Characterisation of potential Holliday junction resolvases in *Trypanosoma brucei*

Author

Alexander Benjamin Lloyd Jones

Supervisors

Dr Paul McKean and Dr Fiona Benson

Submitted for the degree of Master of Science (Research)

to Lancaster University

September 2017

Contents

Figures and Tables

Figures

Tables

Appendix

Acknowledgements

Firstly I would like to thank my supervisors Dr Paul McKean and Dr Fiona Benson for their support, guidance and advice given throughout the course of my MSc by Research project, with a special thanks to Dr Mick Urbaniak for his general recommendations and instruction in proteomics. I am very grateful to Xin Qi for her patience, teaching and her kindness within the lab, and for helping develop my laboratory based skills. For the simplifying my project I am thankful to William McAllister for allowing the use of his previously purified *Tb*FEN1 protein for immunisation and for his assistance with the GEHealthcare AKTA systems. I would also like to thank Dr Corinna Benz for supplying bloodstream *T.brucei* for use in my project.

Many appreciations for the support of technicians throughout the department with special recognitions to Kathryn Lamb and Christine Campbell, for their kindness, support and help. Last but not least, I would like to thank all of the friends present during my Masters degree Alice Carter, James Tollitt and Filipe Duarte; specifically Harriet Steele and Amber Lynch who have continuously taught me throughout my University education.

Declaration

I Alex Jones, declare the contents of this thesis “Characterisation of potential Holliday junction resolvases in *Trypanosoma brucei*” to be my own work. This work has not been submitted in the same form for the award of a higher degree level elsewhere.

Abbreviations

Full Term	Abbreviation
Basic Local Alignment Search Tool	BLAST
Blasticidin S Deaminsase Gene	*bsr*
Blood Brain Barrier	BBB
Bovine Serum Albumin	BSA
Centres for Disease Control and Prevention	CDC
Deoxyribonucleic Acid	DNA
Deoxyribonucleotide triphosphate	dNTP
Double Strand Break	DSB
Double Stranded DNA	dsDNA
Expression Site	ES
Flagellar Attachment Zone	FAZ
Flagellar Pocket	FP
Food and Agriculture Organization of the United Nations	FAO
GAFSIMPAM Epitope	GSP
Haemoglobin	Hb
Haptoglobin	Hp
Hidden Markov Modelling	HMM
Holliday Junction	HJ
Homologous Recombination	HR
Human African Trypanosomiasis	HAT
Immobilised Metal Affinity Chromatography	IMAC
Ion Exchange Chromatography	IEX
Isopropyl β-D-1-Thiogalactopyranoside	IPTG
Luria Broth	LB

Nucleotide Excision Repair	NER
Open Reading Frame	ORF
Phosphate Buffered Saline	PBS
Polymerase Chain Reaction	PCR
PCR Only Tagging	pPOT
Semi Defined Medium-79	SDM-79
Sodium Dodecyl Sulphate	SDS
Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis	SDS-PAGE
Trifluoroacetic Acid	TFA
Tris Acetate-EDTA Buffer	TAE
Trypanosoma brucei Lytic Factor	TLF
Trypanosoma brucei brucei	*T.b.brucei*
Trypanosoma brucei gambiense	*T.b.gambiense*
Trypanosoma brucei rhodesiense	*T.b.rhodesiense*
Untranslated Region	UTR
Variant Surface Glycoproteins	VSG
World Health Organisation	WHO
Yellow Fluorescent Protein	YFP

Abstract

Antigenic variation on the cell surface of the *T. brucei* is essential for the survival of the parasite in the host immune system. Finding mechanisms to stop variant surface glycoprotein (VSG) switching, and hence the parasites ability to evade the host immune system provides a promising opportunity to treat the infection. Holliday junctions are an essential part of homologous recombination, which is used for DNA repair and for the crossover of DNA. Identifying *T. brucei* proteins which act as Holliday junction resolvases are potential candidate targets for inhibiting VSG switching.

In this project, bioinformatics methods were used to identify proteins with homology to human GEN1, a known Holliday junction resolvase. Purification of homologous proteins were carried out from the expression of bacterial cell lines. Parasite cell lines were created for the investigation of the homologous protein localisation and a BioID analysis to interrogate protein-protein interactions.

Bioinformatics analyses identified two proteins with homology to human GEN1, *Tb*FEN1 and *Tb*RAD2. Of these two proteins, *Tb*FEN1 underwent large scale purification and was used for the production of polyclonal antisera, where *Tb*RAD2 was found to require stringent denaturing conditions for purification. Both proteins were seen to localise within the nucleus of the procyclic parasite. Finally, BioID analysis showed protein-protein interactions with *Tb*FEN1 were found within the nucleus and a complex banding pattern of proteins observed.

This work has been able to identify two proteins which potentially act as Holliday junction resolvases in the *T. brucei*. Further investigation of their function and protein interactions will be important to understand whether these provide functional targets for VSG switching.

1 Literature Review

Trypanosoma are a genus of disease causing protozoan parasite. *Trypanosoma brucei (T. brucei),* which is the focus of this work, is the cause of two diseases, Human African Trypanosomiasis (HAT) and Nagana. The disease HAT is classed as a neglected tropical disease, where it is commonly referred to as sleeping sickness given the main symptom of the disease (Kennedy, 2013). Conversely, Nagana infects animals and is a wasting disease in livestock, including cattle (Wells, 1972). Both of these diseases cause severe health and socioeconomic problems for countries circling the equatorial line of Africa. These are some of the poorest and most rural countries in the world (Rutto et al., 2013). Despite evolutionary advances in ape-like mammals to protect against *T. brucei*, the parasites remain resilient to the host immune system, and have evolved to evade these mechanisms (Stijlemans et al., 2016). Understanding the molecular biology of *Trypanosoma*, and the mechanisms used to evade the host immune system, is of paramount importance as we come closer to the World Health Organisation's (WHO) goal for elimination of HAT by 2020 (Franco et al., 2017).

This introduction will give an overview of the epidemiology of *T. brucei*, how it infects and interacts with its host, including the molecular mechanisms used, and finally the current understanding of homologous recombination and the use of Holliday junctions (HJ) for host immune system evasion.

1.1 An introduction to *Trypanosoma brucei*

T. brucei are protozoan parasites that cause human and veterinary diseases, such as Human African trypanosomiasis (HAT), bovine trypanosomiasis (Nagana) and Chagas' disease (Englund et al., 1982). The *T. brucei* species that cause Nagana and HAT are prevalent in central and southern Africa, (see **Figure 1.1)**. The vector for transmission of *T. brucei* is the tsetse fly, endemic to this region (Kennedy, 2013). Not only do these diseases occur in the poorest and most rural areas, the burden of disease affects those mainly of productive age. This consequently decreases a family's ability to generate income, as well as costly medical treatment and a stigmatism for the disease leading to the potential abandonment of those afflicted (Franco et al., 2014); worsening an already desperate economic situation.

There are three sub-species of the *T. brucei* parasite prevalent within 36 sub-Saharan countries in Africa. These species are morphologically identical, yet their genetics, virulence and choice of host vary (WHO, 2016, Englund et al., 1982). *Trypanosoma brucei gambiense* (*T. b. gambiense*) and *Trypanosoma brucei rhodesiense* (*T. b.*

rhodesiense) are both capable of causing disease in humans. However, the two parasites cause symptomatically distinct illnesses, and hence the disease diagnosis are called gambiense HAT and rhodosiense HAT respectively. *T.b gambiense* is only capable of causing disease in humans. The main causes of the Nagana infection in livestock are *Trypanosoma brucei brucei* (*T. b. brucei*) and *T.b.rhodesiense* (Cordon-Obras et al., 2015). These are zoonotic parasites with reservoirs in wild and domesticated animals. Infection of *T. b. rhodosiense* in humans is therefore considered accidental, with relatively low levels of human infection. Gambiense HAT is an anthroponotic disease and subsequently has a large human reservoir (Anderson et al., 2011, Franco et al., 2014).

Sub-species of *T. brucei* are transmitted by the salivary glands of certain species of the *Glossina* genus of the Tsetse fly (Kennedy, 2013). Tsetse flies are the definitive host of these protozoan parasites. With some rare exceptions, trypanosomiasis is only seen where the tsetse fly is present in 36 Sub-Saharan African countries; known as the tsetse belt, where the climate is optimal for *Glossina* survival (Robinson et al., 1997).

Gambiense HAT is the West and Central African variant of trypanosomiasis and is responsible for roughly 97-98% of HAT cases (Franco et al., 2014). The disease caused is less virulent than its East Rhodesiense HAT counterpart. Without treatment both diseases are considered to have a case fatality rate close to 100%, although there have been some reported cases of HAT survivors (Kennedy, 2013). Consequently, the WHO set the target to eliminate Gambiense HAT as a health problem by 2020, which has been developed to a more realistic target of complete interruption of transmission by 2030 (Franco et al., 2017). Annual surveillance of the disease predicts the rate of infection to have decreased by 73% from 25,000 to 7,000 between 2000 and 2012 (Simarro et al., 2015, Simarro et al., 2014). **Figure 1.1** illustrates the prediction and distribution of these cases. It is estimated, however, that for every reported case of HAT there are two cases that go undetected therefore there are estimated to be 20,000 actual cases of HAT annually (WHO, 2013). The rate of infection is relatively low, as there are roughly 70 million people at risk of becoming infected. An estimated 82.2% of these individuals are at risk of being infected with Gambiense HAT in 24 countries and 17.8% at risk of being infected with Rhodesiense HAT in 12 countries (Simarro et al., 2012, WHO, 2016). There is a certain uncertainty to these estimates due to the communities at risk living in highly rural and inaccessible areas.

Nagana plays its own role at debilitating the economic and social potential of these at risk countries. It has been difficult for previous studies to determine the exact infection

rate of Nagana in domesticated and individual animals, because the communities affected are difficult to set up sufficient surveillance systems given their localities. Despite this, the Food and Agriculture Organization (FAO) of the United Nations states that there is a risk of infection to around 50 million heads of cattle, with the risk of around 3 million deaths per year. The economic loss in cattle alone is estimated at US$ 1.0 – 1.2 billion per year, taking into consideration lower birth rates, decrease in milk production and animals being too weak to work. It has been estimated that in tsetse infested lands there is a loss in total agricultural domestic product of US$ 4.75 billion per year (Connor, 1994, FAO, 2016).

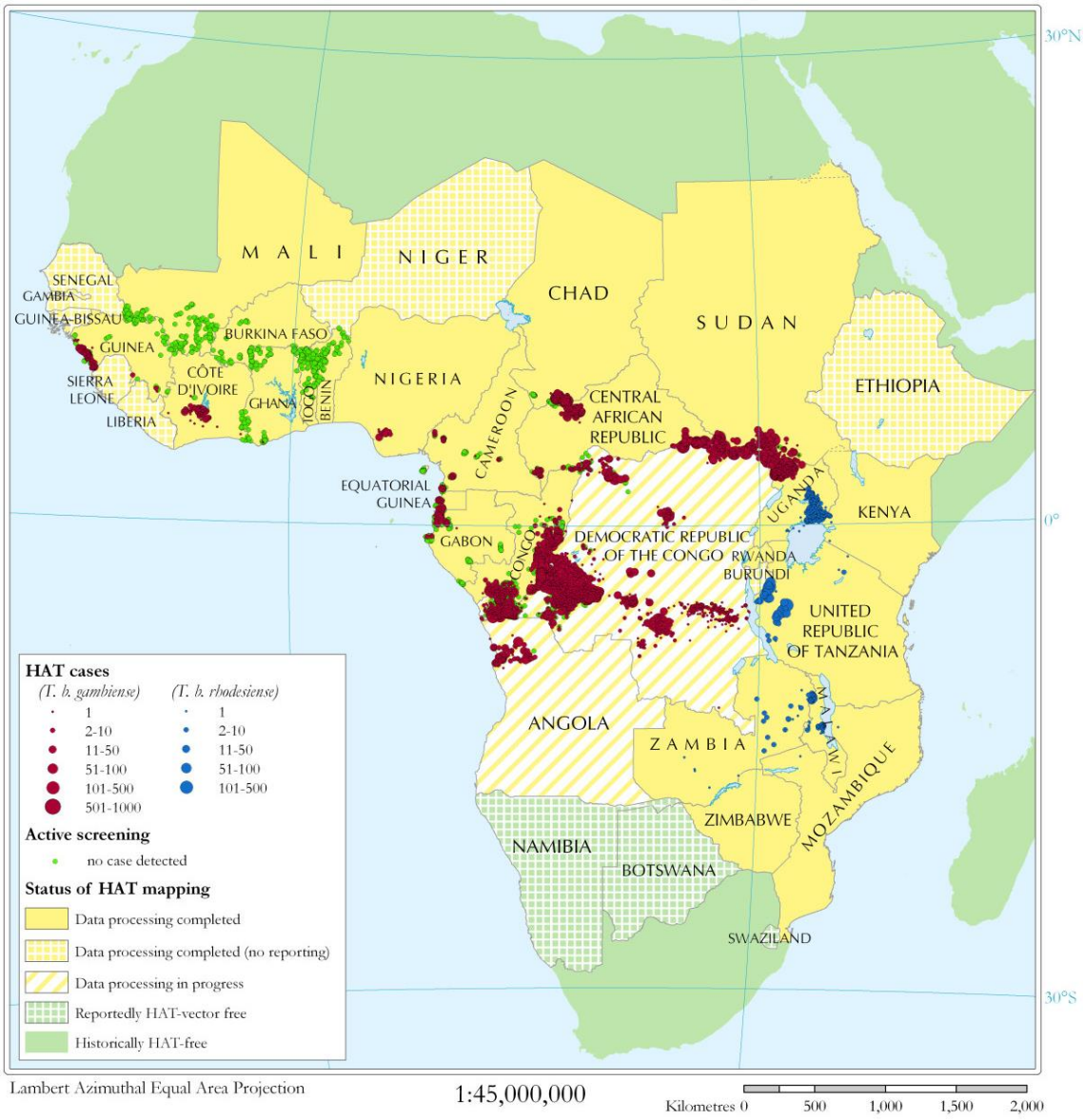


Figure 1.1. Map of HAT distribution. A map to show the representation of *T.b.gambiense* and *T.b.rhodosiense* distribution of infection in Africa, data recorded between 2000 – 2009 following the resolution of the WHO to commit itself to supporting HAT-endemic countries. Taken from (Simarro et al., 2010).

1.2 Cell Structure, Cell Division and Cell Cycle of *Trypanosoma brucei*

The *T. brucei* cell includes, but is not limited to, a single flagellum, nucleus, basal body, an elongated mitochondrion, a kinetoplast that contains the mitochondrial genome and a series of specialised microtubules associated with the flagellum attachment zone (FAZ). As with all eukaryotes, to ensure that viable daughter cells are produced following cytokinesis, these organelles must be replicated and segregated with fidelity (Hammarton et al., 2007, Kohl and Gull, 1998).

The *T. brucei* cell division cycle is illustrated in **Figure 1.2**. This figure shows how the division develops from the FAZ to the division of a new kinetoplast, the subsequent replication of a new nucleus and finally the segregation of the daughter cell's organelles (Hammarton et al., 2007). Understanding the cell cycle and the structure of the parasite as it divides will be of importance for knowing which stage the cell is in under microscopy.

During its life cycle, *T. brucei* undergo structural and morphological changes allowing them to survive and propagate within their hosts. **Figure 1.3** shows this progression of the *T. brucei* life cycle between the Tsetse fly and human host. This life cycle shows the Tsetse fly as the definitive host, as this is where the parasite reproduces and reaches maturity (Gibson et al., 2008). The human is the intermediate or reservoir host, where a small section of the life cycle is completed and the *T. brucei* is rapidly ready for transmission into a new Tsetse fly. From this figure we can see how the parasite moves from different life cycle stages, its localisation within its hosts and prepares for infection.

Tsetse flies inject the metacyclic form of the *T. brucei* during a blood meal. Following injection, the parasites migrate to the lymphatic system and pass into the bloodstream of the host. Here they transform into the slender bloodstream form, once the parasites reach a certain population number they then differentiate into the stumpy form. The stumpy form is initiated by quorum sensing of the stumpy induction factor (Zimmermann et al., 2017). The parasites move from the spinal fluid and migrate across the blood brain barrier (BBB), although the mechanism for this depends on the species of *Trypanosoma*. The life cycle is continued through a second blood meal by an uninfected Tsetse fly, where the fly ingests the bloodstream form of the parasite. In the fly the parasites transform into the procyclic form and migrate through the midgut to the salivary gland, where the parasite is ready to infect again as the metacyclic form (Englund et al., 1982, CDC, 2015).

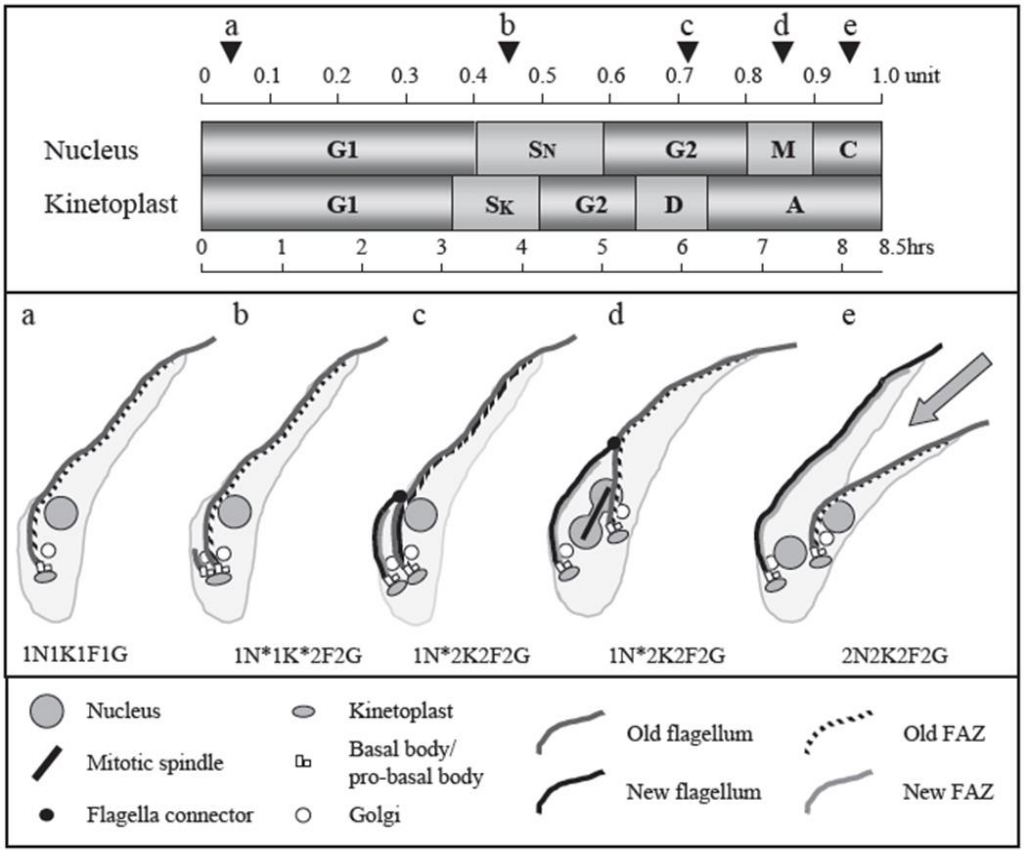


Figure 1.2 Cell structure and division of *T. brucei*. A graphic showing the process of the cell cycle for the parasite *T. brucei*. Comparing the top and bottom panels illustrate the process of change to the organelles and the cell throughout the different stages. Cell cycle stages G1 and G2: Gap phases, S_N and S_K: Synthesis for nucleus and kinetoplast, D: Division of kinetoplast, M: Mitosis of nucleus, A: Kinetoplast segregation and C: Cleavage of the cell. Cell organelles N: Nucleus, K: Kinetoplast, F: Flagellum and G: Golgi, where the number states how many of that organelle is present within the cell. Taken from (Hammarton et al., 2007).

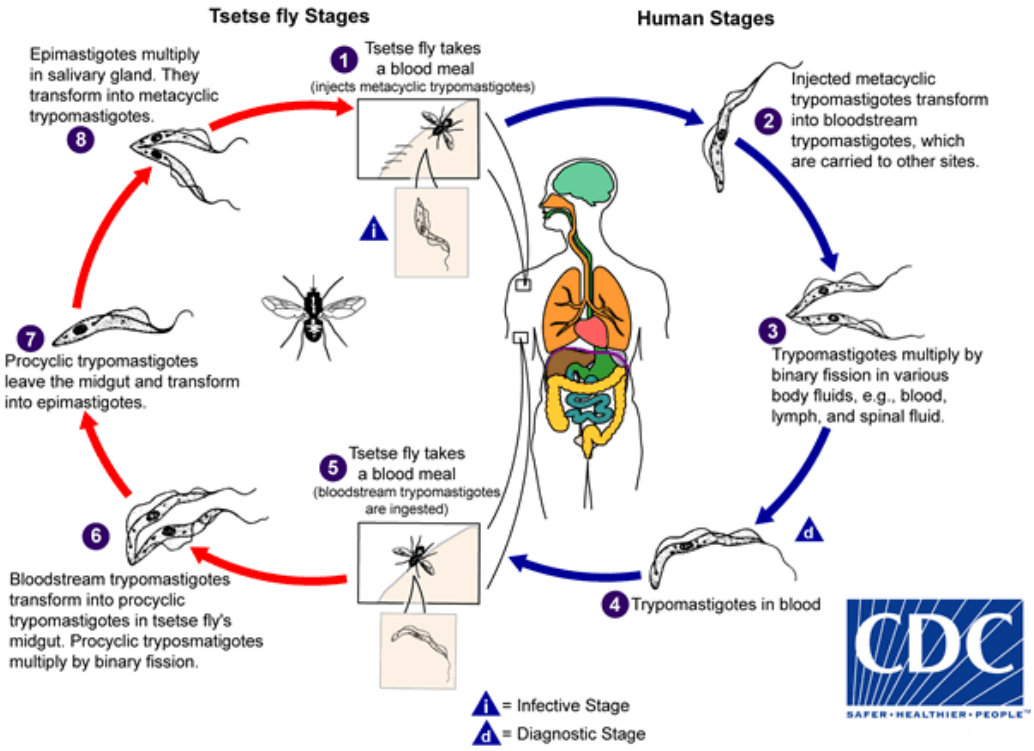


Figure 1.3. Life cycle of *T. brucei*. An illustration to show the life cycle of the parasite *T. brucei* as it migrates between hosts. Stage 1 shows the mammal intermediate host being infected by the metacyclic trypomastigote during a blood meal. In subsequent stages 2-4 the parasite transforms in the bloodstream form and migrates to different areas of the body including the blood, lymph and spinal fluid. In the blood the parasite undergoes replication and can infect new Tsetse flies through a blood meal (stage 5). Stages 6-8 show the parasite undergoing different transformations, here the parasite reaches maturity (Gibson et al., 2008). After 3 weeks, the parasite transforms into the metacyclic form and present in the flies salivary gland ready to infect new mammals. Taken from (CDC, 2015).

1.3 Immune Evasion by *Trypanosoma brucei*

Within the host, *T. brucei* are continually in contact with the host's immune system, leading to a constant arms race between immunity and evasion. The constant threat from both the innate and adaptive immune response of the host has led to the evolution of evasion techniques by the parasite.

1.3.1 Host Innate Immune Response

Many species of primate, including humans, have evolved to have trypanosome lytic factors (TLFs), a type of high density lipoprotein, which acts as part of the innate immune response. TLFs work by using a haptoglobin (Hp)-related protein to help bind to haemoglobin (Hb), a food source for the parasite found within the bloodstream. *T. brucei* are able to ingest Hb from the blood. As TLFs are bound to the Hb, the trypanosome also ingests TLFs with the Hb food source (Bishop et al., 2001, Samanovic et al., 2009, Thomson et al., 2009). TLFs contains apolipoprotein L1 (ApoL1), which is suggested to be the lytic factor. Where it forms pores in the *T. brucei* lysosome membrane, resulting in the depolarization of the lysosome. These osmotic changes are proposed as the cause of the parasite lysis (Perez-Morga et al., 2005, Shiflett et al., 2005). Of the three species of trypanosome being discussed in this report *T.b.gambiense* and *T.b.rhodesiense* have evolved to evade this method of immunity; leaving only *T.b.brucei* susceptive to TLFs (Thomson et al., 2009).

1.3.2 VSG Function

To protect the parasite from both the host's acquired and innate immune responses *T. brucei* sub-species have a variant surface glycoprotein (VSG) coat. This coat contains a variable surface and also protects the invariant surface proteins (Ziegelbauer and Overath, 1993).

VSGs are able to protect the parasite in multiple ways, largely as a result of their structure. The structure allows for tight packaging of the proteins on the surface coat, maximizing the number of them. VSGs are expressed mono-allegically and carry out coat cleansing and VSG switching to alter the VSG expressed (Horn, 2014).

1.3.3 VSG Structure

The VSG has a characteristic structure containing an N-terminal variable region, a large variable region, a conserved region and a C-terminus GPI-anchor. Both the N-terminal signal sequence and the C-terminal GPI anchor have around 20 amino acids, the variable region has approximately 360 amino acids and the conserved region has approximately 100 amino acids (Baral, 2010).

The N-terminal variable region of the VSG is made up of two antiparallel α-helices, considered to be the variable region. It is this variable region which helps to protect the protein from the acquired immune response, through antigenic variation. The conserved region of the protein allows for the dense packing of the VSGs on the surface coat (Rudenko, 2005).

There are several possibilities for why the parasite has a GPI-anchor; they may allow for a higher packing density on the cell surface, the trafficking of VSGs, for recycling of the VSG coat or for the shedding of the VSG coat which requires GPI specific phospholipase C (Lillico et al., 2003).

On the surface of each parasite at least 95% of the external exposed cell consists of VSGs, where there are estimated to be about 10^7 VSGs (Auffret and Turner, 1981) forming a densely packed layer on the parasite. This layer protects the parasites invariant surface proteins from the acquired immune response, as they stand slightly lower than the VSG particles. These invariant proteins have been found to have a similar tertiary structure to the larger VSG, allowing the proteins to pack tightly on the surface (Carrington and Boothroyd, 1996); permitting the parasite to defend against the innate complement pathway immune response (Rudenko, 2005).

1.3.4 Surface Coat Cleansing

Trypanosomes are able to carry out a ‘coat cleaning mechanism’, which is the process in which its VSGs are endocytosed by the flagellar pocket (FP) and then completely replaced. Endocytosis is facilitated by the GPI specific phospholipase C and hydrodynamic flow. Hydrodynamic flow enables the surface coat cleansing mechanism by creating a directional movement of the VSGs bound with host immune products towards the flagellar pocket for removal (Engstler et al., 2004).

By endocytosing the VSGs the parasite is able to remove any host bound proteins, the host proteins are lysed in the cytoplasm. Such host proteins to be lysed are the anti-

VSG antibodies. Once the host has mounted an acquired immune response to the presented VSGs, the parasite will be attacked by anti-VSG antibodies. Bloodstream surface VSGs are believed to have an entire population turnover every 12 minutes (Engstler et al., 2004). Through the process of 'coat cleaning' the parasite is able to defend itself against a low concentration of anti-VSGs, this occurs until a threshold concentration is reached where the parasite succumbs to the immune response (Russo et al., 1993, Engstler et al., 2007). Therefore, allowing for the specific VSG presented to have an increased lifespan.

1.3.5 Monoallelic VSG Presentation

Trypanosomes are monoallelic for VSG presentation, where the trypanosome itself will only express one specific VSG sequence in the variable region at any one time. This limits the number of VSG sequences the host immune system has a response for. To ensure this method works, the majority of the trypanosome population presents the same specific VSG sequence. By most of the population expressing one VSG sequence, the small population that may present an alternate VSG sequence will be protected through the 'herd' effect (Munoz-Jordan et al., 1996). Once the host immune system has been able to raise an effective response to the major VSG presented the trypanosome population will fall. Those trypanosomes presenting the minority VSG sequence will be unaffected by this immune response and will survive to allow for their specific VSG sequence to become the next major sequence presented (Landeira et al., 2009). The hosts acquired immune system will be forced to begin the process once again for the new major sequence being presented, without ever fully eliminating the parasite population.

Through this allelic exclusion resulting in one major VSG sequence being expressed at a time, the parasite population is able to survive multiple acquired immune responses. It is understood that trypanosomes have a library of over 2,000 different VSG sequences that they are able to present (Horn, 2014).

1.3.6 VSG Switching

The major VSG sequence being presented switches as the parasite population rises and then falls in response to the host mounting an acquired immune response. This method allows for extended evasion of the host immune response (Mugnier et al., 2015).

The extensive library of VSG sequences originates from the trypanosome genomic organisation. When the VSG sequence is prepared for expression, it must first be present in the active expression site (ES) as one of the expression site associated genes (ESAGs) (Horn, 2014).

The ES and VSG sequences are held on the telomeric arrays of the parasite chromosomes, allowing for ease of switching of the VSG sequence into the ES. To move into the ES, the VSGs are switched through the process of Homologous Recombination (HR) (Conway et al., 2002). Homologous recombination allows crossover of sections of DNA sequences, allowing a DNA sequence to 'jump' to another chromosome and hence switch. This process usually occurs near the telomeric arrays of chromosomes (Shinohara and Ogawa, 1995).

The trypanosome parasite has a karyotype of 11 megabase chromosomes, three to five intermediate chromosomes and approximately 100 mini-chromosomes (Ogbadoyi et al., 2000). It is the megabase chromosomes and the intermediate chromosomes which hold silent VSGs along with the ES for the VSG sequences (Marcello and Barry, 2007). There are approximately 20 VSG expression sites per trypanosome genome (Hertz-Fowler et al., 2008). The mini-chromosomes do not hold any ES for the VSG sequences, but hold 150-250 silent VSG sequences (Alsford et al., 2009).

Trypanosome populations use three different methods to switch which particular VSG is being transcribed. The most limited mechanism is transcriptional switching, this is where a different ES is activated as the ES to be transcribed. This will only use the VSGs which are currently present in the ESAGs (Rudenko, 1999). To access the VSGs present on other chromosomes the parasite carries out telomere exchange, where by recombination the silent VSGs can be moved into the ES by recombination. For the greatest range of VSG sequences, the parasite can undergo homologous recombination of different VSG sequences to create pseudo-VSG sequences (Horn and Barry, 2005), as depicted in **Figure 1.4**.

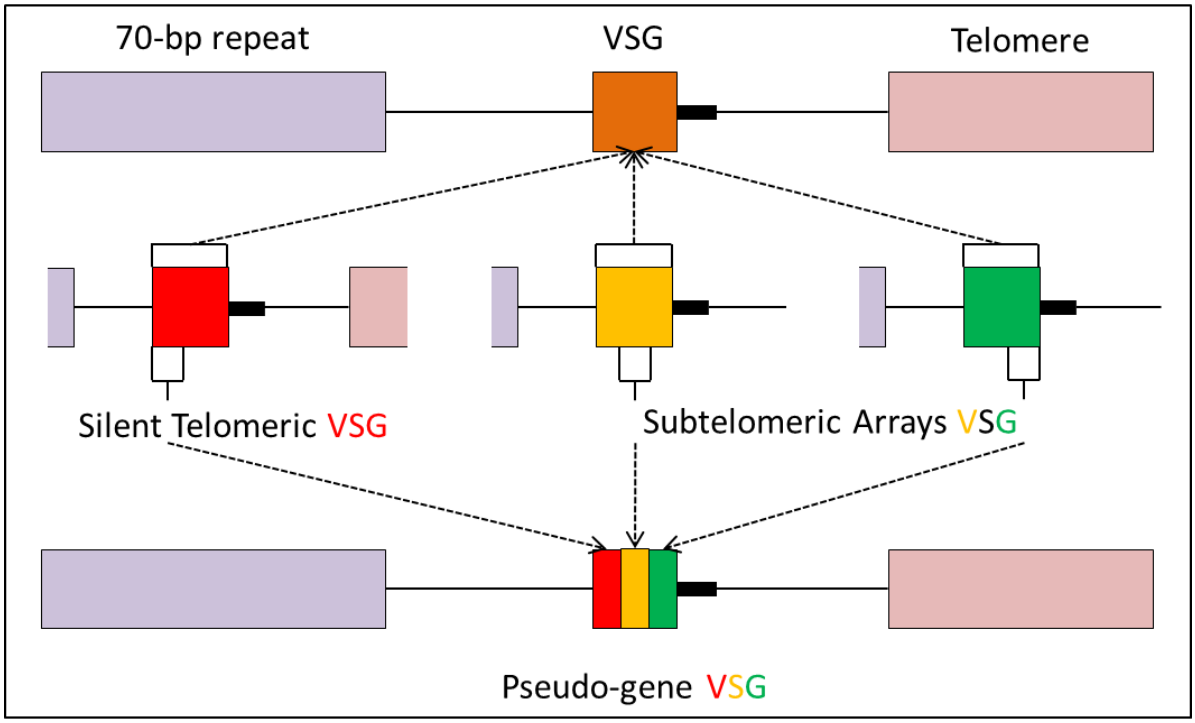


Figure 1.4. Illustration showing the switching of VSGs in the genome. Graphic showing how different VSG genes can be moved from telomeric sites on a chromosome into an expression site of a different chromosome. Pseudo genes made up of multiple different VSG sequence can be made and also added to expression sites. The dotted arrow lines represent homologous recombination.

1.4 Introduction to Homologous Recombination and Holliday Junctions

Homologous recombination is a method of genetic recombination at two identical or nearly identical DNA strands, such as those between homologous chromosomes or more commonly, sister chromatids. Homologous recombination has a critical role in maintaining genome integrity and provides a mechanism to generate diversity. The process is mainly initiated at double strand breaks (DSB) by a meiosis specific endonuclease. Homologous recombination can have specific functions between species, such as VSG switching in *T. brucei* for immune evasion (Heyer et al., 2010, Conway et al., 2002).

The recombination pathway consists of four main stages: 1) homologous pairing 2) exchange of DNA strands 3) template DNA synthesis 4) product generation. The first stage, homologous pairing happens when two chromosomes, one with a DSB, enter close proximity to each other, where one has DSB in the DNA. This allows for exchange of the DNA between participating DNA strands. The chromosome with the double stand break carries out strand invasion of the homologous chromosome. The DNA strands can then be used as a template for DNA synthesis, which is carried out by DNA polymerase. Finally, either dissolution (non-crossover) or resolution (crossover and non-crossover) of the recombination intermediates to generate the DNA repair or a DNA crossover product (Capecchi, 1989, Camerini-Otero and Hsieh, 1995).

The process of HR is initiated when a DSB occurs in the DNA sequence. The chromosome containing the DSB moves into close proximity with a whole homologous chromosome. Resection of the DNA sequence occurs, depending on the organism the DNA resection will either occur at the 5’ or 3’ end of the DNA. As resection occurs at only one end of the DNA strand, this leaves a single strand DNA overhang.

The single strand DNA overhang invades the whole homologous chromosome DNA. As the DNA strand invades, it displaces the previously bound DNA strand creating a displacement loop. The newly invaded DNA strand can now undergo synthesis with DNA polymerase using the homologous chromosome as a template DNA strand. Following DNA synthesis, the invasion strand realigns back onto its chromosome in second end capture.

Following second end capture either one or two HJs are formed. Holliday junctions are cross-stranded sections of DNA composed of four interacting stands, first proposed by Robin Holliday. This structure has been widely accepted to represent the intermediate structure of recombination (Holliday, 1964). To complete the process of HR the HJs need to be resolved, from **Figure 1.5** resolution can occur in different ways.

In **(a)** the DNA strands undergo dissolution, here the displacement loop collapses forming a non-crossover product. Dissolution is the common method used in eukaryotes to create non-crossover products (Wyatt and West, 2014).

Process **(b)** shows two different methods of resolution. Cleavage of the HJs are catalysed by endonucleases called HJ resolvases, this cleavage causes two unconnected DNA duplexes, which can rapidly be re-ligated using DNA ligase. Resolvases work by creating two almost simultaneous nicks in the DNA, which can occur symmetrically or asymmetrically (Wyatt and West, 2014). Depending how the DNA strands of the HJ are resolved, depends on whether a crossover or non-crossover product is formed.

Figure 1.6 depicts the process of symmetrical resolution for a single HJ structure. In this example by cutting vertically a recombinant product is formed, showing the crossover of DNA between chromosomes. Such a crossover of DNA between chromosomes can move different genes between chromosomes. This example of resolving horizontally forms a non-recombinant product, showing non-crossover of DNA between different chromosomes. Non-crossover of DNA between chromosomes can be used as a process of DNA repair (Kowalczykowski et al., 1994).

The process of resolution can produce the products of both crossover and non-crossover products. Humans have two HJ resolvases which have been characterized within the cells, the SLX1-SLX4-MUS81-EME1 (SLX-MUS) protein complex and GEN1 (Sarbajna and West, 2014).

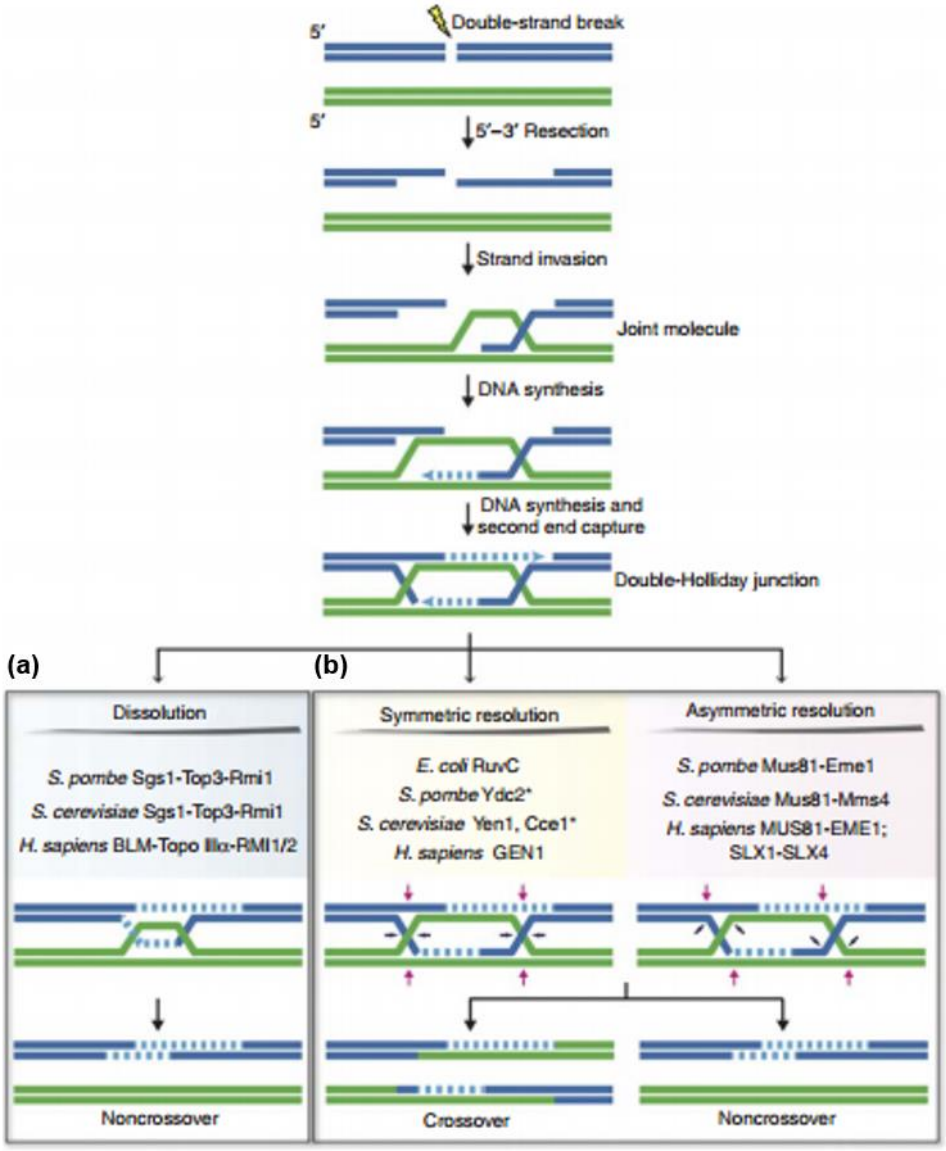


Figure 1.5. Homologous recombination. Schematic showing the process of HR initiated by a double strand break of the DNA. Strand invasion occurs by the resected DNA to form a joint module and a displaced loop. Branch migration of the invaded strand occurs between the two chromosomes, where the displaced loop can anneal to the non-invading resected strand by a process of second end capture. Here the invading strand can replace the donor strand to form a single Holliday junction structure (not shown), or can anneal back to its original damaged DNA to form a double Holliday junction structure. Box (a) shows the process of dissolution, where box (b) shows the process of resolution of the Holliday junctions. Image taken from (Wyatt and West, 2014)

.

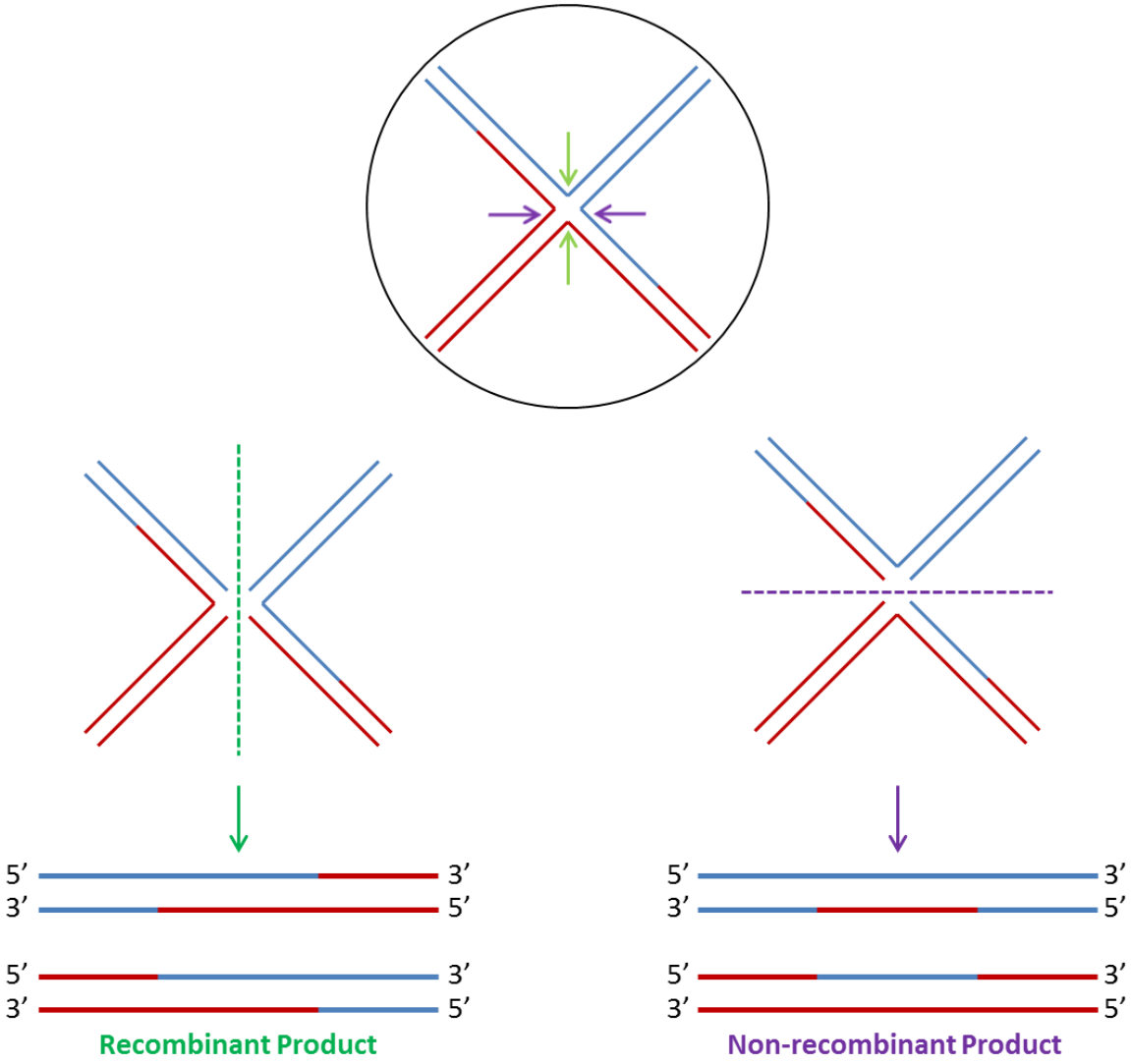


Figure 1.6. Resolution of Holliday junctions. Cartoon to show how the two types of symmetrical resolution forms the crossover recombinant product and the non-crossover non-recombinant product. A process carried out by the HJ resolvase proteins *E. coli* RuvC and *H.sapien* GEN1. Adapted from (Kowalczykowski et al., 1994).

1.5 Endonuclease XPG/RAD2 family of proteins

The XPG/RAD2 protein family is a single stranded structure specific endonuclease functional in nucleotide excision repair (NER) of damage to the DNA from UV radiation (O'Donovan et al., 1994, McCready et al., 2000). The naming of the family comes from the two member proteins Human Xeroderma pigmentosum complementation group G and the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) RAD2. This protein family shares three specific domains; an N-terminal XPG nuclease, an internal XPG nuclease and a helix-hairpin-helix domain (Lieber, 1997). **Figure 1.7** shows the structure similarity between these proteins. XPG has been identified in humans, however RAD2 has not yet been characterised in humans.

