No effect on 5'-flap cleavage activity

TABLE 5.2.1-1 SUMMARY OF OBSERVED EFFECTS ON TBFEN1'S 5'-FLAP CLEAVAGE ACTIVITY FROM THE 13 MUTAGENESIS EXPERIMENTS COMPLETED IN THIS REPORT

EIGHT OF THE MUTANT PROTEINS LOST THE ABILITY TO CLEAVE THE SYNTHETIC 5'-FLAP SUBSTRATE. ONE MUTANT (D237A) SHOWED PARTIAL LOSS OF CLEAVAGE ACTIVITY. THREE MUTANTS (E162A, Q341A AND Q341A + G342A) DID NOT LOSE THEIR ABILITY TO CLEAVE THE 5'-FLAP SUBSTRATE

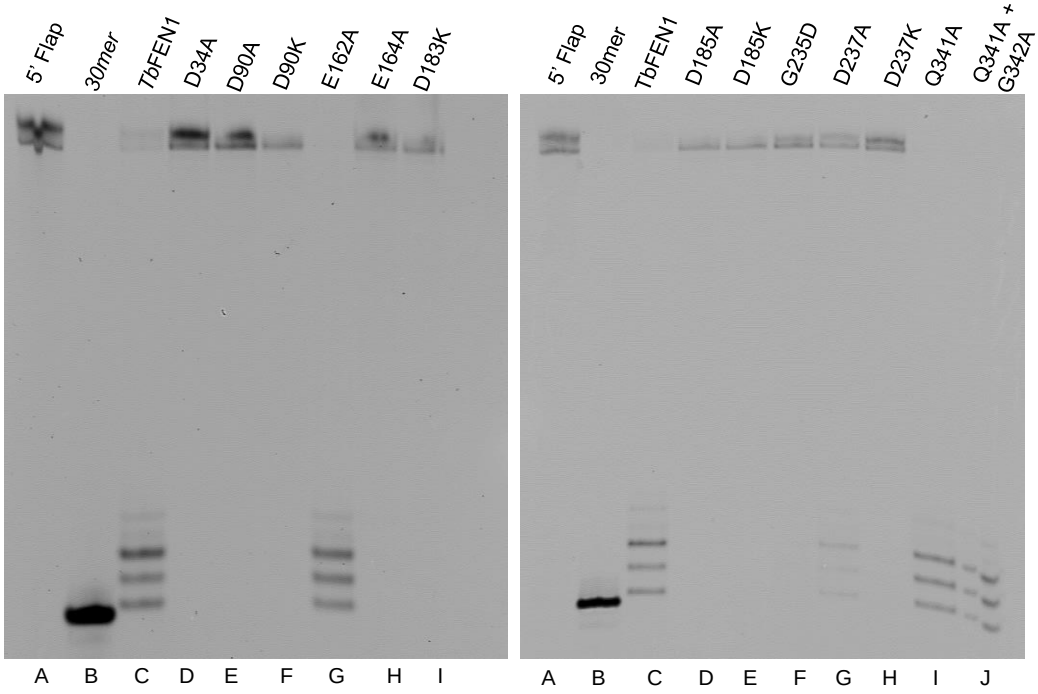


FIGURE 5.2-2 DENATURING UREA-PAGE SHOWING 5'-FLAP CLEAVAGE ACTIVITY OF ALL TBFEN1 MUTANTS SYNTHESISED TO DATE

TWO 12% UREA-PAGE GELS SHOWING THE NUCLEASE ACTIVITY OF ALL THE TBFEN1 MUTANTS SYNTHESISED TO DATE UPON THE SYNTHETIC 5'-FLAP JUNCTION. 1 MM OF EACH ENZYME WAS ADDED TO 200 NM CY5-LABLLED 5'-FLAP JUNCTION IN 2 µL OF 10X CLEAVAGE BUFFER AND MADE UP TO 20 µL WITH DH2O. THE REACTION MIXTURE WAS INCUBATED AT 37°C FOR 20 MINUTES BEFORE ADDITION OF 2 µL OF STOP BUFFER CONTAINING PROTEINASE K, AND INCUBATED FOR A FURTHER 15 MINUTES AT 37°C. THE SAMPLES WERE THEN DILUTED 1:1 WITH FORMAMIDE LOADING BUFFER, AS WELL AS ADDING 10 µL OF LOADING BUFFER TO 2 µL OF 5'-FLAP MARKER AND 4 µL DSDNA MARKER. SAMPLES WERE THEN HEATED TO 95°C FOR 5 MINUTES AND 10 µL RUN ON A 12% UREA-PAGE GEL AT 42 W FOR 2.5 HOURS.

GEL 1: LANE A, 100 MM 5'-FLAP SUBSTRATE. LANE B, 100 MM DSDNA SUBSTRATE. LANE C-I, 100 MM 5'-FLAP SUBSTRATE AND 1 MM ENZYME. LANE C, TBFEN1. LANE D, D34A. LANE E, D90A. LANE F, D90K. LANE G, E162A. LANE H, E164A. LANE I, D183K.

GEL 2: LANE A, 100 MM 5'-FLAP SUBSTRATE. LANE B, 100 MM DSDNA SUBSTRATE. LANE C-J, 100 MM 5'-FLAP SUBSTRATE AND 1 MM ENZYME. LANE C, TBFEN1. LANE D, D185A. LANE E, D185K. LANE F, G235D. LANE G, D237A. LANE H, D237K. LANE I, Q341A. LANE J, Q341A+G342A.

6. LOCALISATION OF *TB*FEN1 IN *TRYPANOSOMA BRUCEI*

As well as studying the nuclease activity of *Tb*FEN1 *in vitro*, procyclic *T. brucei brucei* cells expressing mNeon Green-tagged *Tb*FEN1 were generated to study the localisation of *Tb*FEN1 throughout the cell cycle.

6.1. GENERATION OF *T. BRUCEI* MUTANTS EXPRESSING FLUORESCENT TAGGED *TB*FEN1

A PCR only tagging (pPOT) approach was used to generate cell lines expressing mNeon Green tagged TbFEN1. DNA sequence encoding the mNeon Green open reading frame and a drug resistance cassette was amplified by PCR, using long primers that incorporated 80 nucleotides of homology to TbFEN1 and adjacent UTR. Amplicons were generated to allow either N- or C-terminally tagged TbFEN1 to be generated. PCR amplicons were subsequently purified and used to transfect procyclic T. brucei and recombinant cell lines selected using appropriate drug selection (Figure 6.1-1).

Upon completion of the PCR, the final concentrations were 329.44 µg/µL of the C-terminally tagged and 100.05 µg/µL of the N-terminally tagged, both in 50 µL, giving a final amount of 16.5 µg and 5.0 µg respectively. It had been determined previously to this project that optimal transfection required between 3-5 µg of PCR product (0.7-0.8% w/v), and so the C-terminally tagged construct would need diluting. However after preparing the N-terminally tagged construct there was not enough time left to prepare the second construct.

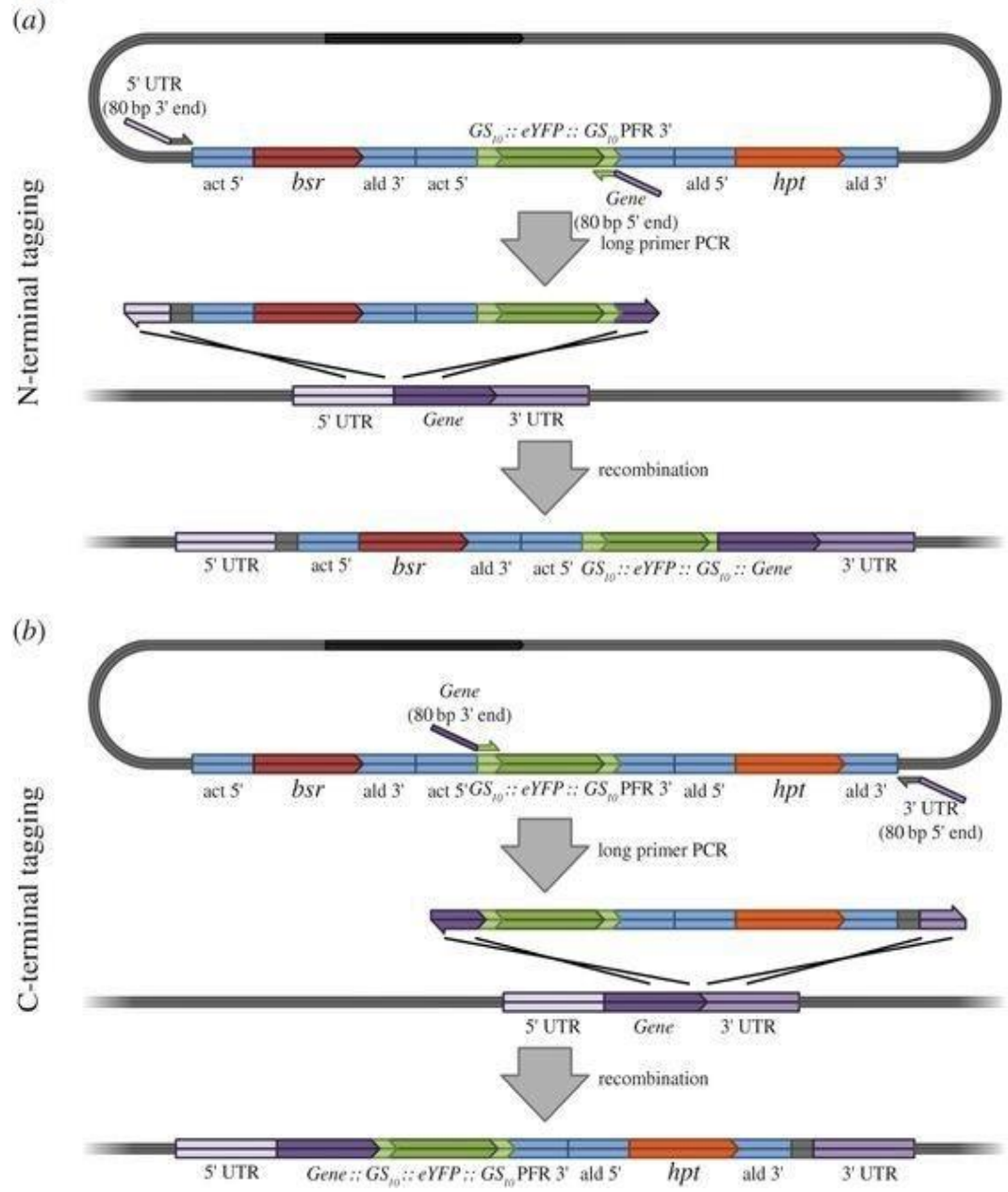


FIGURE 6.1-1 LONG PRIMER PCR TAGGING USING PPOTV4 TO CARRY OUT (A) N-TERMINAL TAGGING (B) C-TERMINAL TAGGING.

A TAILORED PPOT PLASMID, SUCH AS PPOTV4, (WHICH ACT AS TEMPLATES FOR LONG PRIMER PCR, AND CONTAIN AN APPROPRIATE TAG AND RESISTANCE GENE CASSETTE), IS AMPLIFIED USING PCR ALONG WITH LONG PCR PRIMERS WITH AN 80 NUCLEOTIDE 5' OVERHANG WITH HOMOLOGY TO THE TARGET GENE AND ADJACENT UNTRANSLATED REGION (UTR). THIS PRODUCES AN AMPLICON WITH THE TAG FUSED TO THE TARGET GENE, WITH SUCCESSFULLY TRANSFECTED TRYPANOSOMATIDS ABLE TO BE SELECTED FOR USING THE CHOSEN RESISTANCE MARKER.

(DEAN, ET AL., 2015)

6.2. ANALYSIS OF EXPRESSION AND LOCALISATION

The trypanosomes expressing mNG::*Tb*FEN1 were mounted on slides and treated with DAPI and probed with L8C4 to detect PFR2, followed by anti-mouse (TRITC-conjugated) secondary antibodies to detect the L8C4 – highlighting the genetic material and flagellum respectively. The cells were imaged at different stages of the cell cycle, in a 30-layer z-stack used to create a single deconvolved image in each channel (Figure 6.2-1).

In cells that are at the 1K1N and 2K1Nphase of the cell cycle, mNG::*Tb*FEN1 is clearly localised within the nucleolus, appearing as a dark hole in the blue DAPI staining. Interestingly, during nuclear division mNG::*Tb*FEN1 is localised along a line between the two dividing nuclei, perpendicular to the cleavage furrow – presumably bound to DNA that is aligned along the mitotic spindle. By the time that the cells are at the 2K2N stage of the cell cycle and in cytokinesis however, mNG::*Tb*FEN1 has once again localised to the nucleolus of the two daughter nuclei. It should be noted that at no point during the cell cycle, was mNG::*Tb*FEN1 observed to be localised within the kinetoplast.