XPG acts in the NER pathway in two ways. Firstly, it acts to stabilise the transcription factor II H (TFIIH), which co-localises with XPG to the site of DNA damage. Secondly, it cuts the 3' side of DNA damage, working with the XPF-ERCC1 protein, which cuts 5' the side of DNA damage, to remove the damaged section of DNA (Evans et al., 1997). RAD2 acts in a similar manner in yeast, where it cleaves DNA on the 3' side of the damage in the NER pathway (Prakash and Prakash, 2000). **Figure 1.8** illustrates the pathway for XPG, where XPG can be seen bind to the 3' side of the DNA lesion bubble, cleaving the DNA at the 3' junction. The damaged DNA is excised when ERCC1 cleaves at the 5' side, and XPG is no longer required. The single strand of DNA is then repaired with DNA polymerase to form dsDNA with no damaged lesions (Barakat and Tuszynski, 2013).

GEN1 is a human HJ resolvase protein, which was first identified in 2008. It is a member of the XPG/RAD2 family of endonucleases. The function of Human GEN1 is characterised in **Figure 1.6,** GEN1 dimerises at the Holliday junctions where it near simultaneously carries out two symmetrical nicks to produce either a recombinant or non-recombinant product (Ip et al., 2008, Rass et al., 2010). As well as being a HJ resolvase, Human GEN1 shares a function of 5' flap endonuclease activity similar to FEN1 (Rass et al., 2010).

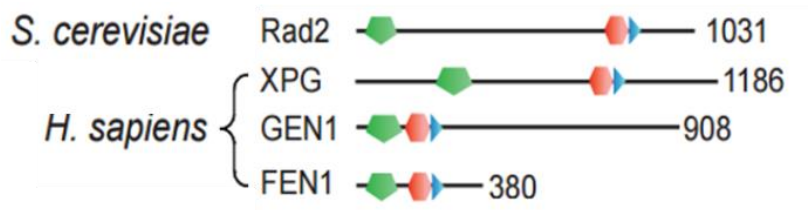


Figure 1.7. XPG/RAD2 domains. Cartoon to show the presence of the three XPG/RAD2 domains of the member proteins *S. cerevisiae* RAD2, Human XPG, Human GEN1 and Human FEN1. Domains are displayed green and red for the N-terminal and internal XPG nucleases and blue for the helix-hairpin-helix motif. Taken from (Ip et al., 2008).

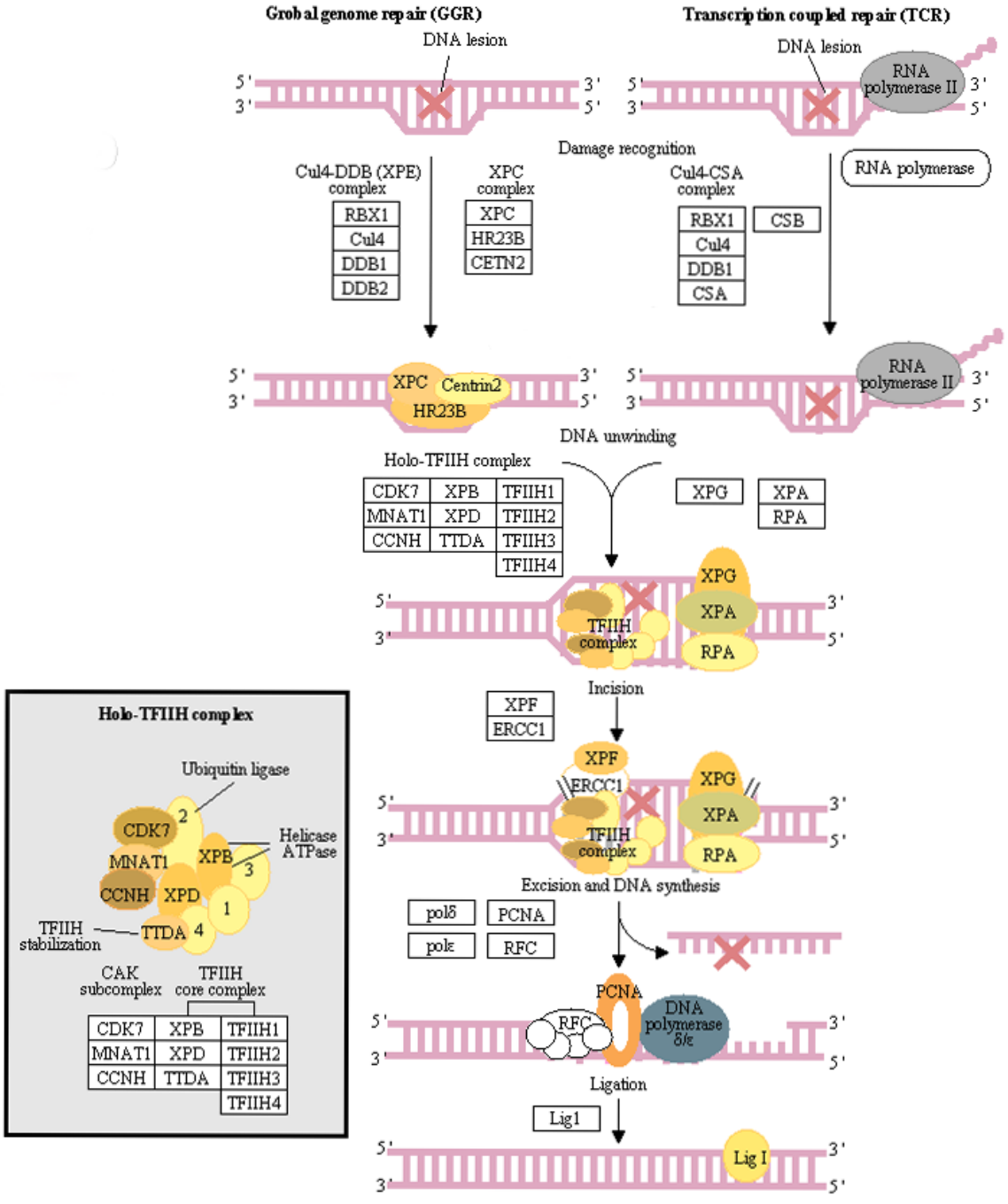


Figure 1.8. Schematic to show the process of Nucleotide Excision Repair. Figure showing the involvement of the XPG protein during nucleotide excision repair. Once the DNA has been unwound XPG and TFIIH bind to the DNA lesion bubble, the TFIIH complex recruits ERCC1. XPG can be seen to nick the 3' side of the DNA lesion bubble where ERCC1 nicks the 5' side. Figure taken from (Barakat and Tuszynski, 2013).

Flap endonuclease protein (FEN1), a member of the XPG/RAD2 family, acts as a 5' flap endonuclease. Flap endonucleases remove single stranded DNA flap structures, such as those from Okazaki fragments created during lag strand DNA replication (Liu et al., 2004). As seen in **Figure 1.9**, DNA polymerase δ or ε displaces the 5' RNA primer during DNA replication. For replication to continue the 5' RNA flap is removed by FEN1 nicking the DNA next to the single-stand overhang to remove the initiator RNA and a portion of DNA (Lieber, 1997, Kao and Bambara, 2003).

Currently there are no proteins which have been characterised to carry out HJ resolvase in the *T. brucei* parasites. If HJ resolvases were found in *T. brucei* parasites, given the functional similarity to human GEN1 we would also expect structural similarities. Through characterisation of orthologs of human GEN1, a HJ resolvase protein may be found in the *T. brucei* and therefore potentially a protein involved in the process of VSG switching.

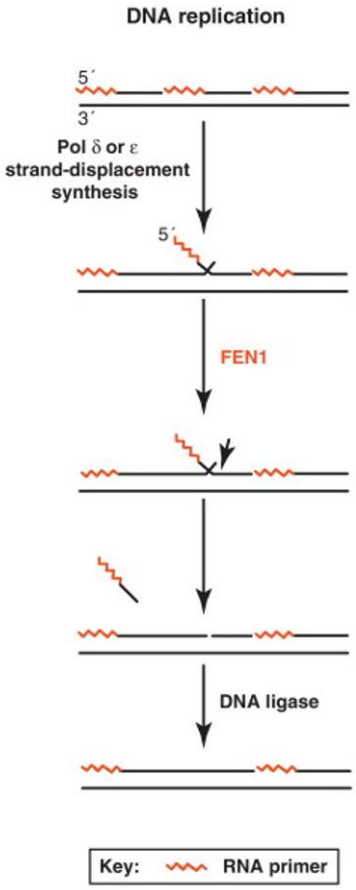


Figure 1.9. Endonuclease activity of FEN1 during DNA replication. Diagram showing the process of DNA replication, where the DNA polymerase displaces the RNA primer to form an RNA-DNA single strand flap. To remove the DNA flap FEN1 nicks the junction of the flap. Taken from (Liu and Wilson, 2012).

1.6 Project Aims

The primary aim of this project is to find proteins within the *T. brucei* parasite which show potential homology to the protein human GEN1. We will carry out a bioinformatic analysis of the proteins to analyse the sequence structures of the trypanosome proteins and compare with human GEN1, a known HJ resolvase protein in humans.

Following identification of trypanosome proteins with homology to the Holliday junction resolvase, human GEN1, sequences will be processed and expressed in bacterial cells for purification. Purified homologous proteins can then be prepared for future experimentation, such as assays to interrogate biochemical function.

Parasite cell lines with tagged homologous proteins will be created to investigate localisation of the proteins within the parasite.

Further parasite cell lines will be produced with a biotinylating marker added to the homologous proteins to investigate potential protein-protein interactions.

2 Materials and Methods

2.1 Buffers, Solutions and Ingredient Providers

2.1.1 Buffers and Solutions

Buffers and media were made using Milli-Q water.

Buffer/Solution	Ingredients
Agarose Gel	Agarose low EEO, used at 0.8% w/v with 1x TAE
Ammonium Persulphate	10% w/v
Biotin	1mM Stock made into cell media, added at 50 μM to 5 x 10^5 cells/ml
Dialysis Buffer	8 M Urea, 0.02% Triton-X, 20 mM Tris, 500 mM Sodium chloride, 10% Glycerol
Dithiothreitol (DTT)	Stock solution made at 1M
DNA Loading Buffer (6x)	0.15% Orange G, 30% Glycerol
Immunoblot BSA Blocking Buffer	1 x PBS, 0.05% Tween 20, 1% w/v BSA
Immunoblot Milk Blocking Buffer	1 x PBS, 0.05% Tween 20, 5% Skimmed Milk Powder
Immunoblot Transfer Buffer (1x)	48mM Tris(hydroxymethyl)amino methane (Tris-HCL), 39 mM Glycine, 0.037% Sodium dodecyl sulphate (SDS), 20% Methanol
Immunoblot Washing Buffer	1 x PBS, 0.025% Tween-20
Immunofluorescence Slide Blocking Buffer	1 x PBS, 0.05% Tween-20, 1% w/v BSA
Immunofluorescence Slide Washing Buffer	1 x PBS, 0.05% Tween-20
Isopropyl β-D-1 thiogalactopyranoside (IPTG)	Stock solution made at 1 M
Lysogeny Broth (LB) Low Salt Agar	15 g per 500 ml
LB Broth	1.25 g per 50 ml
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
PEME 1% NP40	1% Nonident-P40 in 0.1 M piperazine-N,N'-bis(2-

	ethanesulphonic acid) (PIPES), 2 mM ethylene glycol tetraacetic acid (EGTA), 1 mM magnesium sulphate, 0.1 M ethylenediaminetetraacetic acid (EDTA), pH6.9
Protein Purification Native - Immobilised-metal Affinity Chromatography (IMAC)	**Native Lysis Buffer** 0.001% Benzonase, 0.01% Lysozyme in Native Lysis Buffer **Native Binding Buffer 2x** 0.04% Triton-X, 40 mM Tris-HCL (pH 8), 1 M Sodium chloride, 40 mM Imidazole, 20% Glycerol **Native Binding Buffer (Wash)** 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 Denaturing - IMAC	**Denaturing Binding Buffer** 8 M Urea, 0.02% Triton-X, 20 mM Tris-HCL (pH 8), 500 mM Sodium chloride, 20 mM Imidazole, 10% Glycerol **Refolding Buffer (Wash)** 3M Urea, 0.02% Triton-X, 20mM Tris-HCL (pH 8), 500mM Sodium chloride, 20mM Imidazole, 10% Glycerol **Refolding Elution Buffer** 3 M Urea, 0.02% Triton-X, 20 mM Tris-HCL (pH 8), 500mM Sodium chloride, 500 mM Imidazole, 10% Glycerol
Protein Purification Denaturing Guanidine - IMAC	**Guanidine Denaturing Buffer** 6 M Guanidine-Hydrochloride, 0.02% Triton-X, 20 mM Tris-HCL (pH 8), 500 mM Sodium chloride, 10% Glycerol **High Urea Denaturing Wash Buffer** 8 M Urea, 0.02% Triton-X, 20 mM Tris-HCL (pH 8) 500 mM Sodium chloride, 20 mM Imidazole, 10% Glycerol **High Urea Denaturing Elution Buffer** 8 M Urea, 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), 1 mM EDTA, 10% Glycerol **High Salt Buffer** 20 mM Tris-HCL (pH 8.5), 1 mM EDTA, 10% Glycerol, 500 mM Sodium chloride
Semi-Defined Medium-79 (SDM-79 Media) (5L)	SDM-79 powder (5L), 23.8 mM Sodium bicarbonate, pH 7.3. Supplemented with 10% Heat innactivated foetal bovine serum (FBS) and 0.008 M Hemin
Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) – 8% and 10% Resolving Gel	1.5 M Tris-HCL (pH 8.8), 30% Acrylamide mix, 10% SDS, 10% Ammonium persulphate, 0.1% N,N,N', N', -tetramethylethylenediamine (TEMED)
SDS Loading Buffer	100 mM Tris-HCL (pH 6.8), 4% SDS, 0.2% Bromophenol blue, 20% Glycerol, 200 mM DTT
SDS-PAGE 5% Stacking Gel	1 M Tris-HCL (pH 6.8), 30% Acrylamide mix, 10% Ammonium persulphate, 0.1% TEMED
SDS-PAGE Running Buffer (10x)	250 mM Tris-base, 1.92 M Glycine, 35 mM SDS
Tris-Acetic acid EDTA (TAE) Buffer (50x)	2 M Tris, 1M Acetate, 50 mM EDTA, pH 8.6
ZMG Buffer	132 mM Sodium chloride, 8 mM Potassium chloride, 8 mM Disodium hydrogen phosphate, 1.5 mM Potassium dihydrogen phosphate, 0.775 mM Magnesium acetate, 0.063 mM Calcium acetate, pH 7.5

2.1.2 Ingredient Supplier

Ingredient	Provider
30% Acrylamide mix	Sigma-Aldrich
Agarose Low EEO I	Melford
Amersham Hybond-P PVDF Membrane	GE Healthcare
Ammonium Persulphate	Sigma-Aldrich
Biotin	Sigma-Aldrich, Cat – B4639
Bovine Serum Albumin (BSA)	Lyophilised powder – Sigma Aldrich – Cat – A9418 Molecular biology grade solution (conc) – ThermoScientific
Elution Buffer	ThermoScientific – R1263
Dialysis Cassette	Slide-A-Lyzer™ Dialysis Cassette, 3.5K MWCO 0.1ml-0.5ml – ThermoScientific, Cat - 88400
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich
Dithiothreitol (DTT)	Melford
DNA Loading Buffer (6x)	Acros Organics
Doxycycline	Melford
Ethanol	Absolute – FisherScientific
10% Foetal Bovine Serum (FBS)	BioSera
15% FBS	Sigma-Aldrich
GelRed™	10,000x – ThermoScientific
Gene Pulsar® Electroporation Cuvette	BioRad
GeneJET PCR Purification Kit	ThermoScientific – K0701
GeneJET Plasmid Miniprep Kit	ThermoScientific – K0502
GeneRuler™ 1kb DNA Ladder	ThermoScientific
1x Glutamax	Gibco Life Sciences
Glycerol	Melford
Hemin	Sigma-Aldrich
HiTrap™ 1 ml Chelating Column	GE Healthcare – Cat – 17-09203-03
Immobilon™ Western Chemiluminescent HRP Substrate	Millipore
InstantBlue™	Expedeon
Isopropyl β-D-1	Melford

thiogalactopyranoside (IPTG)	
Ligation Buffer	2x Rapid Ligation Buffer (Promega) 10x Ligation Buffer (Promega)
Lysogeny Broth (LB) Low Salt Agar	Melford
LB Broth	Melford
Methanol	FisherScientific
Mono Q™ 5/50 GL	GE Healthcare – Cat – 17-5166-01
0.1% N,N,N', N', -tetramethylethylenediamine (TEMED)	Sigma-Aldrich
PageRuler™ Plus Prestained Protein Ladder (10 – 250 kDa)	ThermoScientific
Paraformaldehyde (PFA) 1x PBS	Sigma
Precision Plus™ Protein All Blue Prestained Protein Standards (10-250 kDa	BioRad
1x Penicillin-Streptomycin	Gibco Life Sciences
Phosphate Buffered Saline (PBS)	(10xPBS) – Melford (20xPBS) – Sigma-Aldrich
Native Lysis Buffer	Qiagen
Ni-NTA Spin Column	Qiagen – Cat 1011712
Semi-Defined Medium (SDM)-79 Powder	Gibco Life Sciences
Super Optimal Broth (SOC)	ThermoScientific
T4 Ligase	Promega
Tuner™ (DE3) pLacI Competent	Novagen
Vectashield® Mounting Medium with DAPI	VectorLabs
XL1-Supercompetent Cells	Agilent Technologies

2.2 Antibiotic, Drug Stocks and *E. coli* strains

2.2.1 Antibiotic and Drug Stocks

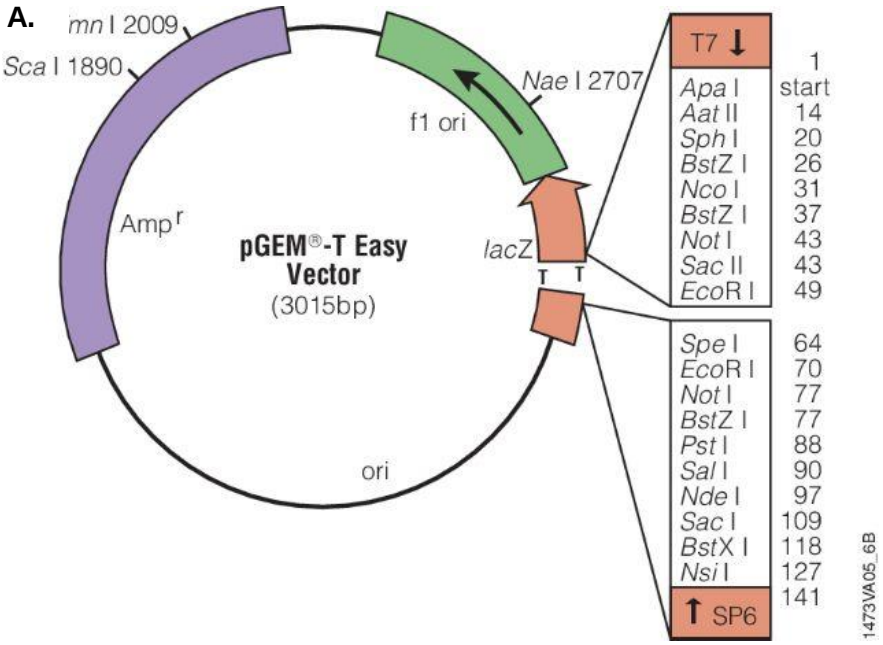
Antibiotic/Drug	**Concentration/*Provider***
Ampicillin	Stock solution made at 100 mg/ml diluted in Milli-Q, *Sigma-Aldrich*, Cat – *A0166*. Used at 100 μg/ml in LB agar and broth
Blasticidin	Stock solution made at 10 mg/ml diluted in Milli-Q, *Millipore*, Cat – *203350*. Used at in 10 μg/ml in trypanosome media.
Chloramphenicol	Stock solution made at 25 mg/ml diluted in absolute Ethanol, *Sigma-Aldrich*, Cat – C0378. Used at 25 μg/ml in LB Agar and 12.5 μg/ml in LB Broth.
Hygromycin	Stock solution made at 25 mg/ml diluted in Milli-Q, *Sigma-Aldrich*, Cat – *H3274*. Used at 25 μg/ml in trypanosome media.
Kanamycin	Stock solution made at 30 mg/ml diluted in Milli-Q, *Sigma-Aldrich*, Cat – *K1377*. Used at 30 μg/ml in LB agar and broth
Phleomycin	Stock solution made at 25 mg/ml diluted in Milli-Q, *Sigma-Aldrich*, Cat – *P9654*. Used at 7.5 μg/ml in trypanosome media.
Puromycin	Stock solution made at 10 mg/ml diluted in Milli-Q, *Sigma-Aldrich*, Cat – *P8833*. Used at 3 μg/ml in trypanosome media.

2.2.2 *E. coli* stains

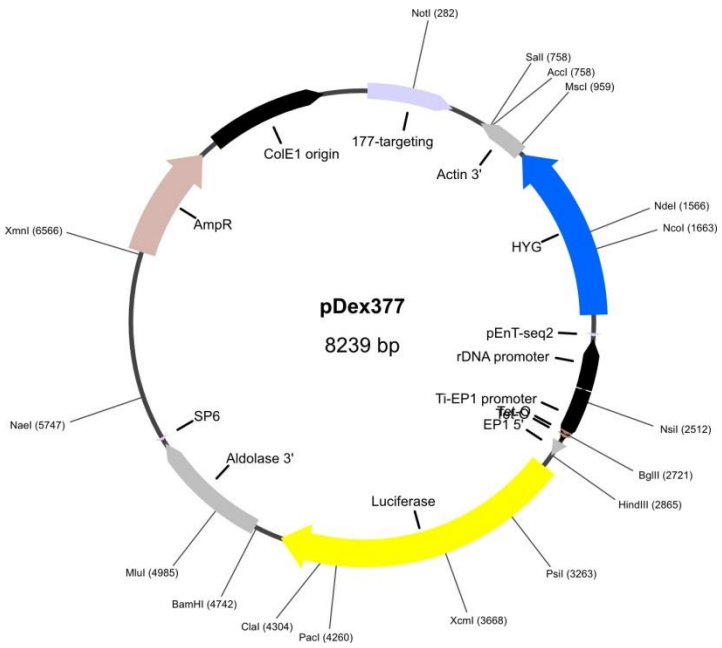
Strain Name	**Full Genotype**
Tuner™ (DE3) pLacI	F– *ompT hsdS*$_B$ (r_{B-} m_{B-}) *gal dcm lacY1*(DE3) pLysS (Cam_R)
XL-1 Blue Competent	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB laclq ZΔM15 Tn10 (Tetr)]

2.3 Vector Maps

pGEM-T Easy

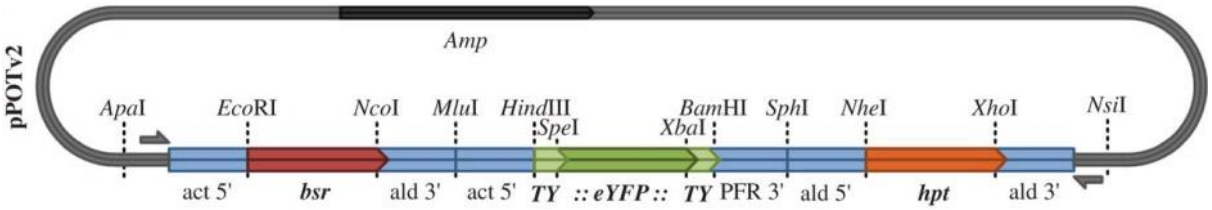


pDEX-377 — Wickstead Lab



pPOT Vector

C.



pET24a - Novagen

D.

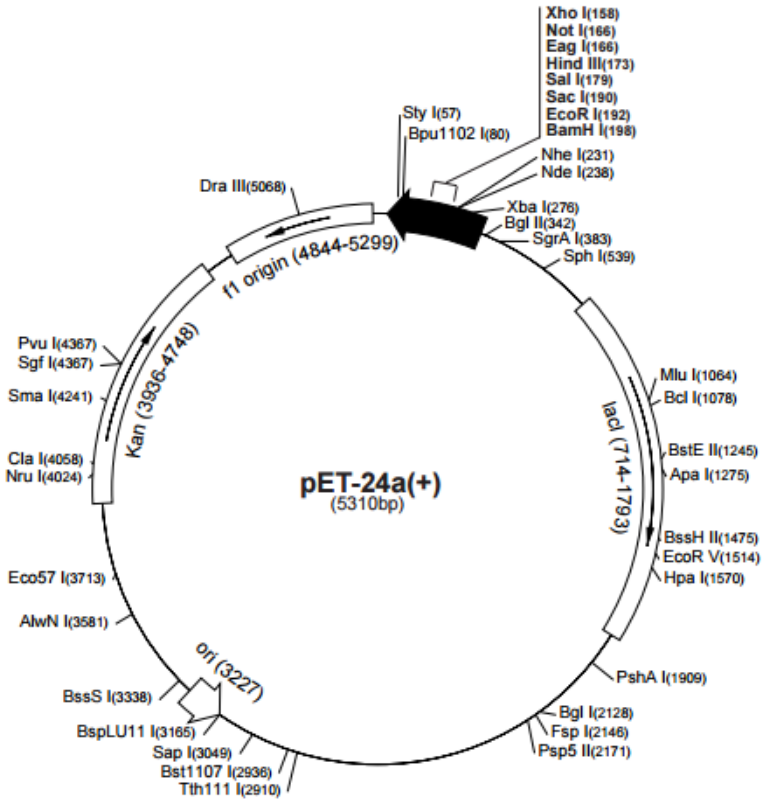


Figure 2.1. Vector Maps. (A) pGEM®-T Easy vector map. This vector was used for the cloning of amplified PCR products, Promega. (B) pDEX377 vector map. Derivatives of this vector were used as a biotinylating and MYC vector, and a GAFSINPAM (GSP) epitope tag vector, (Kelly et al., 2007). (C) PCR only tagging (pPOT) vector v2 vector map. The pPOTv2 vector was used to tag target proteins with a N-terminus YFP tag, (Dean et al., 2015). (D) pET24a(+) vector map. This vector was used for the expression of hexa-histidine tagged proteins, Novagen.

2.4 Antibodies and Affinity Proteins Used

Antibody	Epitope	Published/*Company*	Dilution	Blocking Buffer
Anti c-myc mouse monoclonal	Myc	Ab 18185, *Abcam*	**WB** – 1:1,000 **IF** – 1:200	**WB** – Milk **IF** – BSA
Anti 6x-His tag mouse monoclonal HRP conjugate	6xHis	*Thermo Fisher,* 4E3D10H/E3	**WB** – 1:10,000	**WB** – Milk
Anti KMX-1 mouse monoclonal	β-Tubulin	Courtesy of Dr K. Gull	**WB** – 1:1,000	**WB** – Milk
Anti-TDP-43, clone DB9, mouse monoclonal	GAFSINPAM	*Millipore*	**WB** – 1:10,000 **IF** – 1:200	**WB** – Milk **IF** – BSA
Rabbit 308 pre-immune bleed, TbFEN1 rabbit polyclonal	TbFEN1	*Eurogentec*	**WB** – 1:1,000 **IF** – 1:1,000	**WB** – Milk **IF** – BSA
Rabbit 308 final bleed, TbFEN1 rabbit polyclonal	TbFEN1	*Eurogentec*	**WB** – Range **IF** – Range	**WB** – Milk **IF** – BSA
Rabbit 309 pre-immune bleed, TbFEN1 rabbit polyclonal	TbFEN1	*Eurogentec*	**WB** – 1:1,000 **IF** – 1:1,000	**WB** – Milk **IF** – BSA
Rabbit 309 final bleed, TbFEN1 rabbit polyclonal	TbFEN1	*Eurogentec*	**WB** – Range **IF** – Range	**WB** – Milk **IF** – BSA

Table 2.1. Primary Antibodies. WB - western blotting (immunoblot) and **IF** immunofluorescence.

Antibody	Epitope	Published/*Company*	Dilution	Blocking Buffer
Goat Anti-mouse IgG HRP conjugate		12-349, *Millipore*	**WB** – 1:80,000	**WB** – Milk
Anti-mouse IgG, Rhodamine (TRITC) conjugated		AP124R, *Millipore*	**IF** – 1:200	**IF** – BSA
Anti-rabbit HRP conjugate		P 0217, *DakoCytomation*	**WB** – 1:1,700	**WB** – Milk
Goat anti-rabbit IgG, Alexa Fluor® 568 conjugate		A11036, *Invitrogen*	**IF** – 1:1,000	**IF** – BSA

Table 2.2. Secondary Antibodies. **WB** western blot (immunoblot) and **IF** immunofluorescence.

Protein	Epitope	*Company*	Dilution	Blocking Buffer
Streptavidin HRP conjugate	Biotin	Ab-7403, *Abcam*	**WB** – 1:10,000	**WB** – BSA
Streptavidin TRTIC conjugate	Biotin	Ab 136223, *Abcam*	**IF** – 1:200	**IF** – BSA

Table 2.3. Affinity Protein. **WB** western blot (immunoblot) and **IF** immunofluorescence.

2.5 Polymerase Chain Reaction Primers

Primer ID	Use	Sequence
*Tb*FEN1 *Hind*III F	PCR	5’ AGA AGC TTA TGG GTG TTC TTG GCC TTT CGA AAC TTC 3’
*Tb*FEN1 *Xho*I R	PCR	5’ AGC TCG AGC TTT TTG ACA GCC TTT TTG TGT 3’
*Tb*RAD2 *Hind*III F	PCR	5’ AGA AGC TTA TGG GTG TCC ATG GAC TGT G 3’
*Tb*RAD2 *Nde*I F	PCR	5’ AGC ATA TGG GTG TCC ATG GAC TGT G 3’
*Tb*RAD2 *Xho*I R	PCR	5’ AGC TCG AGC GGA TCA CCC CTG GCT GC 3’
*Tb*RAD2 Muta F	Mutagenesis	5’ GCC ATC GAC GCA AGC ATC TGG ATA GCG CAG 3’
*Tb*RAD2 Muta R	Mutagenesis	5’ CTG CGC TAT CCA GAT GCT TGC GTC GAT GGC 3’
pPOT Tb830 F	pPOT Tag	5’ ATG TCA CAG TTT TTG CTA CTG CCT TTC CTC ACA ATT CTT TCC GGA AAG TCT TTC GTT GAT 3’
pPOT Tb830 R	pPOT Tag	5’ AAA TAA ACT TTC AGC TCT TGT TCC TTA ATG GCG CCA GGA GTT CTG TCG TAC AGA AGC TTC 3’
pPOT Tb11760 F	pPOT Tag	5’ ATA TAC ACA AAT GCA TAT GTG CCG AAG GAG TCA CTT TGA ATG ACT TCT TAT TCG CTT TTT 3’
pPOT Tb11760 R	pPOT Tag	5’ ACG CGT TTG CCT TTC CAA TCG GCG GGT TGC GTA ACC TCA CCG AAG GTG TCG AGG AGG CGC 3’

Table 2.4. Primers. Underlined sequences represent restriction sites, red bases represent bases which will be altered via mutagenesis in the target gene. F and R in primer ID, represent Forward and Reverse primers respectively.

2.6 Equipment Used

Equipment/*Company*	**Uses**
AKTA™ Prime Plus, *GE Healthcare*	Large scale protein purification by IMAC
AKTA™ Purifier HPLC	Large scale protein purification by IEX
ChemiDoc™ MP, *Bio-Rad*	Imaging of coomassie stained SDS-PAGE and chemifluorescence of immunoblots
Gel Doc™ EZ System, *Bio-Rad*	Imaging of Gel Red agarose gels and coomassie stained SDS-PAGE
Deltavision™ Elite, *GE Healthcare*	Fluorescent image capture and deconvolution of immunofluorescence slides
NanoDrop 2000 Spectrophotometer, *ThermoScientific*	Measuring the concentration of dsDNA (260 nm) and purified protein (280 nm)

2.7 Parasite Cell Biology

2.7.1 Parasite Cell Counts

All cell densities were calculated using a haemocytometer and recorded at regular intervals based on the length of the experiment.

2.7.2 Procyclic *T. brucei* Cell Culture

Trypanosome cell cultures underwent passage regularly to maintain the cells in the log phase of growth. Cells were grown to a density of ~1 x 10^7 cells/ml and diluted to 2 x 10^5 – 5 x 10^5 cells/ml. Procyclic cell lines were cultured using SDM-79 composition media in cell culture flasks with standard non-filter screw tops, at 28°C.

2.7.3 Ethanol Precipitation of DNA for Transfection

Plasmids were ethanol precipitated prior to transfection into *T. brucei*. Plasmid vector linearised by digestion with *NotI* (**Methods 2.9.11**) and 3 M sodium acetate (pH 5.2) was added at 1 in 10 of the original digest volume, absolute ethanol was then added at 2.5x of the total volume of digest and sodium acetate. The solution was mixed and DNA left to precipitate at -80°C for 1 - 16 hours.

Precipitated DNA was pelleted at 22,000 x *g* for 50 minutes at 4°C. Supernatant was removed and remaining ethanol allowed to evaporate off inside Class II Microbiology safety cabinet. Once confident all ethanol had evaporated the DNA pellet was washed with 70% ethanol centrifuged to a pellet, dried and then re-suspended in 50 μl sterile elution buffer.

2.7.4 Transfection of procyclic *T. brucei*

For transfection, 3 x 10^7 cells of the cell line to be transfected were pelleted at 400 x *g* for 10 minutes; the supernatant was removed, cells re-suspended in 0.5 ml sterile ZMG buffer, and transferred to a Gene Pulsar® Electroporation Cuvette. Plasmid DNA (3 – 10 μg) was added to the cell suspension and mixed. The cell-DNA suspension was then electroporated using a BTX electroporator (3 pulses 1,700 mV for 100 μs with pulse intervals of 200 μs). Cells were allowed to recover overnight in media without drugs, before being diluted to a cell density of 5 x 10^5 - 1 x 10^6 cells/ml with appropriate drug selection.

2.7.5 Parasite Population Induction

Before beginning an experiment, trypanosomes were cultured in drug free SDM-79 composition media for 48 hours and passed back to a concentration of $5x10^5$ cells/ml. Transgenic cells were then induced with the addition of 1μg/ml doxycycline.

2.7.6 Fluorescence Microscopy

To prepare slides for fluorescence microscopy trypanosome cells were grown in SDM-79 composition media to a density of 1 x 10^7 cells/ml. From the culture, 1 x 10^7 cells were collected and centrifuged at 2,400 x *g* for 3 minutes and the media supernatant removed. The cells were then washed twice by re-suspending the pellet in 1 ml of 1 x PBS and centrifuged at 2,400 x *g* for 3 minutes to pellet the cells. After the final wash, the cells were re-suspended in 1 x PBS at a final concentration of 1 x 10^7 cells/ml and 50 - 100 μl added to wells created on slides with hydrophobic pen. Procyclic cells were left for 15 minutes to allow cells to settle on to the slide before being fixed in ice-cold absolute methanol.

Cytoskeletons were prepared by detergent-extraction by the addition of 1% PEME NP-40 for 25 seconds before fixing in methanol.

For bloodstream cells, a minimum of $2x10^6$ cells were added per slide, with an additional fixing step of 2% PFA for 15 minutes before being added to ice-cold methanol.

Prepared slides were stored in ice-cold methanol at -20°C for a minimum of 10 minutes before probing.

2.7.7 Probing Slides for Fluorescence Microscopy

Cells were re-hydrated in 1 x PBS for 10 minutes, and then incubated for 1 hour in immunofluorescence blocking buffer within a humidified chamber. Once blocking buffer was removed from the slide, primary antibody diluted in blocking buffer (**Table 2.1**) was added at 50 μl per slide, and incubated for a further hour in the humidified chamber. The primary antibody was removed by washing slides three times in immunofluorescence wash buffer, for 5 minutes on each occasion, before addition of secondary antibody diluted in blocking buffer (**Table 2.2**), and incubated for a further hour in the humidified chamber. The secondary antibody was removed by following the same washing procedure as described previously.

A single drop of Vectashield® Mounting Medium with DAPI, was added before a cover slip was placed onto the slide and sealed with nail varnish. Slides were enclosed in foil and stored at 4°C.

2.7.8 Preparation of *T. brucei* Proteins for Immunoblot Analysis

For immunoblot assays, 5 x 10^6 cells were required per lane. The required quantity of cells were pelleted at 400 x *g* for 10 minutes to remove the media supernatant. Cells were then washed twice in 1 x PBS with centrifugation at 2,400 x *g* for 3 minutes. The washed pellet was re-suspended in SDS-loading buffer at a concentration of 5 x 10^5 cells/µl boiled at 100°C for 10 minutes and stored at -80°C.

2.7.9 Preparation of Cell Stabilates

Cell lines were grown to a density of 5 x 10^6 - 1 x 10^7 cells/ml and glycerol added to a final concentration of 10% v/v. Stabilates were collected as 1ml aliquots in twist cap cryogenic tubes and initially stored at -80°C overnight before being moved for long term storage under liquid nitrogen.

Stabilate cells were revived 1:10 in SDM-79 media and cultured under normal conditions. Cells were allowed to recover for at least one subculture, before antibiotics were added for selection.

2.8 Molecular Biology Techniques

2.8.1 Oligonucleotide Primers

Oligonucleotide primers (**Table 2.4**) were designed by Dr Benson and ordered from Invitrogen Life Technologies. Primers were reconstituted to form a stock concentration of 100 μM and stored at -20°C. For use in PCR reactions the stock solutions were diluted in sterile autoclaved Milli-Q water to form a working concentration of 10 μM.

2.8.2 PCR Amplification of Target Genes

High Fidelity PCR was used for the amplification of target genes. Reactions contained

Primers	125 ng
DMSO	256 mM
T. brucei 927 gDNA	50 – 100 ng
PCR nucleotide mix	200 μM (of each dNTP)
Expand™ High Fidelity Polymerase 3.5 U/μl	3.5 U

10X Expand™ High Fidelity Buffer with Magnesium Chloride

The PCR reaction conditions:

Cycle Number	**Temperature (°C)**	**Length (Min:Secs)**	**Cycle Process**	
1	94	05:00	Initial Denaturing	Repeat
2	94	00:30	Denature	X30
3	55 - 62	00:30	Annealing	
4	72	01:30	Elongation	
5	72	05:00	Final Elongation	

Table 2.5. Table showing conditions used for standard PCR reactions. Annealing temperature variance is dependent on the melting point of the primers used.

2.8.3 PCR only Tagging (pPOT)

The process of pPOT tagging by PCR follows the same protocol as **methods 2.9.2**, instead the pPOT vector replaces the gDNA as the template used with a shorter final elongation of 2 minutes.

2.8.4 Purification of PCR products

Purification of the PCR products was carried out using GeneJET™ Purification Kit. Purified products were eluted in 30 - 50 µl of elution buffer and stored at -20°C.

2.8.5 Ligation of PCR products into cloning/expression vectors

PCR products were ligated into linearised cloning/expression vectors **Figure 2.1 A** and **D**, using rapid ligation buffer, 1 - 3 U T4 Ligase, 50 ng vector and PCR product. The amount of DNA insert was calculated to give a 1:3 vector to insert ratio. Ligations were incubated at room temperature for an hour, before being incubated overnight at 16°C.

2.8.6 Transformation into *E. coli*

All transformations used 50 µl of competent *E. coli* strains as received from provider, the *E. coli* cells were slowly thawed on ice. Once thawed 1 µl of the ligation product was added and the mixture left to incubate for 20 minutes on ice. Cells were then heat-shocked at 42°C for 45 seconds and returned to ice for 2 minutes. After 2 minutes on ice, 200 µl of prewarmed LB broth was added. The cells were then incubated at 37°C for 20 minutes and agitated at 200 – 250 rpm, this allowed for recovery prior to plating on antibiotic supplemented LB agar for selection.

After the 2 minutes on ice for *E. coli* strains XL1-Blue supercompetent cells and Tuner™ (DE3) pLacI competent cells, they were instead added to pre-warmed SOC medium and allowed to recover for 90 minutes.

Plates were incubated overnight at 37°C and resistant colonies streaked on new supplemented LB agar to isolate single colonies.

2.8.7 Colony PCR

Transformants were picked using sterile tips, spotted onto labelled LB agar and then used to inoculate a 100 µl aliquot of Milli-Q water. To lyse the *E. coli*, the Milli-Q mix was boiled for 10 minutes and then centrifuged at 11,000 x *g* for 10 minutes. The supernatant, which contains the plasmid of interest, was added to a standard PCR reaction **Table 2.5**, in place of the gDNA.

2.8.8 Purification of plasmid DNA

E. coli colonies were individually picked and inoculated in 5 ml LB broth supplemented with appropriate antibiotics. The bacterial suspension was incubated overnight at 37°C at 200-250 rpm, and cells then pelleted at 2,000 x *g* for 10 minutes. Plasmid DNA was purified using the GeneJet™ Plasmid Miniprep Kit and purified DNA stored at -20°C.

2.8.9 Oligonucleotide Sequencing

Purified recombinant pGEM-T Easy plasmids were sent to Source BioScience at 100 ng/μl for sequencing using standard M13F and M13R primers. For gene Tb927.9.11760 internal primers were also required.

2.8.10 Site-directed Mutagenesis

Site directed mutagenesis was used to change one of the nucleotides within the sequence, whilst ensuring that the amino acid sequence was un-changed. For mutagenesis the QuickChange II Site-Directed Mutagenesis kit was used with 20 ng dsDNA plasmid and the *E. coli* XL1-Blue Supercompetent cells.

2.8.11 Restriction Enzyme Digests

Digests were carried out according to suppliers (Roche) instructions. For double digests, the buffer that was most efficient for both enzymes was used.

2.8.12 Analytical Agarose Gel Electrophoresis

Agarose gels (0.8% w/v in 1 x TAE) containing 1x GelRed™, were cast and placed into gel tanks containing 1 x TAE buffer. DNA samples were mixed with 1x DNA loading buffer and Milli-Q water, and resolved. GeneRuler™ 1kb DNA ladder was used at 0.5 μg as recommended by supplier (ThermoScientific)

2.8.13 Preparative Agarose Gels and Gel Extraction

DNA was loaded onto 0.8% w/v agarose gels which did not include GelRed™. The DNA was resolved by electrophoresis. GeneRuler™ 1kb DNA ladder was used at 0.5 μg as

recommended by supplier (ThermoScientific).Once DNA had been resolved, samples were post stained with GelRed™ for 20-30 minutes under agitation and viewed with UV light. DNA bands were excised from the gel and purified using the GeneJet™ Gel Extraction Kit. Purified DNA was stored at -20°C.

2.8.14 SDS-PAGE

Acrylamide resolving gels (8% or 10%) were cast with a 5% acrylamide stacking gel (Sambrook et al., 1989). Protein samples were mixed with SDS loading buffer and heated to 100°C for 10 minutes. Samples were loaded into wells and resolved at 140 V until the lower MW bands of the PageRuler™ Plus Prestained Protein Ladder (10 – 250 kDa) or Precision Plus™ All Blue Prestained Protein Ladder (10 – 250 kDa), began to run off the gel. Resolved gels, unless being used for immunoblotting, were stained using InstantBlue™.

2.8.15 Transfer of Proteins to Blotting Membrane

Proteins resolved by SDS-PAGE were transferred onto Amersham Hybond-P PVDF Membrane. Firstly, membranes were permeabilised in absolute methanol for 10 seconds, rinsed and soaked with Milli-Q water for 5 minutes, then soaked for a further 5 minutes in western transfer buffer. The membrane, SDS-PAGE, blotting papers and sponges were then loaded into BioRad™ cassettes, and completely submerged in transfer buffer and proteins transferred at 100 V for 60 minutes. If membranes were not to be used straight away they were air dried on blotting paper and stored at 4°C.

2.8.16 Immunoblotting

Membranes that had been stored at 4°C were re-permeabilised in absolute methanol for 10 seconds, and then rinsed and washed in Milli-Q water for 5 minutes, washed twice for 5 minutes in 1 x PBS and then incubated in immunoblot blocking buffer for 1 hour at room temperature or overnight at 4°C under gentle agitation. Primary antibodies were diluted in blocking buffer (see **Table 2.1),** added to the membranes and incubated for an hour at room temperature, or overnight at 4°C, under gentle agitation. After incubation with primary antibody the membrane was washed 3 times for 10 minutes with immunoblot washing buffer. Appropriate secondary antibody was diluted in its appropriate blocking buffer (**Table 2.2)**, and the membrane incubated for an hour at room temperature under gentle agitation before being washed 3 times for 10 minutes with immunoblot washing buffer. Blots were developed

using Immobilon™ western chemiluminescent HRP substrate according to manufacturer's instructions (Millipore).

2.8.17 Exposing of Immunoblots

Signals from the chemiluminescent HRP substrate were exposed using the ChemiDoc™ system. The ChemiDoc™ allowed for prolonged time-period of exposure, along with 'live' visualisation of the exposure.

2.9 Protein Expression and Purification

2.9.1 Small Scale Inductions, Temperature and IPTG Concentration Tests

Small scale expression cultures of 20 – 25 ml LB broth, containing chloramphenicol and kanamycin, were inoculated with 0.2 ml of a fresh overnight culture of the relevant bacterial strain. Cultures were incubated at 37°C with agitation of 150 – 200 rpm, and growth followed by monitoring OD_{600} until the culture had an absorbance of 0.4 - 0.6, indicative of mid-log growth phase.