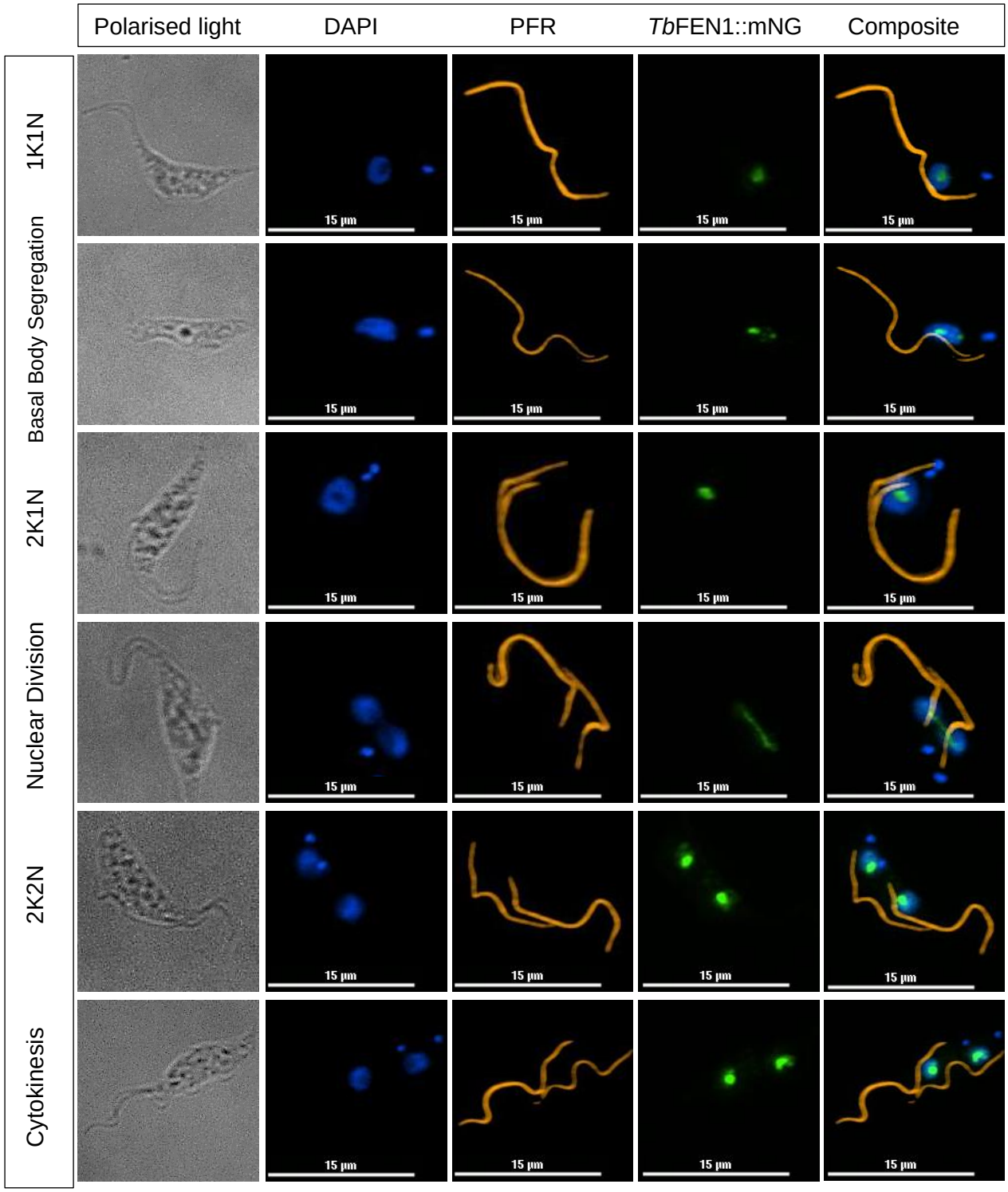


FIGURE 6.2-1 IMMUNOFLUORESCENCE IMAGES SHOWING THE LOCALISATION OF TBFEN1::MNG IN PROCYCLIC FORM T. BRUCEI THROUGH THE CELL CYCLE.

WHOLE CELLS WERE CO-LABELLED WITH THE MONOCLONAL ANTIBODY L8C4, WHICH IS SPECIFIC FOR THE PARAFLAGELLAR ROD PROTEIN PFR2 (ORANGE) AND DAPI (BLUE). CELLS WERE VISUALISED BY DELTAVISION DECONVOLUTION MICROSCOPY. SCALE BAR = 15MM.

7. DISCUSSION

This project aimed to express a novel *T. brucei* endonuclease, *Tb*FEN1, with potential HJ cleavage activity and thus a possible key to understanding the molecular mechanisms behind VSG switching. Firstly, the novel *Tb*FEN1 protein was confirmed to share significant sequence homology with a number of GEN1 and FEN1 orthologs across different species, validating its selection as a potential endonuclease with a role in homologous recombination and VSG switching.

Secondly, *Tb*FEN1, as well as a number of active site mutants, were successfully purified on a large scale using an *E. coli* expression model. Finally, initial studies in this report taken together with previous research carried out in this laboratory, have thrown doubt on *Tb*FEN1's role as a HJ endonuclease as little to no significant activity against Holliday Junctions could be demonstrated at this time, although only a very limited range of conditions were tested. The possibility remains therefore that the optimum conditions required for dimerization and simultaneous nick activity were not identified.

Rather than focus on attempting to optimise for the potential HJ nuclease activity, the possibility was instead considered that *Tb*FEN1 may in fact be a flap-endonuclease. Subsequent experiments therefore focused on studying this 5'-flap nuclease activity in the hopes it would provide a better opportunity to characterise the active site of the putative *Tb*FEN1 enzyme using the mutants purified earlier. These experiments confirmed that *Tb*FEN1 does in fact possess significant 5'-flap cleavage activity akin to that of the GEN1 or FEN1 enzymes. Furthermore, the site-directed mutagenesis carried out targeting potential key amino acids in the *Tb*FEN1 active site were successful in reducing or even wholly eliminating this cleavage activity once more, at least partially validating the comparison to the human GEN1 protein from which the mutagenesis targets were selected.

Meanwhile, cell-cycle localisation studies of the *Tb*FEN1 enzyme *in vivo* showed it following an interesting pattern of localisation, sequestered within the nucleolus until nuclear division whereupon it localised along the mitotic spindle. Furthermore, at no point during the cell cycle was mNG::*Tb*FEN1 observed to localise to the kinetoplast – the network of circular mitochondrial DNA.

7.1. *TB*FEN1 AS A HOLLIDAY JUNCTION RESOLVASE

As stated above, despite sharing significant homology with human GEN1, this study could not find significant HJ cleavage activity for the *Tb*FEN1 enzyme in the time available. Taken together with previous work within this laboratory it would seem increasingly unlikely that *Tb*FEN1 is in fact the sought-after *T. brucei* HJ resolvase. It is possible that the conditions necessary for *Tb*FEN1 dimerization and cleavage were simply not met, as other organism's GEN1 proteins have also been demonstrated to cleave Flap junctions more readily than HJs, such as *D. melanogaster* (Bellendir, et al., 2017). However, it is worth noting that other trypanosome endonucleases have also been identified with potential roles as the parasite's HJ resolvase (Jones, et al., 2017).

7.2. *Tb*FEN1 AS A FLAP ENDONUCLEASE

7.2.1. SEQUENCE ANALYSIS

While the *Tb*FEN1 protein was shown to possess sequence homology to the human GEN1 protein here, it shared far greater homology with human FEN1 as well as other FEN1 enzymes, supporting the hypothesis that the protein of interest is in fact a FEN1 ortholog. FEN1 is traditionally implicated in DNA replication through Okazaki fragment maturation, repair via base excision or non-homologous end joining, and recombination (in cases were homologous DNA molecules have divergent sequences). Therefore, understanding the exact nature of this potential structure specific endonuclease may still prove useful in combating *T. brucei* and sleeping sickness.

7.2.2. 5'-FLAP NUCLEASE ACTIVITY

Having directed study towards charactering the 5'-flap cleavage activity of *Tb*FEN1, a number of nuclease assays were carried out using the purified *Tb*FEN1 protein and a selection of mutants with substitutions made to amino acid residues believed to be key to the enzymes function. These assays were successful in demonstrating *Tb*FEN1's ability to cleave 5'-flap junctions, as well as showing how 9 of the 13 successfully purified mutants were no longer able to cleave 5'-flap junctions, validating those amino acids as indeed key to the enzymes function. While 4 of the mutants still showed some cleavage activity (E162A, D237A, Q341A, and Q341A+G342A), D237A still showed a reduced rate of cleavage compared to wild-type *Tb*FEN1 or the other functional mutants, which would also support its status as a residue with a key functional role. Furthermore, as Q341A and Q341A+G342A were both mutations made to *Tb*FEN1's PCNA-interacting peptide (PIP box), it would be expected that these mutations would not affect the cleavage activity of *Tb*FEN1. Both E162 and D237, meanwhile, are expected to be involved in both magnesium cofactor binding and DNA substrate binding, where all the other targeted residues were implicated in either one role or another (Table 3.1.2-1).

Furthermore, while it was not possible to purify an E162K mutant during the time available for this project, two mutants were prepared with substitutions at position 237. It is noteworthy perhaps that while the substitution of the negatively charged aspartic acid at position 237 for the neutral alanine residue had no effect upon the enzymes activity, the substitution for a basic lysine residue prevented *Tb*FEN1 from cleaving the 5'-flap substrate. Perhaps therefore, while the other active site residues may be more strongly involved in binding DNA or Mg^{2+} such that they can carry the strain resulting from the absence of either E162 or D237 – suggesting these residues only weakly bind - they cannot make up for the significant opposing forces introduced through the addition of the negatively charge lysine. It would certainly be interesting in the future to study the 3D confirmation of *Tb*FEN1 and some of the mutants developed here, to see exactly how they have affected the enzymes structure and function.

Equally interestingly, the observation that it is only the E162 and D237 *Tb*FEN1 mutants that retained some flap cleavage activity is in direct contrast to results from mutagenesis studies of human FEN1, where E158A and D233A *Hs*FEN1 mutants lost both binding and cleavage activity of the flap substrate (Shen, et al., 1996). The same study demonstrated that the mutagenesis D179A of *Hs*FEN1 retained the same activity as the wild-type, while this project found that the *Tb*FEN1 equivalent position, D183, lost 5'-flap cleavage activity, albeit when aspartic acid was replaced with lysine, rather than alanine. This stark difference in results could be due to a misidentification of the key *Tb*FEN1 residues, differences in reaction conditions of the two studies, or simply structural differences between the human and trypanosome FEN1

proteins. It would definitely be worth repeating the experiment with both the original and freshly expressed *Tb*FEN1 mutants to be sure that the results were reliable. If the opportunity does arise in the future to structurally characterise *Tb*FEN1, it would certainly be worth attempting to compare how this data compares with the inferred key residues from this project, or if different residues appear to play a role in the function of *Tb*FEN1 based on its 3D confirmation.

7.2.3. *Localisation of Tb*FEN1 in Procyclic *T. brucei*

The localisation studies demonstrated that mNG::*Tb*FEN1 concentrated within the nucleolus during G1, S and G2 phase, but aligned along the mitotic spindle during mitosis due to elongation of the nucleolus (Ogbadoyi, et al., 2000). This observation is concordant with *T. brucei* protein potentially being a FEN1 ortholog, as previous studies by Guo et al. (2008) showed human FEN1 super-accumulates in the nucleoli of HeLa cells, where it is proposed to have a role in maintaining ribosomal DNA by stabilising rDNA replication forks. However, the study also reported that human FEN1 migrates out of the nucleolus into to the nucleoplasm in response to UV radiation and the ensuing DNA damage, a relocalisation dependent upon phosphorylation of a serine residue at position 187. The study suggests that in HeLa FEN1 acts to rescue stalled replication forks, or that it may simply be relocalised away from the rDNA following damage to arrest replication and allow time for UV damage to be repaired (Guo, et al., 2008). FEN1 is also thought to play an important role in DNA replication, where it is responsible for processing Okazaki fragments on the lagging strand. However, no localisation of our mNG::TbFEN1 protein was observed outside the nucleolus while using the methods and imaging systems detailed in this report, including no observed localisation to the mitochondrial DNA within the kinetoplast. This is in contrast to yeast and mouse FEN1, which has been demonstrated to play a role in mitochondrial DNA maintenance (Kalifa, et al., 2010). The lack of localisation of *Tb*FEN1 protein to the mitochondrial DNA in the kinetoplast would therefore lend credence to its role as a GEN1 resolvase rather than being FEN1 (Kalifa, et al., 2010). There may however be other proteins fulfilling this role in *T. brucei*, which have yet to be identified or characterised. Similarly, after the completion of the practical work for this report, upon further literature studies about protein localisation in *T. brucei,* it has become apparent that the protein translocases typically responsible for the transport of proteins across the out and inner mitochondrial membranes (TOMs and TIMs, respectively) use N-terminal mitochondrial targeting signals to identify proteins for transport. It is possible therefore, that the introduction of our fluorescent tag onto the N-terminus of the *Tb*FEN1 protein, could interrupt this mechanism, and prevent their transport into the mitochondria. While a search of the TrypTag database was carried out to see if there was any existing data on the localisation of *Tb*FEN1 using either tagging method, no successful tagging on either the C or N terminus of the protein had been reported. Further study of *Tb*FEN1 localisation therefore, should prioritise generating a C-terminally tagged protein seeing if any differences are observed between the C-terminally tagged *Tb*FEN1 and the data produced here for mNG::TbFEN1 (Singha, et al., 2012).

Studies of GEN1 enzymes have actually identified that the human protein possesses a nuclear export signal that localises it outside the nucleus until the nuclear membrane breaks down during mitosis. This is presumed to ensure a minimal frequency of crossover events and maintain genome stability until GEN1 is required for processing the persistent HJs prior to mitosis, maintain chromosome integrity. Yen1, in yeast however, is actively impaired in its resolvase activity by phosphorylation, and by Msn5-mediated nuclear export. It is then dephosphorylated by Cdc14 phosphatase and imported back into the nucleus prior to mitosis, since the nuclear envelope remains intact throughout yeast mitosis (Chan & West, 2014). Furthermore, GEN1 has actually been shown to localise to the centrosomes, where it is responsible for maintaining centrosome integrity when it is not utilised resolving HJs. In fact, a

depletion of GEN1 was shown to result in centrosome duplication, leading to supernumerary centrosomes and an arrest of the cell cycle prior to mitosis, and ultimately an increase in apoptosis and multinucleate cells. Furthermore, mutants that had lost their HJ resolvase activity could still associate with centrosomes and maintain their integrity (Gao, et al., 2012) (Sun, et al., 2014).

Since no localisation of the mNG::*TbF*EN1 protein outside the nucleus was observed, this would support the suggestion that it is instead a FEN1 enzyme, despite the lack of localisation to the mitochondrial DNA. It is possible however that mNG::*Tb*FEN1 was present outside the nucleolus but was sufficiently disperse as to be below the threshold of detection by our equipment. If the protein of interest is a FEN1 enzyme, why does it appear to remain sequestered inside the nucleolus during S-phase, when you would expect it to otherwise be assisting in the replication of DNA and processing of Okazaki fragments. Furthermore, *T. brucei* actually possesses no centrosomes and relies on the basal bodies as microtubule organising centres (MOTCs). However these play no role in spindle assembly and instead trypanosomes rely on a chromatin-directed pathway, since they too – like yeast – undergo a 'closed' mitosis (Li, 2012). This would suggest that, unlike human GEN1, *Tb*FEN1 is stored within the nucleus rather than being actively exported, and is instead sequestered in the nucleolus to ensure it does not compete with other non-crossover pathways of DSBR until it is necessary. Since trypanosomes possess no centrosomes there is no need to export the *Tb*FEN1 protein out of the nucleus to maintain them, especially since the lack of nuclear membrane breakdown would mean it would subsequently have to be imported back into the nucleus prior to mitosis.

Having shown that *TbF*EN1 is localised within the nucleolus in a similar manner to FEN1, but was not observed to interact with the chromosomal DNA during S-phase, this calls into question once again whether this enzyme is a FEN1 or GEN1 enzyme. The possibility remains that it is in fact a Holliday junction resolvase, and the conditions of our HJ cleavage assays were not optimised for *Tb*FEN1 activity. Since it is not exported from the nucleus, *Tb*FEN1 may also have other molecular regulatory mechanisms acting upon it, in a similar manner to Yen1, which have yet to be identified and may help determine its true enzymatic activity, e.g. phosphorylation. This may help explain the observed, but insignificant, Holliday junction cleavage demonstrated in this report.

7.3. FURTHER STUDY

The results obtained during the course of this project would suggest that further study is needed before the putative *Tb*FEN1 protein can be ruled out as a potential Holliday junction resolvase for *T. brucei*. Since concluding the laboratory work for this project, further evidence to support the theory that *Tb*FEN1 may require very specific conditions to function as a typical GEN1 enzyme has been identified. Work by Bellendir et al. has shown that GEN1 enzymes can actually show far higher cleavage rates for 5'-flaps than Holliday junctions *in vitro*, and that intact HJs are among the slowest DNA repair intermediates to be repaired by human or drosophila GEN1 (Bellendir, et al., 2017). Interestingly, in 2007 a paper published by Kanai et al. (2007) stated that *D. melanogaster* GEN1 possessed no HJ cleavage activity at all, which has since been shown to be a false statement (Kanai, et al., 2007). It is likely, therefore, that different GEN1 enzymes exhibit varying substrate specificity depending upon their reaction conditions. This project utilised only one buffer composition and only one type of Holliday junction, and so a direction for future study may be to attempt further HJ nuclease assays using

the wild type *Tb*FEN1 enzyme under a variety of conditions and using other branched structures, including nicked HJs and branched structures.