Protein expression was induced with 1 mM IPTG, incubation was continued at 37°C with agitation for a further 3 hours. Samples of 1 ml, were taken prior to addition of IPTG and at intervals of one hour. These samples were centrifuged at 1,000 x *g* for 4 minutes, supernatants removed, and the pellets re-suspended in 100 µl SDS-loading buffer ready for SDS-PAGE, **Methods 2.8.14**. Following the 3 hour incubation period, the remaining bacterial culture was centrifuged at 1,000 x *g* for 10 minutes at 4°C, the supernatant removed and the pellet stored at -80°C.

Different experimental variables were used for the temperature and IPTG concentration. Temperatures 30°C and 37°C were used against 1 mM and 0.1 mM IPTG.

2.9.2 Large scale inductions

As in **Methods 2.9.1**, overnight cultures were prepared and the suspension inoculated in 400 ml fresh LB broth with drugs. Suspensions were incubated at 37°C under agitation and cultured until the bacteria had reached mid-log growth, expression was then induced with 0.1 mM IPTG and cultured for a further 3 hours at 30°C. The resulting bacterial culture was then centrifuged at 1,000 x *g* for 20 minutes at 4°C, supernatant removed and the pellet stored at -80°C.

2.9.3 Small scale lysis under native conditions

Pellets from small scale inductions (**Methods 2.9.1)** were defrosted on ice and 1ml cold native lysis buffer was added. The solution was mixed by pipetting and incubated on ice for 20 minutes. Insoluble material was pelleted by centrifugation at >11,000 x *g* for 15 minutes, the supernatant removed and diluted 1:1 with 2x native binding buffer ready for purification by IMAC

2.9.4 Small scale lysis under denaturing conditions

Pellets from small scale inductions (**Methods 2.9.1)** were defrosted on ice and lysed by addition of 1ml of denaturing buffer (containing either 8 M urea or 6 M guanidine hydrochloride). The solution was mixed, incubated on ice for 20 minutes and then sonicated for 3 x 15 seconds. The lysate was then centrifuged at >11,000 x *g* for 15 minutes and the supernatant diluted 1:1 in the respective denaturing buffer, ready for purification of induced protein.

2.9.5 Large scale lysis under native conditions

Pellets from large scale inductions (**Methods 2.9.2)** were lysed in 16ml native lysis buffer containing 1% protease inhibitor cocktail (EDTA free. The solution was mixed and incubated on ice for 30 minutes, followed by sonication for 3 x 20 seconds. Insoluble material was then pelleted by centrifugation at 66,200 x *g* for an hour at 4°C, the supernatant passed through a 0.45 μm syringe filter and the soluble supernatant diluted 1:1 in 2x native binding buffer.

2.9.6 Spin Column IMAC

Recombinant proteins expressed in small scale inductions were purified using Ni-NTA spin column. These spin columns were used for purification under both native and denaturing lysis methods, with different solutions used for the washing and elution.

Spin columns were equilibrated with binding/denaturing buffer prior to addition of 600 μl of the diluted supernatant obtained from lysis and centrifuged at 260 x *g* for 2 minutes. Material flowing through column was collected and retained for analysis. Spin columns were washed with three 600 μl aliquots of wash buffer, in each case centrifuging at 260 x *g* for 2 minutes, and washes collected and retained for analysis. Bound proteins were then eluted from the columns with three 200 μl aliquots of elution buffer, and eluates collected.

2.9.7 HiTrap™ IMAC

Recombinant proteins expressed in large scale inductions were purified using 1 ml HiTrap™ chelating column, using native buffer conditions at 4°C. Columns were primed with 0.1 M nickel sulphate and excess nickel sulphate remove by washing with Milli-Q water. Columns were then equilibrated with binding buffer and soluble protein injected on to the column. Unbound proteins were removed by washing with an excess of wash buffer. Protein was

then eluted with a gradient of 0 - 100% elution buffer (30 – 500 mM imidazole), collecting 1 ml fractions throughout.

2.9.8 Mono Q™ GL IEX

Recombinant protein were further purified from contaminants by using a Mono Q™ 5/50 GL, anion exchange column. Fractions from IMAC **Methods 2.9.7** which contained the highest concentrations of target protein, were collected and the salt concentration was decreased by diluting 1:10 with low salt IEX buffer. Once the column was equilibrated with binding buffer the diluted protein solution was injected onto the column. Bound proteins were eluted, using a gradient of 0 - 100% high salt IEX buffer (0 – 500 mM NaCl). The flow rate was 1 ml/min, where the flow through was carried out over 38 minutes and the gradient carried out over 25 minutes. The eluate was collected into 1 ml fractions.

3 Identification of *Trypanosoma brucei* proteins with homology to human Holliday junction resolvase protein GEN1

3.1 Chapter Synopsis

To identify putative *T. brucei* orthologues of the human Holliday junction resolvase GEN1bioinformatics analysis was undertaken. This involved comparing the amino acid sequences, secondary and tertiary structures of human GEN1 protein with putative *T. brucei* orthologues.

3.2 Bioinformatic Interrogation of *Tb*FEN1, *Tb*RAD2 and Human GEN1 Proteins

Using the 906 amino-acid sequence of the Human GEN1 protein, a BLASTP search was carried out. The search was performed against the non-redundant protein sequence database for *Trypanosoma brucei brucei* TREU927 (taxid 185431). This search identified four *T. brucei* proteins with various degrees and sequence similarity to human GEN1 (**Table 3.1**).

Protein description	Query Cover (%)	E Value	Identification (%)
Putative Flap Endonuclease FEN1, Tb927.3.830 – *Tb*FEN1	22	$5x10^{-17}$	28
Putative DNA Repair Protein RAD2, Tb927.9.11760 – *Tb*RAD2	22	$2x10^{-11}$	41
Exonuclease	24	0.002	20
Hypothetical Protein (accession n^{o} XM_818129.1)	9	3.2	31

Table 3.1. BLASTP results. Table showing top hit *T. brucei* proteins identified by the Human GEN1 sequence. E value depicts the random background hits to be expected.

Of the four proteins identified, the proteins annotated as *Tb*FEN1 and *Tb*RAD2 were selected for further analysis. The literature review identified that both of these proteins are likely members of the XPG/RAD2 family and may potentially be involved in homologous recombination, hence this decision was deemed appropriate. The exonuclease and hypothetic protein were disregarded because the E values were of low significance indicating their identification may be due to chance. The amino acid sequences of Human GEN1, *Tb*FEN1 and *Tb*RAD2 protein were aligned using the Clustal Omega programme (**Figure 3.1**)

```
GEN1        ------------------------------------------------------------   0
TbFEN1      MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE   60
TbRAD2      MGVHGLWRLLDTFG--E--VTQPADWKGKRVAIDASIWIAQFRSSC---EPGESV-----   48

GEN1        ------------------------------------------------------------   0
TbFEN1      AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHE------   114
TbRAD2      ---EERILEGFFMRILKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHA   105

GEN1        ------------------------------------------------------------   0
TbFEN1      ------------------------------------------------------------   114
TbRAD2      RRLIAAQMSAGLLDVHSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETG   165

GEN1        ------------------------------------------------------------   0
TbFEN1      ------------------------------------------------------------   114
TbRAD2      TILLQPKGRKKRTREVCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTS   225

GEN1        ------------------------------------------------------------   0
TbFEN1      ------------------------------------------------------------   114
TbRAD2      IFMGPRRVAEEVSRALGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSV   285

GEN1        -------------------------------------------------MGVNDLWQILEPV   13
TbFEN1      ------------------------------------------------------------   114
TbRAD2      CEILSSSSCSVIVVDNAIKTDPHAVDAFHHNVSFGKEEESTSDEVEVLSSGDYWSCADND   345

GEN1        KQHIPLRNL----GGKTIAVDLSLWVCEAQTVKK--------------------------   43
TbFEN1      ------------------------------------------------------------   114
TbRAD2      CDDLLSLAASDRTPDTQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSE   405

GEN1        --MMGSVMKP------HLRNLFFRISYLTQMDVKLVFVMEGEPPKLKADVISKRNQSRYG   95
TbFEN1      ----------------------F---------E---KAKEEGDDEAM--EKMSK-------   132
TbRAD2      FNPFGGVVVPSGNLRKDEKEVLLNTSVITSSE---TLETTGIPLKV--PSVSR-------   453
                                  :         :        *    :    :*:

GEN1        SSGKSWSQKTGRSHFKSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGGHVDGCLTND   155
TbFEN1      ------RMV---RVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKAWAVGTED   183
TbRAD2      ------EHVREKQVVPFELLGIVELLDCCGIPYVLSPNEADAQCAFLNEQRVVDAVFTED   507
                             :     :*   *** * :  **:* ** *     . .  *:*

GEN1        GDTFLYGAQTVYRNFTMN-TKDPHVDCYTMSSIKSKLGLDRDALVGLAILLGCDYLPKGV   214
TbFEN1      MDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP-RI   242
TbRAD2      SDVIVHGAPVVLRGFFS---KGRHVVAYRQSDL-LACGVDKVVLVALALLLGCDYAE-GV   562
             *.: .*: .: * :     *   :  *  ..:    *..   :: *.:******    :

GEN1        PGVGKEQALKLIQILKGQSLLQRFNRWNETSCNS--SPQLLVTKKLAHCSVCSHPGSPKD   272
TbFEN1      SGIGPHKAWEGIKK-----------Y---------GS-------LEAFIESLDGTRYVVP   275
TbRAD2      NGLSLLESLHVIAA-----------TWRQTTNSVEGGAEQVRDMLSSWCSAVRRRRIPWG   611
             *:.  :: . *                          .        :  .

GEN1        HERNGCRLCKSDKYCEPHDYEYCCPCEWHRTEHDRQLSEVENNIKKKACCCEGFPFHEVI   332
TbFEN1      EE-----------------FNYK---DARNFFLE-----------------PEVTPGEEI   298
TbRAD2      ED-------------VPLTRFYRNYVKWSTLQLA-----------------DSFPESHVV   641
            .:                  *    .                            .   . :

GEN1        QEFLLNKDKLVKVIRYQRPDLLLFQRFTLEKMEWPNHYACEKLLVLLTHYDMIERKL-GS   391
TbFEN1      DIQFRE------------PDEEGLIKFLVDEKLFSKERVLKGIQR------L--------   332
TbRAD2      DAYFNP------------TVNTDTRPFVCAAPDWTKLRLFASMHG------ILNKKYCGE   683
            :  :                     *      : :      :        :
```

```
GEN1        RNSNQLQPIRIVKTRIRNGVHCFEIEWEKPEHYAMEDKQ----HGEFALLTIEEESLFEA   447
TbFEN1      -------------------------------RDALTKKTQGRLDQFFTITKPQKQVNSEA   361
TbRAD2      RLENAQ-------------RECQ---RRQPPSGDPADSAQRRLTDFFSPLPNRERVIFRK   727
                                            ..       *:    .:.   .

GEN1        AYPEIVAVYQKQKLEIKGKKQKRIKPKENNLPEPDEVMSFQSHMTLKPTCEIFHKQNSKL   507
TbFEN1      STAGTKRNR--GAVALPGVLQRKSS-----SGHKKAVKK---------------------   393
TbRAD2      QPPKF---S--E--ALS---YLRAA-----RGDP--------------------------   746
                         :      :         .

GEN1        NSGISPDPTLPQESISASLNSLLLPKNTPCLNAQEQFMSSLRPLAIQQIKAVSKSLISES   567
TbFEN1      ------------------------------------------------------------   393
TbRAD2      ------------------------------------------------------------   746


GEN1        SQPNTSSHNISVIADLHLSTIDWEGTSFSNSPAIQRNTFSHDLKSEVESELSAIPDGFEN   627
TbFEN1      ------------------------------------------------------------   393
TbRAD2      ------------------------------------------------------------   746


GEN1        IPEQLSCESERYTANIKKVLDEDSDGISPEEHLLSGITDLCLQDLPLKERIFTKLSYPQD   687
TbFEN1      ------------------------------------------------------------   393
TbRAD2      ------------------------------------------------------------   746


GEN1        NLQPDVNLKTLSILSVKESCIANSGSDCTSHLSKDLPGIPLQNESRDSKILKGDQLLQED   747
TbFEN1      ------------------------------------------------------------   393
TbRAD2      ------------------------------------------------------------   746


GEN1        YKVNTSVPYSVSNTVVKTCNVRPPNTALDHSRKVDMQTTRKILMKKSVCLDRHSSDEQSA   807
TbFEN1      ------------------------------------------------------------   393
TbRAD2      ------------------------------------------------------------   746


GEN1        PVFGKAKYTTQRMKHSSQKHNSSHFKESGHNKLSSPKIHIKETEQCVRSYETAENEESCF   867
TbFEN1      ------------------------------------------------------------   393
TbRAD2      ------------------------------------------------------------   746


GEN1        PDSTKSSLSSLQCHKKENNSGTCLDSPLPLRQRLKLRFQST     908
TbFEN1      ------------------------------------------    393
TbRAD2      ------------------------------------------    746
```

Figure 3.1. Multiple sequence alignment. Sequence alignment of Human GEN1, *Tb*FEN1 and *Tb*RAD2 to show the basic similarity in sequence between the three protein sequences. Result obtained by EMBL-EBI Clustal Omega (Goujon et al., 2010, Sievers et al., 2011, McWilliam et al., 2013)

As discussed in **Chapter 1.5,** and shown in **Figure 1.7,** human GEN 1, FEN1 and RAD2 have three conserved domains that are characteristic of the XPG/RAD2 family. To confirm identity and explore in greater detail the sequence similarity between the *T. brucei* and human protein sequences, a reciprocal BLASTP analysis was undertaken using the *T. brucei* proteins as the query sequence.

As seen in **Figure 3.2**, it was confirmed that GEN1 and *Tb*FEN1 have highly-conserved regions at the N-terminus of the protein sequence, between 1-350 aa for both proteins. Contrastingly, *Tb*RAD2 has two discrete regions of domain structure at 1-125aa and 450-700aa. From the figure, it appears that both GEN1 and *Tb*FEN1 proteins have their active sites near the N-terminus. Along with these active sites there are DNA and metal binding sites. Therefore, it can be inferred that interaction with the DNA and the proteins functions with DNA occurs here. In contract, *Tb*RAD2's DNA and metal binding sites appear closer to the C-terminus, indicating the location of the active sites is at the opposite terminus to *Tb*FEN1.

The BLASTP identified DNA and metal binding sites in both the human GEN1 and *Tb*FEN1 at similar positions in the protein. The similarity between the human and *T. brucei* proteins indicate potential conserved functional regions between the two proteins. In both *T. brucei* proteins, putative binding sites were identified in the Helix-3 turn-Helix (H-3TH) regions, suggesting a potential functional role at this region. To explore these relationships further a more rigorous pairwise alignment tool Matcher EMBOSS was used to compare the amino acid sequence of the two trypanosome proteins directly to human GEN1.

The Matcher software found one region of partially conserved sequence alignment between each of the trypanosome genes and human GEN1 (**Figure 3.3**). Matcher EMBOSS identified the amino acid region 52-222 in human GEN1 as conferring the greatest sequence similarity with the amino acid region 67-250 in *Tb*FEN1, at 44.9% sequence similarity. When comparing human GEN1 with *Tb*RAD2, the analysis identified the amino acid region 118-226 in human GEN1 as having the greatest sequence similarity (59.6%) with the region 470-574 in *Tb*RAD2.

These results support the output from BLASTP in **Figure 3.2**, where the regions of similarity are evident before and at the beginning of the H-3TH domain region. This further supports the suggestion of functional similarity being related to this region. The H-3TH region is part of the secondary structure of the proteins. As a result of these analyses, it was important to investigate the secondary protein structure.

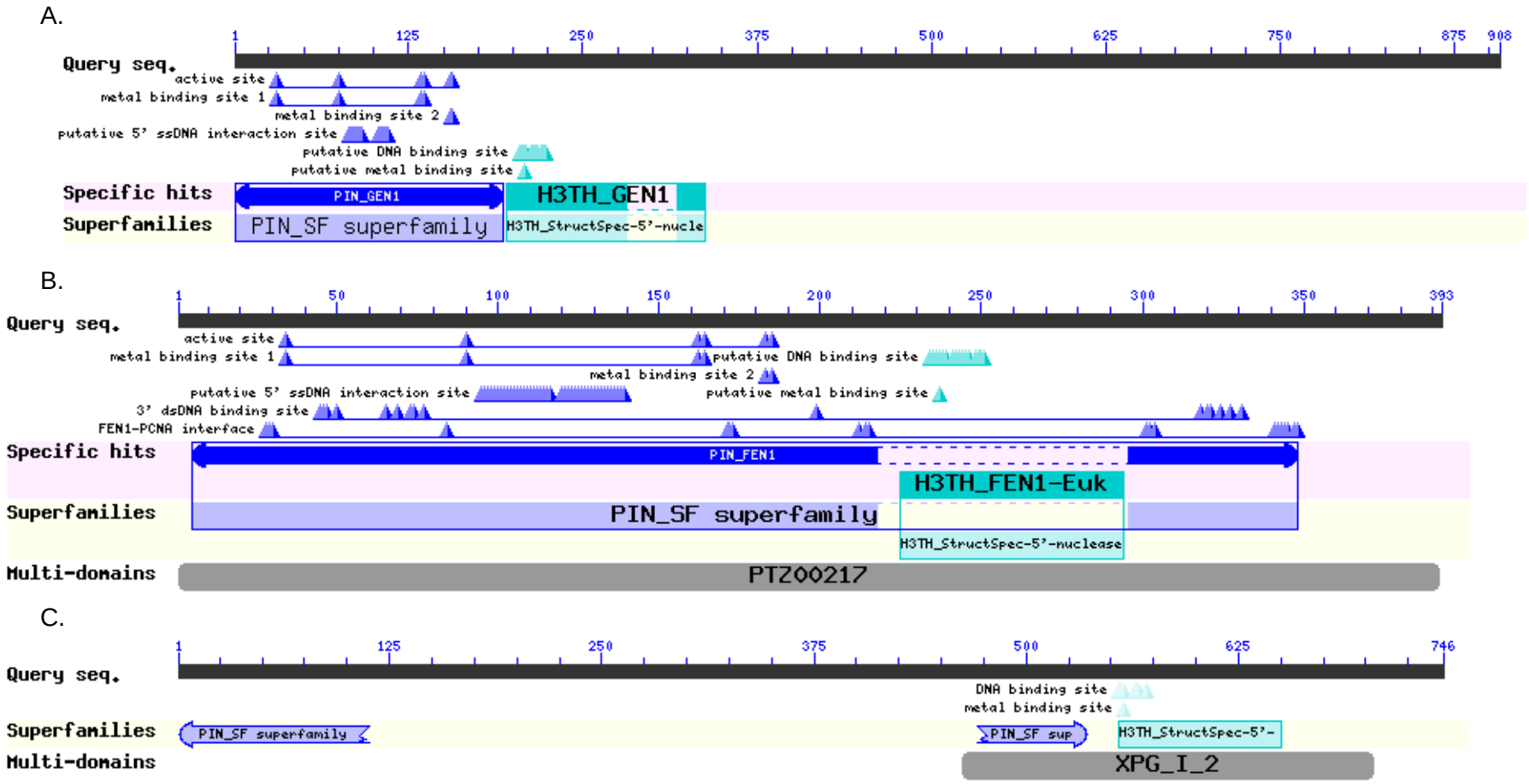


Figure 3.2. Conserved amino-acid regions. Schematics showing the conserved regions of the proteins of interest (A) human GEN1, (B) *Tb*FEN1 and (C) *Tb*RAD2. Results obtained by BLASTP using NCBI Conserved Domain tool (Marchler-Bauer and Bryant, 2004, Marchler-Bauer et al., 2009, Marchler-Bauer et al., 2011, Marchler-Bauer et al., 2015)

GEN1-*Tb*FEN1

```
GEN1              52 HLRNLFFRISYLTQMDVKLVFVMEGEPPKLKADVISKRNQS------RYG     95
                     ||..:|||...:....::.::|.:|:||.|||..:..|.|.      .:.
TbFEN1            67 HLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFE    116

GEN1              96 SSGKSWSQKTGRSHFKSVLR-------ECLHMLECLGIPWVQAAGEAEAM    138
                     .:.:....:.....|.::|       |...:|..:|||.|||..||||.
TbFEN1           117 KAKEEGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQ    166

GEN1             139 CAYLNAGGHVDGCLTNDGDTFLYGAQTVYRNFTMN-TKDPHVDCYTMSSI    187
                     ||.|.........|.|.|...:|::.:.|:.|.. .|...:..|.:..|
TbFEN1           167 CAELVKKNKAWAVGTEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEI    216

GEN1             188 KSKLGLDRDALVGLAILLGCDYLPKGVPGVGKEQA    222
                     ....|......:.|.|||||||:|: :.|:|..:|
TbFEN1           217 LEASGFSMQQFIDLCILLGCDYVPR-ISGIGPHKA    250
```

GEN1-*Tb*RAD2

```
GEN1             118 LHMLECLGIPWVQAAGEAEAMCAYLNAGGHVDGCLTNDGDTFLYGAQTVY    167
                     :.:|:|.|||:|.:..||:|.||:||....||...|.|.|..::||..|.
TbRAD2           470 VELLDCCGIPYVLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVL    519

GEN1             168 RNFTMNTKDPHVDCYTMSSIKSKLGLDRDALVGLAILLGCDYLPKGVPGV    217
                     |.|.  :|..||..|..|.:.: .|:|:..||.||:||||||. :||.|:
TbRAD2           520 RGFF--SKGRHVVAYRQSDLLA-CGVDKVVLVALALLLGCDYA-EGVNGL    565

GEN1             218 GKEQALKLI    226
                     ...::|.:|
TbRAD2           566 SLLESLHVI    574
```

Sequence to GEN1	Length	Identity (%)	Similarity (%)	Gaps (%)	Score
*Tb*FEN1	185	28.6	44.9	8.1	149
*Tb*RAD2	109	41.3	59.6	3.7	184

Figure 3.3. Pairwise sequence alignment. A sequence alignment using the more rigorous alignment tool Matcher, which identifies local similarities based on the Lalign tool. Results obtained from EMBL-EBI and EMBOSS tool Matcher, (Rice et al., 2000, McWilliam et al., 2013, Li et al., 2015). From the Matcher analysis only partial sequences were found to be of similar alignment.

3.3 Secondary Structure Comparisons

To gain a greater confidence that the proteins of interest may be GEN1 orthologues, the secondary structures were predicted. The sections observed were part-sequences which showed conserved domains, as seen in **Figures 3.2** and **3.3**.

Structures were predicted using PHYRE^2, which produces structural models of the proteins of interest based on their alignment to known protein sequences/structures (Kelley et al., 2015). **Figures 3.4** and **3.5** show the predictions of the secondary structures human GEN1 against *Tb*FEN1 and *Tb*RAD2 respectively.

Comparing the predicted secondary structures of human GEN1 to *Tb*FEN1, it can be observed that there is clearly structural similarity, with the only section of difference between the proteins' being in the region of amino acids 83-124 and 98-154 of the human GEN1 and *Tb*FEN1 sequences respectively. For human GEN1, there is one long helix, whereas for *Tb*FEN1 this same section is split into three. At the C-terminus of the three helices in the stated region there is an area of weakened 'SS confidence', suggesting a lack in confidence of the three apparent helices. These three helices could be associated with gaps seen in the matched sequence shown in **Figure 3.3**, which occur between amino acids 108-113 and 137-143 in the *Tb*FEN1 sequence. Comparing directly to the secondary structures, 108-113aa is part of a helix where 137-143aa is part turn and part helix.

At the end of both sequences in **Figure 3.4**, the conserved amino acid domains appear in a pattern of 6 conserved regions, followed by another 6 regions, finally followed by 4 conserved regions, at 204-223aa and 324-351aa for human GEN1 and *Tb*FEN1 respectively. This is consistent with **Figure 3.2** in location and in pattern, where figure B identifies a conserved region at the end of the sequence, where the regions are close together with two of equal size followed by a slightly shorter region, likely another 6-6-4 pattern. Appearing as putative DNA binding sites with an alternative domain, the alternative domain in human GEN1 appears as a metal binding site as shown in **Figure 3.2**. Trypanosome FEN1 protein instead shows this alternative domain to be a putative 5' ssDNA interaction site instead of the metal binding site as expected from **Figure 3.2**.

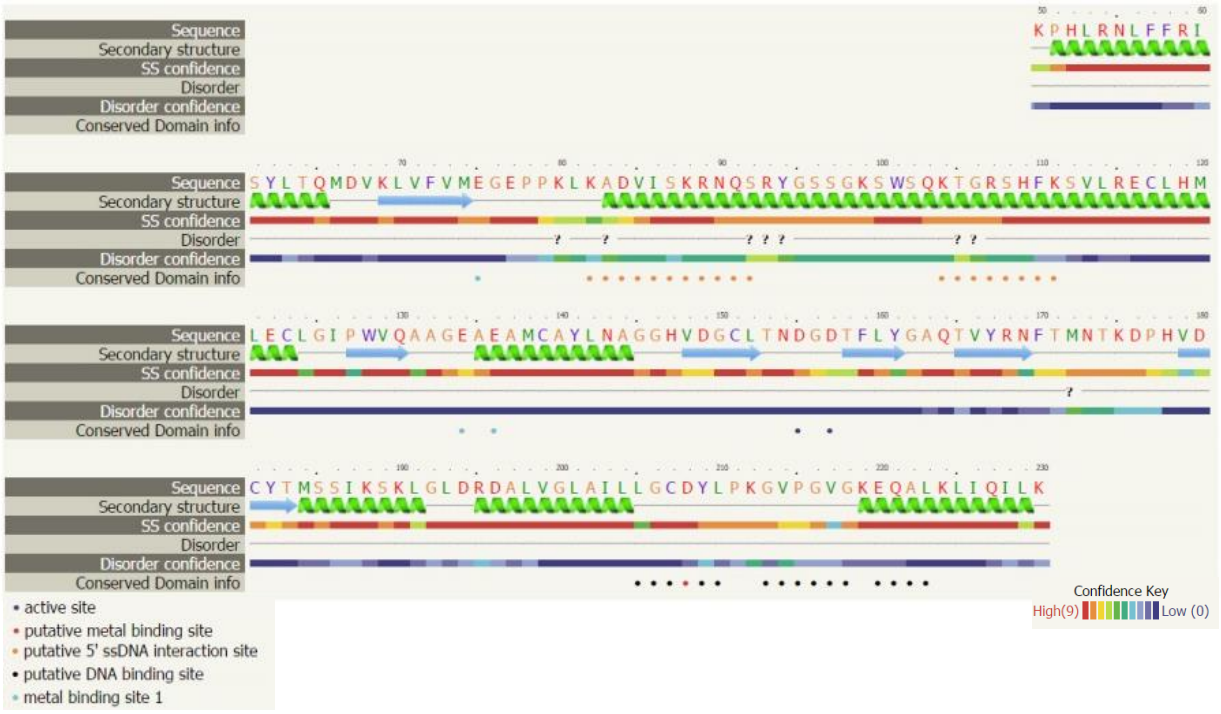
Figure 3.5 compares the predicted secondary structures for human GEN1 and *Tb*RAD2. It can be seen from the figure that the two secondary structure sequences are almost identical. Where only the predicted conserved domains do not appear to match. As described previously, human GEN1's has a 6-6-4 pattern of amino acid domains, whereas

*Tb*RAD2 shows a conserved sequence of 2-1-3-1 amino acid domains. For *Tb*RAD2 these domains match **Figure 3.2** where the majority are DNA binding sites and one metal binding site. Therefore, the expected domain roles for *Tb*RAD2 match the human GEN1 domains although they do not appear in the same pattern.

The matching secondary structures from **Figures 3.4** and **3.5** provide a greater confidence that the trypanosome proteins may potentially have a similar function or role as that of human GEN1.

Although the predicted secondary structures appear to be similar, further analysis was carried out to predict the tertiary structures. Sections of the secondary structures not considered in this analysis, and different amino acid sequences, may cause the trypanosome proteins to have different tertiary structures.

A. Human GEN1, Covering protein sequence 50 – 230.



B. *Tb*FEN1, Covering protein sequence 61 - 260

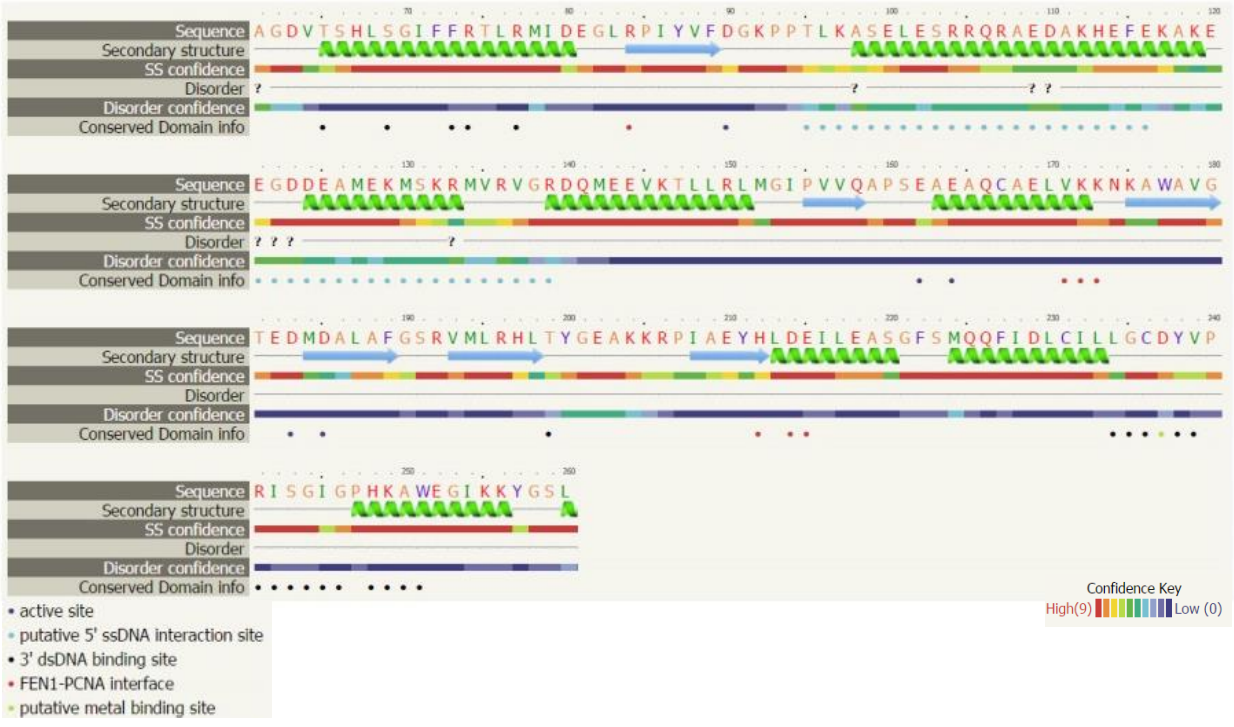
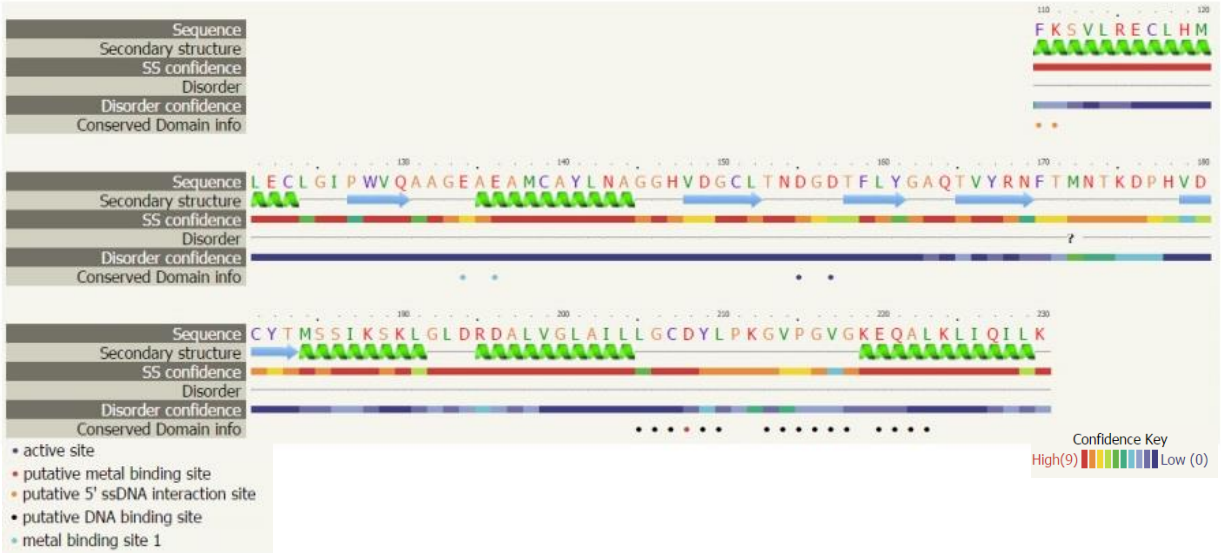


Figure 3.4. Predicted secondary structures of human GEN1 and *Tb*FEN1. Comparison of secondary structures predicted within the area of interest found by Matcher, with ranging confidence and predictions of conserved domains. Results obtained with PHYRE[2] (Kelley et al., 2015).

A. Human GEN1, Covering protein sequence 110 - 230



B. *Tb*RAD2, Covering protein sequence 468 – 580

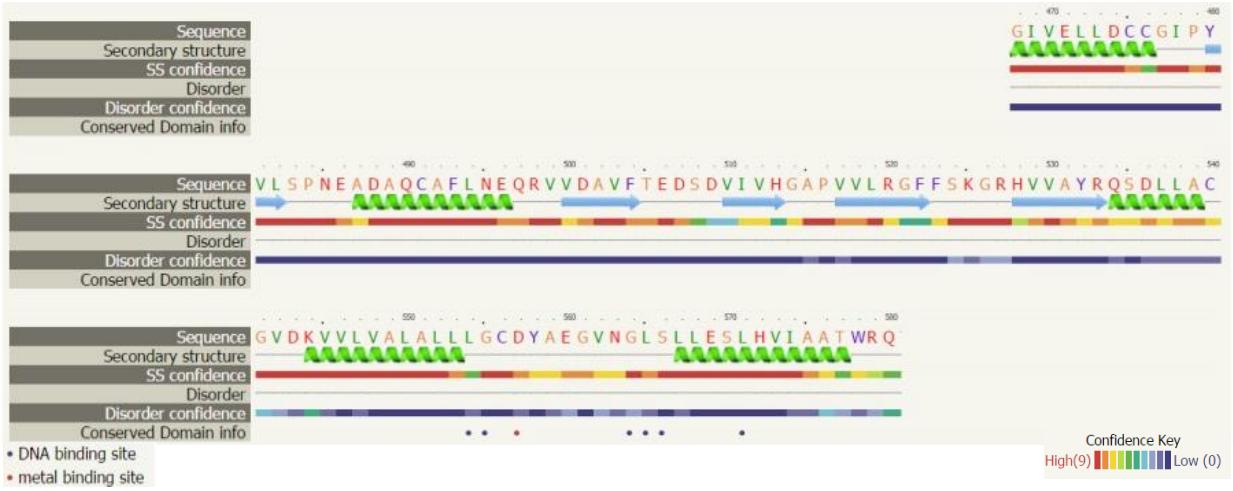


Figure 3.5. Predicted secondary structures of human GEN1 and *Tb*RAD2. Comparison of secondary structures predicted within the area of interest found by Matcher, with ranging confidence and predictions of conserved domains. Results obtained with PHYRE[2] (Kelley et al., 2015).

3.4 Tertiary Structure Comparisons

Tertiary structures were predicted to analyse how the interactions of the underlying amino acids and changes in the secondary structures might affect the proteins 3-dimensional character. Although this analysis does not specifically tell us about the functionality of the protein, similar predicted structures to the GEN1 protein support evidence that they may be orthologues.

As well as predicting secondary structures, PHYRE^2 can also be used to predict a potential tertiary structure for the protein of interest. PHYRE^2 has both Normal and Intensive prediction parameters; normal prediction parameters compare hidden Markov modelling (HMM) of the sequence of interest against HMM of known proteins to find areas of homology; Intensive prediction parameters use these results to build a tertiary structure *ab initio* based on the confidence of this HMM-HMM comparison. The method of *ab initio* comparison can be highly inaccurate so these intensive models can only be taken as a guide of what the protein may look like.

For the human GEN1 protein sequence PHYRE^2 was unable to predict a viable tertiary structure. The structure predicted simply showed few secondary structures and only turns. Therefore, it could not be meaningfully used to compare against the trypanosome proteins. Instead the programme was able to search for already resolved GEN1 protein structures. This identified the crystal structure of the eukaryotic *Chaetomium thermophilum* (*C. thermophilum*) GEN1 bound to DNA to be found at a resolution of 2.5 Å; as seen in **Figure 3.6 A** and **B**. To view the comparison of the tertiary structures of GEN1 to the *T. brucei* proteins of interest, the fungal eukaryotic GEN1 was used. The programme also found the resolved crystal structure for human FEN1 shown in **C** of **Figure 3.6**, however no resolved protein structure for RAD2 was identified to be used for comparison purposes.

Figure A shows how the fungal GEN1 interacts with DNA in a pair like manner. This suggests that GEN1, acts in a pair when carrying out Homologous Recombination. To understand this interaction further it was important to look at how the protein and DNA bind together.

Figure B and **Figure C** focus on how the different proteins interact and bind with DNA. By comparing the two figures, it can be seen that the fungal GEN1 and human FEN1 proteins bind with DNA slightly differently. From these figures, it appears that the human FEN1 protein binds in a more physical manner than the fungal GEN1. This is seen by the way the

FEN1 wraps round the DNA double helix, relative to the GEN1 which binds to the side. This difference could potentially be due to FEN1's smaller size or that it acts singularly. The wrapping of the protein may also affect its function with the DNA relative to GEN1.

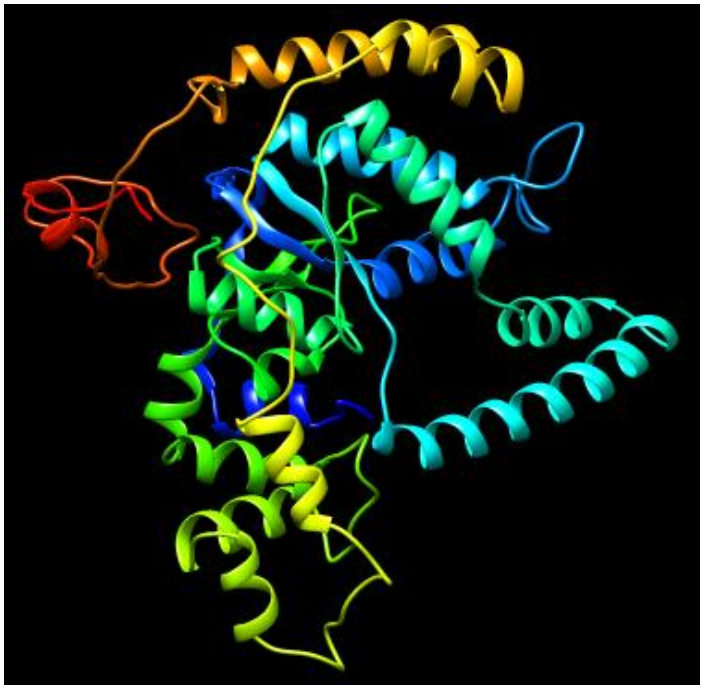
Figures D and **E** show the predicted tertiary structures of the trypanosome proteins *Tb*FEN1 and *Tb*RAD2. The *Tb*FEN1 structure shares similar morphology to the human FEN1 structure, which can easily be seen from **Figure C**. Although the predicted version appears to contain longer turns between the helices and beta-strands. The similar structures suggest *Tb*FEN1 would interact with DNA in a similar manner to human FEN1.

A. *C. thermophilum* GEN1 Chains A and B bound to DNA

B. *C. thermophilum* GEN1 Chain A bound to DNA

C. Human FEN1 bound with DNA

D. *Tb*FEN1 Predicted Structure



E. *Tb*RAD2 Predicted Structure

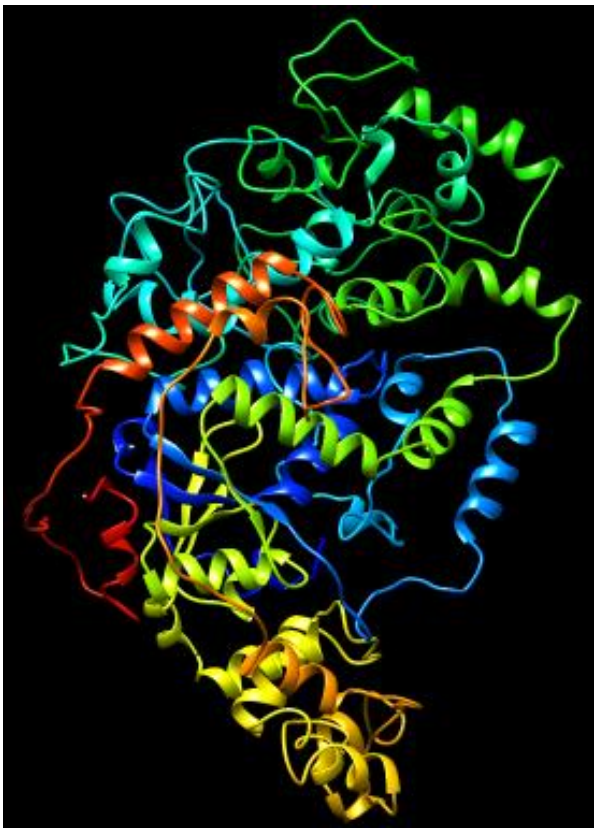


Figure 3.6. Tertiary structures. Ribbon models of tertiary structures for (A) *C. thermophilum* GEN1 Chains A and B, PDB: 5CO8. (Liu et al., 2015) (B) *C. thermophilum* GEN1 Chain A, PDB: 5CNQ. (Liu et al., 2015) (C) human FEN1, PDB: 3Q8L. (Tsutakawa et al., 2011) (D) *Tb*FEN1, PHYRE^2 predicted. (Kelley et al., 2015) (E) *Tb*RAD2, PHYRE^2 predicted. (Kelley et al., 2015). Images A, B and C show proteins bound to a DNA molecule, shown by the colour magenta. Rainbow coloured images B, C, D and E show the protein sequence from the N-terminus to the C-terminus, by the changing from blue to red. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). (Pettersen et al., 2004).

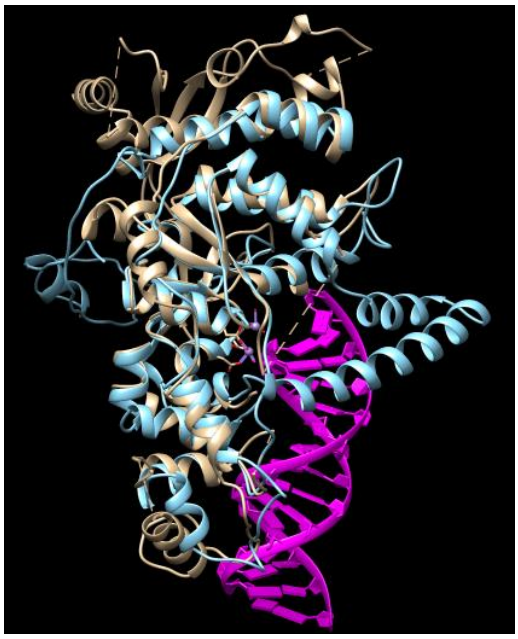
To obtain a clearer visual impression of the similarity between the tertiary structures, the *T. brucei* proteins were superimposed on to the fungal GEN1 protein. To superimpose the proteins the UCSF Chimera MatchMaker tool was used; shown in **Figure 3.7** for *Tb*FEN1 and *Tb*RAD2 respectively.

From **Figure 3.7 A**, it can clearly be seen that there is more structural similarity between the fungal GEN1 and *Tb*FEN1 than could originally been interpreted from **Figure 3.6**. This is seen by the many overlapping helices, strands and turns. Where the fungal GEN1 is shown bound to DNA, the predicted trypanosome structures are not based on being bound to DNA. As expected from the individual tertiary structures, there is a distinct difference between structures at the point where the *Tb*FEN1 would wrap around the DNA. It can be assumed that upon *Tb*FEN1 binding to DNA the two structures could have a closer resemblance due to conformational change.

In **Figure 3.7 B,** it can clearly be seen that it is the C-terminus of the protein which has structural similarity to the fungal GEN1, seen by the overlapping helices and strands. From this information, it appears that it is the C-terminus of the protein which interacts with DNA. From this comparison, it can now be seen that the protein appears as if it will wrap around the DNA in an equivalent manner as the *Tb*FEN1 would.

The superimposed images, provides clearer evidence that all three proteins have structural homology, although it cannot provide a clear answer as to whether these proteins have a similar function in homologous recombination to GEN1 it may suggest an orthologous relationship.

A. Chain A *C. thermophilum* GEN1 with *Tb*FEN1



B. Chain A *C. thermophilum* GEN1 with *Tb*RAD2

Figure 3.7. Superimposed tertiary structures. Predicted Ribbon tertiary structures (A) *Tb*FEN1 (light blue) and (B) *Tb*RAD2 (light blue) (Kelley et al., 2015), superimposed on *C. thermophilum* GEN1 (beige) chain A, PDB: 5CNQ (Liu et al., 2015). Both images show GEN1 bound to DNA, coloured in magenta. Made using UCSF Chimera MatchMaker tool, Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). (Pettersen et al., 2004).

3.5. Chapter Summary

The results reported in this chapter identify two *T. brucei* proteins with similar sequence and structure to the human GEN1 protein. A comparison of the amino acid sequences, domain architectures, as well as predicted secondary and tertiary structures, were used to predict whether these proteins were potential orthologues of the human protein.

From this bioinformatics analysis, it was observed that both *T. brucei* proteins show a high degree of similarity to the human GEN1 in their primary amino acid sequence, along with similar observed secondary and tertiary structures sequences and the domain architectures.

Given the structural similarities of the trypanosome proteins, and that the human GEN1 protein is a member of this XPG/RAD2 family, the trypanosome proteins are likely candidates for also being members of this family.

The structural similarities suggest the *T. brucei* proteins may function in a similar manner to the human GEN1, where direct experimental analysis of the *T. brucei* proteins is carried in future chapters to explore this.

4 Expression and Purification of *T. brucei* Holliday Junction Resolvase-like Proteins

4.1 Chapter Synopsis

For any future biochemical experimentation, the *Tb*FEN1 and *Tb*RAD2 proteins would need to be expressed and purified. Therefore, to see whether these proteins could be expressed as recombinant fusion proteins, *E. coli* exploratory expression and purification experiments were carried out.

4.2 Large Scale Purification of His-Tagged *Tb*FEN1

So that the *Tb*FEN1 purified protein could be meaningfully used for potential future experiments, a large amount of the protein would be required.

The plasmid used was pJGO4 (Jennifer Owen, unpublished) and is a derivative of pET24a in which the *Tb*FEN1 open reading frame has been inserted via *Nde*I and *Xho*I sites to generate a recombinant plasmid in which *Tb*FEN1 is expressed with a C-terminal 6xHis tag. Expression and purification conditions had previously been optimised (McAllister, unpublished).

Bacteria containing recombinant plasmid pJGO4 were induced using the optimum expression conditions determined in previous small scale purification (0.1 mM IPTG at 30°C). Bacteria were lysed under native buffer conditions and the insoluble material removed by centrifugation.

Soluble proteins were injected into a Hi-Trap™ Ni-NTA column, using an AKTA Prime Plus system unbound proteins were removed by washing with a low imidazole buffer, and bound proteins eluted with a 20 mM to 500 mM imidazole gradient in binding buffer. The absorbance of material eluting from the column was monitored by the system, and fractions with elevated absorbance (indicative of proteins in eluate) were analysed by SDS-PAGE.

Figure 4.1 shows SDS-PAGE analysis of the purification fractions from IMAC purification, the SDS-PAGE results indicate that the 48 kDa *Tb*FEN1 protein is present in fractions 18-27, at 150-300 mM imidazole concentrations. Where the fractions 21-25 contained the highest concentration of *Tb*FEN1 protein, at 200-275 mM. Further analysis indicates the

presence of a range of other *E. coli* proteins that remain as contaminants at this stage of the purification.

Elution fractions shown in **Figure 4.1** were visually assessed to identify which fractions contained the greatest amount of *Tb*FEN1 protein, in combination with the least amount of contaminants present. From this it was decided to use take fractions 23-25 forward for IEX purification. These samples were pooled and diluted with IEX binding buffer to reduce the salt concentration and then injected onto the Mono-Q™ column. Using the the AKTA Purifier the unbound proteins were washed away with a low salt buffer, and the bound proteins eluted with a 0 mM to 500 mM salt gradient.

Figure 4.3 shows SDS-PAGE analysis of the fractions obtained following IEX purification. Analysis of elution fractions indicates the presence of the 48kDa *Tb*FEN1 protein in fractions 4-13, at 125-350 mM sodium chloride. With the highest concentration of the *Tb*FEN1 protein in fractions 6-8, at 175-225 mM sodium chloride.

From fraction 8 onwards, additional protein bands are visible, indicating the presence of contaminating *E. coli* proteins in these fractions. From this analysis, the TbFEN1 was determined to be most pure in fractions 4-7, and present in the highest concentration in fractions 6 and 7. Fractions 6 and 7 were retained and used in further experiments.

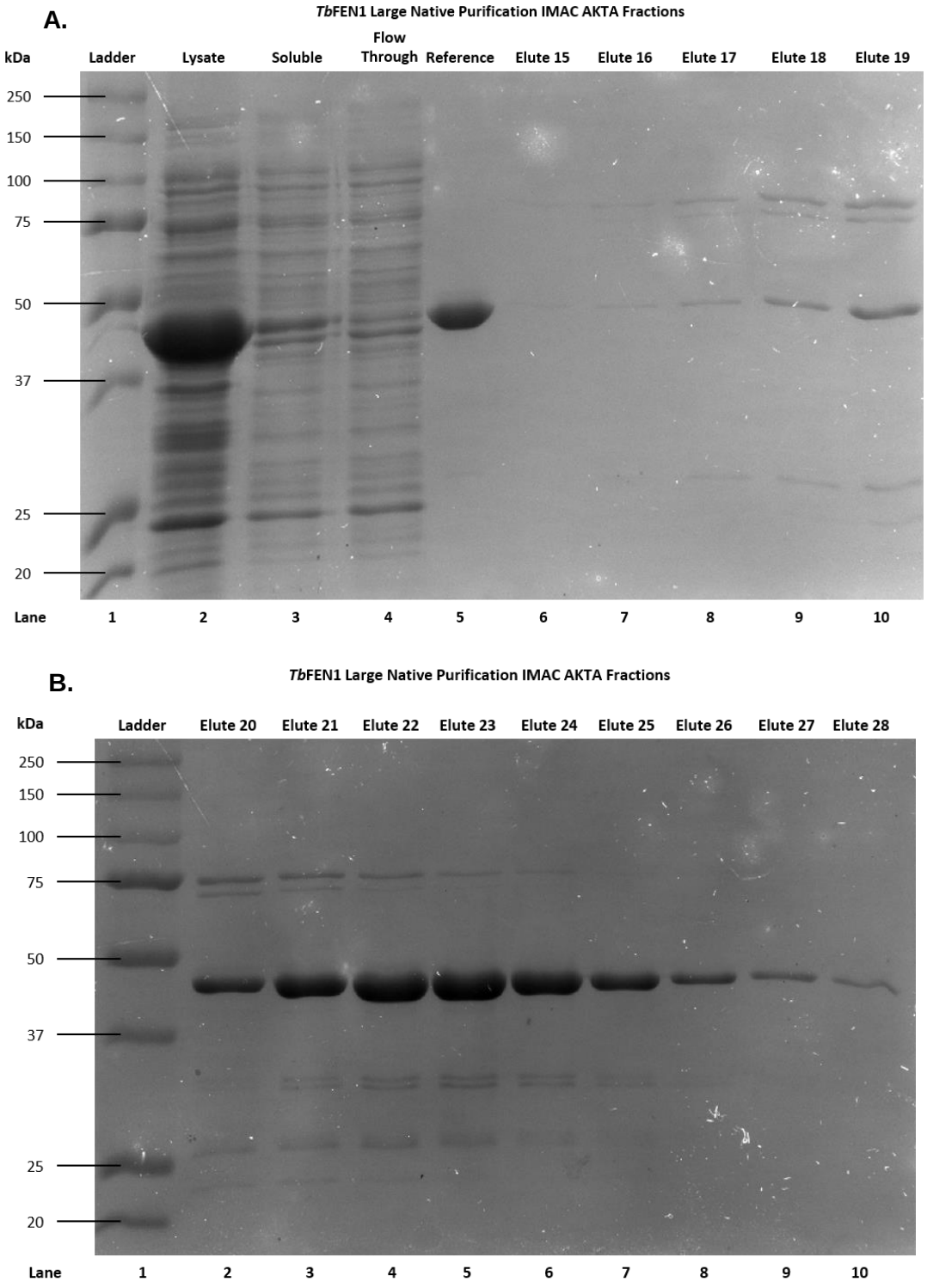


Figure 4.1. SDS-PAGE analysis of fractions obtained following purification by IMAC. 12% SDS-PAGEs showing the eluted fractions of IMAC purification as imidazole concentration was increased. (a) Shows the lysate, soluble, flow through and *Tb*FEN1 reference fractions for comparison against elutes 15-19, (b) Shows elutes 20-28. The fractions eluted correspond with the fractions in **Figure 4.2**. Ladder used – Precision plus protein™ All blue prestained proteins standards 10-250kDa protein ladder, BioRad.

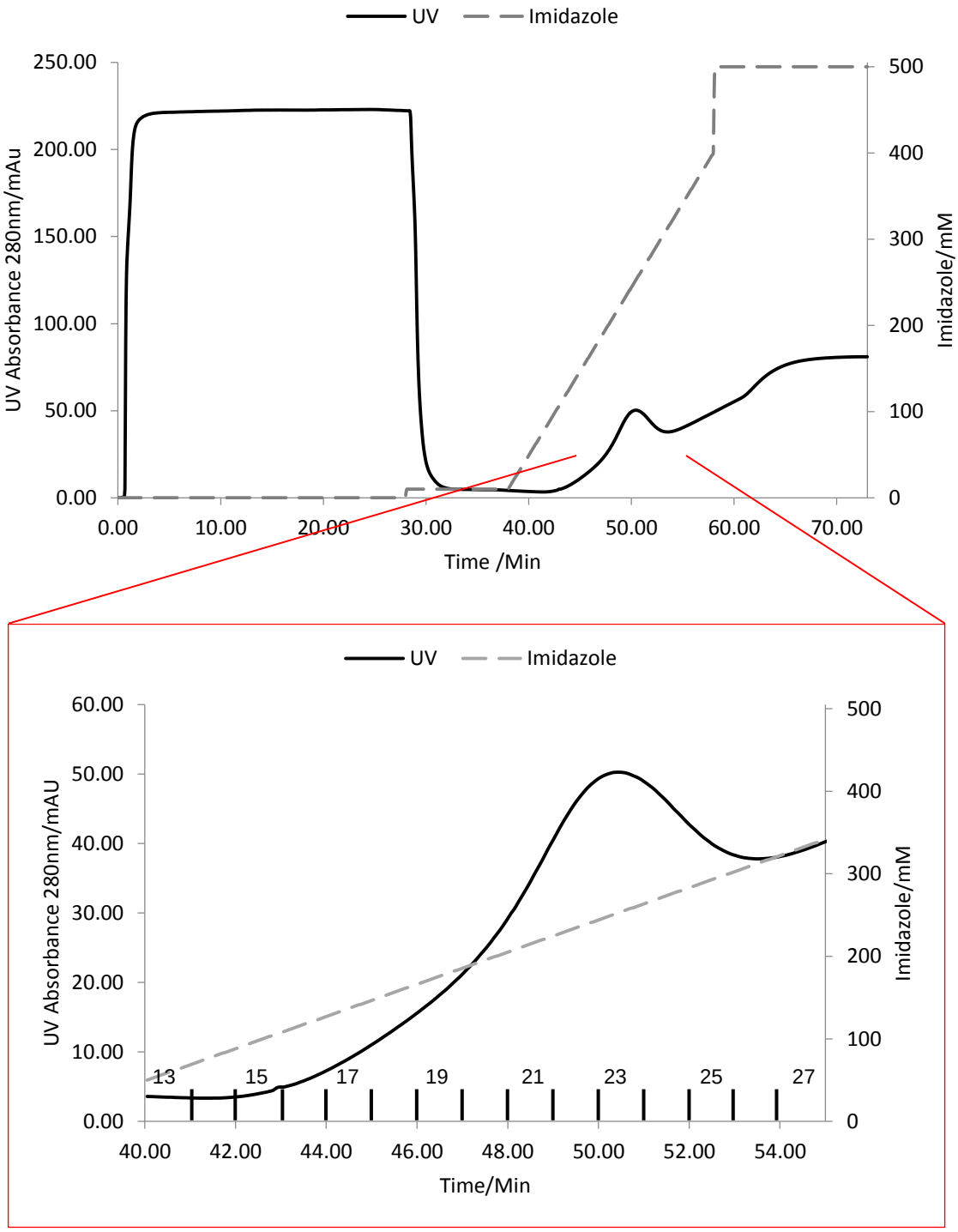


Figure 4.2. UV Absorbance of Large Scale *Tb*FEN1 IMAC Purification Curves. Graphs to show the rate of protein elution against the concentration imidazole with time recorded. A more precise 'red' graph was used to show the protein peak and points at which eluted fractions were collected, these fractions correspond with those in **Figure 4.1**.

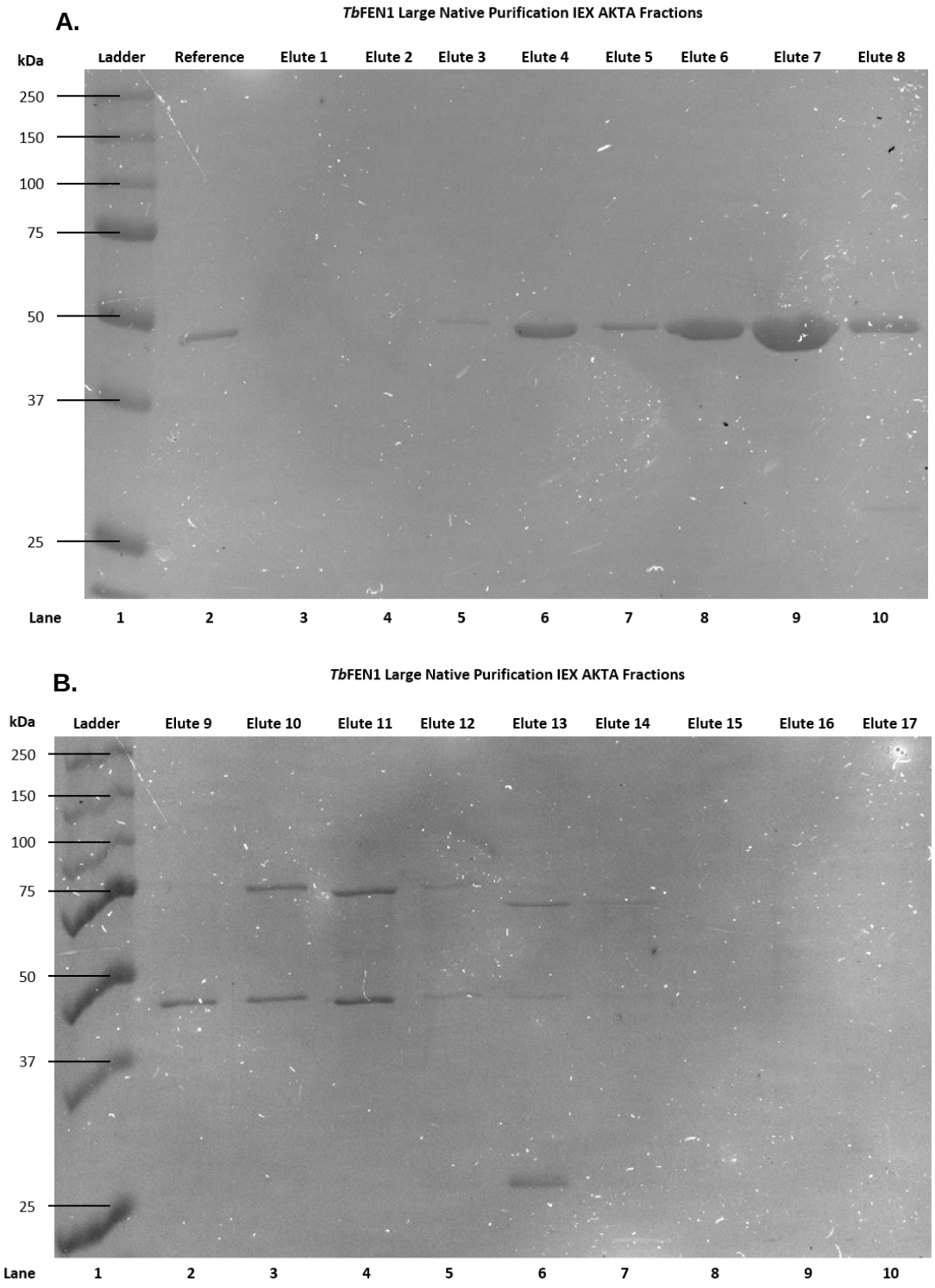


Figure 4.3. SDS-PAGE analysis of fractions obtained following purification *Tb*FEN1 on a Mono-Q anion-exchange column. 12% SDS-PAGEs to show the eluted fractions collected in IEX purification as NaCl concentration was increased. (a) shows the reference *Tb*FEN1 protein for comparison against the elutes 1-8, (B) shows the elutes 9-10. The fractions collected correspond with those collected in **Figure 4.4**. Ladder used – Precision plus protein™ All blue prestained proteins standards 10-250kDa protein ladder, BioRad.

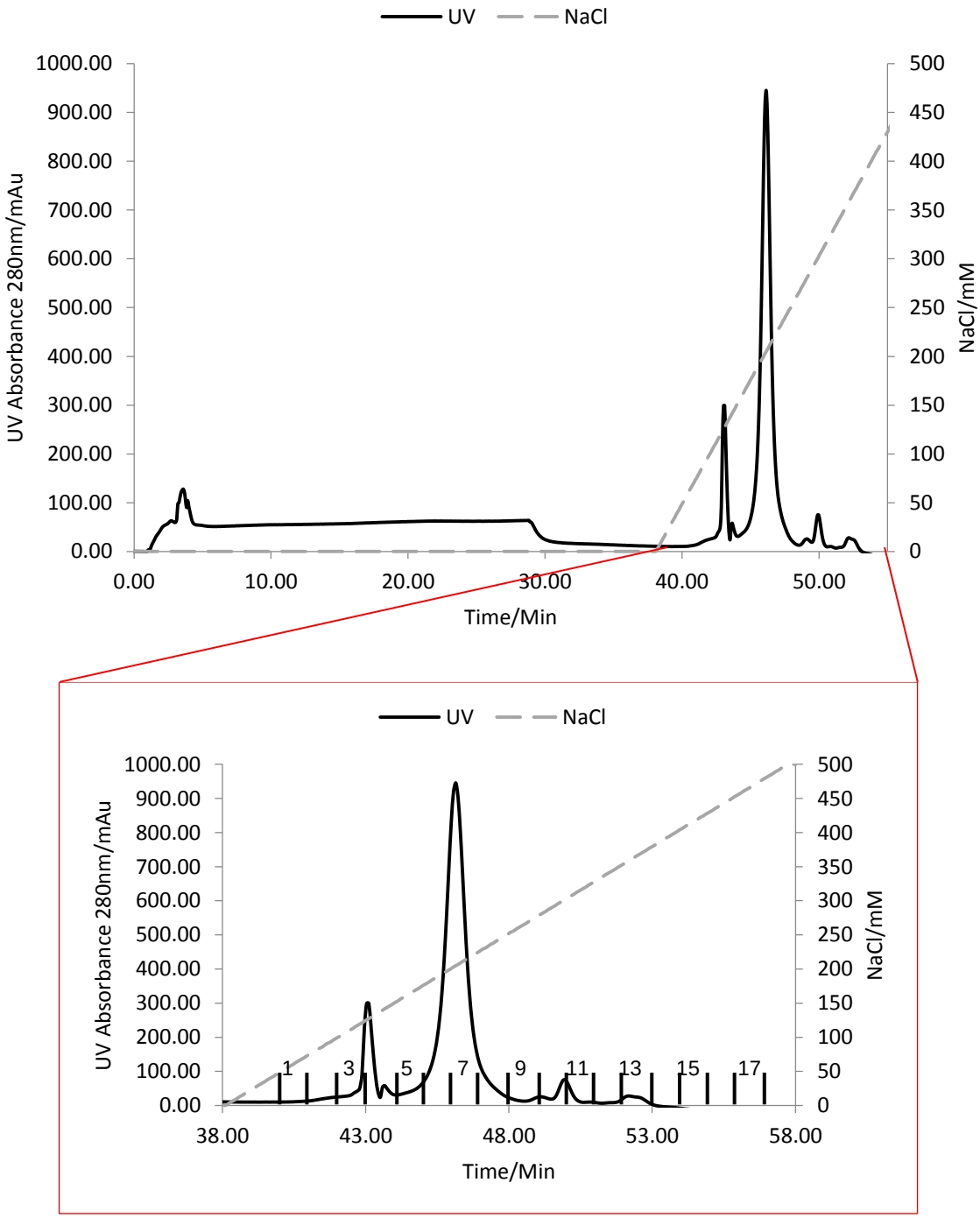


Figure 4.4. UV absorbance trace of eluate from Mono-Q column loaded with pooled fractions (23-25) from IMAC. Graphs to show the rate of protein elution against the concentration NaCl with time recorded. A more precise 'red' graph was used to show the protein peaks and points at which eluted fractions were collected, these fractions correspond with those in **Figure 4.3**.

4.3 Generation of pET24a::*Tb*RAD2 clone

The *Tb*RAD2 gene was PCR amplified from *T. bruce*i 427 genomic DNA and purified for ligation into the cloning vector pGEM-T Easy.

Figure 4.5 compares the *Tb*RAD2 PCR product to *Nde*I/*Xho*I digested putative pGEM-T Easy *Tb*RAD2 positive plasmids (lanes 3, 5, 7 and 8), to ensure an insert was present prior to sequencing. The figure also shows that the PCR product is marginally larger than the DNA inserts released from recombinant plasmids. This was expected as the *Tb*RAD2 sequence contained an internal *Nde*I site (at bp 96-101).

DNA from two recombinant plasmids (lanes 3 and 5 in **Figure 4.5)** were selected and sent for nucleotide sequencing. The resulting nucleotide sequences were conceptually translated into amino acid sequence and compared to *Tb*RAD2 sequence from *T. brucei* strains 427 and 927. Homologous sequences were identified in both strains, however, the 427 strain did not have a complete sequence for *Tb*RAD2 at the time of my analysis amino acids which were not defined are annotated by an X in **Appendix 1**.

The sub-cloning strategy for the *Tb*RAD2 is to incorporate the full *Tb*RAD2 open reading frame (ORF) into the pET24a vector, so that a 6xHis tag can be added to the C-terminus of the protein. For the sequence to be inserted into this vector unidirectionally, the 5' *Nde*I and 3' *Xho*I restriction sites have been added to the ORF. As there is an internal *Nde*I site this causes the ORF to be digested and therefore needed to be removed by single-site mutagenesis. The aim of the single-site mutagenesis, was to change the sequence coding for the *Nde*I site but maintain the amino acid coded for. Therefore, the adenine residue within the restriction site was replaced with cytosine. This meant the codon sequence was mutated from an ATA to ATC; this codon still coded for isoleucine, but the recognition sequence for *Nde*I was removed from within the gene.

To ensure the *Nde*I site within the gene had been successfully removed, the *Tb*RAD2 gene was released from the pGEM-T Easy vector by digestion with *Nde*I and *Xho*I and resolved on an agarose gel (**Figure 4.6**). This DNA insert was larger than that derived from the original plasmid, indicating that the internal *Nde*I site had been successfully removed. This was confirmed by nucleotide sequencing

The mutated *Tb*RAD2 sequence was digested form the pGEM-T Easy vector using *Nde*I and *Xho*I, the DNA insert was then isolated and ligated into the pET24a vector. The pET24a vector was transformed into supercompetent *E. coli*, where kanamycin resistant

transformants were selected. Following ligation into pET24a and transformation into *E. coli*, plasmid DNA was prepared from representative transformants and digested with *Nde*I and *Xho*I as seen in **Figure 4.7**. From this figure, it is evident that a DNA band is visible at approximately 2,200bp in both the *Tb*RAD2 PCR product lane and the digested pET24a::*Tb*RAD2 plasmid lane; indicating that *Tb*RAD2 had inserted correctly into the pET24a vector. This recombinant plasmid will now be referred to as pAJ100 throughout.

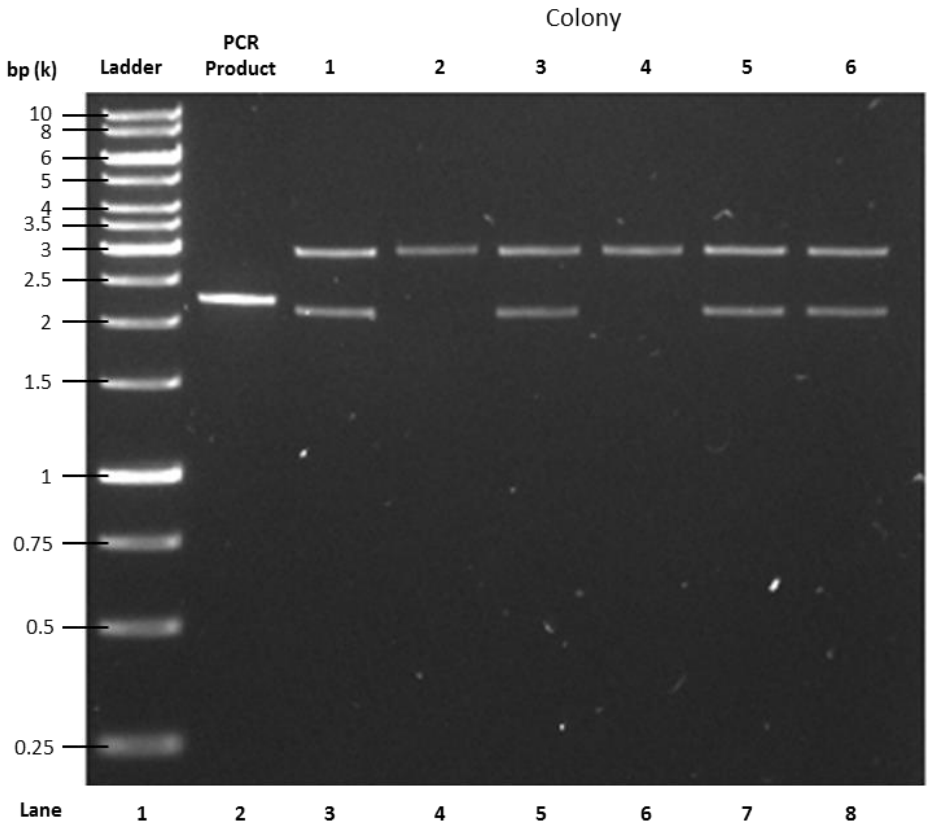


Figure 4.5. Agarose gel analysis of putative pGEM-T Easy *Tb*RAD2 recombinant plasmids. A 0.8% agarose gel showing six putative pGEM-T Easy RAD2 clones digested with *Nde*I and *Xho*I. The amplified PCR product *Tb*RAD2, Lane 2, for digest comparison. Ladder used – GeneRuler™ 1kb DNA ladder, ThermoScientific.

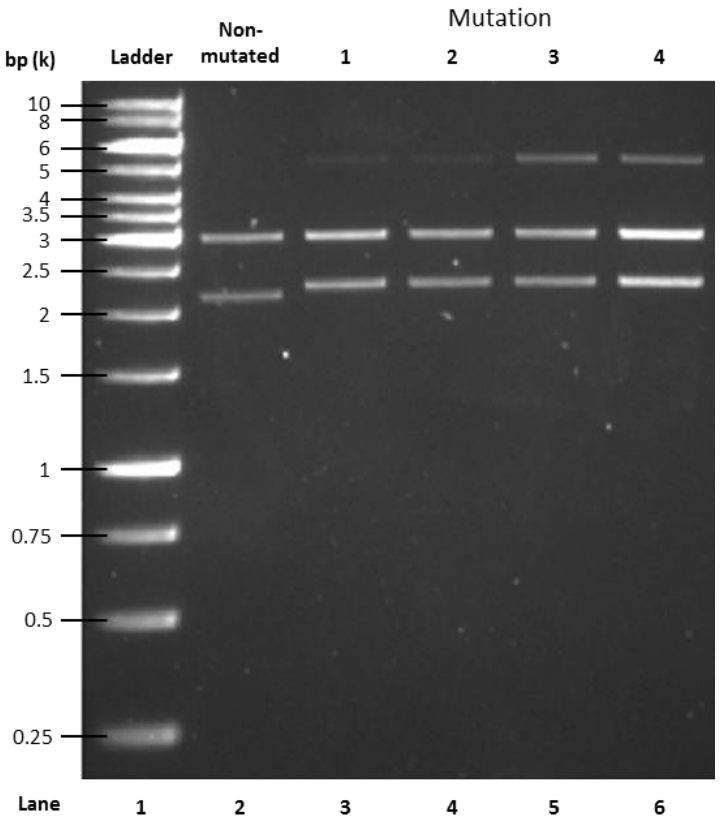


Figure 4.6. Agarose gel analysis of mutated putative pGEM-T Easy*Tb*RAD2 recombinant plasmids. A 0.8% agarose gel showing four mutated putative pGEM-T Easy RAD2 clones digested with *Nde*I and *Xho*I. With a non-mutated version of the vector for comparison, Lane 2. Ladder used – GeneRuler™ 1kb DNA ladder, ThermoScientific.

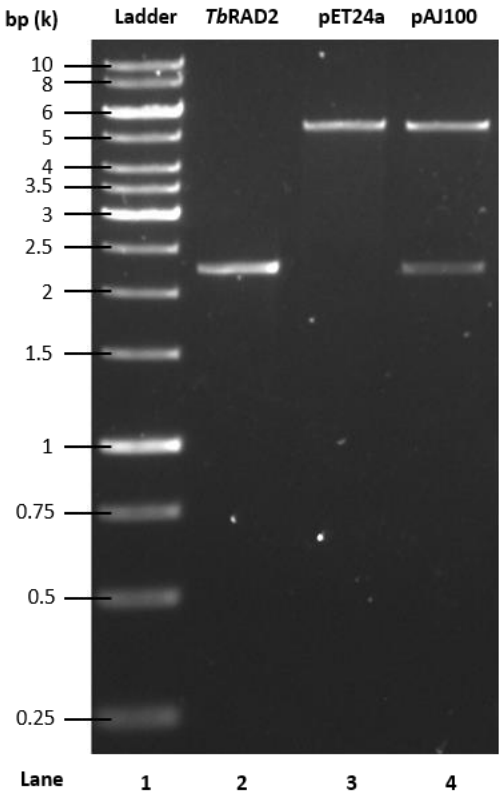


Figure 4.7. Agarose gel analysis of putative pET24a::*Tb*RAD2 recombinant plasmid.
A 0.8% agarose gel showing pET24a::*Tb*RAD2 digested with digested with *Nde*I and *Xho*I. Resolved with the PCR product putative *Tb*RAD2 and the pre-linearised pET24a for comparison, lanes 2 and 3 respectively. Ladder used – GeneRuler™ 1kb DNA ladder, ThermoScientific.

4.4 Optimisation of Conditions for Expression of His-tagged *Tb*RAD2

Prior to expression, the pAJ100 recombinant vector was transformed into the Tuner™ pLacI bacterial strain. Tuner™ bacterial strains are a mutant of BL21 with a lacZY deletion. With this deletion, the bacterial population will carry out a uniform entry of IPTG, which allows for a concentration dependent level of induction. Therefore, the concentration of IPTG can be adjusted to view how this affects protein expression, where the best conditions for the production for 6xHis-tagged *Tb*RAD2 were optimised.

The level of expression for *Tb*RAD2 protein was checked under the conditions of 0.1mM or 1mM IPTG at 30°C or 37°C. **Figure 4.8** shows the resolved proteins expressed in the Tuner™ pLacI strain under these conditions. Under all conditions, but most clearly at 37°C, there was a visible band at 110 kDa that increased in intensity over the expression period (lanes 2-5 and 7-10).

This suggests the induced protein has an apparent molecular weight of 110 kDa on SDS-PAGE, compared with a molecular weight of 83.2 kDa predicted on the basis of the primary amino acid sequence using the ExPASy pI/Mw tool.

Further analysis of the *Tb*RAD2 expression results were carried out using an anti-His antibody and visualised using an immunoblot. The 6xHis tag added to the C-terminus of the *Tb*RAD2 protein, by the pET24a vector, is detected by the antibody and is visualised in **Figure 4.9**. Both **A** and **B** show the presence of only one band which is expressed at 110 kDa, over the period of induction (lanes 2-5 and 7-10) the 110 kDa band increases in intensity. This complements the results seen in **Figure 4.8** and provides greater certainty that the band is representing the protein *Tb*RAD2, as only the *Tb*RAD2 gene present in the pET24a vector should be being expressed with a 6xHis tag following induction of the cell line.

In both figures, it can clearly be seen that there is a greater level of expression of the *Tb*RAD2 protein at 37°C (**B**) than at 30°C (**A**), shown by the thicker bands at 110 kDa. In **Figure 4.8 A**, at 30° the 110 kDa banding appears to be stronger for induction with 1 mM IPTG (lanes 7-10) than induction with 0.1 mM (lanes 2-4), this pattern is replicated in **B** at 37°C. For the immunoblots seen in **Figure 4.9 A**, for the 0.1 mM induction the His-tagged band can only be seen in lane 5 following a three hour induction period. For the 1 mM induction, the band can be seen in lanes 8 and 9 following incubation for two hours, for the third hour the band disappears which appears to be a blotting error. However, in **Figure 4.9**

B there is clear banding in all induction lanes, where induction with 1 mM IPTG (lanes 7-10) appears thicker than induction with 0.1 mM.

From these results, the optimum induction conditions for the expression of *Tb*RAD2 were found to be 1 mM IPTG at 37°C. These conditions were taken forward for the preparation of the pAJ100 cell line for *Tb*RAD2 purification.

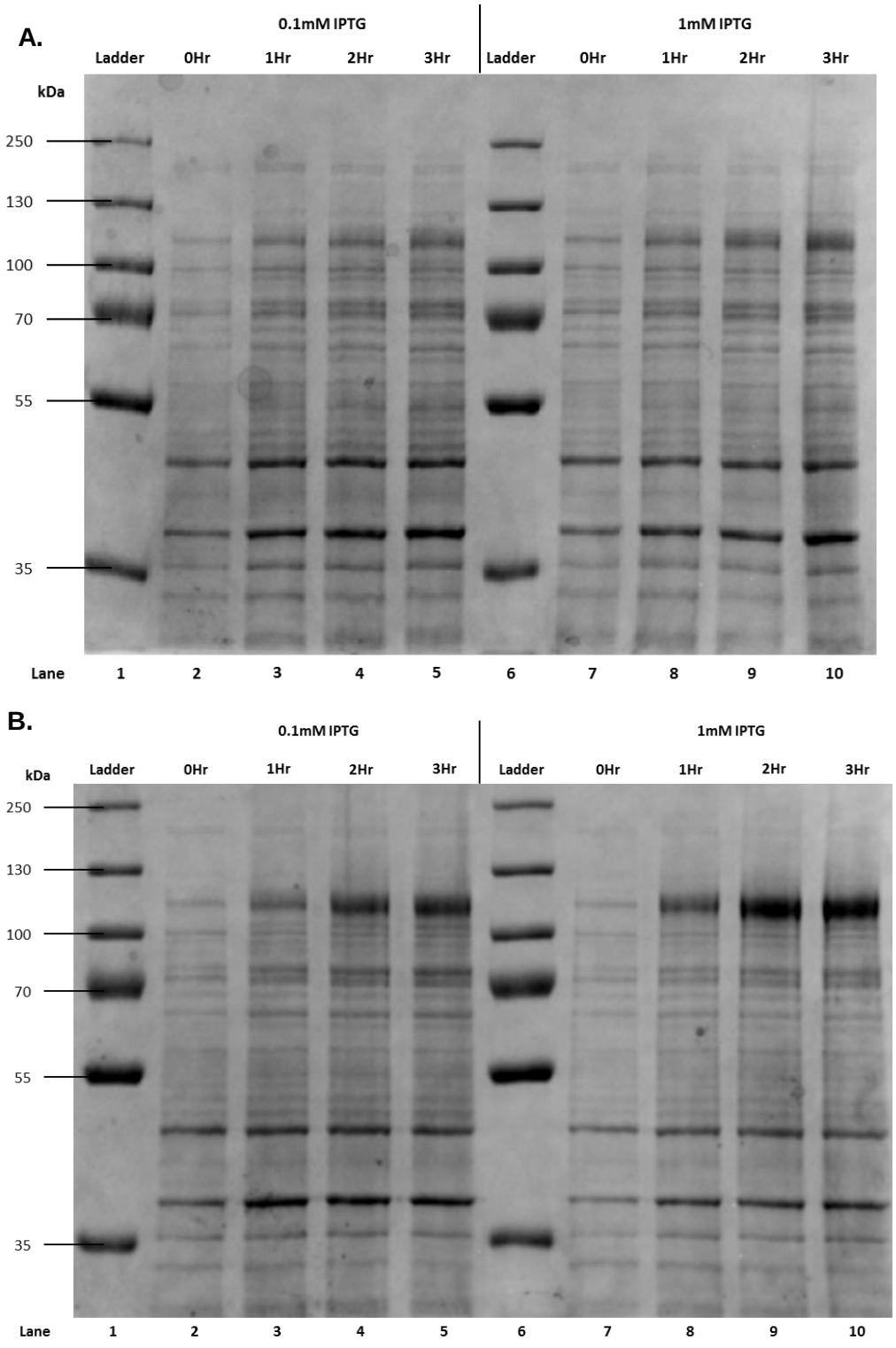


Figure 4.8. SDS-PAGE analysis of expression of *Tb*RAD2 Protein under Different Conditions. Resolved 8% SDS-PAGE to show the amount of increased protein expression under different induction conditions, (A) incubated at 30°C after induction and (B) incubated at 37°C once induced. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

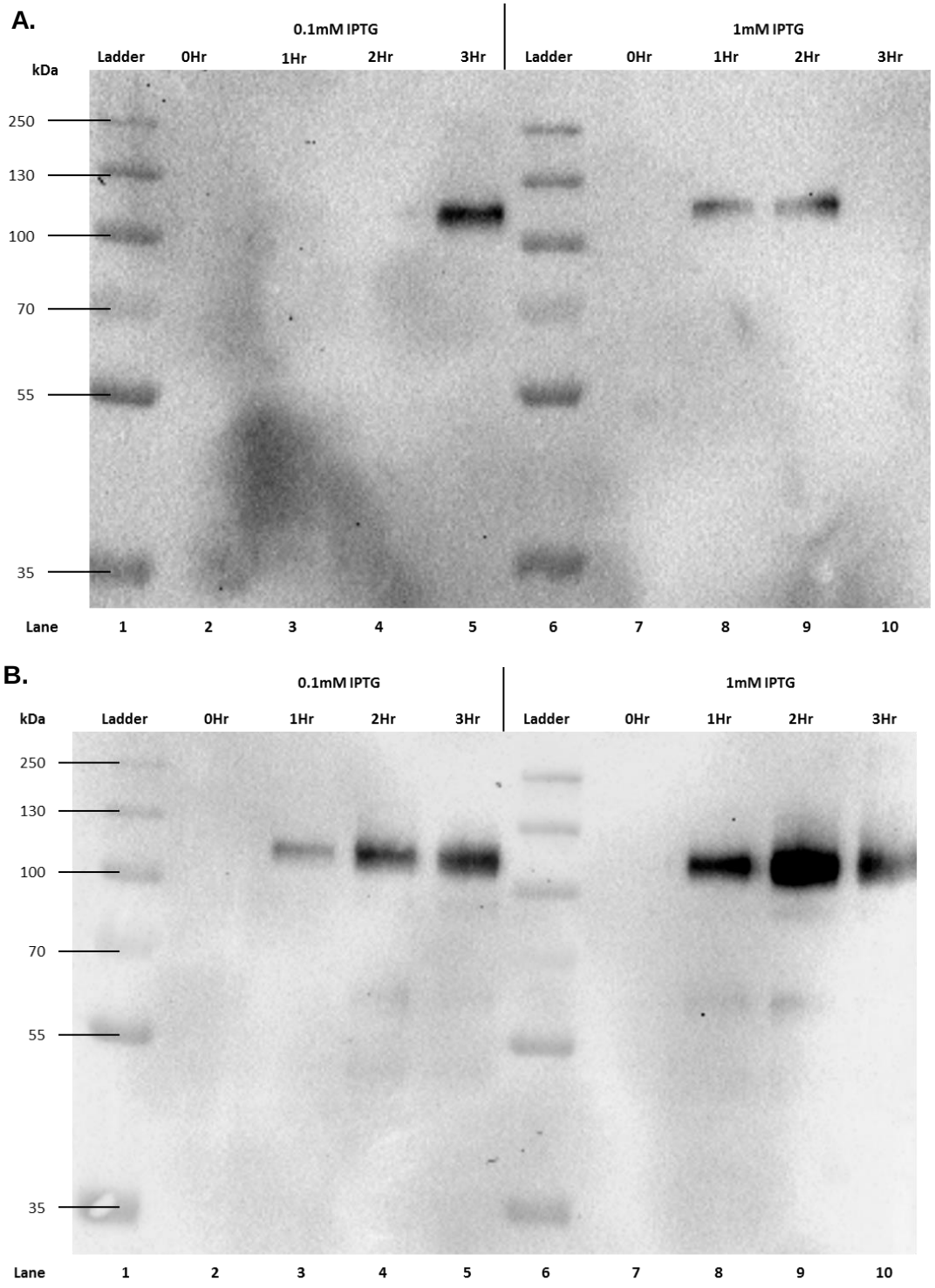


Figure 4.9. Anti-His *Tb*RAD2 Induction Immunoblot. A resolved 8% SDS-PAGE transferred and immunoblotted with antibody Anti-His, showing the amount of expression under different induction conditions. (A) incubated at 30°C after induction and (B) incubated at 37°C once induced. Exposure for (A) was 450 seconds and (B) for 484 seconds. Imaging of the ladder was merged with the immunoblot for reference, ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

4.5 Purification of His-Tagged *Tb*RAD2

Following the identification of the optimum expression conditions for pAJ100 the 6xHis-tagged *Tb*RAD2 was purified using a Ni-NTA spin column to remove *E. coli* contaminants. The *Tb*RAD2 was initially purified under native conditions, however, as seen in lane 3 of **Figure 4.10** there was insufficient soluble protein to purify. This preliminary data showed that the His-tagged *Tb*RAD2 was insoluble under native conditions.

A further purification experiment was carried out using 8 M urea denaturing conditions. As seen in **Figure 4.11** there is a large level of protein at 110 kDa present in the lysate fraction, but this again appears to be lost in the soluble fraction. This indicates that the induced protein was predominantly insoluble in 8 M urea. Therefore, a stronger 6 M guanidine hydrochloride denaturing condition was used to lyse the induced pAJ100 *E. coli* cell line.

Figure 4.12 shows the purification results of *Tb*RAD2 under 6 M guanidine hydrochloride denaturing lysis conditions. The 6 M guanidine hydrochloride is a very strong denaturant which breaks the internal bonds of the protein and causes solubility of hydrophobic molecules of the protein. It can be seen that the induced 110 kDa protein is present in the soluble fraction. Therefore, the 6 M guanidine hydrochloride lysis buffer was able solubilise the protein. The 110 kDa protein appears to become purified in the eluate fractions following the use of the IMAC column, these bands appear with a weak intensity which may be due to some of the 110 kDa protein being lost in the flow through fraction.

Once the small scale purifications were complete, larger scale purifications could be planned. However, due to the stringent denaturing conditions required for *Tb*RAD2. It was decided it would be unsuitable to carry out large scale purification.