The possibility that *Tb*FEN1's activity is controlled by molecular modifications that may not have been carried over in the gene we expressed in *E. coli* should also be considered. Certainly human FEN1 has been demonstrated to be a target for phosphorylation, methylation and acetylation *in vivo*, however these all act to downregulate its activity (Zheng, et al., 2011). Therefore, a possible avenue for future research would be to produce crystal structures of the *Tb*FEN1 protein and analyse its final post-translational structure, along with any potential modifications. This would have the added benefit of providing an opportunity to study the exact confirmation of the *Tb*FEN1 active site, and how similar it is to either GEN1 or FEN1 homologs. Similarly, mass spectrometry is another very useful tool for analysing covalent modifications of proteins, as specific shifts in mass and specific fragments can easily be attributed to different types of modification, e.g. phosphorylation or acetylation. Indeed it is even possible to quantify changes in the amount of post-translational modification occurring, or to study the spatial and temporal regulation of these modifications (Witze, et al., 2007). These methods could then perhaps be used to see if the post-translational modifications of *Tb*FEN1 change in response to DNA damage or other stressors.

If *Tb*FEN1 is indeed as vitally important to *T. brucei*'s DNA repair or VSG switching mechanisms as proposed here, then another avenue for future research might be to use RNA interference (RNAi) to inhibit the expression or translation of the *Tb*FEN1 gene. First, a double-stranded RNA sequence would be generated that is complementary to the *Tb*FEN1 gene sequence, synthesised, and introduced it into the trypanosome cell. From here it would be recognised as exogenous and cleaved into many small 20-25 bp fragments, called small interfering RNAs (siRNAs), by the dicer protein. These siRNAs are integrated into the RNA-induced silencing complex (RISC) and subsequently anneal to the target mRNA, causing it to be cleaved by RISC and thus no longer useable as a template for translation (Ahlquist, 2002). Using this technique it would be possible to supress the *Tb*FEN1 gene and observe how this effected mechanisms such as DNA repair (by artificially introducing DNA damage using UV for example) or VSG switching (by observing changes in expression of a fluorescently tagged VSG in a population). If it was therefore, crucial to DNA repair then the cells should not be able to survive DNA damage, and similarly, if *Tb*FEN1 is crucial to VSG switching, then a decrease in the rate of switching compared to a control population should be observed.

Regardless of determining if *Tb*FEN1 is able to cleave HJs, its role as a 5'-flap endonuclease still makes it an important target in *T. brucei*'s DNA repair machinery. Having identified a number of mutagenesis products that lead to a loss of 5'-flap cleavage activity, as well as several that maintain it, a possible avenue for future research would be to transfect a *T. brucei* cell line to express one of these mutants to observe its effect upon cell viability and if it does affect VSG switching rates. The PIP-box mutants, while still able to cleave 5'-flaps, would also be an interesting target for *in vivo* localisation studies, since the PIP-box is thought to play a role in targeting the enzyme to where it needs to be. Regardless of whether or not those residues targeted here indeed play key roles in substrate and metal binding, as this project supposes, the experiments carried out here have been successful in demonstrating 7 residues which are key to the 5'-flap cleavage activity of *Tb*FEN1. That may be because they do have roles in substrate binding and cleavage, or it could simply be that changing these residues drastically effects the 3D confirmation of the *Tb*FEN1 protein. With this information, it may be possible to begin studying compounds that may interact with these different residues to inhibit *Tb*FEN1s activity, and thus work towards developing a drug. Through the previously mentioned RNA interference studies it may be possible to demonstrate that *Tb*FEN1 is fundamental for *T. brucei*'s viability

within the human host. Equally, if further repeats of the work carried out here can demonstrate that, for example, D183 is entirely key for *Tb*FEN1's activity, but that *Hs*FEN1's D179 residue is not fundamentally key for its function, then this may prove an especially valuable target. This would potentially mean that potential drug compounds that interfere with this residue would successfully inhibit *Tb*FEN1 without interfering with *Hs*FEN1. As for further study into *Tb*FEN1's potential HJ-resolvase activity, the next best step might be to carry out high-throughput screening of different buffer conditions for the cleavage assay. Perhaps using a microplate system with different conditions and a fluorescence polarisation assay, by adding the an appropriately labelled HJ substrate and the *Tb*FEN1 protein to each well and observing the increase in depolarization of the fluorescence signal to determine cleavage (McWhirter, et al., 2013). This would allow the testing of huge numbers of different conditions (you can get 96 well plates, 384 well plates, or even 1536 well plates) in a single assay, and thus this may provide an avenue to say once and for all if *Tb*FEN1 can cleave HJs or not.

This project has demonstrates that the putative *Tb*FEN1 protein has clear 5'-flap activity, localises to the nucleolus and that at least some of the residues demonstrated in the past to be key to the function of other known FEN1 proteins are equally key to this trypanosome protein. It has highlighted the need to increase our understanding of the molecular mechanisms that underpin *T.* brucei's pathogenicity, as well as identifying some potential future avenues of research to do so. Finally, with 7 residues (D34, D90, E164, D183, D185, G235, and D237) having been demonstrated as being important to *Tb*FEN1 substrate cleavage, this may provide a stepping off point for future research into developing a drug to combat HAT and Nagana.

8. APPENDICES

Sample	Conc. (µg/ml)
GEN1 G6	4280.282
GEN1 G7	10172.43
GEN1 G8	3058.564
E162A A15	3508.521
E162A B15	7936.889
E162A B14	4482.477
GEN1 C7	555.326
E164A A15	1067.935
E164A B15	1062.239
E164A B14	888.5215
D90A A15	1697.304
D90A B15	1540.674
D90A B14	1102.108
D185A D2	3967.021
D185A D1	3827.477
D185A E1	2221.304
D237A A14	296.1738
D237A A15	1497.956
D237A B15	452.8042
D34A A15	877.1302
D34A B15	458.4999
D34A B14	327.4999
D237K A13	617.9781
D237K A14	1703
D237K A15	1113.5
Q341A A14	515.4564
Q341A A15	2543.108
Q341A B15	991.0432
G342A A14	253.4565
G342A A15	2545.956
G342A B15	1170.456
D183K B15	1916.586

D185K A15	2739.608
D90K A15	797.3911
G235D B13	435.7173

9. REFERENCES

(Zephyris), R. W., 2007. *File:Holliday junction coloured.png.* [Online]
Available at: https://commons.wikimedia.org/wiki/File:Holliday_junction_coloured.png
[Accessed 30 March 2020].

Ahlquist, P., 2002. RNA-Dependent RNA Polymerases, Virus and RNA Silencing. *Science,* 296(5571), pp. 1270-1273.

Aksoy, S. et al., 2017. Human African trypanosomiasis control: Achievements and challenges. *PLoS Neglected Tropical Diseases,* 11(4), p. e0005454.

Bartossek, T. et al., 2017. Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. *Nature Microbiology,* 2(11), pp. 1523-1532.

Bärtsch, S., Kang, L. E. & Symington, L. S., 2000. RAD51 is Required for the Repair of Plasmid Double-Stranded DNA Gaps from Either Plasmid or Chromosomal Templates. *Molecular and Cellulcar Biology,* 20(4), pp. 1194-1205.

Bellendir, S. P. et al., 2017. Substrate preference of Gen endonucleases highlights the importance of branched structures as DNA damage repair intermediates. *Nucleic Acids research,* 45(9), pp. 5333-5348.

Bellendir, S. P. et al., 2017. Substrate preference of Gen endonucleases highlights the importance of branched structures as DNA repair intermediates. *Nucleic Acids Research,* 45(9), pp. 5333-5348.

Blum, M. L. et al., 1993. A Structural Motif in the Variant Surface Glycoproteins of Trypanosoma brucei. *Nature,* 362(6321), pp. 603-609.

Boothroyd, C. E. et al., 2009. A yeast-endonuclease-generated DNA break induces anitgenic switching in Trypanosoma brucei. *Nature,* 459(7244), pp. 278-281.

Burri, C., 2010. Chemotherapy against human African trypanosomiasis: Is there a road to success?. *Parasitology,* 137(14), pp. 1987-1994.

Büsher, P., Cecchi, G., Jamonneau, V. & Priotto, G., 2017. Human African Trypanosomiasis. *The Lancet,* 390(10110), pp. 2397-2409.

Carrington, M. et al., 1991. Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. *Journal of Molecular Biology,* 221(3), pp. 823-835.

CDC, 2012. *Parasites - African Trypanosomiasis (also known as Sleeping Sickness). Prevention & Control.* [Online]
Available at: https://www.cdc.gov/parasites/sleepingsickness/prevent.html
[Accessed 26 March 2020].

Centers for Disease Control and Prevention, 2012. *Sleeping Sickness, Epidemiology & Risk Factors.* [Online]
Available at: https://www.cdc.gov/parasites/sleepingsickness/epi.html

Centers for Disease Control and Prevention, 2018. *Sleeping Sickness Biology.* [Online]
Available at: https://www.cdc.gov/parasites/sleepingsickness/biology.html
[Accessed 27 September 2018].

Cestari, I. & Stuart, K., 2018. Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes. *Current Genomics,* 19(2), pp. 119-132.

Chan, Y. W. & West, S., 2015. GEN1 promotes Holliday junction resolution by a coordinated nick and counter-nick mechanism. *Nucleic Acids Research,* 43(22), pp. 10882-10892.

Chan, Y. W. & West, S. C., 2014. Spatial control of the GEN1 Holliday junction resolvase ensures genome stability. *Nature Communications,* Volume 5, p. 4844.

Chaves, I. et al., 1999. Control of variant surface glycoprotein gene-expression sites in Trypanosoma brucei. *The EMBO Journal,* 18(17), pp. 4846-4855.

Checchi, F. et al., 2008. Estimates of the duration fo the early and late stage of gambiense sleeping sickness. *BMC Infectious Diseases,* 8(16).

Colavito, S., Prakash, R. & Sung, P., 2010. Promotion and regulation of homologous recombination by DNA helicases. *Methods,* 51(3), pp. 329-335.

Connolly, B. et al., 1991. Resolution of Holliday junctions in vitro requires the Escherichia coli ruvC gene product. *Proceedings of the National Academy of Sciences of the USA,* 88(14), pp. 6063-6067.

Constantinou, A. & West, S. C., 2004. Holliday Junction Branch Migration and Resolution Assays. *Methods in Molecular Biology,* Volume 262, pp. 239-253.

Cox, M. M. & Battista, J. R., 2005. DEINOCOCCUS RADIODURANS - THE CONSUMMATE SURVIVOR. *Nature Reviews Microbiology,* 3(11), pp. 882-892.

Dean, S. et al., 2015. A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biology, The Royal Society,* 5(1), p. 140197.

Dunderdale, H. J. et al., 1991. Formation and resolution of recombination intermediates by E. coliRecA and RuvC proteins. *Nature,* 354(6354), pp. 506-510.

Engstler, M. et al., 2007. Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes. *Cell,* 131(3), pp. 505-515.

Ford, L. B., 2007. civil conflict and sleeping sickness in Africa in general and Uganda in particular. *Conflict anf Health,* p. 1:6.

Franco, J. R. & Priotto, G., 2018. *Mapping the risk of human African trypanosomiasis.* [Online]
Available at: http://www.who.int/trypanosomiasis_african/country/risk_AFRO/en/

Franco, J. R., Simarro, P. P., Diarra, A. & Jannin, J. G., 2014. Epidemiology of human African trypanosomiasis. *Clinical epidemiology,* pp. 257-275.

Frank, G. et al., 1998. Partial Functional Deficiency of E160D Flap Endonuclease-1 Mutantin Vitro and in Vivo Is Due to Defective Cleavage of DNA Substrates. *Journal of Biological Chemistry,* 273(49), pp. 33064-33072.

Gao, M. et al., 2012. A Novel Role of Human Holliday Junction Resolvase GEN1 in the Maintenance of Centrosome Integrity. *PLoS One,* 7(11), p. e49687.

Gao, M. et al., 2012. A Novel Role of Human Holliday Junction Resolvase GEN1 in the Maintenance of Centrosome Integrity. *Public Library of Science One,* 7(11), p. e49687.

García-Luis, J. & Machín, F., 2014. Mus81-Mms4 and Yen1 resolve a novel anaphase bridge formed by noncanonical Holliday junctions. *Nature Communications,* Volume 5, p. 5652.

Garner, E. et al., 2013. Human GNE1 and the SLX4-associated nucleases MUS81 and SLX1 are essential for the resolution of replication-induced Holliday Junctions. *Cell Reports,* 5(1), pp. 207-215.

Gary, R. et al., 1999. A Novel Role in DNA Metabolism for the Binding of Fen1/Rad27 to PCNA and Implications for Genetic Risk. *Molecular and Cellular Biology,* 19(8), pp. 5373-5382.

Genois, M.-M.et al., 2014. DNA Repair Pathways in Trypanosomatids: from DNA Repair to Drug Resistance. *American Society for Microbiology: Microbiology and Molecular Biology Reviews,* 78(1), pp. 40-73.

Glover, L., Alsford, S. & Horn, D., 2013. DNA Break Site at Fragile Subtelomeres Determines Probability and Mechanism of Antigenic Variation in African Trypanosomes. *PLoS Pathogens,* 9(3), p. e1003260.

Gorecka, K. M., Komorowska, W. & Nowotny, M., 2013. Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Research,* 41(21), pp. 9945-9955.

Guo, Z. et al., 2008. Nuclear Localization and Dynamic Roles of Flap Endonuclease 1 in Ribosomal DNA Replication and Damage Repair. *American Society for Microbiology: Molecular and Cellular Biology,* 28(13), pp. 4310-4319.

Gupta, R. C., Bazemore, R., Golub, E. I. & Radding, C. M., 1997. Activities of human recombination protein Rad51. *Proceedings of the National Academy of Sciences of the USA,* 94(2), pp. 463-468.

Haber, J. E., Ira, G., Malkova, A. & Sugawara, N., 2004. Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. *Philosophical Transactions of the Royal Society B Biological Sciences,* 359(1441), pp. 79-86.

Hammarton, T. C., 2007. Cell cycle regulation in Trypanosoma brucei. *Molecular and Biochemical Parasitology,* 153(1), pp. 1-8.

Henneke, G., Friedrich-Heineken, E. & Hubscher, U., 2003. Flap Endonuclease 1: a novel tumour suppressor protein. *Trends in Biochemical Sciences,* 28(7), pp. 384-390.

Hertz-Fowler, C. et al., 2008. Telomeric Expression sites Are Highly Conserved in Trypanosoma brucei. *PLOS One,* 3(10), p. e3527.

Heyer, W.-D., 2004. Recombination: Holliday Junction Resolution and Crossover Formation. *Current Biology,* 14(2), pp. 56-58.

Holliday, R., 1964. A mechanism for gene conversion in fungi. *Genetics Research,* 5(2), pp. 282-304.

Horn, D., 2014. Antigenic Variation in African Trypanosomes. *Molecular and Biochemical Parasitology,* 195(2), pp. 123-129.

Hosfield, D. J., Mol, C. D., Shen, B. & Tainer, J. A., 1998. Structure of the DNA Repair and Replication Endonuclease and Exonuclease FEN-1: Coupling DNA and PCNA Binding to FEN-1 Activity. *Cell,* 95(1), pp. 135-146.

Hutchinson, O. C. et al., 2003. VSG structure: similar N-terminal domains can form functional VSGs with different types of C-terminal domain. *Molecular and Biochemical Parasitology,* 130(2), pp. 127-131.

Hwang, K. Y., Baek, K., Kim, H.-Y. & Cho, Y., 1998. The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nature Structural & Molecular Biology,* Volume 5, pp. 707-713.

Ip, S. C. et al., 2008. Identification of Holliday junction resolvases from humans and yeast. *Nature,* Volume 456, pp. 357-361.

Ishikawa, G. et al., 2004. DmGEN, a novel RAD2 family endo-exonuclease from Drosophila melanogaster. *Nucleic acids research,* 32(21), pp. 6251-6259.