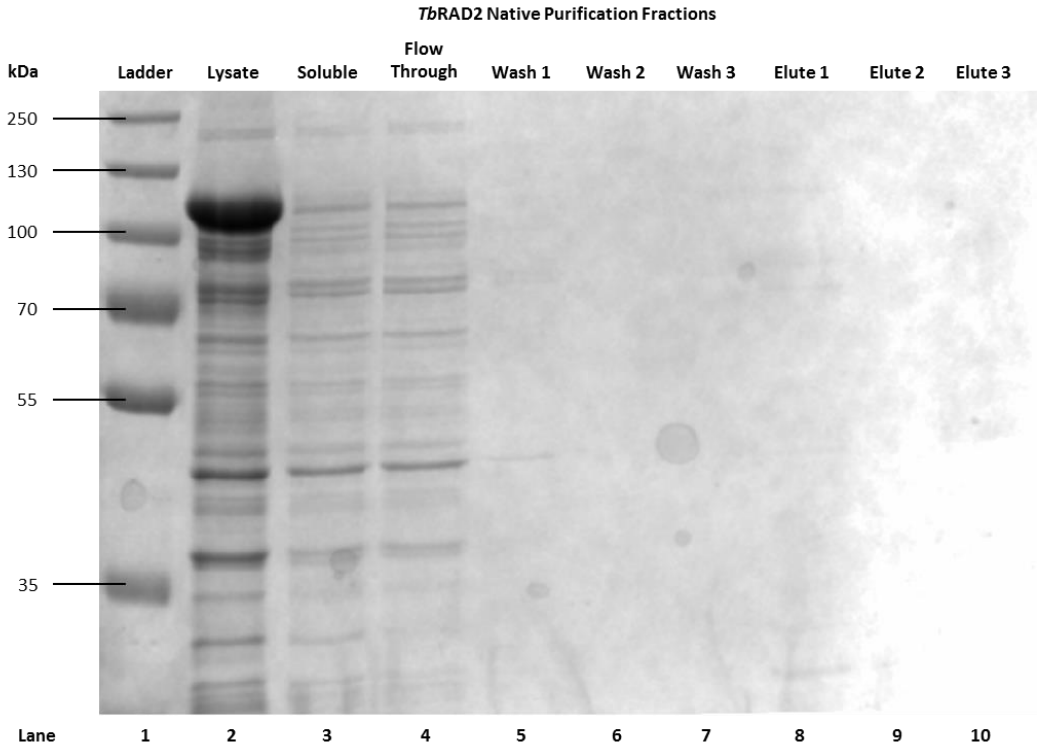


Figure 4.10. SDS-PAGE analysis if native purification of *Tb*RAD2. Resolved 8% SDS-PAGE showing the purification of *Tb*RAD2 by Ni-NTA spin column under native conditions. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

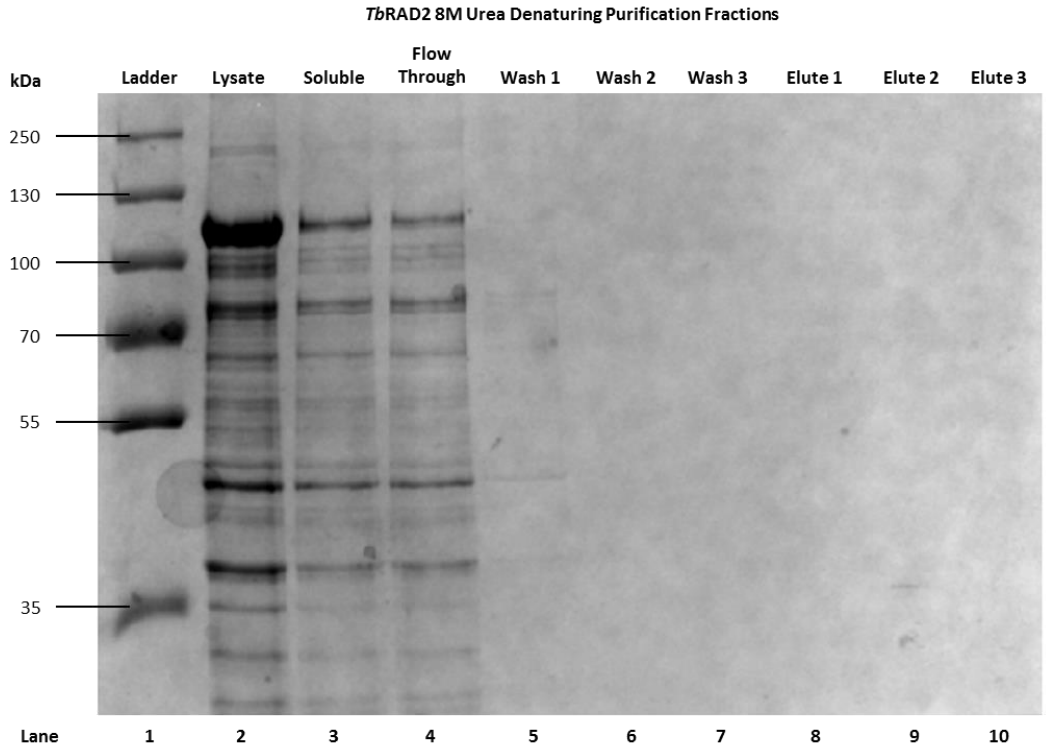


Figure 4.11. SDS-PAGE to Show 8M Urea Denaturing Purification *Tb*RAD2. Resolved 8% SDS-PAGE showing the purification of *Tb*RAD2 by Ni-NTA spin column, under 8M Urea denaturing conditions. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

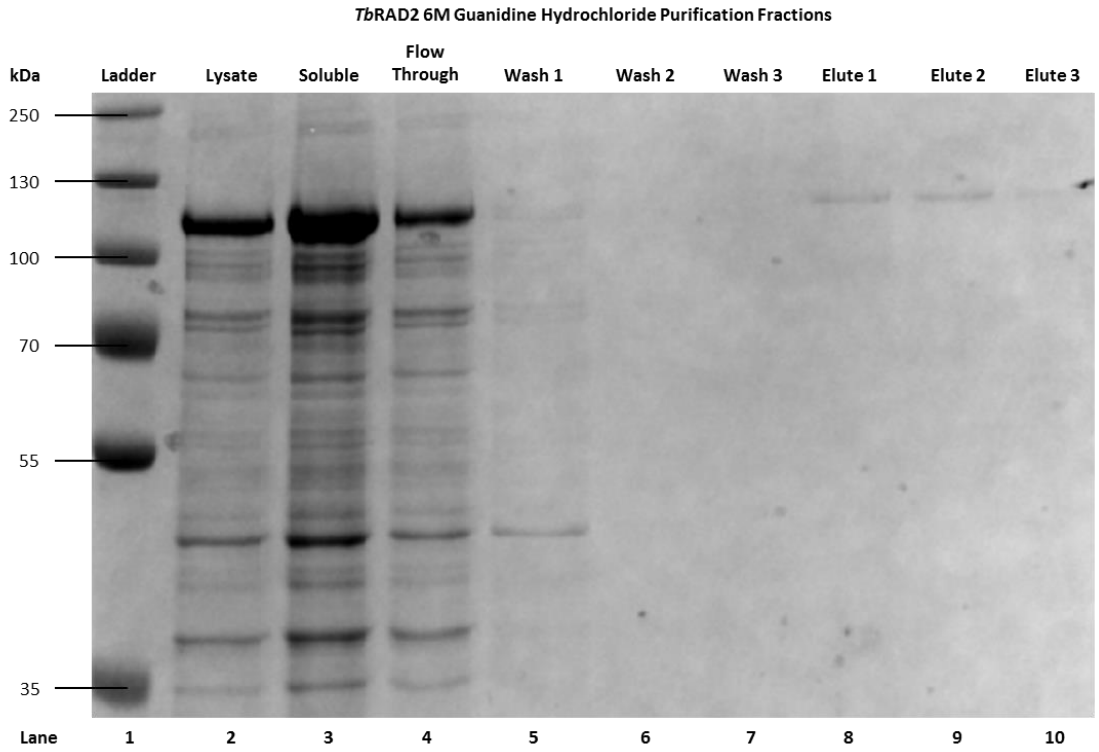


Figure 4.12. SDS-PAGE to Show 6M Guanidine Chloride Denaturing Purification of *Tb*RAD2. Resolved 8% SDS-PAGE showing the purification of *Tb*RAD2 by Ni-NTA spin column, under 6M guanidine hydrochloride denaturing conditions. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

4.6 Anti-*Tb*FEN1 Antibody Production

In order to generate antibodies to *Tb*FEN1, an aliquot of protein purified from a large scale purification, was used to generate polyclonal antibodies.

The protein was sent to Eurogentec for anti-sera production using their 'Speedy 28-day Polyclonal Antibody Programme'. Two rabbits were used for antisera production, each was injected with approximately 200 μg of recombinant *Tb*FEN1 0, 7, 10 and 18 days following commencement of the procedure. Three bleeds were collected, including a pre-inoculation bleed prior to exposure to the *Tb*FEN1 antigen, and further bleeds at 18 and 28 days post exposure.

4.7 Characterisation of anti-*Tb*FEN1 antisera

Polyclonal antisera were raised against *Tb*FEN1, the antisera was raised to investigate interactions between co-expressed proteins in *E. coli* as well as to investigate the expression and localisation of the protein in procyclic and bloodstream trypanosome cells.

In order to characterize the polyclonal antibodies raised against the *Tb*FEN1 protein, a series of immunoblot experiments were carried out to test specificity and sensitivity of antisera. These tests were carried out against *E. coli* extracts containing expressed *Tb*FEN1 and purified *Tb*FEN1 protein. Parallel experiments were performed with pre-inoculation antisera. A range of concentrations were used for both the *E. coli* extract and recombinant protein immunoblots.

Specificity of the antisera was investigated using an immunoblot of the extracts prepared from the induction of *Tb*FEN1 in *E. coli*. Fractions were obtained from IMAC purification of the His-tagged *Tb*FEN1. An SDS-PAGE of small scale *Tb*FEN1 purification is seen in **Figure 4.13**, it can clearly be seen that a band at 48 kDa is purified from the other proteins. Where all contaminant proteins are seen in the cell lysate, soluble and flow-through fractions, and the 48 kDa band can be seen to be mostly purified in the lanes showing the eluates. It is expected that this is the His-tagged *Tb*FEN1 protein as it has been purified using the nickel column and is similar to the predicted 6xHis-tagged weight of 45.1 kDa, as calculated using the ExPASY pl/Mw tool.

Figure 4.14 shows the 1:5,000 of the final bleed antisera immunoblots for rabbit 308 (**A**) and 309 (**B**) against extracts taken during the purification of His-tagged *Tb*FEN1 from

induced *E. coli*. For both gels it can clearly be seen that there is a band at 48 kDa being purified from the cell lysate to the eluates, as seen in **Figure 4.13**. Although the detection of the 48 kDa band does not prove that the antisera has direct specificity to *Tb*FEN1, it does show that both rabbit antisera are able to detect the *Tb*FEN1 protein.

Along with the detection of the 48 kDa band in **Figure 4.14**, there is also clear detection of smaller bands than 48 kDa in lanes 2, 3, 4 and 8. For the whole cell lysate fraction in lane 2, there is greatest complex banding pattern, it is likely that these smaller bands are insoluble degradation products of the recombinant protein.

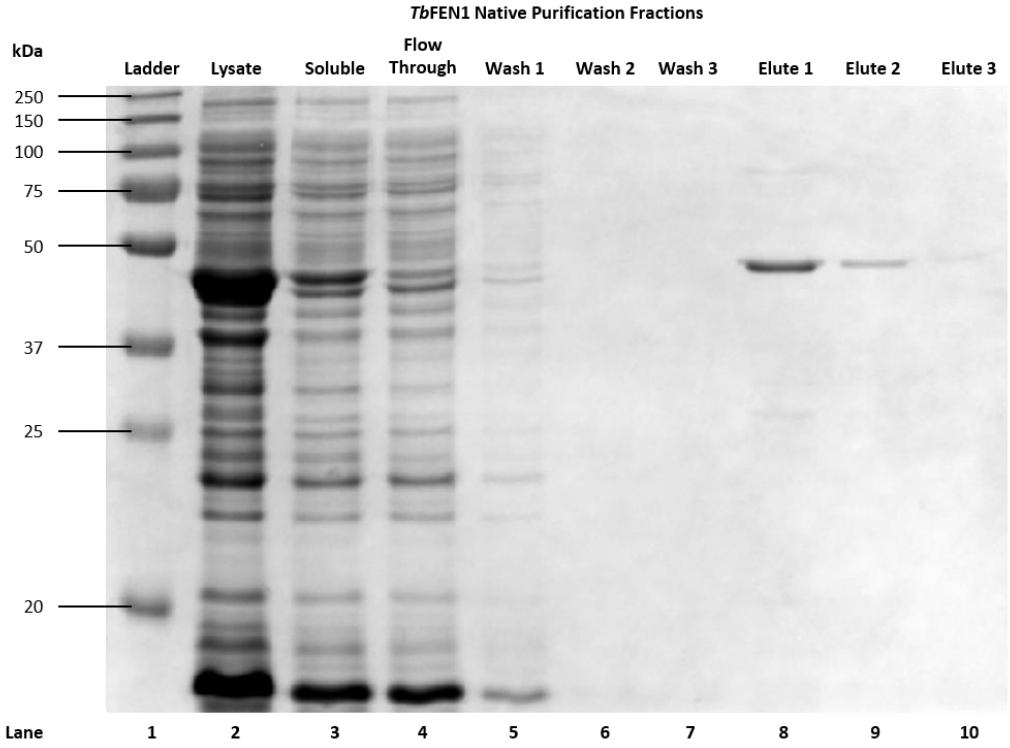


Figure 4.13. SDS-PAGE analysis of native IMAC purification of *Tb*FEN1 Resolved 12% SDS-PAGE showing the purification of *Tb*FEN1 by Ni-NTA spin column under native conditions. Ladder used – Precision plus protein™ All blue prestained proteins standards 10-250kDa protein ladder, BioRad.

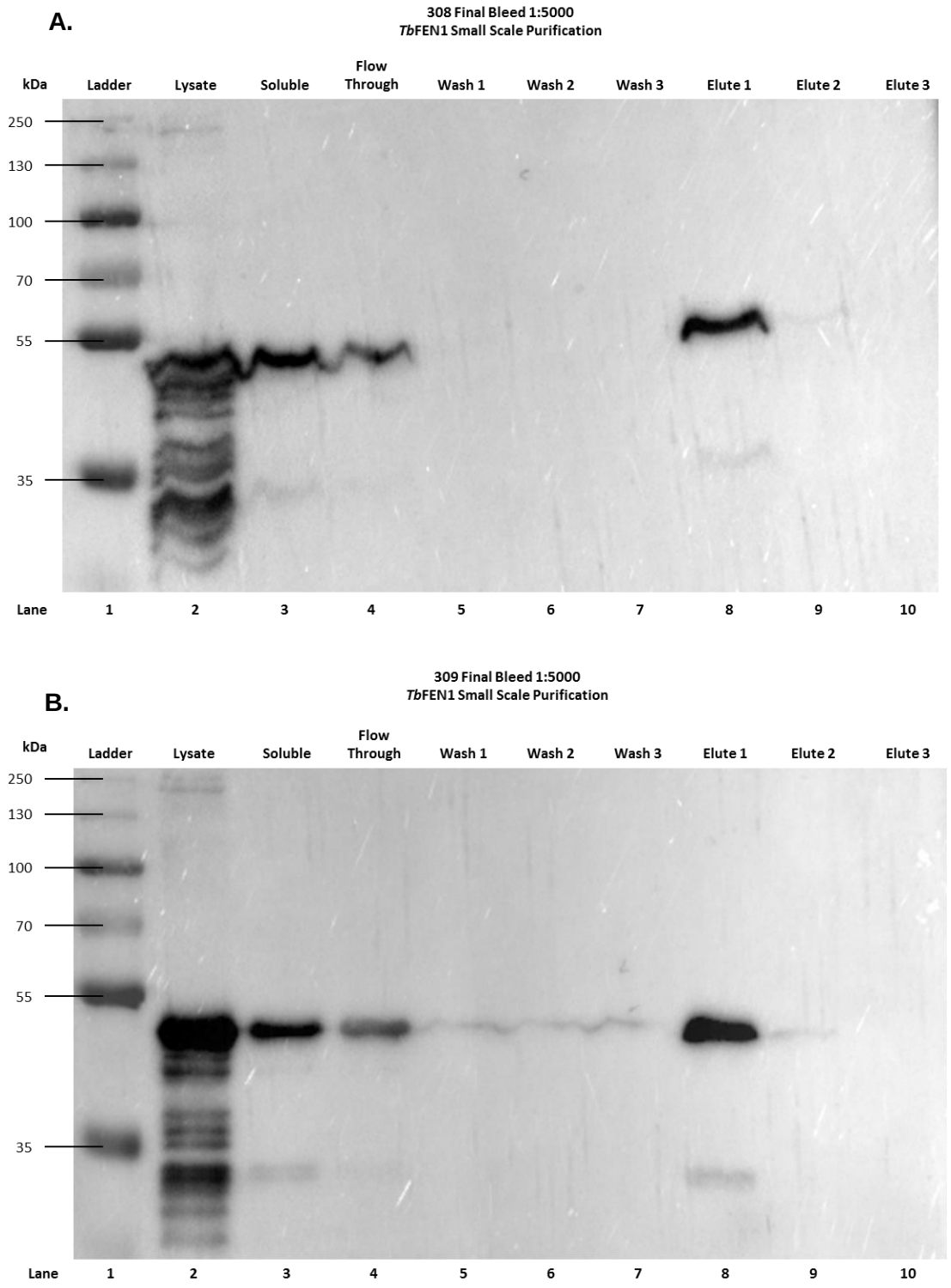


Figure 4.14. Immunoblot identification of rabbit 308 and 309 antisera against fractions collected during purification of *Tb*FEN1. Developed immunoblots using the 1:5,000 final bleed polyclonal anti-sera immunisation rabbit 308 (a) and rabbit 309 (b). Immunoblots were resolved with fractions from the small scale purification of the *Tb*FEN1 protein from induced *E. coli* extract. Blots were developed for 10 and 11 seconds respectively. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

Figure 4.15 shows the 1:5000 final bleed antisera immunoblot against different concentrations of the purified recombinant protein, for rabbits 308 and 309. In the figure, it can be seen that there is recognition of a protein at 48 kDa from 5 µg to 10 ng. Therefore, from this result both antisera have a detection limit of 10 ng of *Tb*FEN1 presence, on an immunoblot.

Again, in **Figure 4.15** bands can be seen at a smaller size than 48 kDa. In comparison **Figure 4.14**, these cannot be due to insoluble degradation as the recombinant protein was originally purified from a soluble extract. These degradation products are generated in the latter stages of purification or as biproduct of storage.

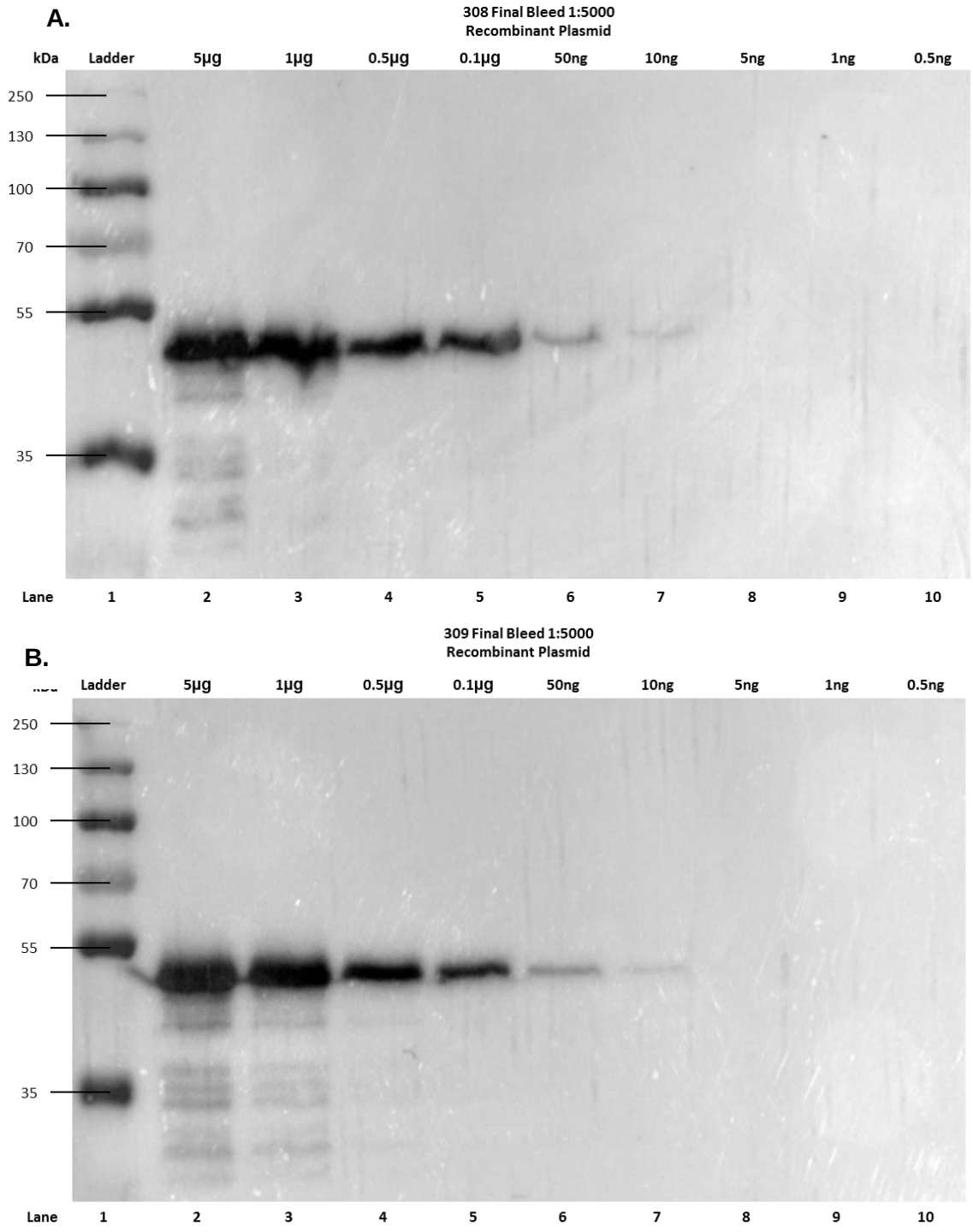


Figure 4.15. Immunoblot to show rabbit 308 and 309 antisera against different concentrations of recombinant *Tb*FEN1 Developed immunoblots using the 1:5,000 final bleed polyclonal antisera immunisation of rabbit 308 (a) and rabbit 309 (b), against different concentrations of the recombinant protein *Tb*FEN1. Immunoblots were resolved with the purified recombinant protein at the concentrations 5µg, 1µg, 0.5µg, 0.1µg, 50ng, 10ng, 5ng, 1ng and 0.5ng. These blots were developed for 6 seconds and 3 seconds respectively. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

4.8 Chapter Summary

In this chapter the optimum expression conditions for *Tb*FEN1 and *Tb*RAD2 were used to express the protein ready for purification. The *Tb*FEN1 protein was purified under native conditions, where *Tb*RAD2 was found to be purified under stringent denaturing conditions. As a result, large scale purification was carried out for *Tb*FEN1 only.

Polyclonal antisera was produced for *Tb*FEN1, ready for localisation experiments. Immunoblots were used to interrogate the specificity and sensitivity of the antisera to *Tb*FEN1. The antisera was able to detect the *Tb*FEN1 protein amongst bacterial contaminants and detect the recombinant protein to an amount of 10 ng.

5 Localisation of *Tb*FEN1 and *Tb*RAD2 in *T. brucei*

5.1 Chapter Synopsis

The intracellular localisation of *Tb*FEN1 and *Tb*RAD2 in *T. brucei* was investigated using multiple approaches; (i) expression of *Tb*FEN1::YFP and *Tb*RAD2::YFP fusion proteins from endogenous gene loci, (ii) inducible over-expression of *Tb*FEN1 and *Tb*RAD2 fused to a small epitope tag; GSP, and (iii) polyclonal antisera specific for *Tb*FEN1.

5.2 Localisation of YFP-Tagged *Tb*FEN1 and *Tb*RAD2

To investigate the localisation of *Tb*FEN1 and *Tb*RAD2, the open reading frames (ORF) encoding the two proteins were tagged using a PCR only tagging (pPOT) approach.

As seen in **Figure 5.1** the DNA sequence encoding YFP and a blasticidin cassette were PCR amplified from the pPOT vector using long (~100bp) primers that incorporated a 5' overhang of 80 nucleotides with homology to the target gene and its adjacent untranslated region (UTR). The resulting amplicon was then used to transfect *T. brucei*, the amplicon combines into the genome in such a way that the YFP tag is fused in frame with the target gene. Recombinant cells were then selected using blasticidin.

For N-terminal tagging of genes, the forward primer is complementary to the last 80bp of the 5' untranslated region (UTR) of the gene, where the reverse primer is complementary to the first 80bp of the ORF of the reverse complement (Dean et al., 2015).

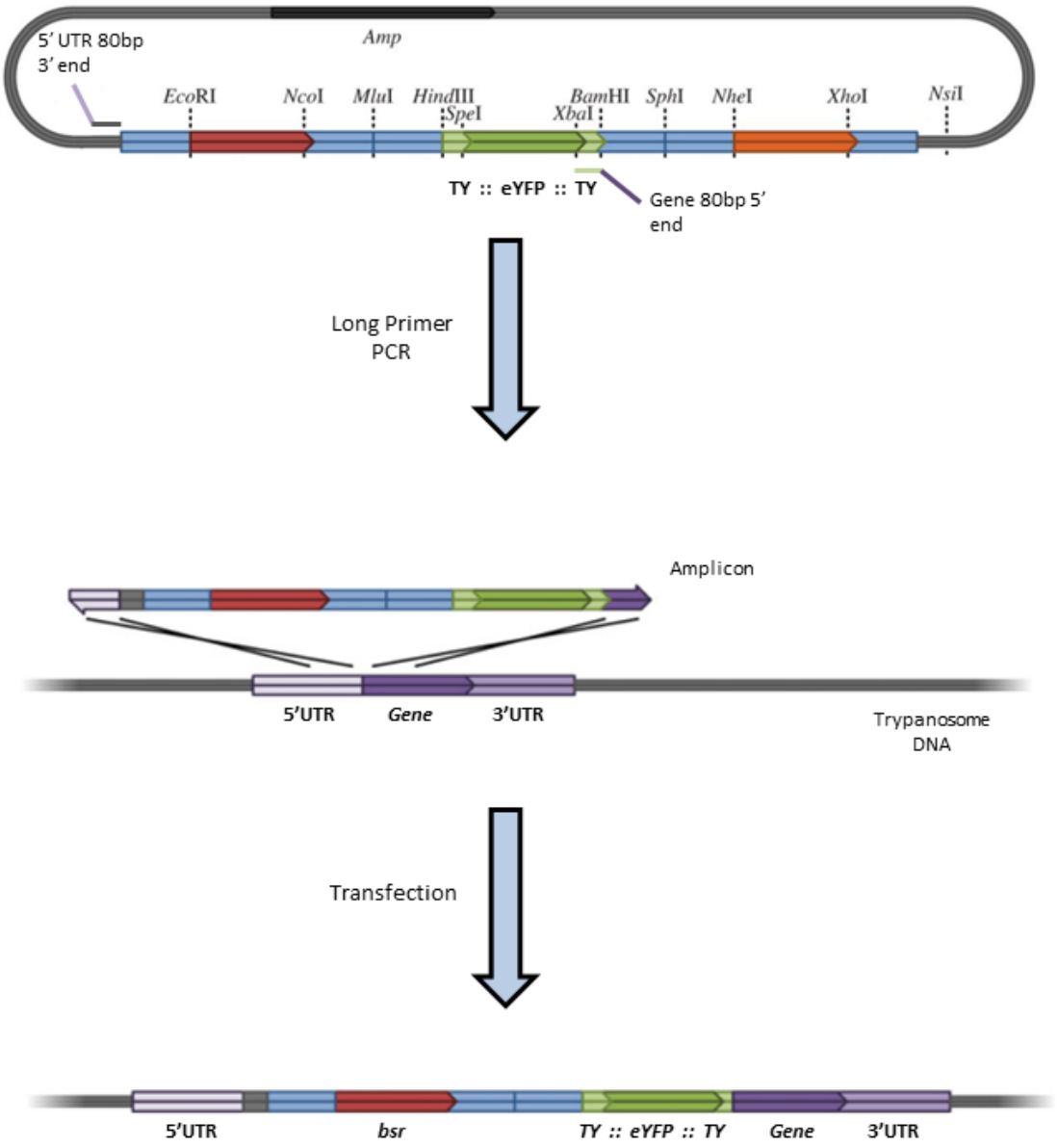


Figure 5.1. Process of pPOT Tagging. Schematic showing the process of pPOT tagging of target genes in the trypanosome genome. Based on (Dean et al., 2015).

Fluorescence microscopy was used to investigate the localisation of YFP-tagged *Tb*FEN1 and *Tb*RAD2. **Figures 5.2** and **5.3** show the results of this specific YFP tagging, and how the localisations of the fusion proteins change over the course of the cell cycle.

*Tb*FEN1: YFP::*Tb*FEN1 can be clearly visualised within the nucleus throughout the cell cycle. The intensity of the signal is not uniform throughout the nucleus and appears to have a specific point of intensity (**Figure 5.2**). The YFP::*Tb*FEN1 signal appears to elongate as the cells move from 1N1K to late 1N2K. As the *T. brucei* cells come to the end of the telophase stage of replication, a 'strand' of YFP::*Tb*FEN1 fluorescence can be seen connecting the two nuclei together. As the YFP-tagged *Tb*FEN1 signal elongates during mitosis there is a clear connecting strand between the two nuclei, *Tb*FEN1 appears to be associating with DNA that is being segregated along the mitotic spindles.

*Tb*RAD2: The YFP::*Tb*RAD2 fluorescence signal is much weaker (**Figure 5.3**) than that observed for YFP::*Tb*FEN1. Despite the weaker signal, the YFP tagged *Tb*RAD2 protein appears to be present throughout the cell cycle and can be observed within both the nucleus and a distinct foci within the cytoplasm positioned between the nucleus and kinetoplast. Similar to *Tb*FEN1, the nuclear YFP::*Tb*RAD2 fluorescence signal becomes elongated as the cell prepares for mitosis. The presence of the YFP::*Tb*RAD2 protein in both the nuclei and cytoplasm suggests that the recombinant protein may not be being correctly targeted within the cell.

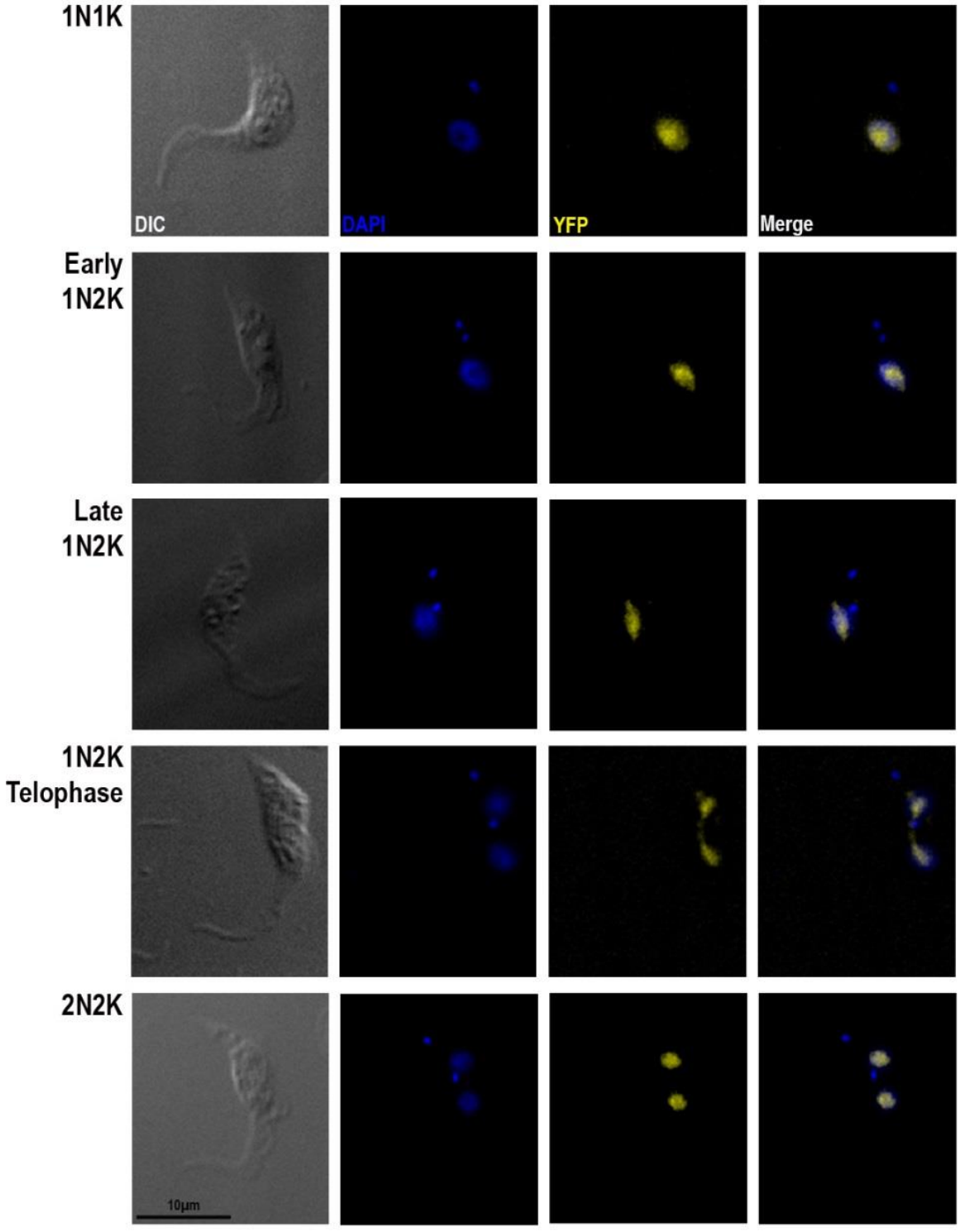


Figure 5.2. Localisation of YFP::*Tb*FEN1 in *T. brucei*. Fluorescent microscopy images of procyclic cells expressing YFP::*Tb*FEN1 protein at different stages of the cell cycle. Captured using Deltavison Denconvolution microscopy. Scale bar - 10µm.

.

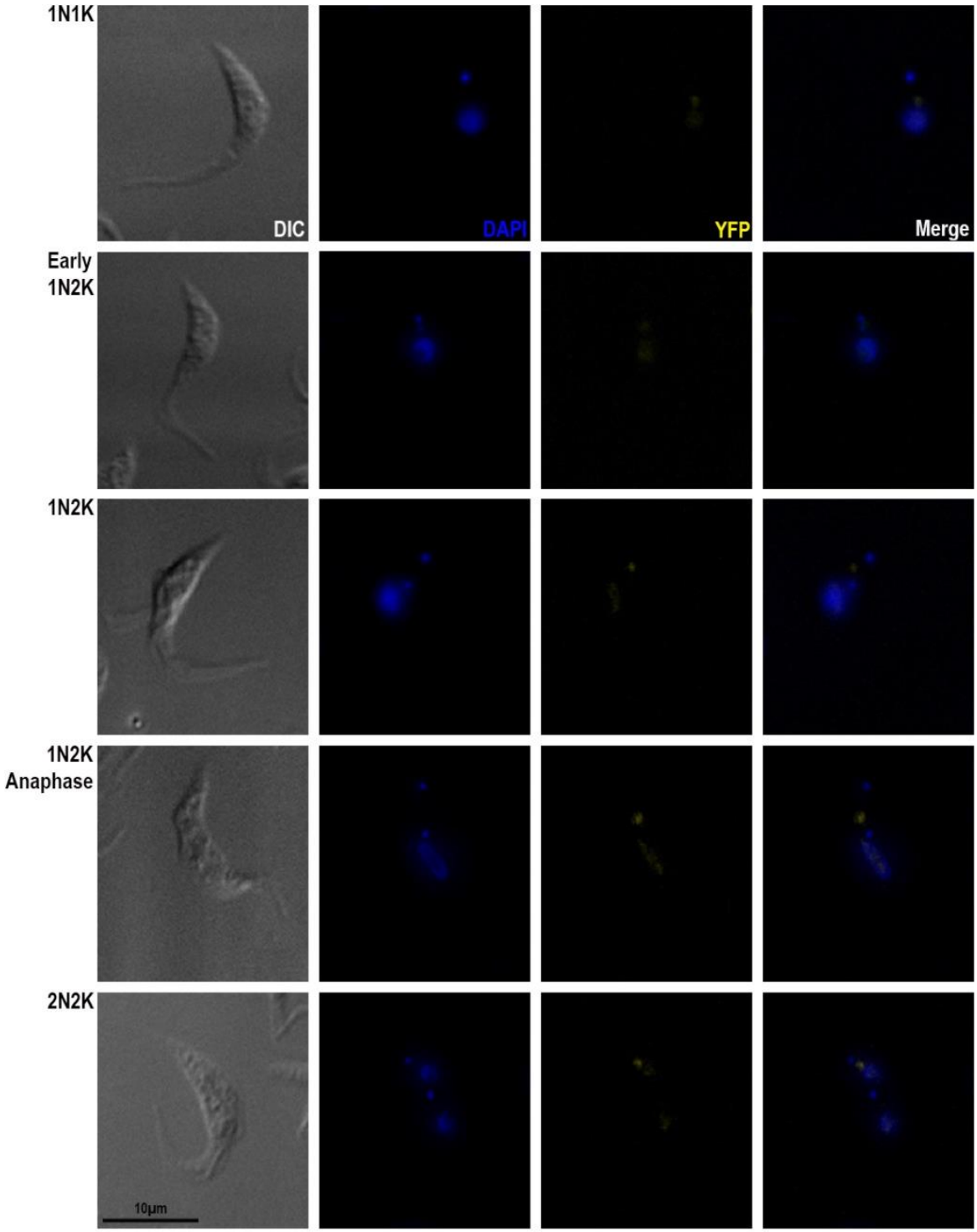


Figure 5.3. Localisation of YFP::*Tb*RAD2 in *T. brucei*. Fluorescent microscopy images of procyclic cells expressing YFP::*Tb*RAD2 protein at different stages of the cell cycle. Captured using Deltavison Denconvolution microscopy. Scale bar - 10µm.

5.3 Localisation of GSP epitope tagged *Tb*FEN1 and *Tb*RAD2

Further analysis of the localisation of the parasite proteins was carried out by expressing both proteins fused to a small epitope tag (a nine amino acid sequence coding for the amino acids GAFSIMPAM (GSP)) at the C-terminus. The epitope can be detected using anti-GSP monoclonal antibody DB9.

A small epitope tag was used as it could potentially present a more native localisation pattern, compared to the localisation observed for the larger YFP tagged proteins used previously. The epitope tag was also encoded at the C-terminus instead of the N-terminus (as was the case for the YFP tagged proteins), allowing for investigation into how this may also affect targeting of the proteins within the cell. It was hoped that that the *Tb*RAD2::GSP fusion protein may be targeted to a more precise location within the cell than that observed for the YFP tagged protein. The fusion proteins were also over-expressed by induction, whereas the YFP tagged proteins were expressed from endogenous loci. It was anticipated that by over-expressing the proteins a stronger fluorescence pattern would be seen.

To create the GSP fusion proteins, *Tb*FEN1 and *Tb*RAD2 ORFs were amplified by PCR (**Figure 5.4 A** and **B)**; the primers used added a 5' *Hind*III and 3' *Xho*I restriction site to facilitate cloning of the ORFs into the *T. brucei* expression vector pDEX377. The PCR products were initially ligated into the vector pGEM-T Easy and positive transformants identified by restriction digestion with *Hind*III and *Xho*I; an example of positive pGEMT Easy clones for both *Tb*FEN1 and *Tb*RAD2 are shown in **Figure 5.4 C** and **D**. These plasmids were sent for nucleotide sequence analysis and the conceptual translation of both sequences compared to wild type sequences available at GeneDB (**Figure 5.5**). The alignment of the cloned sequences with the wild type sequence reveals that the *Tb*FEN1 sequence is identical to the Tb 927 genome reference strain (**A**) sequence, however, the *Tb*RAD2 sequence in the pGEMT clone differed from both Tb 927 and Tb 427 sequences(**B** and **C**).

DNA sequences encoding both *Tb*FEN1 and *Tb*RAD2 were digested from the pGEM-T Easy plasmids and subcloned into pDEX377_GSP in which the ORFs were fused in frame to the nine amino acid sequence GSP tag. Positive transformants were again identified by restriction digestion, with the excised fragments migrating with the predicted molecular sizes for *Tb*FEN1 and *Tb*RAD2 (~1,182bp and ~2,241 respectively) (**Figure 5.4 F**).

The resulting plasmids were prepared for transfection into *T. brucei* by linearising with NotI. It can be seen in **Figure 5.4 G** that the recombinant vectors were completely digested and as expected the *Tb*RAD2::GSP DNA fragment was larger than *Tb*FEN1::GSP. Digests were ethanol precipitated and transfected into *T. brucei* and drug resistant cell lines established. Cells were then induced with doxycycline to enable expression of *Tb*FEN1::GSP and *Tb*RAD2::GSP fusion proteins production and cells prepared for immunofluorescence microscopy. **Figure 5.6** and **5.7** show the localisation of *Tb*FEN1::GSP and *Tb*RAD2::GSP respectively.

The inducible expression of *Tb*FEN1::GSP in procyclic cells (**Figure 5.6**) identified a similar nuclear localisation pattern to that observed for YFP::*Tb*FEN1. *Tb*FEN1::GSP localisation was restricted to the nucleus and during mitosis exhibited the same elongated localisation along the axis of the mitotic spindle between dividing nuclei. However, in comparison to YFP::TbFEN1, which exhibited consistent level of expression during the cell cycle, *Tb*FEN1::GSP expression appeared to be elevated in nuclei during mitosis. The increased level of fluorescence of the *Tb*FEN1 protein in the nucleus during mitosis, indicates that the protein is being recruited and therefore may have a functional role during nuclear replication.

Localisation of the *Tb*RAD2::GSP fusion protein (**Figure 5.7**) failed to provide consistent results, with only weak signal detected in the nucleus. Although there was evidence for weak expression of *Tb*RAD2::GSP in the cell lines, the protein appeared to be diffusely present throughout the cytoplasm. This suggests that even following over-expression targeting of the protein was inefficient.

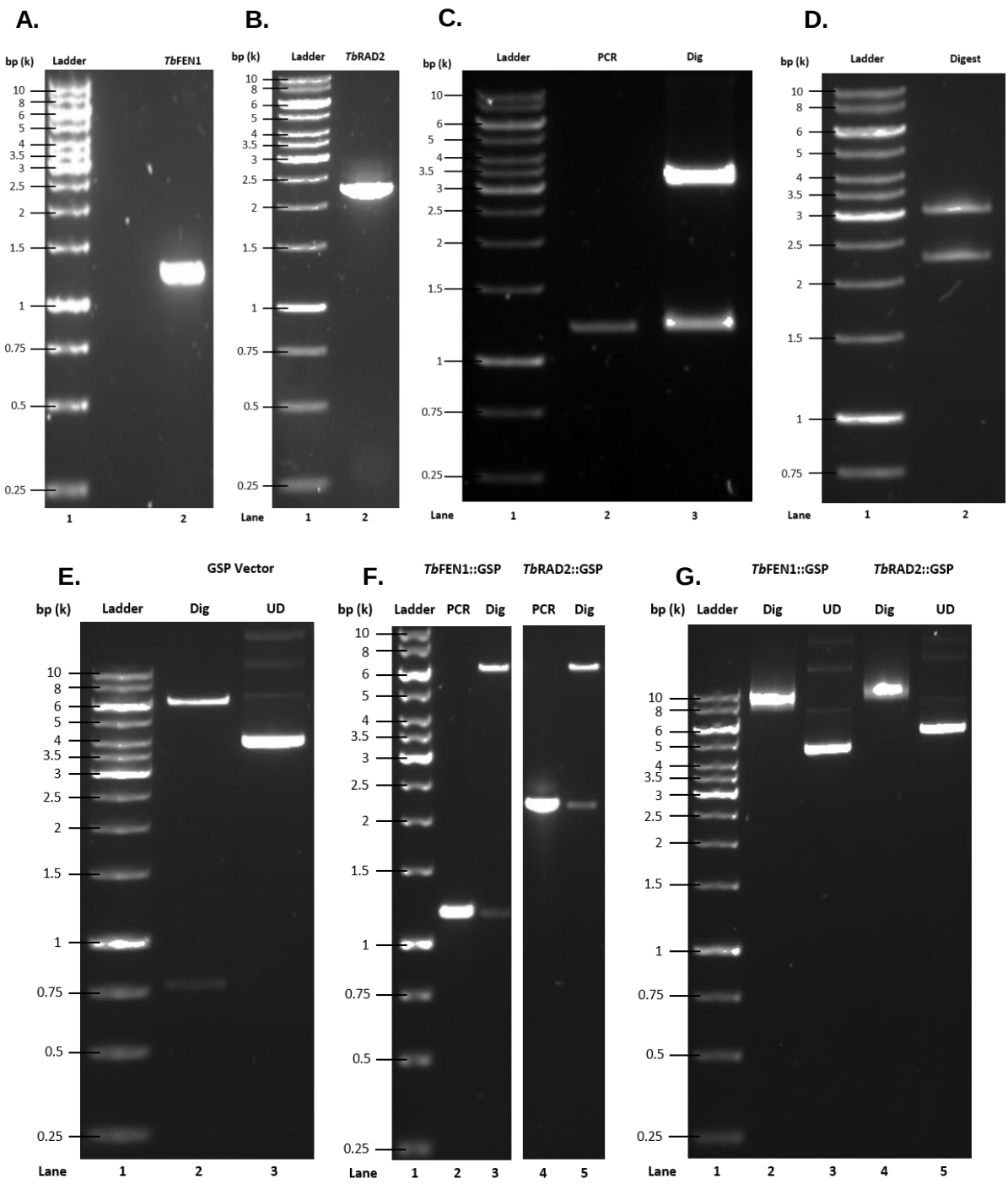


Figure 5.4. Agarose gels showing *Tb*FEN1::GSP and *Tb*RAD2::GSP preparation for transfection. Resolved 0.8% agarose gels showing different stages of preparation for the transfection of *Tb*FEN1::GSP and *Tb*RAD2::GSP into *T. brucei*. (A) Shows the size the amplified PCR product *Tb*FEN1 with a 5' *Hind*III and 3; *Xho*I, (B) Shows the size of the amplified PCR product *Tb*RAD2 with a 5' *Hind*III and a 3' *Xho*I, (C) *Tb*FEN1 PCR product compared to digested (Dig) pGEM-T Easy::*Tb*FEN1 with *Hind*III and *Xho*I of the pGEM-T Easy vector, (D) pGEM-T Easy::*Tb*RAD2 digested with *Hind*III and *Xho*I (E) Preparation of the GSP vector for ligation, with digestion by *Hind*III and *Xho*I (F) Comparison of PCR product with the trypanosome gene and GSP ligated recombinant vector, (G) Complete linearisation of recombinant vector with *Not*I compared to UD. Ladder used – GeneRuler™ 1kb DNA ladder, ThermoScientific.