Ishikawa, G. et al., 2004. DmGEN, a novel RAD2 family endo-exonuclease from Drosophila melanogaster. *Nucleic Acids Research,* 32(21), pp. 6251-6259.

Iwasaki, H. et al., 1989. Overproduction, Purification, and ATPase Activity of the Escherichia coli RuvB Protein Involved in DNA Repair. *Journal of Bacteriology,* 171(10), pp. 5276-5280.

Iwasaki, H., Takahagi, M., Nakata, A. & Shinagawa, H., 1992. Escherichia coli RuvA and RuvB proteins specifically interact with Holliday junctions and promote branch migration. *Genes and Development,* 6(11), pp. 2214-2220.

Iwasaki, H. et al., 1991. Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. *The EMBO Journal,* 10(13), pp. 4381-4389.

Jamonneau, V. et al., 2015. Accuracy of Individual Rapid Tests for Serodiagnosis of Gambiense Sleeping Sickness in West Africa. *PLOS Neglected Tropical Diseases,* 9(2), p. e0003480.

Jehi, S. E., Wu, F. & Li, B., 2014. Trypanosoma brucei TIF2 suppresses VSG switching by maintaining subtelomere integrity. *Cell Research,* 24(7), pp. 870-885.

Jones, A. B. L., McKean, P. & Benson, F. E., 2017. *Characterisation of potential Holliday junction resolvases in Trypanosoma brucei,* Lancaster: Lancaster University.

Kalifa, L. et al., 2010. Evidence for a Role of FEN1 in Maintaining Mitochondrial DNA. *DNA Repair,* 8(10), pp. 1242-1249.

Kanai, Y. et al., 2007. DmGEN shows a flap endonuclease activity, cleaving the blocked-flap structure and model replication fork. *The FEBS Journal,* 274(15), pp. 3914-3927.

Keeney, S., 2011. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dynamics and Stability,* 1(2), pp. 81-123.

Kennedy, P. G., 2004. Human African Trypanosomiasis of the CNS: Current Issues and Challenges. *Journal of Clinical Investigation,* 113(4), pp. 496-504.

Langousis, G. & Hill, K. L., 2014. Motility and more: the flagellum of Trypanosoma brucei. *Nature Reviews Microbiology,* 12(7), pp. 505-518.

Larsen, E. et al., 2003. Proliferation Failure and Gamma Radiation Sensitivity of Fen1 Null Mutant Mice at the Blastocyst Stage. *Molecular and Cellular Biology,* 23(15), pp. 5346-5353.

Lee, S.-H.et al., 2015. Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife,* Volume 4, p. e12256.

Li, B., 2015. DNA Double-Strand Breaks and Telomeres Play Important Roles in Trypanosoma brucei Antigenic Variation. *Eukaryotic Cell,* 14(3), pp. 196-205.

Lindahl, T. & Barnes, D. E., 2000. Repair of Endogenous DNA Damage. *Cold Springs Harbor Symposia on Quantitative Biology,* Volume 65, pp. 127-133.

Li, Z., 2012. Regulation of the Cell Division cycle in Trypanosoma brucei. *American Society for Microbiology: Eukaryotic Cell,* 11(10), pp. 1180-1190.

Lobo, I. & Shaw, K., 2008. Discovery and Types of Genetic Linkage. *Nature Education,* 1(1), p. 139.

Lodish, H. et al., 2004. In: *Molecular Biology of the Cell (5th Ed.).* New York: WH Freeman, p. 963.

Lundkvist, G. B., Kristensson, K. & Bentivoglio, M., 2004. Why Trypanosomes Cause Sleeping Sickness. *APS Physiology,* 19(4), pp. 198-206.

Mansfield, J. M. & Paulnock, D. M., 2005. Regulation of innate and acquired immunity in African trypanosomiasis. *Parasite Immunology,* 27(10-11), pp. 361-371.

Maree, J. P. & Patterton, H. G., 2014. The epigenome of Trypanosoma brucei: a regulatory interface to an unconventional transcriptional machine. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms,* 1839(9), pp. 743-750.

Matos, J. et al., 2011. Regulatory Control of the Resolution of DNA Recombination Intermediates during Meiosis and Mitosis. *Cell,* 147(1), pp. 158-172.

Matovu, E. et al., 2017. Serological tests for gambiense human African trypanosomiasis detect antibodies in cattle. *Parasites and Vectors,* 10(1), p. 546.

Matthews, K. R., Ellis, J. R. & Paterou, A., 2004. Molecular regulation of the life cycle of African trypanosomes. *Trends in Parasitology,* 20(1), pp. 40-47.

McAllister, W. B. & Benson, F. E., 2015. *The purification of a novel trypanosome protein as a potential GEN1 Holliday junction resolvase ortholog,* Lancaster: Lancaster University.

McMahill, M. S., Sham, C. W. & Bishop, D. K., 2007. Synthesis-Dependent Strand Annealing in Meiosis. *PLOS Biology,* 5(11), p. e299.

McWhirter, C. et al., 2013. Development of a High-Throughput Fluorescence Polarization DNA Cleavage Assay for the Identification of FEN1 Inhibitors. *SAGE Journals,* 18(5), pp. 567-575.

MD, P. P. G. K., 2013. Clinical Features, Diagnosis, and Treatment of Human African Trypanosomiasis (Sleeping Sickness). *The Lancet Neurology,* 12(2), pp. 186-194.

Molyneux, D., Ndung'u, J. & Maudlin, I., 2010. Controlling Sleeping Sickness—"When Will They Ever Learn?". *PLoS Neglected Tropical Diseases,* 4(5), p. e609.

Mugnier, M. R., Stebbins, C. E. & Papavasiliou, F. N., 2016. Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma brucei. *PLOS Pathogens,* 12(9).

Navarro, M. & Gull, K., 2001. A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature,* Volume 414, pp. 759-763.

Noblett, R. A. & Benson, F. E., 2017. *Characterisation of the Trypanosoma brucei GEN1 protein: a FLAP endonuclease or a Holliday junction resolvase,* Lancaster: Lancaster University.

Ogbadoyi, E. et al., 2000. Architecture of the Trypanasoma brucei nucleus during interphase and mitosis. *Chromasoma,* Volume 108, pp. 501-513.

Ohshima, K., Kang, S., Larson, J. E. & Wells, R. D., 1996. TTA-TAA Triplet Repeats in Plasmids Form a Non-H Bonded Structure. *Journal of Biological Chemistry,* 271(28), pp. 16784-16791.

Oxford Academic, 2014. *Youtube - Holliday Junction Resolution.* [Online] Available at: https://www.youtube.com/watch?v=MvnWxN81Qps [Accessed 31 10 2018].

Parenteau, J. & Wellinger, R. J., 1999. Accumulation of Single-Stranded DNA and Destabilization of Telomeric Repeats in Yeast Mutant Strains Carrying a Deletion of RAD27. *Molecular and Cellular Biology,* 19(6), pp. 4143-4152.

Parenteau, J. & Wellinger, R. J., 2002. Differential processing of leading- and lagging-strand ends at Saccharomyces cerevisiae telomeres revealed by the absence of Rad27p nuclease. *Genetics,* 162(4), pp. 1583-1594.

Parker, M. & Kingori, P., 2016. Good and Bad Research Collaborations: Researchers' Views on Science and Ethics in Global Health Research. *PLoS ONE,* 11(10), p. e0163579.

Parsons, C. A., Tsaneva, I., Lloyd, R. G. & West, S. C., 1992. Interaction of Escherichia coli RuvA and RuvB proteins with synthetic Holliday junctions. *Proceedings of the National Academy of Sciences of the USA,* Volume 89, pp. 5452-5456.

Parsons, C. A. & West, S. C., 1993. Formation of a RuvAB-Holliday Junction Complex in Vitro. *Journal of Molecular Biology,* 232(2), pp. 397-405.

Pays, E., 2005. Regulation of antigen gene expression in Trypanosoma brucei. *Trends in Parasitology,* 21(11), pp. 517-520.