A. *Tb*927.3.830 – Putative FEN1, *Trypanosoma brucei* TREU927

```
TbFEN1      MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE  60
            MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE
927 FEN1    MGVLGLSKLLYDRTPGAIKEQELKVYFGRRIAIDASMAVYQFVIAMKGFQEGQSVELTNE  60

TbFEN1      AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE  120
            AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE
927 FEN1    AGDVTSHLSGIFFRTLRMIDEGLRPIYVFDGKPPTLKASELESRRQRAEDAKHEFEKAKE  120

TbFEN1      EGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKAWAVG  180
            EGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKAWAVG
927 FEN1    EGDDEAMEKMSKRMVRVGRDQMEEVKTLLRLMGIPVVQAPSEAEAQCAELVKKNKAWAVG  180

TbFEN1      TEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP  240
            TEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP
927 FEN1    TEDMDALAFGSRVMLRHLTYGEAKKRPIAEYHLDEILEASGFSMQQFIDLCILLGCDYVP  240

TbFEN1      RISGIGPHKAWEGIKKYGSLEAFIESLDGTRYVVPEEFNYKDARNFFLEPEVTPGEEIDI  300
            RISGIGPHKAWEGIKKYGSLEAFIESLDGTRYVVPEEFNYKDARNFFLEPEVTPGEEIDI
927 FEN1    RISGIGPHKAWEGIKKYGSLEAFIESLDGTRYVVPEEFNYKDARNFFLEPEVTPGEEIDI  300

TbFEN1      QFREPDEEGLIKFLVDEKLFSKERVLKGIQRLRDALTKKTQGRLDQFFTITKPQKQVNSE  360
            QFREPDEEGLIKFLVDEKLFSKERVLKGIQRLRDALTKKTQGRLDQFFTITKPQKQVNSE
927 FEN1    QFREPDEEGLIKFLVDEKLFSKERVLKGIQRLRDALTKKTQGRLDQFFTITKPQKQVNSE  360

TbFEN1      ASTAGTKRNRGAVALPGVLQRKSSSGHKKAVKK  393
            ASTAGTKRNRGAVALPGVLQRKSSSGHKKAVKK
927 FEN1    ASTAGTKRNRGAVALPGVLQRKSSSGHKKAVKK  393
```

B. *Tb*427tmp.211.2870 – Putative 5'*Hind*III RAD2, *Trypanosoma brucei* Lister strain 427

```
TbRad2      MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60
            MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI
427 Rad2    MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60

TbRad2      LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120
            LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV
427 Rad2    LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120

TbRad2      HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE  180
            HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKR RE
427 Rad2    HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRXRE  180

TbRad2      VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240
            VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA
427 Rad2    VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240

TbRad2      LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300
            LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD
427 Rad2    LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300

TbRad2      NAIKSDPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD  360
            NAIK DPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD
427 Rad2    NAIKXDPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD  360

TbRad2      TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420
            TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR
427 Rad2    TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420

TbRad2      KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480
            KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY
427 Rad2    KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480

TbRad2      VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540
            VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC
427 Rad2    VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540

TbRad2      GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600
            GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC
427 Rad2    GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600

TbRad2      SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA  660
            SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCA
427 Rad2    SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAX  660

TbRad2      PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720
            PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR
427 Rad2    PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720

TbRad2      ERVIFRKQPPKFSEALSYLRAARGDP  746
            ERVIFRKQPPKFSEALSYLRAARGDP
427 Rad2    ERVIFRKQPPKFSEALSYLRAARGDP  746
```

C. *Tb*927.9.11760 – Putative 5'*Hind*III RAD2, *Trypanosoma brucei* TREU927

```
TbRad2      MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60
            MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI
927 Rad2    MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60

TbRad2      LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120
            LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV
927 Rad2    LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120

TbRad2      HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE  180
            HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE
927 Rad2    HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE  180

TbRad2      VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240
            VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA
927 Rad2    VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240

TbRad2      LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300
            LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD
927 Rad2    LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300

TbRad2      NAIKSDPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD  360
            NAIK+DPH VDAFHHNVSFGKEEESTSDEVEVLS GDYWSCADNDCDDLLSLAASDRTPD
927 Rad2    NAIKTDPHAVDAFHHNVSFGKEEESTSDEVEVLSSGDYWSCADNDCDDLLSLAASDRTPD  360

TbRad2      TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420
            TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR
927 Rad2    TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420

TbRad2      KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480
            KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY
927 Rad2    KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480

TbRad2      VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540
            VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC
927 Rad2    VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540

TbRad2      GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600
            GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC
927 Rad2    GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600

TbRad2      SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA  660
            SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA
927 Rad2    SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA  660

TbRad2      PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720
            PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR
927 Rad2    PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720

TbRad2      ERVIFRKQPPKFSEALSYLRAARGDP  746
            ERVIFRKQPPKFSEALSYLRAARGDP
927 Rad2    ERVIFRKQPPKFSEALSYLRAARGDP  746
```

Figure 5.5. Amplified Sequence Comparison of *Tb*FEN1 and *Tb*RAD2. Alignment results of the sequenced *Tb*FEN1 and *Tb*RAD2 protein against known sequences (A) *T. brucei* TREU 927 (B) *T. brucei* lister strain 427 and (C) *T. brucei* TREU 927. Where red amino-acids are not the same between the two compared sequences, green amino-acids are the same between the compared sequences but not between alignments B and C. The X symbol shows were the amino-acid is not specific in the known sequence. Results obtained by EMBL-EBI's Clustal Omega (Goujon et al., 2010, Sievers et al., 2011, McWilliam et al., 2013).

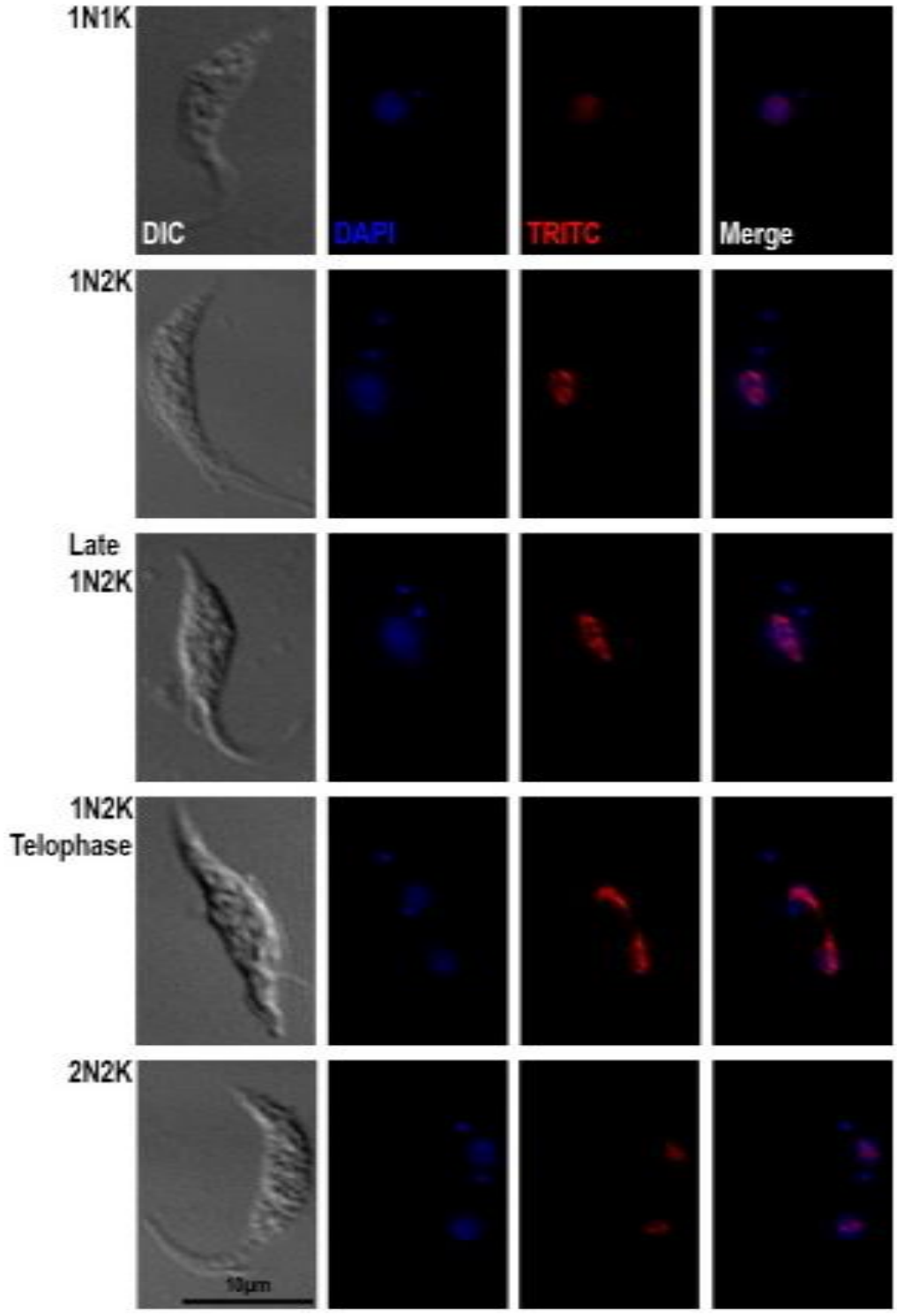


Figure 5.6. Localisation of *Tb*FEN1::GSP in procyclic *T. brucei.* Fluorescent microscopy images of *Tb*FEN1::GSP protein at different stages of the cell cycle. The anti-GSP antibody DB9 and rhodamine conjugated anti-mouse IgG secondary antibody were used to detect *Tb*FEN1::GSP protein, images were captured using the Deltavison Denconvolution microscopy. Scale bar - 10µm.

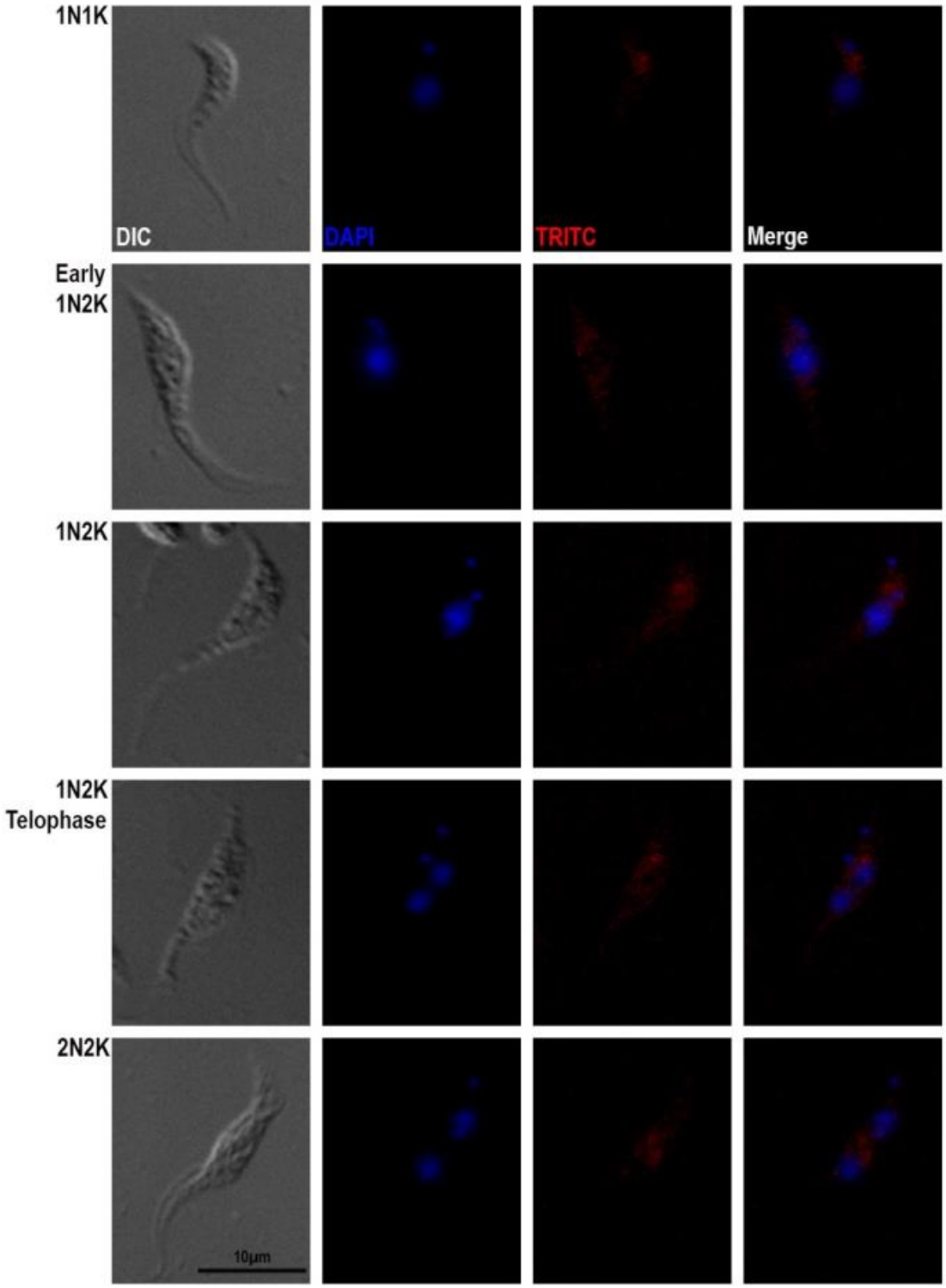


Figure 5.7. Localisation of *Tb*RAD2::GSP in *T. brucei.* Fluorescent microscopy images of the labelled *Tb*RAD2::GSP protein at different stages of the cell cycle. The anti-GSP antibody DB9 and rhodamine conjugated anti-mouse IgG secondary antibody were used to detect *Tb*RAD2::GSP protein, images were captured using the Deltavison Denconvolution microscopy. Scale bar - 10µm.

5.4 Validation of the anti-*Tb*FEN1 polyclonal antibody

The *Tb*FEN1 antibody described in **Chapter 4** was used to examine *T. brucei* cells prior to carrying out immunofluorescence experiments, immunoblot analysis was carried out to investigate the antibody's specificity to native *Tb*FEN1 in parasite cell lines.

Figures 5.8 and **5.9** show the immunoblot results for rabbits 308 and 309 respectively, using immunoblots resolved with different strains of *T. brucei*. Strains used were procyclic 427, PTP and cells expressing *Tb*FEN1::GSP and also bloodstream 427 cells. The 308 pre-immune antiserum produced a non-specific complex pattern of interaction with parasite proteins for each *T. brucei* strain. In the final bleed at both dilutions of 1:1,000 and 1:5,000, the non-specific banding pattern was reduced and a band corresponding to the TbFEN1 protein was clearly detected. In contrast to antiserum from rabbit 308, the 309 pre-immune antiserum showed no non-specific reactivity to *T. brucei* proteins. However, the final bleed at both 1:1,000 and 1:5,000 also showed strong reactivity to a protein band of the predicted MW for *Tb*FEN1, along with proteins of larger molecular mass. Although detection of native *Tb*FEN1 was weak, final bleed antisera from both rabbits showed strong reactivity towards over-expressed epitope tagged *Tb*FEN1::GSP protein.

Having demonstrated that the polyclonal antiserum could detect both native *Tb*FEN1 and *Tb*FEN1::GSP in parasite lysates, the anti-sera was used to localise native and over-expressed *Tb*FEN1::GSP by immunofluorescence microscopy.

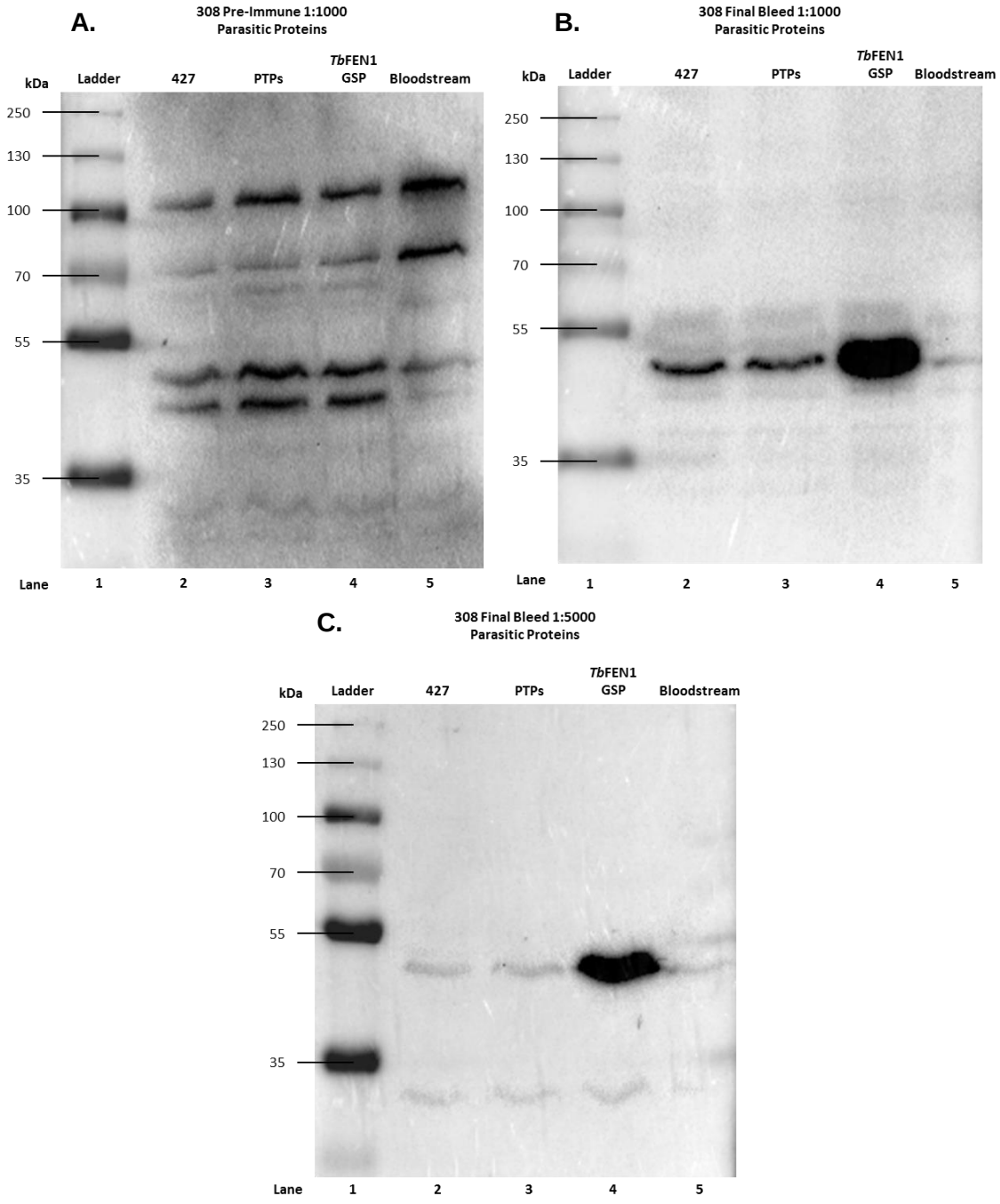


Figure 5.8. Rabbit 308 Parasite Protein Immunoblot. Developed immunoblots using the polyclonal anti-sera immunisation against the protein *Tb*FEN1 in Rabbit 308, immunoblots were resolved with the parasitic proteins from the *T. brucei* strains procyclic 427, procyclic PTPs, induced *Tb*FEN1 and bloodstream 427. Blots (A) used the rabbit pre-immune sera, (B) diluted 1:1,000 final bleed sera and (C) diluted 1:5,000 final bleed, blots were developed for 200 seconds, 100 seconds and 300 seconds respectively. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

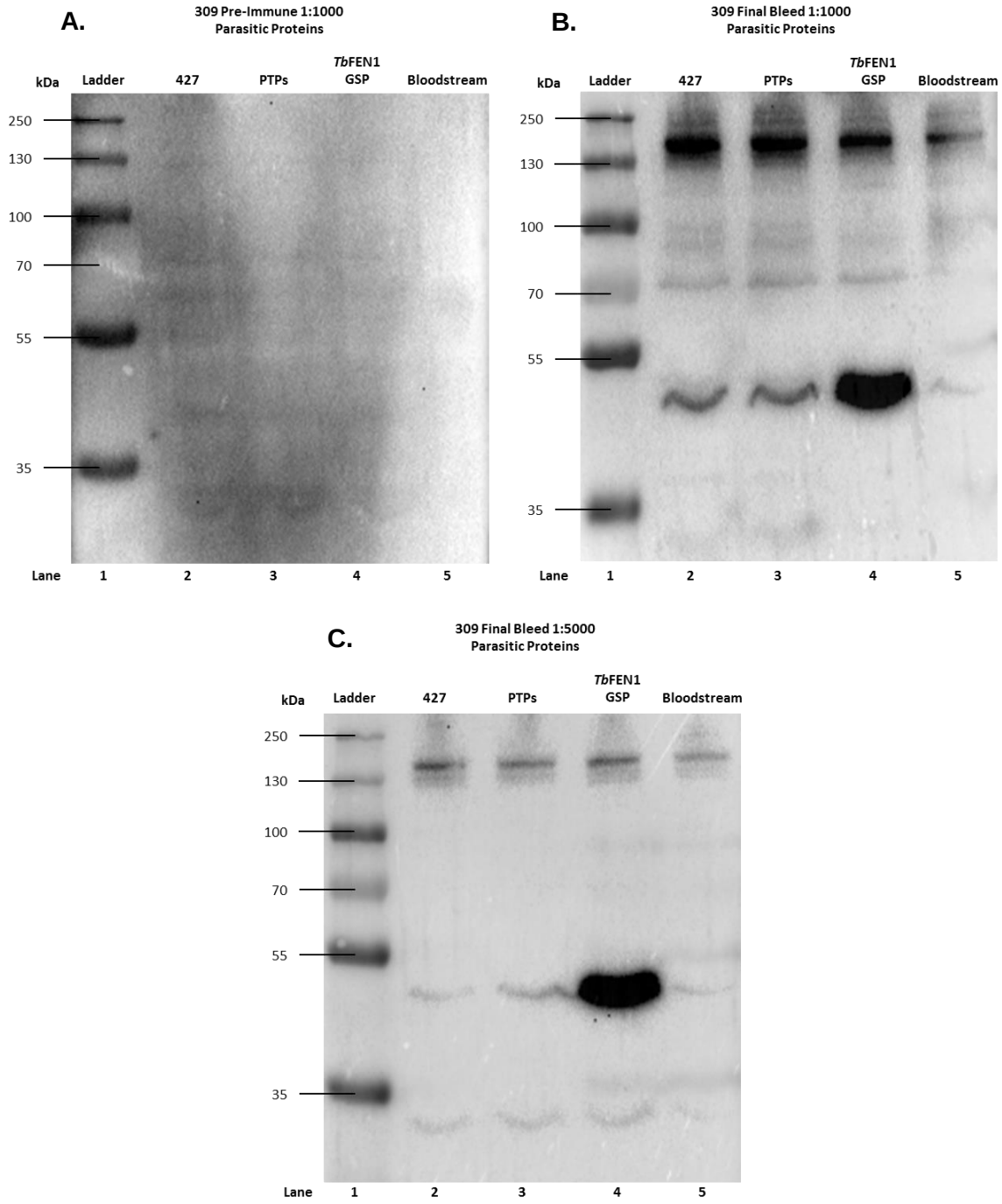


Figure 5.9. Rabbit 309 Parasite Protein Immunoblot. Developed immunoblots using the polyclonal anti-sera immunisation against the protein *Tb*FEN1 in Rabbit 309, immunoblots were resolved with the parasitic proteins from the *T. brucei* strains procyclic 427, procyclic PTPs, induced *Tb*FEN1 and bloodstream 427. Blots (A) used the rabbit pre-immune sera, (B) diluted 1:1,000 final bleed sera and (C) diluted 1:5,000 final bleed, blots were developed for 200 seconds, 100 seconds and 500 seconds respectively. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

5.5 Localisation of *Tb*FEN1 using anti-*Tb*FEN1 antiserum

The localisation of *Tb*FEN1 was investigated using anti-*Tb*FEN1 antibodies, and as a control the antisera was also used to probe cells that were overexpressing *Tb*FEN1::GSP.

Since the immunoblotting analysis had shown that rabbit 308 showed a high degree of non-specific recognition against *T. brucei* proteins, experiments were undertaken to optimise the dilution of the antisera used in the immunofluorescence experiments. As the antisera is diluted it is expected that non-specific reactivity in the serum will be reduced and that the stronger antigen specific reactivity will be seen.

Figures 5.10 and **5.11** show the immunofluorescence results for rabbits 308 and 309 against procyclic 427 cells, where pre-immune sera was used at 1:1,000 dilution and antisera from the final bleed at 1:5,000. The results of additional dilutions are presented in the **Appendix 2-5**.

As seen in **Figure 5.10** and **Figure 5.11**, the 308 and 309 pre-immune bleeds don't provide a clear localisation pattern to any parasitic proteins. For both 308 and 309 final bleed antisera there was again no clear specific reactivity to the parasite proteins, indicating the expression of *Tb*FEN1 is below the level of detection for both antisera. In the **Appendix 2** and **4** both dilutions at 1:200 show final bleed reactivity for both rabbits, however, there is no distinct specificity.

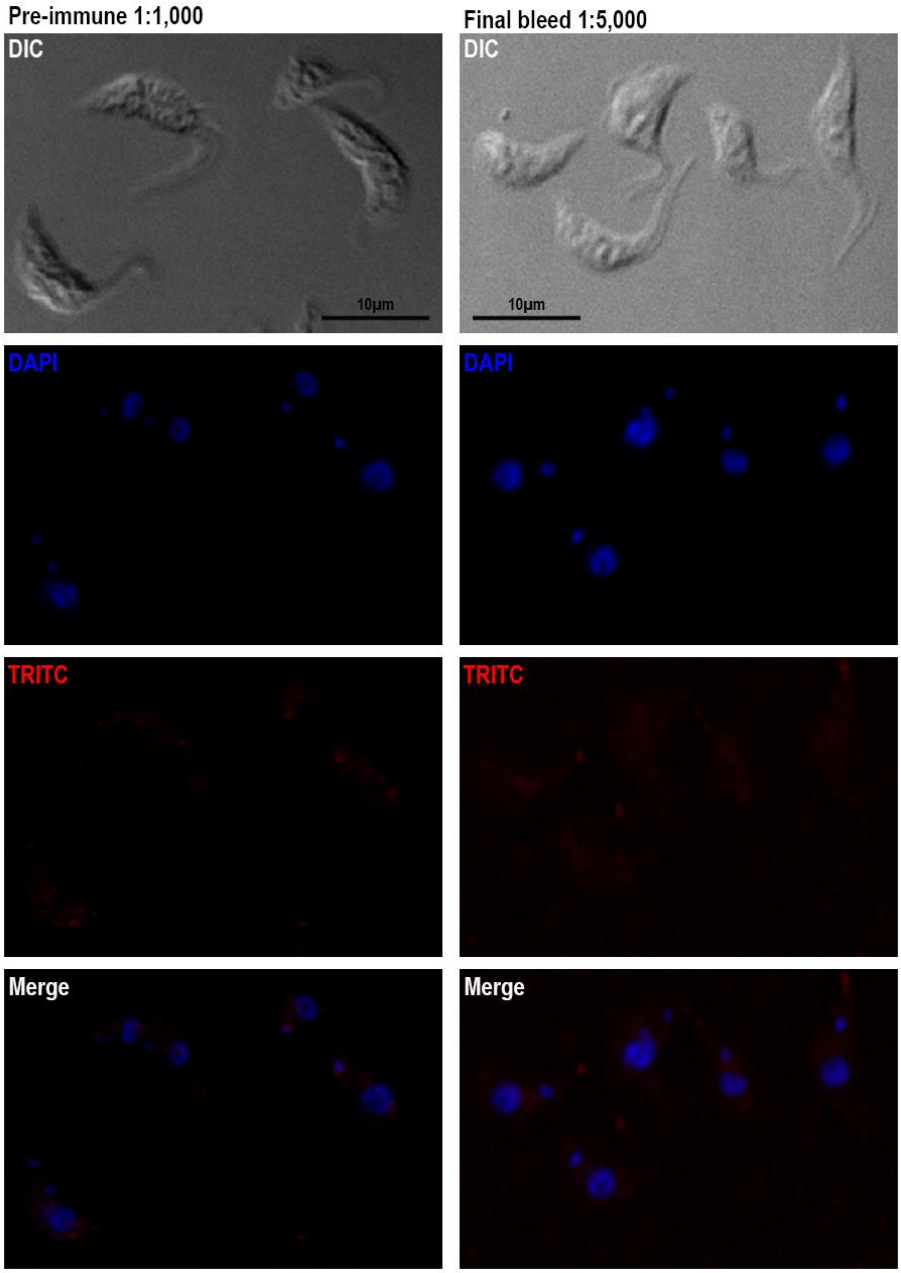


Figure 5.10. Immunofluorescence microscopy images for rabbit 308 pre-bleed and final bleed anti-sera against native *Tb*FEN1 in the 427 parasite cell line. Fluorescent microscopy panels showing the affinity of the pre-immune and post immunisation final bleed anti-sera of rabbit 308 against the native *T. brucei* protein FEN1, in the procyclic 427 parasite strain. The pre-bleed anti-sera was used at a concentration of 1:1,000, where the final bleed was used at a concentration of 1:5,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were captured using the Deltavison Denconvolution microscopy. Scale bar - 10µm.

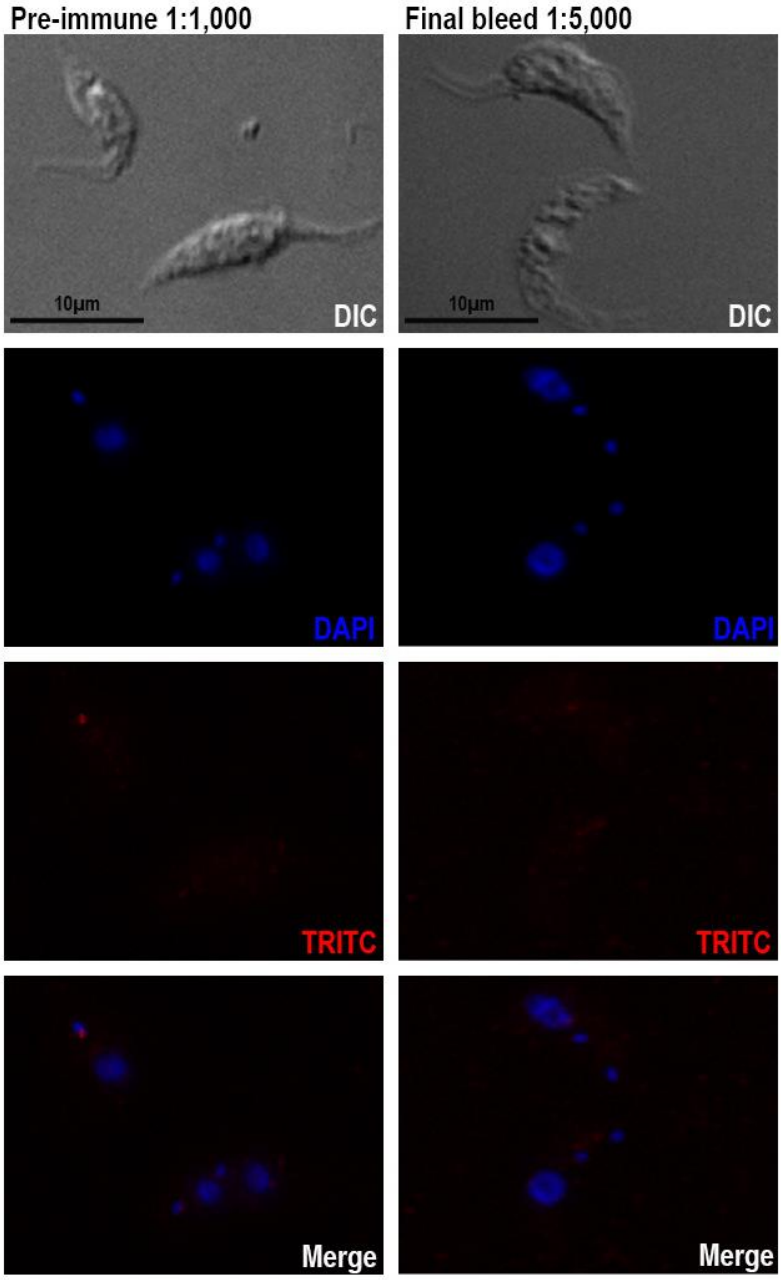


Figure 5.11. Immunofluorescence microscopy images for rabbit 309 pre-bleed and final bleed anti-sera against native *Tb*FEN1 in the 427 parasite cell line. Fluorescent microscopy panels showing the affinity of the pre-immune and post immunisation final bleed anti-sera of rabbit 309 against the native *T. brucei* protein FEN1, in the procyclic 427 parasite strain. The pre-bleed anti-sera was used at a concentration of 1:1,000, where the final bleed was used at a concentration of 1:5,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were captured using the Deltavison Denconvolution microscopy. Scale bar - 10µm.

As the anti-*Tb*FEN1 antisera had failed to convincingly detect the expression or localisation of native *Tb*FEN1 protein, the cell line over-expressing *Tb*FEN1::GSP was analysed. From **Figure 5.12** and **5.13** it can be seen that the 308 and 309 pre-immune anti-sera do not provide any detection to the parasite proteins.

However, *Tb*FEN1::GSP expression was clearly detected by both the 308 and 309 antisera. As anticipated (based on previous immunofluorescence results using the anti-GSP monoclonal antibody) *Tb*FEN1::GSP was localised to the nucleus and the signal appeared to elongate during mitosis.

The detection and nuclear localisation of *Tb*FEN1::GSP confirmed that antisera from both rabbits could detect *Tb*FEN1 and that lack of detection of native protein reflected an issue of sensitivity, i.e. the polyclonal antisera could only detect *Tb*FEN1 once the level of expression of the protein had been dramatically increased.

Despite the unconvincing immunoblot results observed in **Figures 5.10** and **5.11**, immunofluorescence analysis was also undertaken on bloodstream cells but as seen in **Appendix 6-13** both rabbit 308 and 309 antisera only gave non-specific binding within bloodstream form cells.

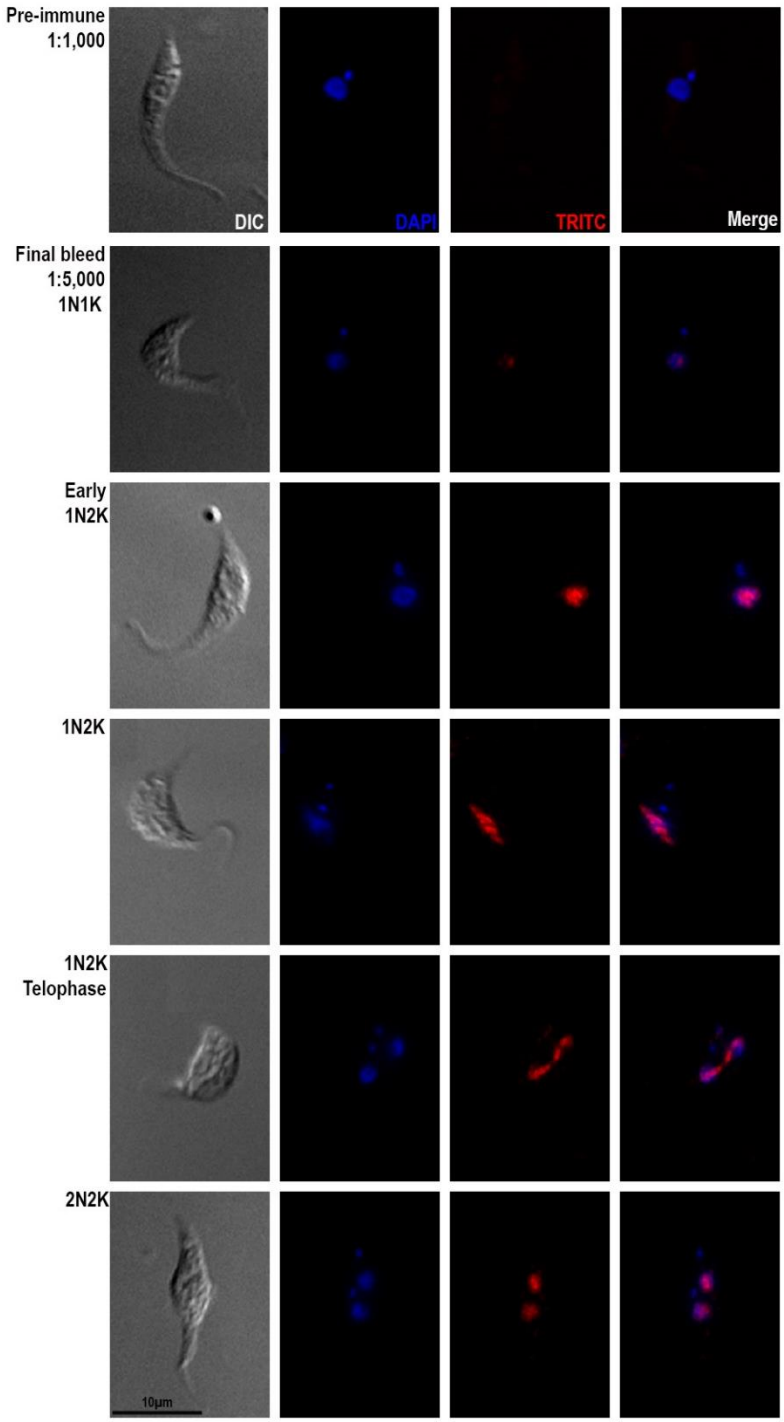


Figure 5.12. Immunofluorescence microscopy images for rabbit 308 pre-bleed and final bleed anti-sera against *Tb*FEN1 in the overexpressed *Tb*FEN1::GSP cell line. Fluorescent microscopy panels showing the affinity of the pre-immune and post immunisation final bleed anti-sera of rabbit 308 against the 24 hour induced *Tb*FEN1::GSP cell line. The pre-immune anti-sera was used at a concentration of 1:1,000 where the final bleed was used at a concentration of 1:5,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were captured using the Deltavison Denconvolution microscopy. Scale bar - 10µm.

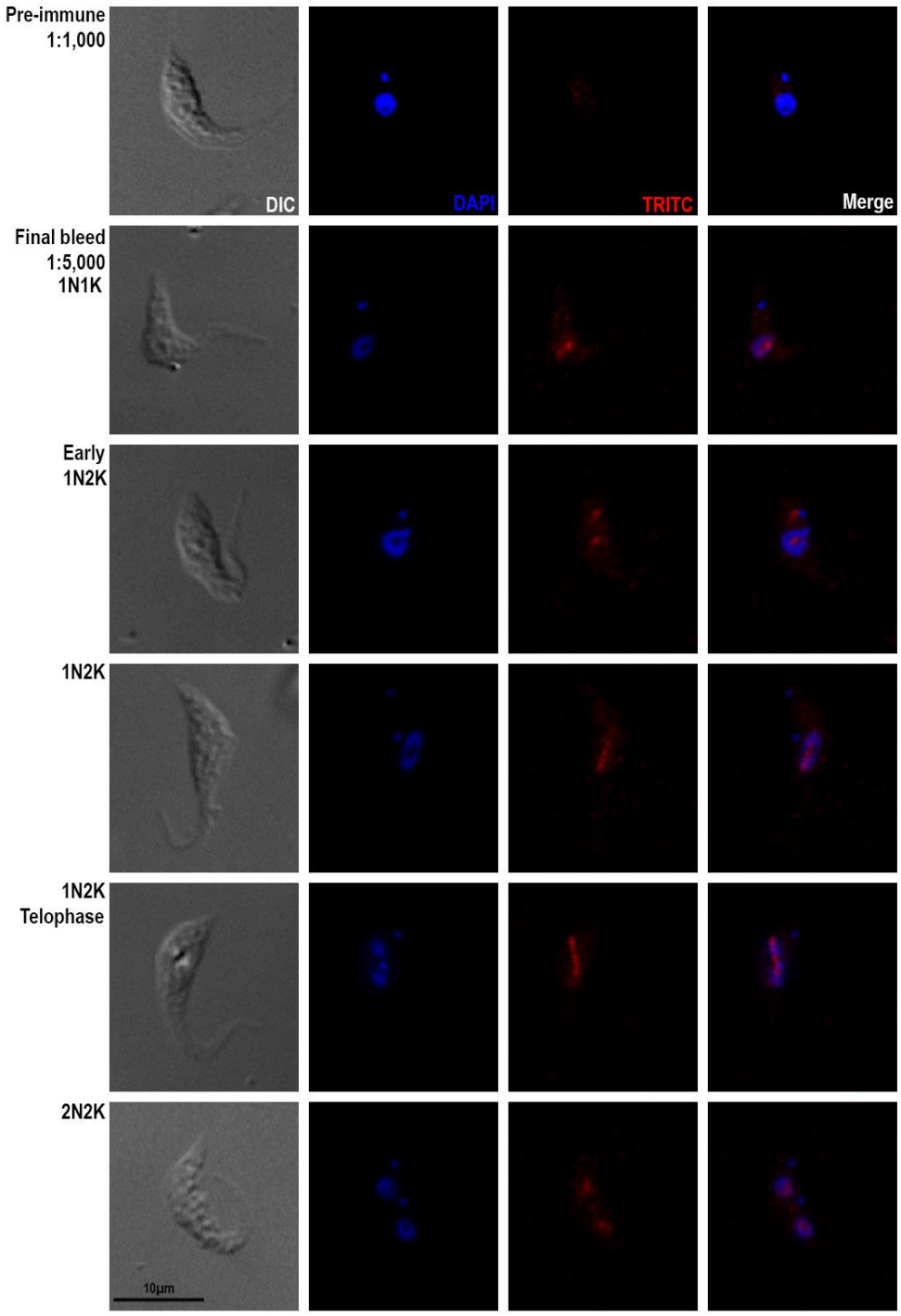


Figure 5.13. Immunofluorescence microscopy images for rabbit 309 pre-bleed and final bleed anti-sera against *Tb*FEN1 in the overexpressed *Tb*FEN1::GSP cell line.
Fluorescent microscopy panels showing the affinity of the pre-immune and post immunisation final bleed anti-sera of rabbit 309 against the 24 hour induced *Tb*FEN1::GSP cell line. The pre-immune anti-sera was used at a concentration of 1:1,000 where the final bleed was used at a concentration of 1:5,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were captured using the Deltavison Denconvolution microscopy. Scale bar - 10µm.

5.5 Chapter Summary

The results presented in this chapter demonstrate that *Tb*FEN1 localises to nucleus of procyclic *T. brucei*. This was most clearly demonstrated by the localisation of the YFP::*Tb*FEN1 and *Tb*FEN1::GSP tagged proteins. Although anti-*Tb*FEN1 polyclonal anti-sera clearly recognises *Tb*FEN1 in cells overexpressing *Tb*FEN1::GSP protein, the antisera was unable to detect native *Tb*FEN1 protein in immunofluorescence experiment.

The YFP::*Tb*RAD2 was weakly detected within the nucleus, but predominantly appears as a discrete focus within the cytoplasm. Expression of *Tb*RAD2 as a C-terminal GSP fusion also failed to convincingly, show any specific localisation within the cell and so no further experimentation was carried out on *Tb*RAD2 in this project.

6 Use of BioID to investigate *Tb*FEN1 Protein-Protein Interactions

6.1 Chapter Synopsis

To investigate proteins that might interact with, or come into close proximity to, *Tb*FEN1 a procyclic cell line was established to express *Tb*FEN1 fused to the bacterial biotin ligase BirA (*Tb*FEN1::MYC::BirA*). This approach (referred to as BioID) enables the proteins which are biotinylated by the fusion protein to be identified using streptavidin. The intention here was to visualise the interacting proteins.

6.2 Generation of a procyclic cell line expressing *Tb*FEN1::MYC::BirA*

The *Tb*FEN1 ORF was amplified by PCR (**Figure 5.4 A)** ligated into pGEMT and the DNA insert sequenced**.** Following verification of the wild type sequence the ORF was subcloned into the pDEX377::MYC::BirA* vector and transfected into the procyclic PTP cell line. The resulting cell line was induced to express *Tb*FEN1::MYC::BirA* and cell density monitored to establish if over-expression of the fusion protein affected cell growth (**Figure 6.1**). Over 120 hours of induction and biotinylation there was no observable effect on cell growth in induced cells compared to the uninduced cell line. It can be assumed therefore that the induction of *Tb*FEN::MYC::BirA* does not cause a dominant negative phenotype.

Cells were prepared for immunofluorescence microscopy 24 hours post-induction. The cells were probed with anti-myc antibody to detect the expression and localisation of the *Tb*FEN1::MYC::BirA* protein and probed with FITC conjugated streptavidin to visualise biotinylated proteins within the cells (**Figure 6.3**)

Figure 6.3 shows the localisation of the *Tb*FEN1::MYC::BirA* protein and resulting biotinylated proteins over the course of the cell cycle. Similarly to previous localisation results, it can be seen that the *Tb*FEN1::MYC::BirA* protein localises to the nucleus. As expected, biotinylated proteins also localise within the nucleus throughout the cell cycle. The MYC-tagged protein shows distinct segregation to either side of the nucleus prior to mitosis in the 1N2K stage, with less obvious spindle microtubule localisation during mitosis. The biotinylated proteins do not appear to show any mitotic spindle localisation.

Immunoblots were carried out on lysates prepared from *Tb*FEN1::MYC::BirA* expressing cells and the parental non-transfected PTP cell line (**Figure 6.4**). The addition of MYC::BirA* to the *Tb*FEN1 protein increases the size of the protein to 80.1 kDa; a protein

of this approximate size was detected using anti-myc monoclonal (**Figure 6.4A)**. In **Figure 6.4B** the streptavidin was used to reveal the complexity of biotinylated proteins that were induced following induction of *Tb*FEN1::MYC::BirA* .

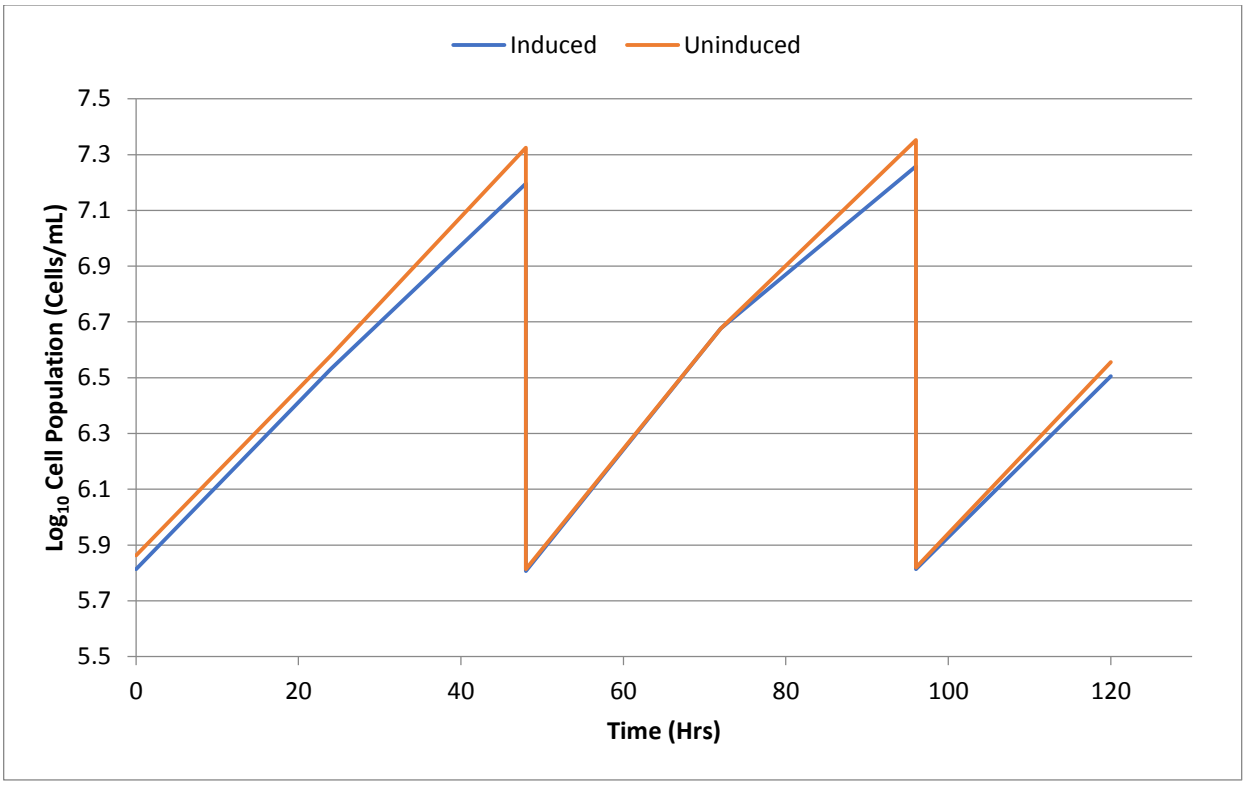


Figure 6.1. Growth Curve of the *Tb*FEN1::MYC::BirA* Induced and Uninduced Cell Lines. Graph to compare the growth rates of the induced and uninduced *Tb*FEN1::MYC::BirA* cell lines, where 48 and 96 hours show point of passage back. The induced cell line was induced with doxycyline and biotin at 0 hours. Cell population is shown in a $\log_{10}$ format for graph simplicity.

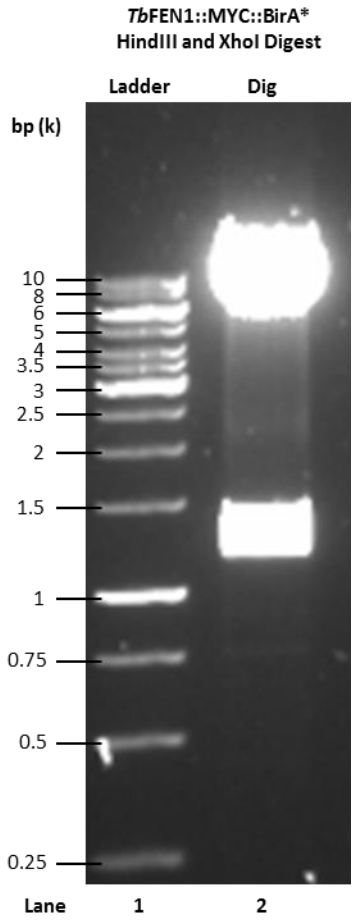


Figure 6.2. Recombinant Plasmid *Tb*FEN1::MYC::BirA* *Hind*III and *Xho*I Digest. An overloaded 0.8% gel for the digested recombinant plasmid *Tb*FEN1::MYC::BirA*. The plasmid was digested with the restriction enzymes *Hind*III and *Xho*I to show the separation of the *Tb*FEN1 ~ 1,200bp and the vector ~8,000bp. Ladder used – GeneRuler™ 1kb DNA ladder, ThermoScientific.

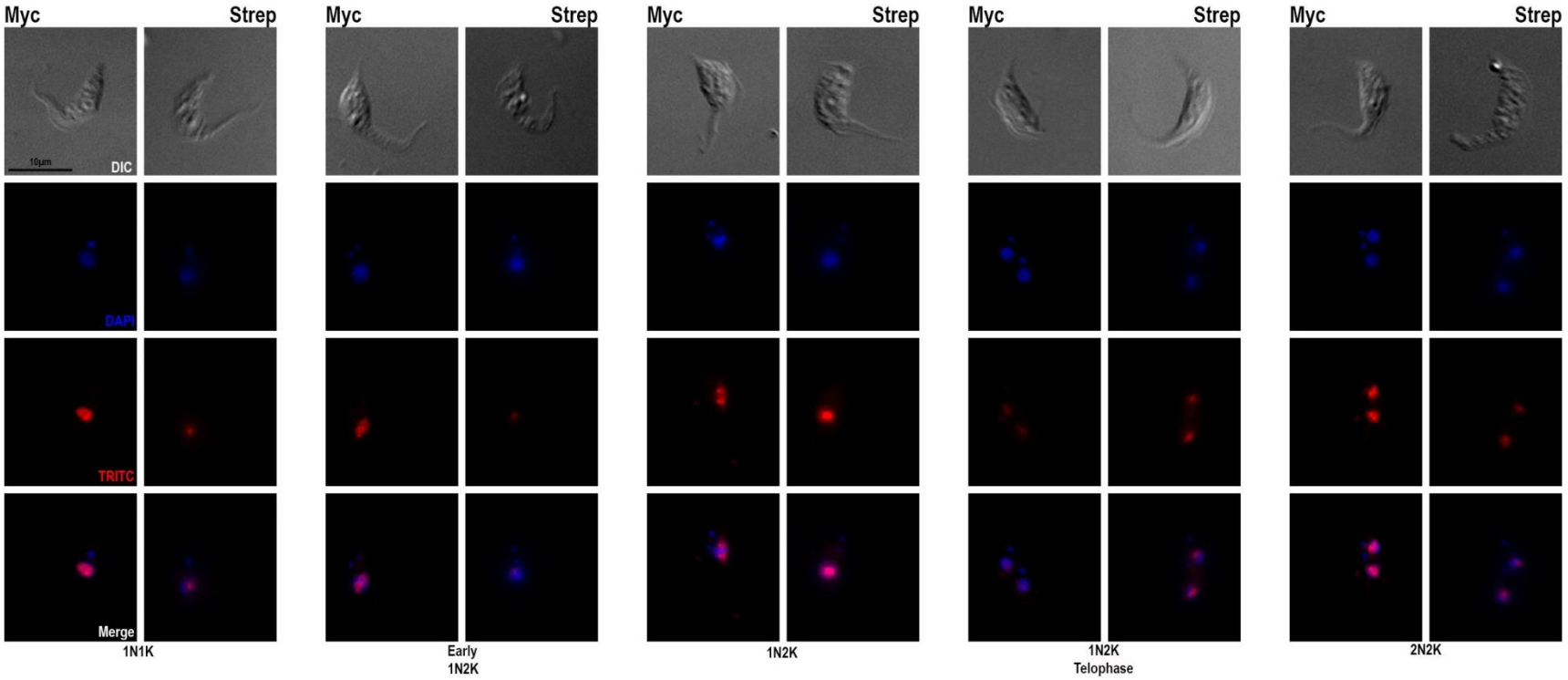
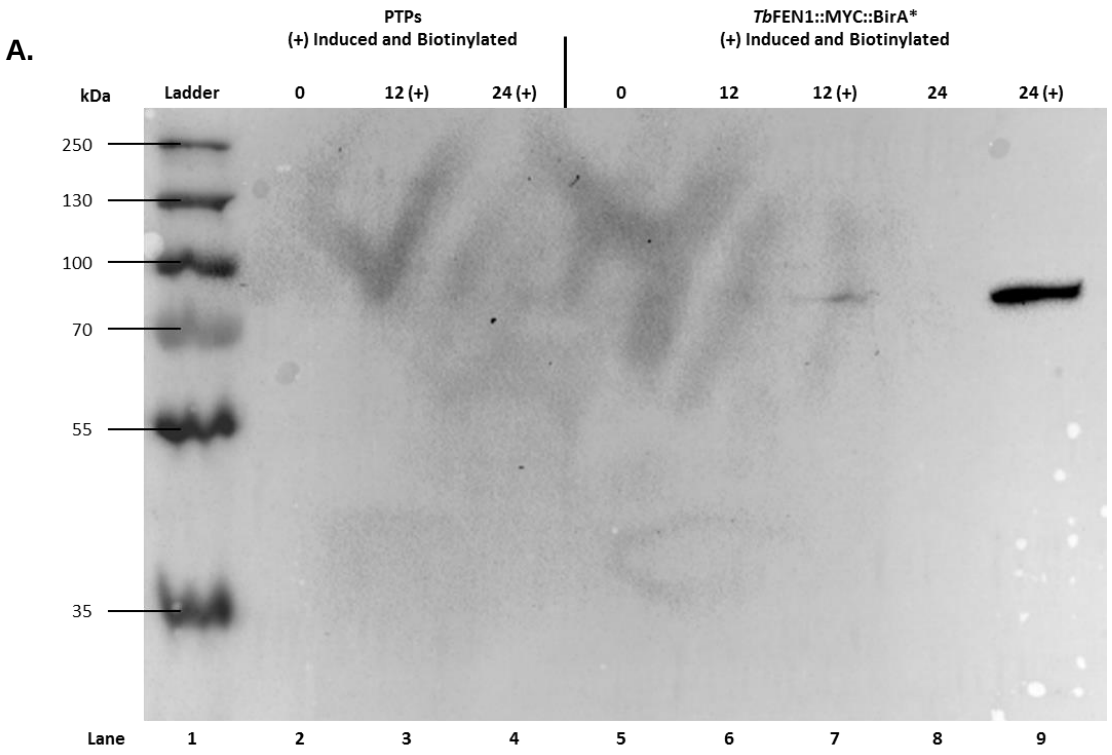
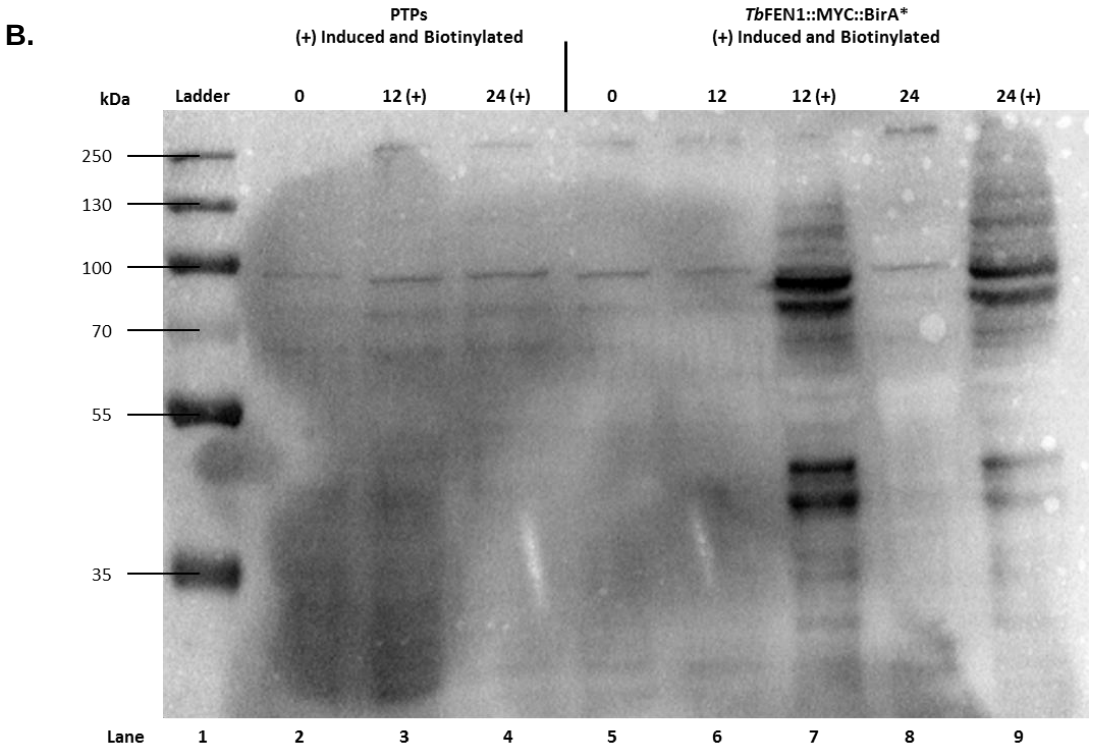


Figure 6.3. Localisation Microscopy After 24 Hours Induction and Biotinylation of *Tb*FEN1::MYC::BirA*. Fluorescent microscopy to show localisation of the *Tb*FEN1 protein with a MYC tag and the localisation of biotinylated interacting proteins using streptavidin, shown by the Myc and Strep sections respectively. Cells were prepared for microscopy after 24 hours of induction and biotinylation. Images were taken using the Deltavision Deconvolution Microscopy.


A.
PTPs
(+) Induced and Biotinylated
TbFEN1::MYC::BirA*
(+) Induced and Biotinylated
kDa
Ladder
0
12 (+)
24 (+)
0
12
12 (+)
24
24 (+)
250
130
100
70
55
35
Lane
1
2
3
4
5
6
7
8
9


B.
PTPs
(+) Induced and Biotinylated
TbFEN1::MYC::BirA*
(+) Induced and Biotinylated
kDa
Ladder
0
12 (+)
24 (+)
0
12
12 (+)
24
24 (+)
250
130
100
70
55
35
Lane
1
2
3
4
5
6
7
8
9

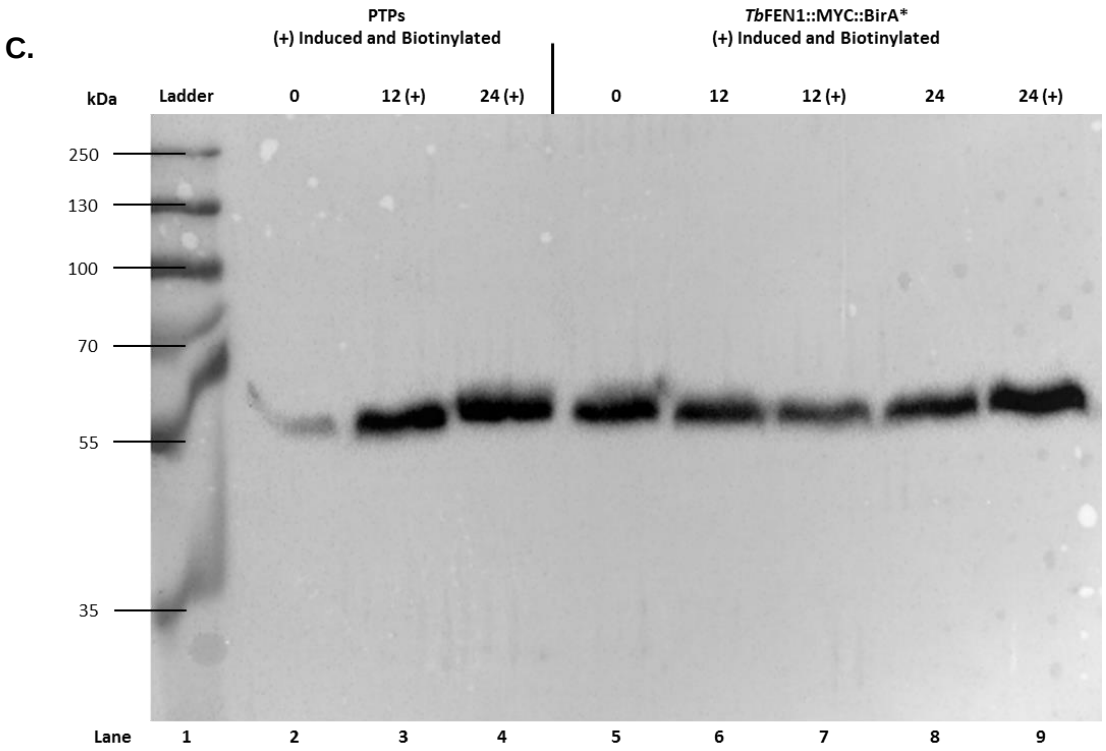


Figure 6.4. Immunoblots of PTPs and Transfected *Tb*FEN1::MYC::BirA* PTPs. Resolved 0.8% immunoblots to compare induced and biotinylated non-transfected PTPs to uninduced and induced *Tb*FEN1::MYC::BirA*. (A) Uses anti c-myc to observe the appearance of the induced protein over 0 to 24 hours, (B) Used the antibody streptavidin to show the appearance of biotinylated proteins over 0 to 24 hours, (C) Is a positive control using the antibody KMX-1 to bind to β-tubulin. Numbers show the time the cell line was cultured for in hours, where the + sign shows whether the cell line was induced and biotinylated at 0 hours. Ladder used – PageRuler™ Plus pre-stained 10-250kDa protein ladder, ThermoScientific.

6.3 Chapter Summary

Indirect immunofluorescence assays demonstrated nuclear localisation for the *Tb*FEN1::MYC::BirA* fusion protein. This also further indicates that expression of *Tb*FEN1 as a fusion to BirA* results in the biotinylation of nuclear proteins. Western blotting confirmed this and also showed that a number of *T. brucei* proteins are biotinylated by *Tb*FEN1::MYC::BirA*. The molecular identity of these proteins was not resolved during this project.

7 Discussion

7.1 What has been achieved by this study?

This study fully or partially accomplished four main objectives in relation to two proteins that are potential *T. brucei* homologs of the human GEN1 protein.

Bioinformatics analysis demonstrated that two proteins encoded in the *T. brucei* genome had sequence and structural homology to human GEN1.

The *Tb*FEN1 protein was expressed and purified on a large scale and *Tb*FEN1 specific polyclonal antisera produced. Although*Tb*RAD2 was expressed, and could be purified under strong denaturing conditions, a large scale purification of this protein was not achieved.

Intracellular localisation experiments of both YFP and GSP-tagged versions of *Tb*FEN1 identified localisation to the nucleus. Although *Tb*FEN1 specific polyclonal antisera was unable detect native *Tb*FEN1 expression it was able to detect overexpressed *Tb*FEN1::GSP and confirmed a nuclear localisation. However, the localisation of tagged versions of the *Tb*RAD2 protein were unconvincing.

A cell line was generated to allow identification of *Tb*FEN1 interacting/proximal proteins. A procyclic cell line expressing TbFEN1::BirA* was established and following induction of this BirA* tagged protein a complex banding pattern was detected on immunoblots using FITC conjugated streptavidin, suggesting biotinylation of *Tb*FEN1 interacting/near neighbour proteins. These biotinylated proteins were also shown to be present within the nucleus, as predicted from previous localisation of tagged versions of the *Tb*FEN1protein.

7.2 Relevance to antigenic variation

One of the key questions that needs to be answered to develop treatments for HAT, is understanding the process of antigenic switching and its subsequent effects on the trypanosome surface. Antigenic variation is carried out by activating different expression sites, or by switching of the VSG sequences on chromosome telomeres to expression sites. Switching is known to occur by homologous recombination (Horn, 2014). As described previously, human GEN1 is a protein known to be involved in homologous recombination, by acting as a HJ resolvase (Ip et al., 2008). The process of homologous recombination is important for gene conversion events, which allows switching to take place.

Bioinformatic analysis identified that *T. brucei* encodes two proteins with sequence and structural homology to human GEN1. These proteins were annotated in GeneDB as *Tb*FEN1

and *Tb*RAD2. These designations were based on their potential function as a flap endonuclease for *Tb*FEN1, and as a member of the DNA repair protein superfamily XPG/RAD2 for *Tb*RAD2. However, there is currently no biochemical characterisation to support these designations. As such, these are only primary indications for the protein designations. It cannot be certain these designations are correct until appropriate biochemical assays have been carried out.

The structural homology of *T. brucei* proteins identified in the bioinformatic analysis leads to the suggestion that these proteins likely share functionality, to the human GEN1. Although this was not covered in the scope of this project it provides an opportunity for future analysis. If these proteins are shown to be of functional importance in homologous recombination, and thus antigenic switching, this provides a potential therapeutic target.

Our project was able to investigate the localisation of the proteins in the parasite. Identifying localisation of the *T. brucei* proteins to the nucleus provides further reassurances that the proteins are involved in DNA-associated processes, such as homologous recombination.

Inhibitors of human FEN1 have been identified previously in relation to cancer biology (Deshmukh et al., 2017). In this case, the inhibitors interacted directly with the human FEN1 protein, rather than the DNA, resulting in anti-proliferative activity. The structural homology identified in bioinformatic analysis of the human and *T. brucei* proteins mean it may be possible to repurpose these inhibitors to act on the *T. brucei* proteins. This would rely on the active sites of the inhibitors having close to 100% homology to one another. Although, the structural analysis presented here, and work on the *Tc*FEN1 protein in *Trypanosoma cruzi* (*T. cruzi*) have identified differences in sequence homology to FEN1 proteins, 100% homology may not be possible (Ponce et al., 2017). However, if the active sites did share sequence homology and if further experimentation could identify functional similarities, the re-purposing of the inhibitors may result in the same anti-proliferative activity of the *T. brucei* cell line, and halt cellular replication. If inhibition is replicated, further analysis could investigate whether the inhibitors affect localisation, repeating the localisation analysis in Chapter 5 with the addition of inhibitors.

Work has recently been carried out investigating similar questions of the FEN1 protein in *T. Cruzi* species (Ponce et al., 2017). This study identified nuclear localisation of the *Tc*FEN1 protein, indicative of involvement in DNA activity, as identified here with the *Tb*FEN1 protein. As part of this study, *Tc*FEN1 was attributed as being crucial in repairing damage caused by oxidative stress. In addition to targeting the *Tb*FEN1 protein to halt antigenic switching, this suggests inhibiting *Tb*FEN1 may reduce the parasites ability to repair DNA damage caused by oxidative stress. When in the hosts bloodstream, the *T. brucei* parasites are susceptible

to high levels of reactive oxygen species, released from neutrophils as part of the host immune response, and thus the induction of oxidative stress and DNA damage (van Eeden et al., 1999). Being able to replicate these *Tc*FEN1 results in the *Tb*FEN1 protein will be important for understanding whether this is a possible downstream effect of targeting *Tb*FEN1. The YFP and GSP tagged cell lines generated could be used to visualise inhibition.

The native presence of the protein *Tb*FEN1 seems to decrease when the parasite cell is in the bloodstream phase in comparison to the procyclic stage of the cell cycle. If the protein were to be involved in the switching of VSGs to evade the host, it would be expected that this protein would be highly expressed in the bloodstream form comparative to the procyclic form. As it is the bloodstream form which presents VSGs to the host and subsequently is able to switch using HR. It is possible, this result in Chapter 5, is due to a blotting error or investigative error. The *Tb*FEN1 anti-sera may not have been evenly spread across the immunoblot membrane, making it look like there is less *Tb*FEN1 in this cell line. It would be important to replicate this analysis to identify whether this reduced concentration is true or due to error.

The process of hydrodynamic flow is crucial to the survival of *T. brucei* in the bloodstream, working in conjunction with VSG switching. Hydrodynamic flow allows the parasite to endocytose the external VSG proteins through the flagellar pocket, inside the cytoplasm any attached host antibodies will become removed and lysed. The VSGs are then presented back on the surface, antibody free. As identified in Markus Engstler's work on hydrodynamic flow, there is a threshold concentration of anti-VSG immunoglobulins the host immune system needs to reach to overcome the parasite (Engstler et al., 2007). The parasite is only capable of recycling the VSG-Ig complexes at low concentrations of immunoglobulins. If *Tb*FEN1 is identified to have a functional role in the homologous recombination, and subsequently VSG switching, inhibiting *Tb*FEN1 will limit *T. brucei*'s ability to undergo VSG switching. Both mechanisms must work in conjunction with one another for prolonged evasion of the host immune system. By blocking antigenic switching, the *T. brucei* will no longer have a repertoire of a few thousand VSGs, meaning the host will need to mount fewer immune responses as the remaining VSG repertoire is exhausted.

7.3 Localisation and Protein-protein interactions

Understanding where a protein localised throughout the cell cycle is important but does not provide insight into its potential function. Human GEN1 has previously been observed to interact with centrosomes and microtubules in mammalian cells where the protein also acts as a control for centrosome copy number, a process which acts separately to Holliday

junction resolvase function (Gao et al., 2012). Conversely, the human FEN1 appears to only localise in the nucleus of mammalian cells at replication forks for DNA repair and replication (Guo et al., 2008). From the localisation results in chapter 5, there is an indication of interaction of *Tb*FEN1 with the microtubules, where fluorescence of *Tb*FEN1 appears to show a strand between the two nuclei at the end of mitosis. However, this may be due to interaction with the nuclear DNA, where the cell lines should be compared to antibodies which are known to interact with the microtubules.

Recruitment of the protein during the cell cycle can also provide understanding of the potential protein function. Human GEN1 protein is functional during S/G2 phase of the cell cycle, where most dissolution of HJs occurs; with a peak accumulation in the M-phase, for centrosome interaction and mitotic stability (Matos and West, 2014, Matos et al., 2011). However, in the human FEN1 protein there should be degradation after the S phase has been complete, this process of degradation is considered crucial for the progression of the cells' cycle (Guo et al., 2012). For the localisation to the nucleus of YFP::*Tb*FEN1and *Tb*FEN1::GSP there are no distinctive fluctuations in recruitment of the protein throughout the cell cycle. We would expect protein accumulation signals for *Tb*FEN1 in either late 1N1K to mitosis for a protein like GEN1, or a large decrease in protein signal after early 1N2K for a protein like FEN1. Greater recruitment at cell cycle phases would indicate greater activity of the protein, furthering our understanding of its function.

The BioID analysis in Chapter 6 was unable to definitively identify any protein-protein interactions of the *Tb*FEN1 protein. Given knowledge of human FEN1 interaction with PCNA and the identification of it in *T. brucei* interactions would be expected to occur (Sakurai et al., 2005). PCNA is crucial to the DNA replication, acting with DNA polymerases, including FEN1. In humans, PCNA interacts with FEN1 at branched flaps and nicked substrates, stimulating FEN1 activity (Li et al., 1995, Craggs et al., 2014).

PCNA has previously been characterised in *T. brucei* (Kaufmann et al., 2012), where it is clearly detected in the S-phase of the cell cycle; the same phase PCNA is active in human DNA replication (Kirsten and Ulrich, 2002). In *T. brucei,* PCNA has been identified to be about 32 kDa in size. Regulation of *Tb*PCNA is crucial for normal proliferation and replication of the bloodstream *T.* brucei, where overexpression stops proliferation providing opportunities for targeted interventions (Valenciano et al., 2015). Protein-protein Interaction experiments of *Tb*FEN1 identified a band at approximately 30 kDa on the immunoblots. It is possible this band is identifying *Tb*FEN1 interaction with *Tb*PCNA. The bioinformatics analysis completed in this project identified a FEN1-PCNA interface as part of the conserved domains in the *Tb*FEN1 gene, adding further support for suggested PCNA interaction.

However, our analysis does not provide confirmation of this interaction, which would require experimental confirmation. This could be carried out by purifying biotinylated proteins and carrying out SDS-PAGE and an immunoblot probe with anti-PCNA antibodies. If the anti-PCNA detects the presence of the *Tb*PCNA protein then it would indicate a *Tb*PCNA-*Tb*FEN1 interaction.

Intervening on PCNA to disrupt the cell cycle has previously been proposed in cancer biology in humans, as well as in *T. brucei* as a therapeutic target (Müller et al., 2013, Valenciano et al., 2015). If future studies confirmed the *Tb*FEN1 gene forms interactions with *Tb*PCNA, this could provide a potential future target which may disrupt the cell cycle.

7.4 *Tb*RAD2 protein

Much of the work presented here focusses on *Tb*FEN1 due to difficulties in denaturing the much larger protein, *Tb*RAD2. This does not mean however, that *Tb*RAD2 may not be of functional importance in VSG switching. The protein *Tb*RAD2 shared greater sequence similarity to the human GEN1 protein than the *Tb*FEN1 protein. Through visual comparisons of the sections of the secondary structures, the human GEN1 and *Tb*RAD2 shared almost 100% homology. There is also evidence of human GEN1 being difficult to solubilise (Rass et al., 2010), as experienced with the *Tb*RAD2. This additional sequence similarity suggests *Tb*RAD2 may share greater functional similarity with human GEN1, and hence as a HJ resolvase, than *Tb*FEN1.

7.5 Critique of methods and results

7.5.1 Protein expression

Localisation experiments with the *Tb*FEN1 polyclonal anti-sera were completed with native expression levels of the parasite protein and were unable to provide specificity at the range of concentrations used. This indicates that although the anti-sera can detect *Tb*FEN1 as low as 10 ng on an immunoblot, native expression within the parasite is below the detectable limit for the sensitivity of the protein. The overexpressed GSP cell line was able to be detected by the antisera.

When considering the tagged proteins, the YFP tag was added in sequence to the endogenous loci, though it is possible that this may not have been a native level of expression. For both YFP tagged proteins the level of expression was constant throughout

the cell cycle, where *Tb*FEN1 was strong and *Tb*RAD2 weak. Placing the tag on the C-terminus of *Tb*RAD2 may provide stronger expression.

GSP experiments were carried out using a vector requiring an induction process, this process would have overexpressed the trypanosome protein, therefore the localisation may not be native to the protein. Despite this overexpression, protein expression of *Tb*RAD::GSP remained weak. Conversely, for *Tb*FEN1 the level of expression appeared to increase during cell replication. The GSP sequence was added to the C-terminus and would have subsequently lost 3'UTR control, where the level of expression would not have been controlled. It would be important to carry out experiments to investigate whether N-terminal tagging of GSP causes the level of protein expression to change.

7.5.2 YFP and GSP tag location in *T. brucei* proteins

For the *Tb*FEN1 protein, clear localisation was seen in the nucleus in tagging experiments with YFP and GSP.

When considering the *Tb*RAD2 protein, there was weak evidence of the protein being localised to the nucleus using YFP tagging, but no specific localisation of the GSP tagged *Tb*RAD2 protein. The YFP tag was added to the N-terminus, conversely the GSP epitope was present on the C-terminus. It is possible, the difference in visualisation is due to these tag locations. A GSP tag was used because it is a smaller 9 amino acid epitope, which would be expected to affect localisation less than the larger YFP tag. A potential future experiment would be to place GSP at the N-terminus and look again at localisation patterns. This would identify whether it was the different tags or different locations of the tags which affected expression and/or localisation.

7.5.3 Expression of T*b*RAD2

Although *TbRAD2* had to be purified under stringent denaturing conditions, it would have still been possible to carry out large scale purification. However, purifying a denatured protein may be of little use if the protein could not be refolded to a functional state. As a result, any antibodies created to the protein, could be detected on an immunoblot, but may not be detected within the parasite. Alternatively, the use of different vectors and/or *E. coli* bacterial strains could be investigated to produce a soluble protein suitable for biochemical studies.

7.5.4 Identification of *Tb*FEN1 interacting proteins

Although preliminary experiments demonstrated that induction of TbFEN1::BirA* led to the biotinylation of a number of proteins within the nucleus of *T. brucei*, the identity of these proteins was not established. When investigating protein-protein interactions of *TbFEN1*, only a banding pattern of interacting proteins and the localisation of these interactions were

investigated. The banding pattern produced provides little information as to what the proteins are. Whilst it is possible to compare the sizes of the banding patterns with known *T. brucei* proteins, it does not specifically identify the protein. Further experimentation to interrogate these bands is therefore important for understanding what protein-protein interactions are being identified.

7.6 Suggestions for Future Work

7.6.1 Biochemical characterisation of *Tb*FEN1

Activity of the purified *Tb*FEN1 can be detected by endonuclease assay. The *Tb*FEN1 is expected to carry out nuclease activity, such as the HJ resolvase activity seen in human GEN1. Assays would need to be carried out to investigate the nuclease activity of *Tb*FEN1 with branched oligonucleotides. Firstly, the oligonucleotides would need to be produced using partially complementary sequence to produce branched DNA substrate. The nuclease assay involves incubating the *Tb*FEN1 protein with the produced synthetic substrates. Nuclease activity can be detected by the protein which will remove the branched chains, which can be detected as by the change in weight of the DNA. Following this, the proteins optimum conditions will need to be found along with the proteins Michealis constant and turnover number. Further experimentation against a wide range of oligonucleotide structures can then be carried out; such as HJ, replication forks and 5' flaps at different protein concentrations. This will provide information to whether the protein is interacting with the structure due to saturation. By finding different types of structures the protein is able to interact with, a picture of the *Tb*FEN1's functions can begin to be put together (Wright et al., 2011).

7.6.2 Antisera affinity purification

From the purified *Tb*FEN1, polyclonal antisera was produced. The native localisation experiments identified that the antisera did not have the specificity to *Tb*FEN1 at high antisera concentrations as it interacted with other parasite proteins. At low antisera concentration, the antisera did not have the sensitivity to the low-level expression of native *Tb*FEN1. To view the native level of expression of *Tb*FEN1, the anti-*Tb*FEN1 antibody in the antisera will need to be purified out. This can be done by using the previously purified *Tb*FEN1 protein to carry out affinity purification to increase the antisera sensitivity and specificity.

7.6.3 Future experiments for anti-*Tb*RAD2

Cell lines expressing *Tb*RAD2 with YFP and GSP tags were created. These could be used to investigate localisation with a *Tb*RAD2 antibody, should effective purification of the protein be carried out and a *Tb*RAD2 antisera be produced. This would provide support for future functionality experiments if this protein is found to localise to the nucleus.

7.6.4 RNAi experiments

The *Tb*FEN1 antibody and YFP-tagged protein cell line could be used to investigate RNAi induced depletion of *Tb*FEN1. It would not be possible to use the GSP cell line because it will require induction of the cell line to express the tagged protein. It is possible to create RNAi sequences specific to *TbFEN1* which would neutralize *Tb*FEN1 mRNA (Wang et al., 2000), and hence production of the protein. Using the YFP cell line the depletion of the *Tb*FEN1 protein can be observed over time. If it is observed that the *Tb*FEN1 protein is decreasing in the YFP tagged cell line, both the wild type and YFP tagged cell line could be observed for a negative phenotype. If the *Tb*FEN1 antibody were to be purified, depletion may potentially be visualized for the wild type *T. brucei*.

7.6.5 Investigating protein-protein interactions

To investigate the protein-protein interactions identified with *Tb*FEN1, SILAC analysis and mass spectrometry could be carried out. The SILAC process involves growing multiple parasite cell lines in media of different amino acid molecular weights; resulting in relative varying protein weights between cell lines, such as: Light, Medium and Heavy. A BirA* marker on the *Tb*FEN1 protein, adds a biotin molecule to proteins which come within a 20-30 nm radius of it. Proteins which are biotinylated in the BioID process are potential interacting proteins (Roux et al., 2013). Biotinylated proteins can then be separated from other parasite proteins via streptavidin, leaving only the proteins which have interacted with the protein of interest, through the SILAC cell lines, there will be three different weights for the same proteins present. These proteins can then undergo mass spectrometry, where the results can be compared to the *Tb*FEN1 protein library to identify the interacting proteins. This provides a stronger indication of the potential functions of the *TbFEN1*. This method has previously been carried out successfully in mammalian cells and in *T. brucei* cell lines to investigate protein-protein interactions (Morriswood et al., 2013, Wang and Huang, 2014, Emmott and Goodfellow, 2014). This method provides a straight forward approach to fully identifying protein-protein interactions, and could specifically be used following this work to investigate whether there is a true interaction between *Tb*FEN1 and *Tb*PCNA.

Bibliography

ALSFORD, S., HORN, D. & GLOVER, L. 2009. DNA breaks as triggers for antigenic variation in African trypanosomes. *Genome Biology,* 10**,** 223.

ANDERSON, N. E., MUBANGA, J., FEVRE, E. M., PICOZZI, K., EISLER, M. C., THOMAS, R. & WELBURN, S. C. 2011. Characterisation of the wildlife reservoir community for human and animal trypanosomiasis in the Luangwa Valley, Zambia. *PLoS Negl Trop Dis,* 5**,** e1211.

AUFFRET, C. A. & TURNER, M. J. 1981. Variant specific antigens of Trypanosoma brucei exist in solution as glycoprotein dimers. *Biochemical Journal,* 193**,** 647-650.

BARAKAT, K. & TUSZYNSKI, J. 2013. *Nucleotide Excision Repair Inhibitors: Still a Long Way to Go, New Research Directions in DNA Repair*, InTech.

BARAL, T. N. 2010. Immunobiology of African trypanosomes: need of alternative interventions. *J Biomed Biotechnol,* 2010**,** 389153.

BISHOP, J. R., SHIMAMURA, M. & HAJDUK, S. L. 2001. Insight into the mechanism of trypanosome lytic factor-1 killing of Trypanosoma brucei brucei. *Mol Biochem Parasitol,* 118**,** 33-40.

CAMERINI-OTERO, R. D. & HSIEH, P. 1995. Homologous recombination proteins in prokaryotes and eukaryotes. *Annual review of genetics,* 29**,** 509-552.

CAPECCHI, M. R. 1989. Altering the genome by homologous recombination. *Science,* 244**,** 1288-1292.

CARRINGTON, M. & BOOTHROYD, J. 1996. Implications of conserved structural motifs in disparate trypanosome surface proteins. *Mol Biochem Parasitol,* 81**,** 119-26.

CDC. 2015. *African Trypanosomiasis* [Online]. Available: http://www.cdc.gov/parasites/sleepingsickness/biology.html [Accessed 2016].

CONNOR, R. J. 1994. The impact of nagana. *Onderstepoort J Vet Res,* 61**,** 379-83.

CONWAY, C., PROUDFOOT, C., BURTON, P., BARRY, J. D. & MCCULLOCH, R. 2002. Two pathways of homologous recombination in Trypanosoma brucei. *Mol Microbiol,* 45**,** 1687-700.

CORDON-OBRAS, C., RODRIGUEZ, Y. F., FERNANDEZ-MARTINEZ, A., CANO, J., NDONG-MABALE, N., NCOGO-ADA, P., NDONGO-ASUMU, P., APARICIO, P., NAVARRO, M., BENITO, A. & BART, J. M. 2015. Molecular evidence of a Trypanosoma brucei gambiense sylvatic cycle in the human african trypanosomiasis foci of Equatorial Guinea. *Front Microbiol,* 6**,** 765.

CRAGGS, T. D., HUTTON, R. D., BRENLLA, A., WHITE, M. F. & PENEDO, J. C. 2014. Single-molecule characterization of Fen1 and Fen1/PCNA complexes acting on flap substrates. *Nucleic Acids Research,* 42**,** 1857-1872.

DEAN, S., SUNTER, J., WHEELER, R. J., HODKINSON, I., GLUENZ, E. & GULL, K. 2015. A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biology,* 5.

DESHMUKH, A. L., CHANDRA, S., SINGH, D. K., SIDDIQI, M. I. & BANERJEE, D. 2017. Identification of human flap endonuclease 1 (FEN1) inhibitors using a machine learning based consensus virtual screening. *Mol Biosyst,* 13**,** 1630-1639.

EMMOTT, E. & GOODFELLOW, I. 2014. Identification of Protein Interaction Partners in Mammalian Cells Using SILAC-immunoprecipitation Quantitative Proteomics. *Journal of Visualized Experiments : JoVE***,** 51656.

ENGLUND, P. T., HAJDUK, S. L. & MARINI, J. C. 1982. The molecular biology of trypanosomes. *Annu Rev Biochem,* 51**,** 695-726.

ENGSTLER, M., PFOHL, T., HERMINGHAUS, S., BOSHART, M., WIEGERTJES, G., HEDDERGOTT, N. & OVERATH, P. 2007. Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell,* 131**,** 505-15.

ENGSTLER, M., THILO, L., WEISE, F., GRUNFELDER, C. G., SCHWARZ, H., BOSHART, M. & OVERATH, P. 2004. Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei. *J Cell Sci,* 117**,** 1105-15.

EVANS, E., FELLOWS, J., COFFER, A. & WOOD, R. D. 1997. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *The EMBO journal,* 16**,** 625-638.

FAO. 2016. *Programme Against African Trypanosomosis (PAAT)* [Online]. Food and Agricultural Organizaton of the United Nations. Available: http://www.fao.org/ag/againfo/programmes/en/paat/disease.html [Accessed June 2016].

FRANCO, J. R., CECCHI, G., PRIOTTO, G., PAONE, M., DIARRA, A., GROUT, L., MATTIOLI, R. C. & ARGAW, D. 2017. Monitoring the elimination of human African trypanosomiasis: Update to 2014. *PLoS Negl Trop Dis,* 11**,** e0005585.

FRANCO, J. R., SIMARRO, P. P., DIARRA, A. & JANNIN, J. G. 2014. Epidemiology of human African trypanosomiasis. *Clin Epidemiol,* 6**,** 257-75.
GAO, M., RENDTLEW DANIELSEN, J., WEI, L. Z., ZHOU, D. P., XU, Q., LI, M. M., WANG, Z. Q., TONG, W. M. & YANG, Y. G. 2012. A novel role of human holliday junction resolvase GEN1 in the maintenance of centrosome integrity. *PLoS One,* 7**,** e49687.
GIBSON, W., PEACOCK, L., FERRIS, V., WILLIAMS, K. & BAILEY, M. 2008. The use of yellow fluorescent hybrids to indicate mating in Trypanosoma brucei. *Parasit Vectors,* 1**,** 4.
GOUJON, M., MCWILLIAM, H., LI, W., VALENTIN, F., SQUIZZATO, S., PAERN, J. & LOPEZ, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res,* 38**,** W695-9.
GUO, Z., KANJANAPANGKA, J., LIU, N., LIU, S., LIU, C., WU, Z., WANG, Y., LOH, T., KOWOLIK, C., JAMSEN, J., ZHOU, M., TRUONG, K., CHEN, Y., ZHENG, L. & SHEN, B. 2012. Sequential posttranslational modifications program FEN1 degradation during cell-cycle progression. *Mol Cell,* 47**,** 444-56.
GUO, Z., QIAN, L., LIU, R., DAI, H., ZHOU, M., ZHENG, L. & SHEN, B. 2008. Nucleolar localization and dynamic roles of flap endonuclease 1 in ribosomal DNA replication and damage repair. *Mol Cell Biol,* 28**,** 4310-9.
HAMMARTON, T., WICKSTEAD, B. & MCKEAN, P., G. 2007. Cell Structure, Cell Division and Cell Cycle. *In:* BARRY, D., MCCULLOCH, R., MOTTRAM, J. & ACOSTA-SERRANO, A. (ed.) *Trypanosomes: After the Genome.* Wymondham: Horizon Bioscience.
HERTZ-FOWLER, C., FIGUEIREDO, L. M., QUAIL, M. A., BECKER, M., JACKSON, A., BASON, N., BROOKS, K., CHURCHER, C., FAHKRO, S., GOODHEAD, I., HEATH, P., KARTVELISHVILI, M., MUNGALL, K., HARRIS, D., HAUSER, H., SANDERS, M., SAUNDERS, D., SEEGER, K., SHARP, S., TAYLOR, J. E., WALKER, D., WHITE, B., YOUNG, R., CROSS, G. A., RUDENKO, G., BARRY, J. D., LOUIS, E. J. & BERRIMAN, M. 2008. Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS One,* 3**,** e3527.
HEYER, W. D., EHMSEN, K. T. & LIU, J. 2010. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet,* 44**,** 113-39.
HOLLIDAY, R. 1964. A mechanism for gene conversion in fungi.
HORN, D. 2014. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol,* 195**,** 123-9.
HORN, D. & BARRY, J. D. 2005. The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome Res,* 13.
IP, S. C., RASS, U., BLANCO, M. G., FLYNN, H. R., SKEHEL, J. M. & WEST, S. C. 2008. Identification of Holliday junction resolvases from humans and yeast. *Nature,* 456**,** 357-61.
KAO, H. I. & BAMBARA, R. A. 2003. The protein components and mechanism of eukaryotic Okazaki fragment maturation. *Crit Rev Biochem Mol Biol,* 38**,** 433-52.
KAUFMANN, D., GASSEN, A., MAISER, A., LEONHARDT, H. & JANZEN, C. J. 2012. Regulation and spatial organization of PCNA in Trypanosoma brucei. *Biochemical and Biophysical Research Communications,* 419**,** 698-702.
KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols,* 10**,** 845-858.
KENNEDY, P. G. 2013. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet Neurol,* 12**,** 186-94.
KIRSTEN, D. & ULRICH, H. 2002. Colocalization of human Rad17 and PCNA in late S phase of the cell cycle upon replication block. *Oncogene,* 21**,** 7710.
KOHL, L. & GULL, K. 1998. Molecular architecture of the trypanosome cytoskeleton. *Mol Biochem Parasitol,* 93**,** 1-9.
KOWALCZYKOWSKI, S. C., DIXON, D. A., EGGLESTON, A. K., LAUDER, S. D. & REHRAUER, W. M. 1994. Biochemistry of homologous recombination in Escherichia coli. *Microbiological reviews,* 58**,** 401-465.
LANDEIRA, D., BART, J. M., VAN TYNE, D. & NAVARRO, M. 2009. Cohesin regulates VSG monoallelic expression in trypanosomes. *J Cell Biol,* 186**,** 243-54.
LI, W., COWLEY, A., ULUDAG, M., GUR, T., MCWILLIAM, H., SQUIZZATO, S., PARK, Y. M., BUSO, N. & LOPEZ, R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res,* 43**,** W580-4.
LI, X., LI, J., HARRINGTON, J., LIEBER, M. R. & BURGERS, P. M. 1995. Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *J Biol Chem,* 270**,** 22109-12.