Pearce, F., 2000. Inventing Africa. *New Scientist,* p. 30.

Portman, N. & Gull, K., 2010. The paraflagellar rod of kinetoplastid parasites: From structure to components and function. *International Journal for Parasitology,* 40(2), pp. 135-148.

Punatar, R. S. et al., 2017. Resolution of Single and Double Holliday Junction Recombination Intermediates by GEN1. *Proceedings of the National Academy of Sciences of the USA,* 114(3), pp. 443-450.

Qiu, J., Li, X., Frank, G. & Shen, B., 2001. Cell Cycle-dependent and DNA Damage-inducible Nuclear Localization of FEN-1 Nuclease Is Consistent with Its Dual Functions in DNA Replication and Repair. *Journal of Biological Chemistry,* 276(7), pp. 4901-4908.

Qui, J., Bimston, D. N., Partikian, A. & Shen, B., 2002. Arginine Residues 47 and 70 of Human Flap Endonuclease-1 Are Involved in DNA Substrate Interactions and Cleavage Site Determination. *Journal of Biological Chemistry,* 277(27), pp. 24659-24666.

Rass, U. et al., 2010. Mechanisms of Holliday junction reolsution by the human GEN1 protein. *Genes & Development,* 24(14), pp. 1559-1569.

Reagan, M. S., Pittenger, C., Siede, W. & Friedberg, E. C., 1995. Characterization of a Mutant Strain of Saccharomyces cerevisiae with a Deletion of the RAD27 Gene, a Structural Homolog of the RAD2 Nucleotide Excision Repair Gene. *Journal of Bacteriology,* 177(2), pp. 364-371.

Rudenko, G., 2011. African trypanosomes: the genome and adaptions for immune evasion. *Essays in Biochemistry,* Volume 51, pp. 47-62.

Saharia, A. et al., 2008. Flap Endonuclease 1 Contributes to Telomere Stability. *Current Biology,* 18(7), pp. 496-500.

Sakofsky, C. J., Ayyar, S. & Malkova, A., 2012. Break-Induced Replication and Genome Stability. *Biomolecules,* 2(4), pp. 483-504.

Sarbajna, S. & West, S. C., 2014. Holliday junction processing enzymes as guardians of genome stability. *Trends in Biochemical Sciences,* 39(9), pp. 409-419.

Schmid-Hempel, P., 2009. Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Philosophical Transactions of the Royal Scoiety B: Biological Sciences,* 364(1513), pp. 85-98.

Schwede, A., Macleod, O. J., MacGregor, P. & Carrington, M., 2015. How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins. *PLoS Pathogens,* 11(12).

Sharm, R. et al., 2008. Asymmetric Cell Division as a Route to Reduction in Cell Length and Change in Cell Morphology in Trypanosomes. *Protist,* 159(1), pp. 137-151.

Sharples, G. J. & Lloyd, R. G., 1991. Resolution of Holliday Junctions in Escherichia coli: Identification of the ruvC Gene Product as a 19-Kilodalton Protein. *Journal of Bacteriology,* 173(23), pp. 7711-7715.

Shen, B., Nolan, J. P., Sklar, L. A. & Park, M. S., 1996. Essential Amino Acids for Substrate Binding and Catalysis of Human Flap Endonuclease 1. *Journal of Biological Chemistry,* 271(16), pp. 9173-9176.

Shiba, T., Iwasaki, H., Nakata, A. & Shinagawa, H., 1993. Escherichia coli RuvA and RuvB proteins involved in recombination repair: physical properties and interactions with DNA. *Molecular and General Genetics MGG,* 237(3), pp. 395-399.

Sima, N., McLaughlin, E. J., Hutchinson, S. & Glover, L., 2019. Escaping the immune system by DNA repair and recombination in African trypanosomes. *Open Biology,* 9(11), p. 190182.

Simarro, P. P. et al., 2012. Update on field use of the available drugs for the chemotherapy of human African trypanosomiasis. *Parasitology,* 139(7), pp. 842-846.

Singha, U. K. et al., 2012. Protein Translocase of Mitochondrial Inner Membrane in Trypanosoma brucei. *Journal of Biological Chemistry,* 287(18), p. 14480–14493.

Stodola, J. L. & Burgers, P. M., 2016. Resolving individual steps of Okazaki fragment maturation at msec time-scale. *Nature Structural & Molecular Biology,* 23(5), pp. 402-408.

Storici, F. et al., 2002. The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair. *The EMBO Journal,* 21(21), pp. 5930-5942.

Sun, L. et al., 2014. Expression and Localization of GEN1 in Mouse Mammary Epithelial Cells. *Journal of Biochemical and Molecular Toxicology,* 28(10), pp. 450-455.

Swuec, P. & Costa, A., 2014. Molecular mechanism of double Holliday junction dissolution. *Cell and Bioscience,* 4(36).

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W., 1983. The double-strand-break repair model for recombination. *Cell,* 33(1), pp. 25-35.

Taylor, J. E. & Rudenko, G., 2006. Switching Trypanosome coats: what's in the wardrobe?. *Trends in Genetics,* 22(11), pp. 614-620.

Tsaneva, I. R., Illing, G., Lloyd, R. G. & West, S. C., 1992. Purification and properties of the RuvA and RuvB proteins of Escherichia coli. *Molecular and General Genetics MGG,* 235(1), pp. 1-10.

Tsaneva, I. R., Müller, B. & West, S. C., 1992. ATP-dependent branch migration of holliday junctions promoted by the RuvA and RuvB proteins of E. coli. *Cell,* 69(7), pp. 1171-1180.

Tsutakawa, S. E. et al., 2011. Human Flap Endonuclease Structures, DNA Double Base Flipping and a Unified Understanding of the FEN1 Superfamily. *Cell,* 145(2), pp. 198-211.

van Gool, A. J., Hajibagheri, N. M., Stasiak, A. & West, S. C., 1999. Assembly of the Escherichia coli RuvABC resolvasome directs the orientation of Holliday junction resolution. *Genes and Development,* 13(14), pp. 1861-1870.

Vickerman, K., 1985. Developmental Cycles and Biology of pathogenic trypanosomes. *British Medical Bulletin,* 41(2), pp. 105-114.

Vilenchik, M. M. & Knudson, A. G., 2003. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 100(22), pp. 12871-12876.

Vink, C., Rudenko, G. & Seifert, H., 2012. Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiology Reviews,* 36(5), pp. 917-948.

Wechsler, T., Newman, S. & West, S. C., 2011. Abberant chromosome morphology in human cells defective for Holliday junction resolution. *Nature,* 471(7340), pp. 642-646.

WHO, 2018. *Trypanosomiasis, human African Fact Sheet.* [Online]
Available at: http://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness)

Witze, E. S., Old, W. M., Resing, K. A. & Ahn, N. G., 2007. Mapping protein post-translational modifications with mass spectrometry. *Nature Methods,* 4(10), pp. 798-806.

Wyatt, H. D., Sarbajna, S., Matos, J. & West, S. C., 2013. Coordinated Actions of SLX1-SLX4 and MUS81-EME1 for Holliday Junction Resolution in Human Cells. *Molecular Cell,* 52(2), pp. 234-247.

Wyatt, H. D. & West, S. C., 2014. Holliday Junction Resolvases. *Cold Spring Harbor Perspectives in Biology,* 6(9), p. a023192.

Yuan, J. & Chen, J., 2010. MRE11-RAD50-NBS1 Complex Dictates DNA Repair Independent of H2AX. *Journal of Biological Chemistry,* 285(2), pp. 1097-1104.

Zafeiriou, D. et al., 2001. Xeroderma Pigmentosum Group G with Severe Neurological Involvement and Features of Cockayne Syndrome in Infancy. *Pediatiric Research,* Volume 49, pp. 407-412.

Zheng, L. et al., 2011. Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Research,* 39(3), pp. 781-794.