LIEBER, M. R. 1997. The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays,* 19**,** 233-40.

LILLICO, S., FIELD, M. C., BLUNDELL, P., COOMBS, G. H. & MOTTRAM, J. C. 2003. Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. *Mol Biol Cell,* 14**,** 1182-94.

LIU, Y., FREEMAN, A. D., DECLAIS, A. C., WILSON, T. J., GARTNER, A. & LILLEY, D. M. 2015. Crystal Structure of a Eukaryotic GEN1 Resolving Enzyme Bound to DNA. *Cell Rep,* 13**,** 2565-75.

LIU, Y., KAO, H. I. & BAMBARA, R. A. 2004. Flap endonuclease 1: a central component of DNA metabolism. *Annu Rev Biochem,* 73**,** 589-615.

LIU, Y. & WILSON, S. H. 2012. DNA base excision repair: a mechanism of trinucleotide repeat expansion. *Trends Biochem Sci,* 37**,** 162-72.

MARCELLO, L. & BARRY, J. D. 2007. Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res,* 17**,** 1344-52.

MARCHLER-BAUER, A., ANDERSON, J. B., CHITSAZ, F., DERBYSHIRE, M. K., DEWEESE-SCOTT, C., FONG, J. H., GEER, L. Y., GEER, R. C., GONZALES, N. R., GWADZ, M., HE, S., HURWITZ, D. I., JACKSON, J. D., KE, Z., LANCZYCKI, C. J., LIEBERT, C. A., LIU, C., LU, F., LU, S., MARCHLER, G. H., MULLOKANDOV, M., SONG, J. S., TASNEEM, A., THANKI, N., YAMASHITA, R. A., ZHANG, D., ZHANG, N. & BRYANT, S. H. 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res,* 37**,** D205-10.

MARCHLER-BAUER, A. & BRYANT, S. H. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res,* 32**,** W327-31.

MARCHLER-BAUER, A., DERBYSHIRE, M. K., GONZALES, N. R., LU, S., CHITSAZ, F., GEER, L. Y., GEER, R. C., HE, J., GWADZ, M., HURWITZ, D. I., LANCZYCKI, C. J., LU, F., MARCHLER, G. H., SONG, J. S., THANKI, N., WANG, Z., YAMASHITA, R. A., ZHANG, D., ZHENG, C. & BRYANT, S. H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res,* 43**,** D222-6.

MARCHLER-BAUER, A., LU, S., ANDERSON, J. B., CHITSAZ, F., DERBYSHIRE, M. K., DEWEESE-SCOTT, C., FONG, J. H., GEER, L. Y., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., JACKSON, J. D., KE, Z., LANCZYCKI, C. J., LU, F., MARCHLER, G. H., MULLOKANDOV, M., OMELCHENKO, M. V., ROBERTSON, C. L., SONG, J. S., THANKI, N., YAMASHITA, R. A., ZHANG, D., ZHANG, N., ZHENG, C. & BRYANT, S. H. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res,* 39**,** D225-9.

MATOS, J., BLANCO, M. G., MASLEN, S., SKEHEL, J. M. & WEST, S. C. 2011. Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. *Cell,* 147**,** 158-72.

MATOS, J. & WEST, S. C. 2014. Holliday junction resolution: regulation in space and time. *DNA Repair (Amst),* 19**,** 176-81.

MCCREADY, S. J., OSMAN, F. & YASUI, A. 2000. Repair of UV damage in the fission yeast Schizosaccharomyces pombe. *Mutat Res,* 451**,** 197-210.

MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res,* 41**,** W597-600.

MORRISWOOD, B., HAVLICEK, K., DEMMEL, L., YAVUZ, S., SEALEY-CARDONA, M., VIDILASERIS, K., ANRATHER, D., KOSTAN, J., DJINOVIĆ-CARUGO, K., ROUX, K. J. & WARREN, G. 2013. Novel Bilobe Components in Trypanosoma brucei Identified Using Proximity-Dependent Biotinylation. *Eukaryotic Cell,* 12**,** 356-367.

MUGNIER, M. R., CROSS, G. A. & PAPAVASILIOU, F. N. 2015. The in vivo dynamics of antigenic variation in Trypanosoma brucei. *Science,* 347**,** 1470-3.

MÜLLER, R., MISUND, K., HOLIEN, T., BACHKE, S., GILLJAM, K. M., VÅTSVEEN, T. K., RØ, T. B., BELLACCHIO, E., SUNDAN, A. & OTTERLEI, M. 2013. Targeting Proliferating Cell Nuclear Antigen and Its Protein Interactions Induces Apoptosis in Multiple Myeloma Cells. *PLOS ONE,* 8**,** e70430.

MUNOZ-JORDAN, J. L., DAVIES, K. P. & CROSS, G. A. 1996. Stable expression of mosaic coats of variant surface glycoproteins in Trypanosoma brucei. *Science,* 272**,** 1795-7.

O'DONOVAN, A., DAVIES, A. A., MOGGS, J. G., WEST, S. C. & WOOD, R. D. 1994. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature,* 371**,** 432-5.

OGBADOYI, E., ERSFELD, K., ROBINSON, D., SHERWIN, T. & GULL, K. 2000. Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. *Chromosoma,* 108**,** 501-13.

PEREZ-MORGA, D., VANHOLLEBEKE, B., PATURIAUX-HANOCQ, F., NOLAN, D. P., LINS, L., HOMBLE, F., VANHAMME, L., TEBABI, P., PAYS, A., POELVOORDE, P., JACQUET, A., BRASSEUR, R. & PAYS, E. 2005. Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science,* 309**,** 469-72.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* 25**,** 1605-12.

PONCE, I., ALDUNATE, C., VALENZUELA, L., SEPÚLVEDA, S., GARRIDO, G., KEMMERLING, U., CABRERA, G. & GALANTI, N. 2017. A Flap Endonuclease (TcFEN1) Is Involved in Trypanosoma cruzi Cell Proliferation, DNA Repair, and Parasite Survival. *Journal of Cellular Biochemistry,* 118**,** 1722-1732.

PRAKASH, S. & PRAKASH, L. 2000. Nucleotide excision repair in yeast. *Mutat Res,* 451**,** 13-24.

RASS, U., COMPTON, S. A., MATOS, J., SINGLETON, M. R., IP, S. C., BLANCO, M. G., GRIFFITH, J. D. & WEST, S. C. 2010. Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes Dev,* 24**,** 1559-69.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.

ROBINSON, T., ROGERS, D. & WILLIAMS, B. 1997. Mapping tsetse habitat suitability in the common fly belt of southern Africa using multivariate analysis of climate and remotely sensed vegetation data. *Med Vet Entomol,* 11**,** 235-45.

ROUX, K. J., KIM, D. I. & BURKE, B. 2013. BioID: a screen for protein-protein interactions. *Curr Protoc Protein Sci,* 74**,** Unit 19.23.

RUDENKO, G. 1999. Mechanisms mediating antigenic variation in Trypanosoma brucei. *Mem Inst Oswaldo Cruz,* 94**,** 235-7.

RUDENKO, G. 2005. Maintaining the protective variant surface glycoprotein coat of African trypanosomes. *Biochem Soc Trans,* 33**,** 981-2.

RUSSO, D. C., GRAB, D. J., LONSDALE-ECCLES, J. D., SHAW, M. K. & WILLIAMS, D. J. 1993. Directional movement of variable surface glycoprotein-antibody complexes in Trypanosoma brucei. *Eur J Cell Biol,* 62**,** 432-41.

RUTTO, J. J., OSANO, O., THURANIRA, E. G., KURGAT, R. K. & ODENYO, V. A. 2013. Socio-economic and cultural determinants of human african trypanosomiasis at the Kenya - Uganda transboundary. *PLoS Negl Trop Dis,* 7**,** e2186.

SAKURAI, S., KITANO, K., YAMAGUCHI, H., HAMADA, K., OKADA, K., FUKUDA, K., UCHIDA, M., OHTSUKA, E., MORIOKA, H. & HAKOSHIMA, T. 2005. Structural basis for recruitment of human flap endonuclease 1 to PCNA. *The EMBO Journal,* 24**,** 683-693.

SAMANOVIC, M., MOLINA-PORTELA, M. P., CHESSLER, A. D., BURLEIGH, B. A. & RAPER, J. 2009. Trypanosome lytic factor, an antimicrobial high-density lipoprotein, ameliorates Leishmania infection. *PLoS Pathog,* 5**,** e1000276.

SAMBROOK, J., FRITSCH, E. & MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual,* New York, Cold Spring Harbour Laboratory Press.

SARBAJNA, S. & WEST, S. C. 2014. Holliday junction processing enzymes as guardians of genome stability. *Trends Biochem Sci,* 39**,** 409-19.

SHIFLETT, A. M., BISHOP, J. R., PAHWA, A. & HAJDUK, S. L. 2005. Human high density lipoproteins are platforms for the assembly of multi-component innate immune complexes. *J Biol Chem,* 280**,** 32578-85.

SHINOHARA, A. & OGAWA, T. 1995. Homologous recombination and the roles of double-strand breaks. *Trends in biochemical sciences,* 20**,** 387-391.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SODING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol,* 7**,** 539.

SIMARRO, P. P., CECCHI, G., FRANCO, J. R., PAONE, M., DIARRA, A., PRIOTTO, G., MATTIOLI, R. C. & JANNIN, J. G. 2015. Monitoring the Progress towards the Elimination of Gambiense Human African Trypanosomiasis. *PLoS Negl Trop Dis,* 9**,** e0003785.

SIMARRO, P. P., CECCHI, G., FRANCO, J. R., PAONE, M., DIARRA, A., RUIZ-POSTIGO, J. A., FEVRE, E. M., MATTIOLI, R. C. & JANNIN, J. G. 2012. Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl Trop Dis,* 6**,** e1859.

SIMARRO, P. P., CECCHI, G., FRANCO, J. R., PAONE, M., DIARRA, A., RUIZ-POSTIGO, J. A., MATTIOLI, R. C. & JANNIN, J. G. 2014. Mapping the capacities of fixed health facilities to cover people at risk of gambiense human African trypanosomiasis. *Int J Health Geogr,* 13**,** 4.

SIMARRO, P. P., CECCHI, G., PAONE, M., FRANCO, J. R., DIARRA, A., RUIZ, J. A., FEVRE, E. M., COURTIN, F., MATTIOLI, R. C. & JANNIN, J. G. 2010. The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *Int J Health Geogr,* 9**,** 57.

STIJLEMANS, B., CALJON, G., VAN DEN ABBEELE, J., VAN GINDERACHTER, J. A., MAGEZ, S. & DE TREZ, C. 2016. Immune Evasion Strategies of Trypanosoma brucei within the Mammalian Host: Progression to Pathogenicity. *Front Immunol,* 7**,** 233.

THOMSON, R., SAMANOVIC, M. & RAPER, J. 2009. Activity of trypanosome lytic factor: a novel component of innate immunity. *Future Microbiol,* 4**,** 789-96.

TSUTAKAWA, S. E., CLASSEN, S., CHAPADOS, B. R., ARVAI, A. S., FINGER, L. D., GUENTHER, G., TOMLINSON, C. G., THOMPSON, P., SARKER, A. H., SHEN, B., COOPER, P. K., GRASBY, J. A. & TAINER, J. A. 2011. Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell,* 145**,** 198-211.

VALENCIANO, A. L., RAMSEY, A. C. & MACKEY, Z. B. 2015. Deviating the level of proliferating cell nuclear antigen in Trypanosoma brucei elicits distinct mechanisms for inhibiting proliferation and cell cycle progression. *Cell Cycle,* 14**,** 674-688.

VAN EEDEN, S. F., KLUT, M. E., WALKER, B. A. & HOGG, J. C. 1999. The use of flow cytometry to measure neutrophil function. *J Immunol Methods,* 232**,** 23-43.

WANG, X. & HUANG, L. 2014. Defining Dynamic Protein Interactions Using SILAC-Based Quantitative Mass Spectrometry. *Methods in molecular biology (Clifton, N.J.),* 1188**,** 191-205.

WANG, Z., MORRIS, J. C., DREW, M. E. & ENGLUND, P. T. 2000. Inhibition of Trypanosoma brucei gene expression by RNA interference using an integratable vector with opposing T7 promoters. *Journal of Biological Chemistry,* 275**,** 40174-40179.

WELLS, E. A. 1972. The importance of mechanical transmission in the epidemiology of nagana: a review. *Trop Anim Health Prod,* 4**,** 74-88.

WHO 2013. Control and Surveillance of Human African Trypanosomiasis. Report of a WHO Expert Committee. . *WHO Technical Report Series 984.* Geneva, Switzerland.

WHO. 2016. *Trypanosomiasis, Human African (Sleeping Sickness)* [Online]. World Health Organisation. Available: http://www.who.int/mediacentre/factsheets/fs259/en/ [Accessed June 2016].

WRIGHT, W. D., EHMSEN, K. T. & HEYER, W. D. 2011. Assays for structure-selective DNA endonucleases. *Methods Mol Biol,* 745**,** 345-62.

WYATT, H. D. & WEST, S. C. 2014. Holliday junction resolvases. *Cold Spring Harb Perspect Biol,* 6**,** a023192.

ZIEGELBAUER, K. & OVERATH, P. 1993. Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei. *Infect Immun,* 61**,** 4540-5.

ZIMMERMANN, H., SUBOTA, I., BATRAM, C., KRAMER, S., JANZEN, C. J., JONES, N. G. & ENGSTLER, M. 2017. A quorum sensing-independent path to stumpy development in Trypanosoma brucei. *PLoS Pathog,* 13**,** e1006324.

Appendix

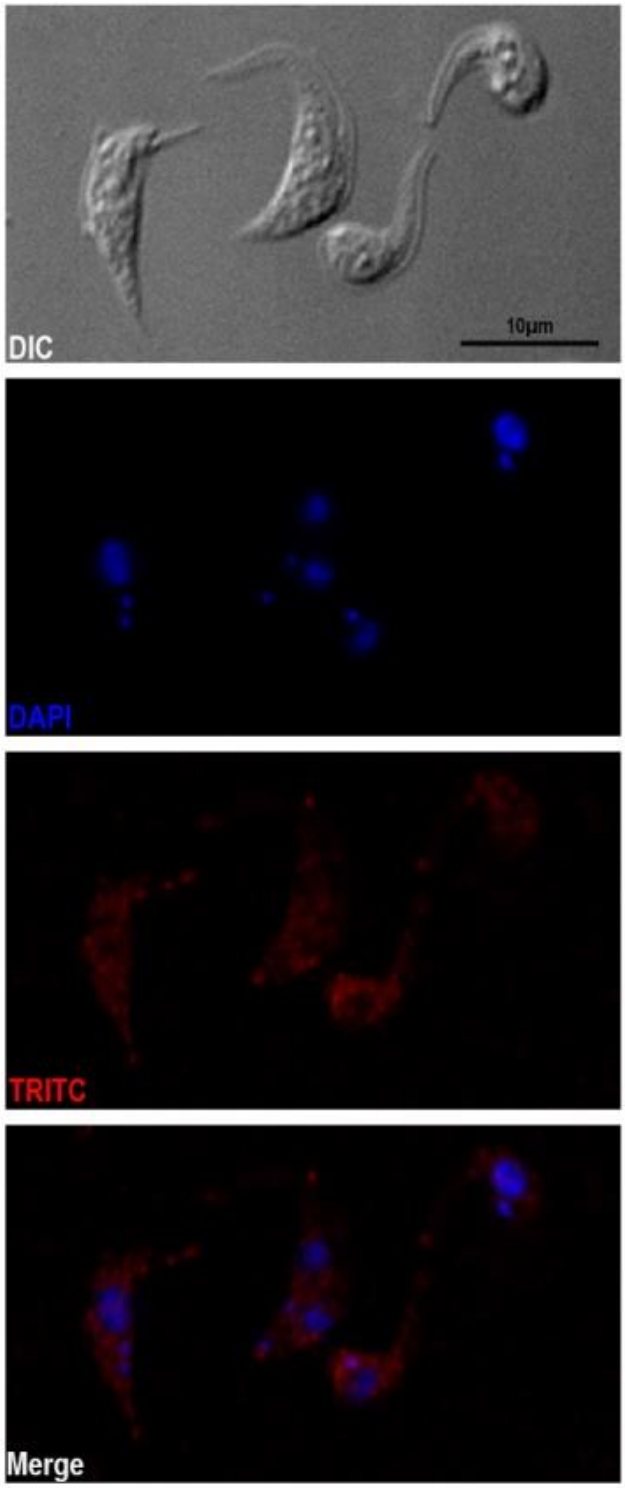
A. *Tb*427tmp.211.2870 – Putative 5'*Nde*I RAD2, *Trypanosoma brucei* Lister strain 427

```
TbRAD2       MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60
             MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI
427 RAD2     MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60

TbRAD2       LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120
             LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV
Sbjct  61    LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120

TbRAD2       HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE  180
             HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKR RE
427 RAD2     HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRXRE  180

TbRAD2       VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240
             VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA
427 RAD2     VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240

TbRAD2       LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300
             LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD
427 RAD2     LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300

TbRAD2       NAIKSDPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD  360
             NAIK DPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD
427 RAD2     NAIKXDPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD  360

TbRAD2       TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420
             TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR
427 RAD2     TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420

TbRAD2       KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480
             KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY
427 RAD2     KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480

TbRAD2       VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540
             VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC
427 RAD2     VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540

TbRAD2       GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600
             GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC
427 RAD2     GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600

TbRAD2       SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA  660
             SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCA
427 RAD2     SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAX  660

TbRAD2       PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720
             PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR
427 RAD2     PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720

TbRAD2       ERVIFRKQPPKFSEALSYLRAARGDP  746
             ERVIFRKQPPKFSEALSYLRAARGDP
427 RAD2     ERVIFRKQPPKFSEALSYLRAARGDP  746
```

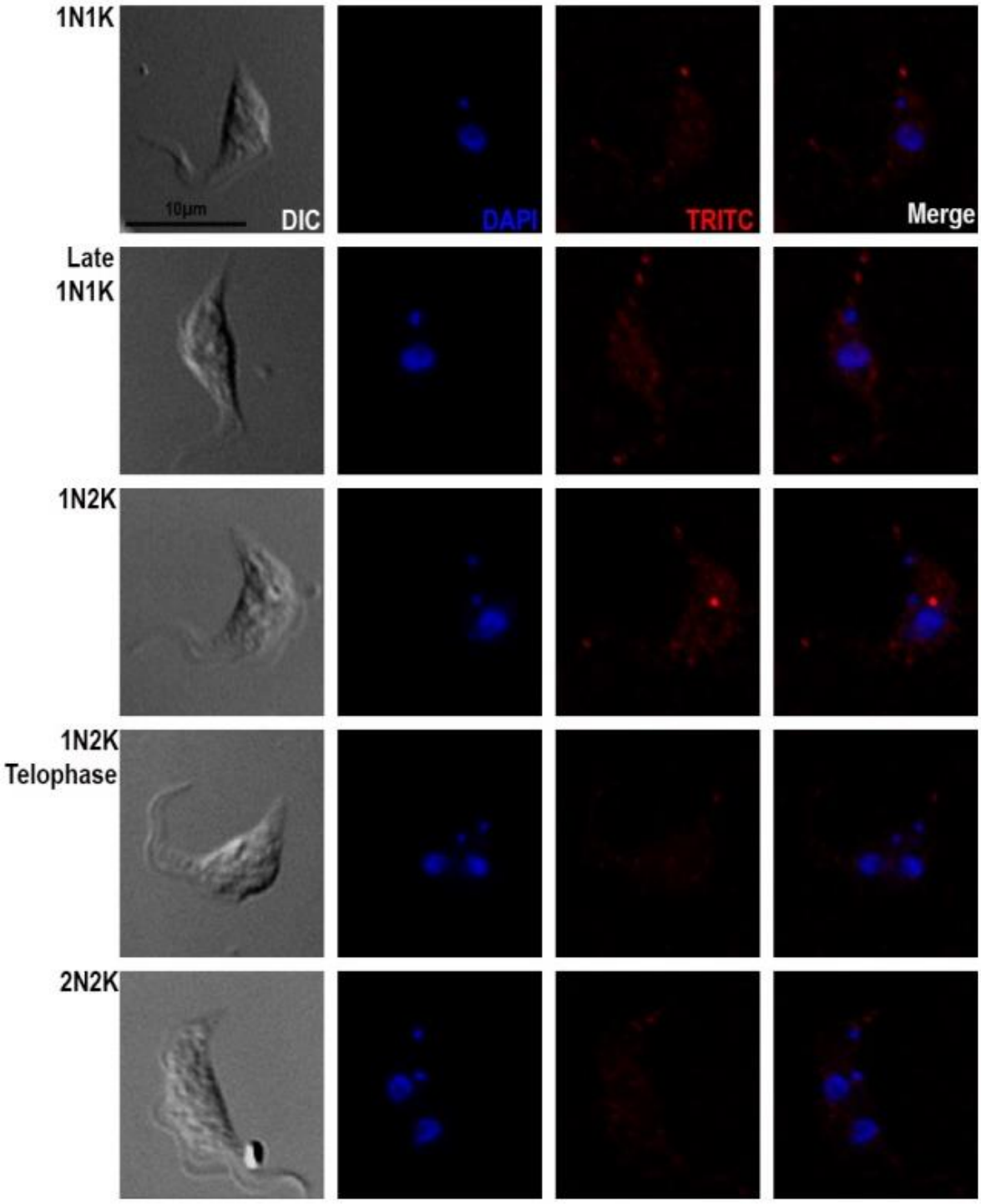
B. *Tb*927.9.11760 – Putative 5'*Nde*I RAD2, *Trypanosoma brucei* TREU927

```
TbRAD2      MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60
            MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI
927 RAD2    MGVHGLWRLLDTFGEVTQPADWKGKRVAIDASIWIAQFRSSCEPGESVEERILEGFFMRI  60

TbRAD2      LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120
            LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV
927 RAD2    LKLLFYGIEPIFVFDGPSTMSKRAEQRRRAQHREALEQAMVTRHARRLIAAQMSAGLLDV  120

TbRAD2      HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE  180
            HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE
927 RAD2    HSLPRKYRSPGSGKKLQKPLRQSLPPTDLLHDVDEDVGESCVETGTILLQPKGRKKRTRE  180

TbRAD2      VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240
            VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA
927 RAD2    VCLAPEVVSRSLTHSFLSEAEIFLEQRKTFEKFHENNRLAYTSTSIFMGPRRVAEEVSRA  240

TbRAD2      LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300
            LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD
927 RAD2    LGGATRGEAESIQGSSAGNSSSSSVLVEGVGSAAIVVEEECGDSVCEILSSSSCSVIVVD  300

TbRAD2      NAIKSDPHTVDAFHHNVSFGKEEESTSDEVEVLSRGDYWSCADNDCDDLLSLAASDRTPD  360
            NAIK+DPH VDAFHHNVSFGKEEESTSDEVEVLS GDYWSCADNDCDDLLSLAASDRTPD
927 RAD2    NAIKTDPHAVDAFHHNVSFGKEEESTSDEVEVLSSGDYWSCADNDCDDLLSLAASDRTPD  360

TbRAD2      TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420
            TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR
927 RAD2    TQCNDSTHLWYPGTQLLGGLGSADDGGIVDESRDNCTETSCGLSEFNPFGGVVVPSGNLR  420

TbRAD2      KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480
            KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY
927 RAD2    KDEKEVLLNTSVITSSETLETTGIPLKVPSVSREHVREKQVVPFELLGIVELLDCCGIPY  480

TbRAD2      VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540
            VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC
927 RAD2    VLSPNEADAQCAFLNEQRVVDAVFTEDSDVIVHGAPVVLRGFFSKGRHVVAYRQSDLLAC  540

TbRAD2      GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600
            GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC
927 RAD2    GVDKVVLVALALLLGCDYAEGVNGLSLLESLHVIAATWRQTTNSVEGGAEQVRDMLSSWC  600

TbRAD2      SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA  660
            SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA
927 RAD2    SAVRRRRIPWGEDVPLTRFYRNYVKWSTLQLADSFPESHVVDAYFNPTVNTDTRPFVCAA  660

TbRAD2      PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720
            PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR
927 RAD2    PDWTKLRLFASMHGILNKKYCGERLENAQRECQRRQPPSGDPADSAQRRLTDFFSPLPNR  720

TbRAD2      ERVIFRKQPPKFSEALSYLRAARGDP  746
            ERVIFRKQPPKFSEALSYLRAARGDP
927 RAD2    ERVIFRKQPPKFSEALSYLRAARGDP  746
```

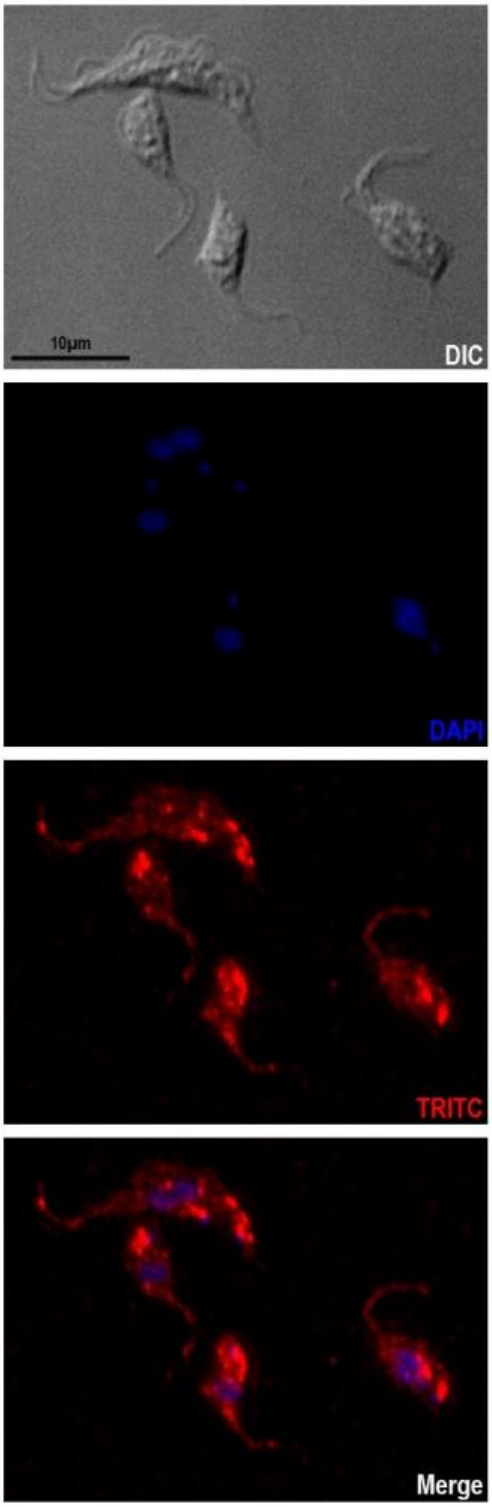
Appendix 1. Amplified Sequence Comparison of *Tb*RAD2. Alignment results of the sequenced *Tb*FEN1 and *Tb*RAD2 protein against known sequences (A) *T. brucei* TREU 927 (B) *T. brucei* lister strain 427 and (C) *T. brucei* TREU 927. Where red amino-acids are not the same between the two compared sequences, green amino-acids are the same between the compared sequences but not between alignments B and C. The X symbol shows were the amino-acid is not specific in the known sequence. Results obtained by EMBL-EBI's Clustal Omega (Goujon et al., 2010, Sievers et al., 2011, McWilliam et al., 2013).



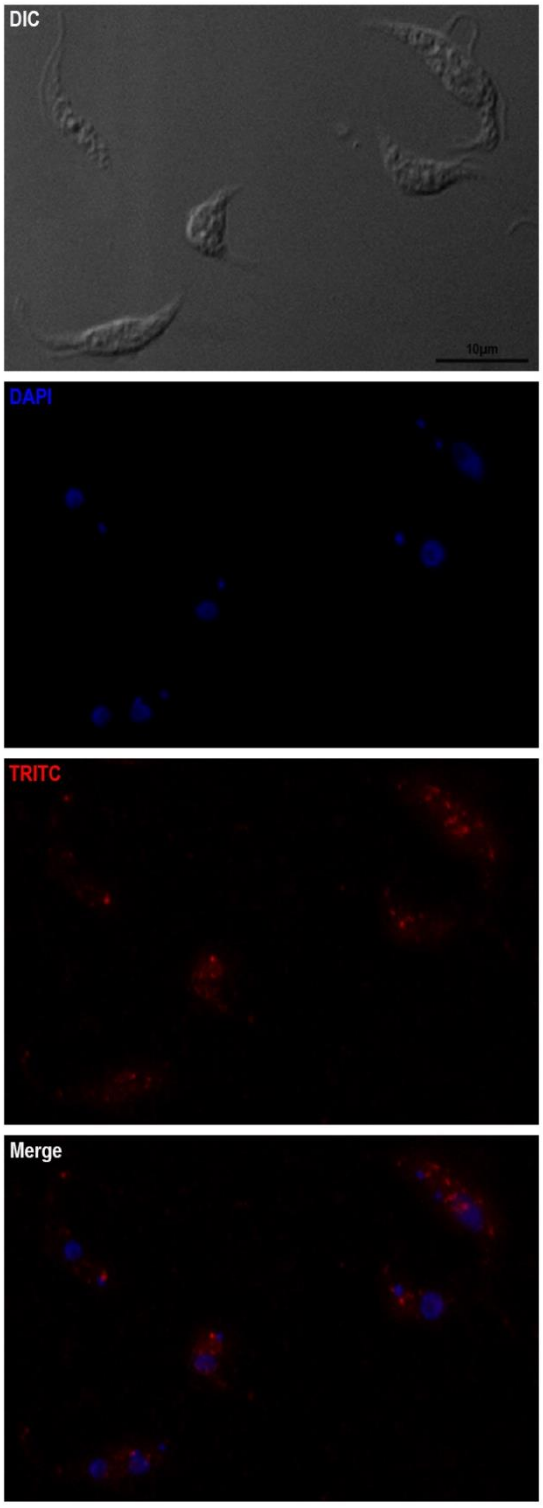
Appendix 2. Final bleed 308 1:200 anti-sera against *T. brucei* procyclic 427. Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 308 to the native T. brucei protein FEN1, in the procyclic 427 parasite strain. The anti-sera was used at a concentration of 1:200 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



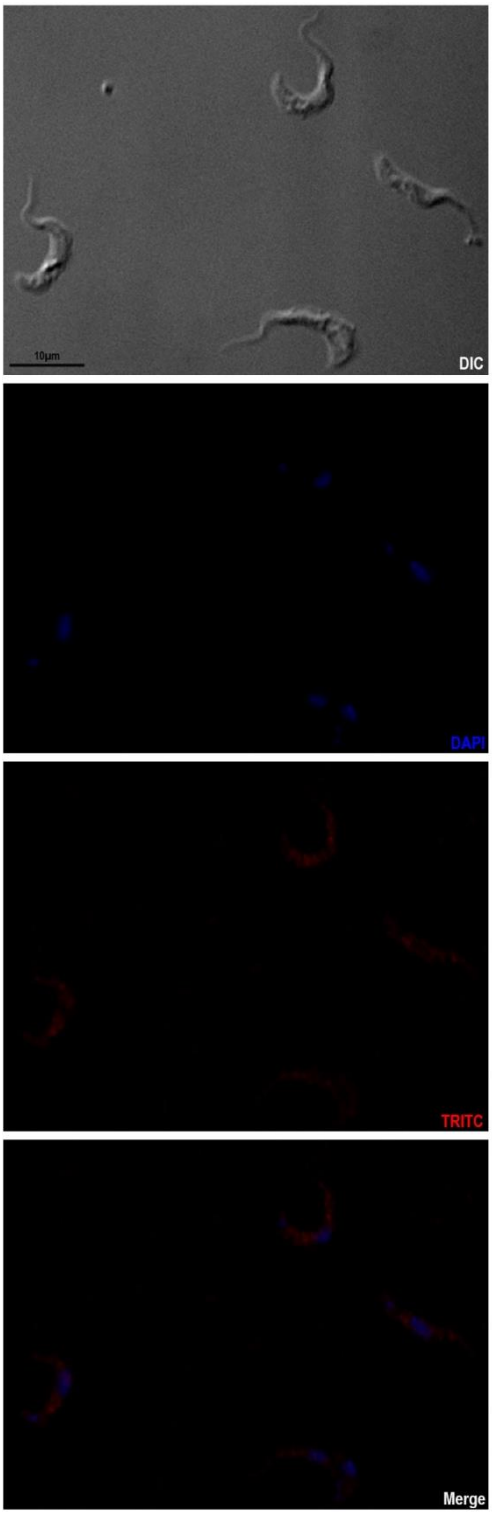
Appendix 3. Final bleed 308 1:1,000 anti-sera against *T. brucei* procyclic 427.
Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 308 to the native *T. brucei* protein FEN1, in the procyclic 427 parasite strain. The anti-sera was used at a concentration of 1:1,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



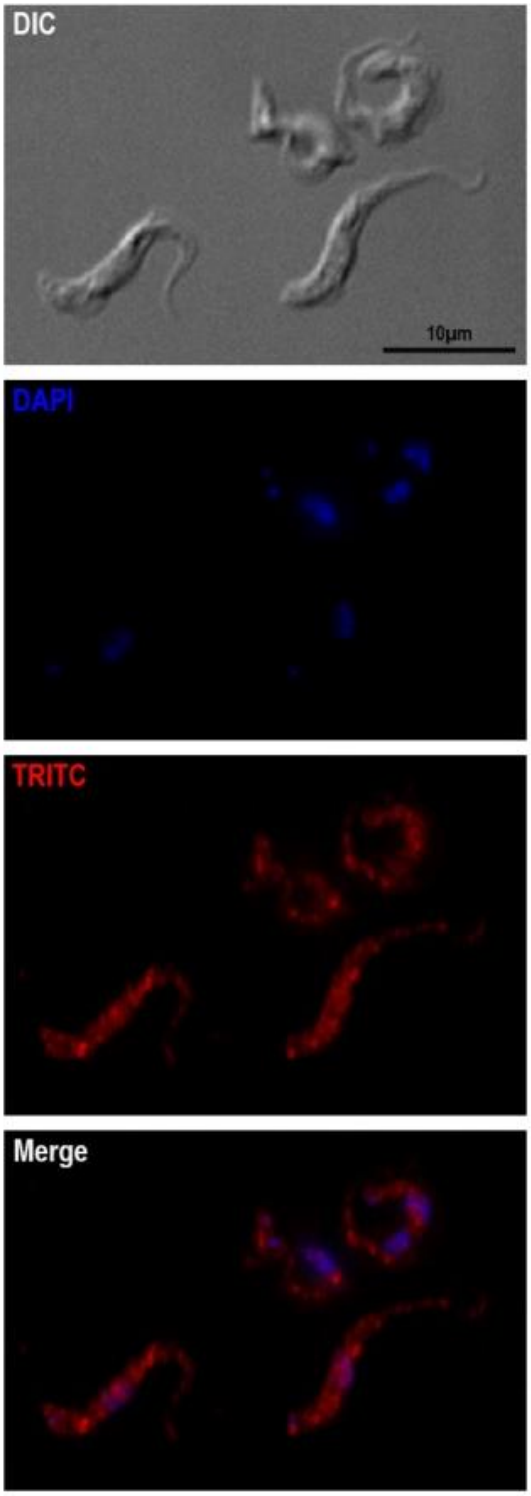
Appendix 4. Final bleed 309 1:200 anti-sera against *T. brucei* procyclic 427. Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 309 to the native *T. brucei* protein FEN1, in the procyclic 427 parasite strain. The anti-sera was used at a concentration of 1:200 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



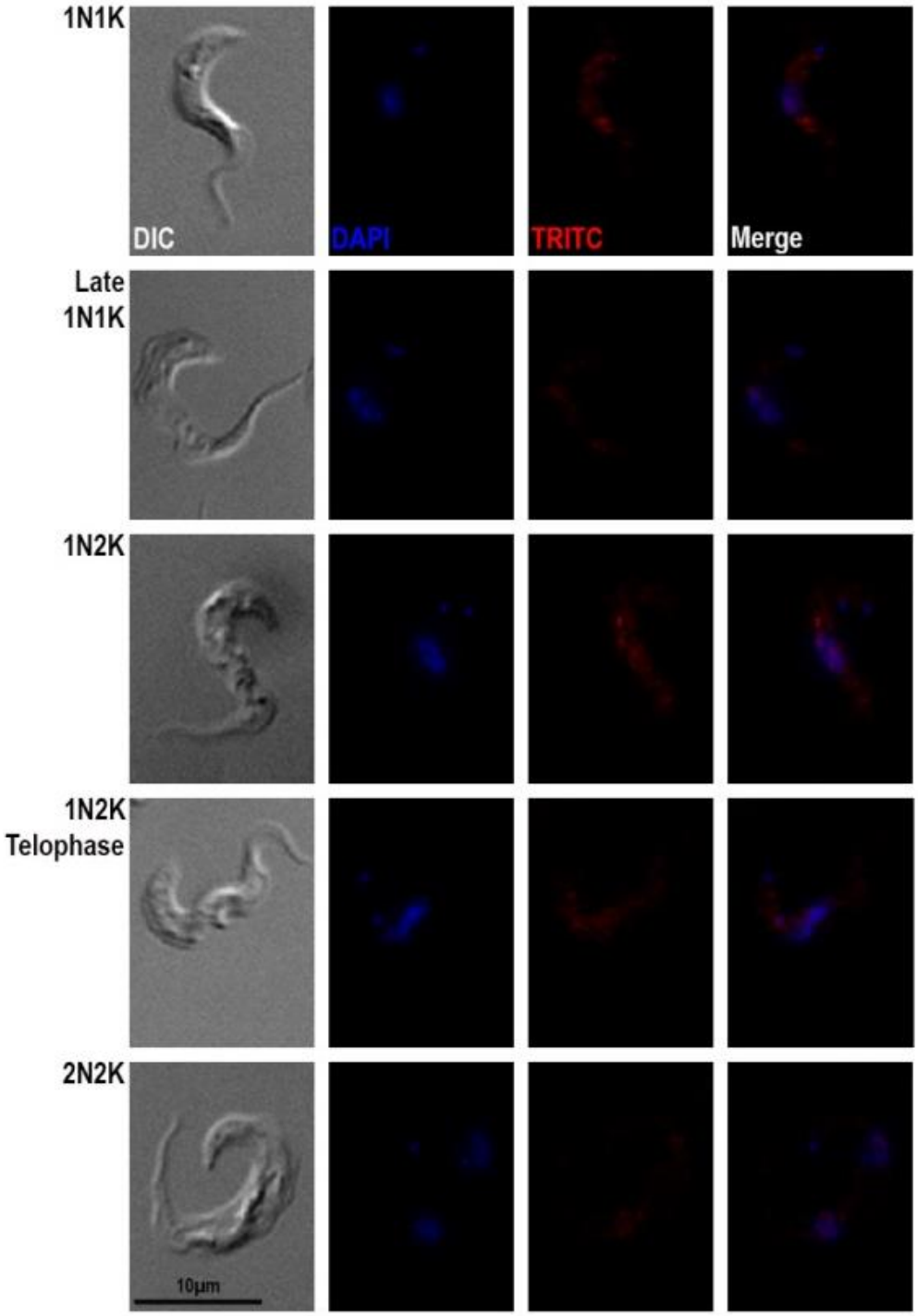
Appendix 5. Final bleed 309 1:1,000 anti-sera against *T. brucei* procyclic 427.
Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 309 to the native *T. brucei* protein FEN1, in the procyclic 427 parasite strain. The anti-sera was used at a concentration of 1:1,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



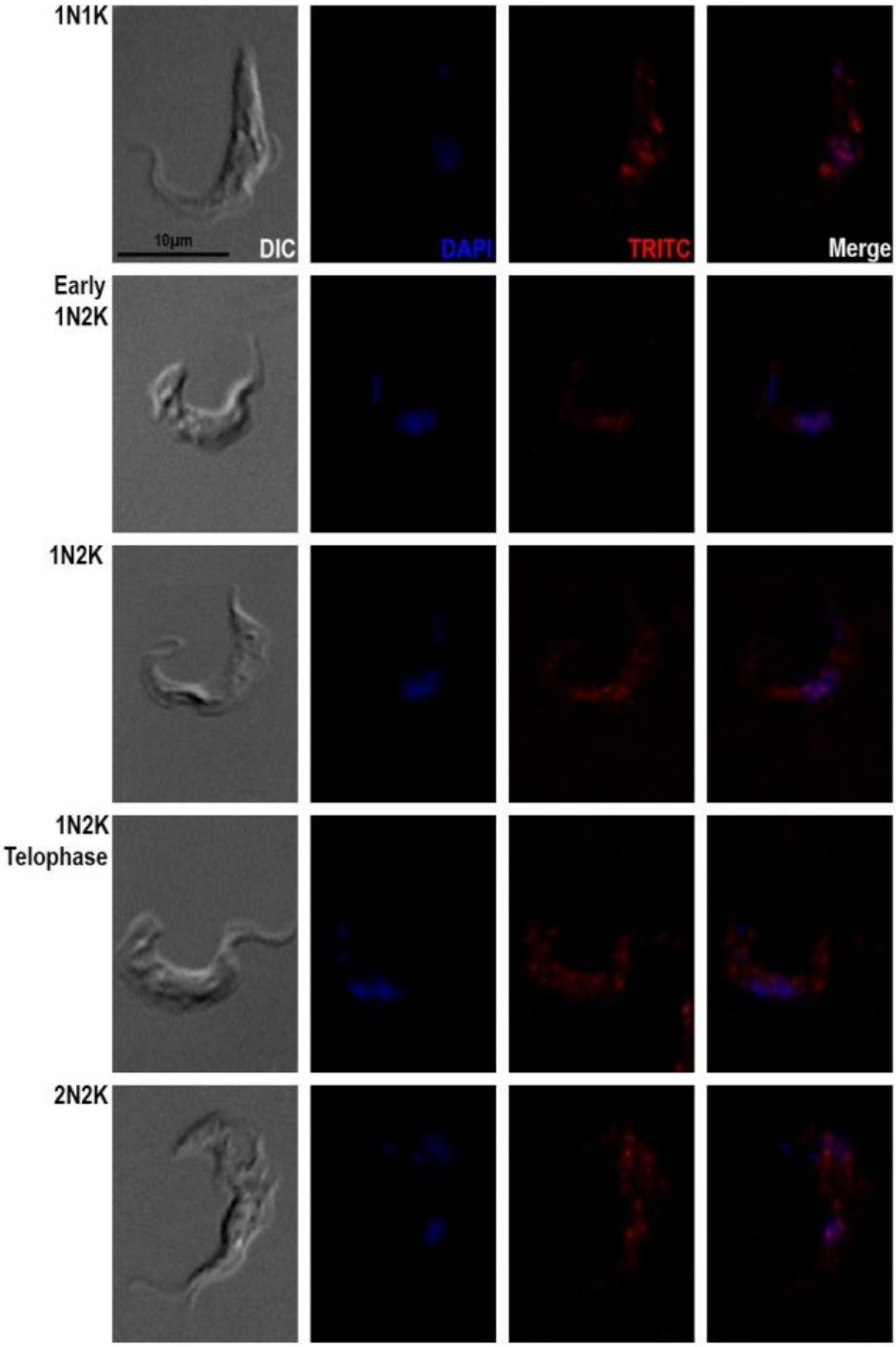
Appendix 6. Pre-immune 308 1:1,000 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy panels showing the affinity of the pre-immune bleed anti-sera of rabbit 308 to the native *T. brucei* protein FEN1, in the bloodstream 427 parasite strain. The anti-sera was used at a concentration of 1:1,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



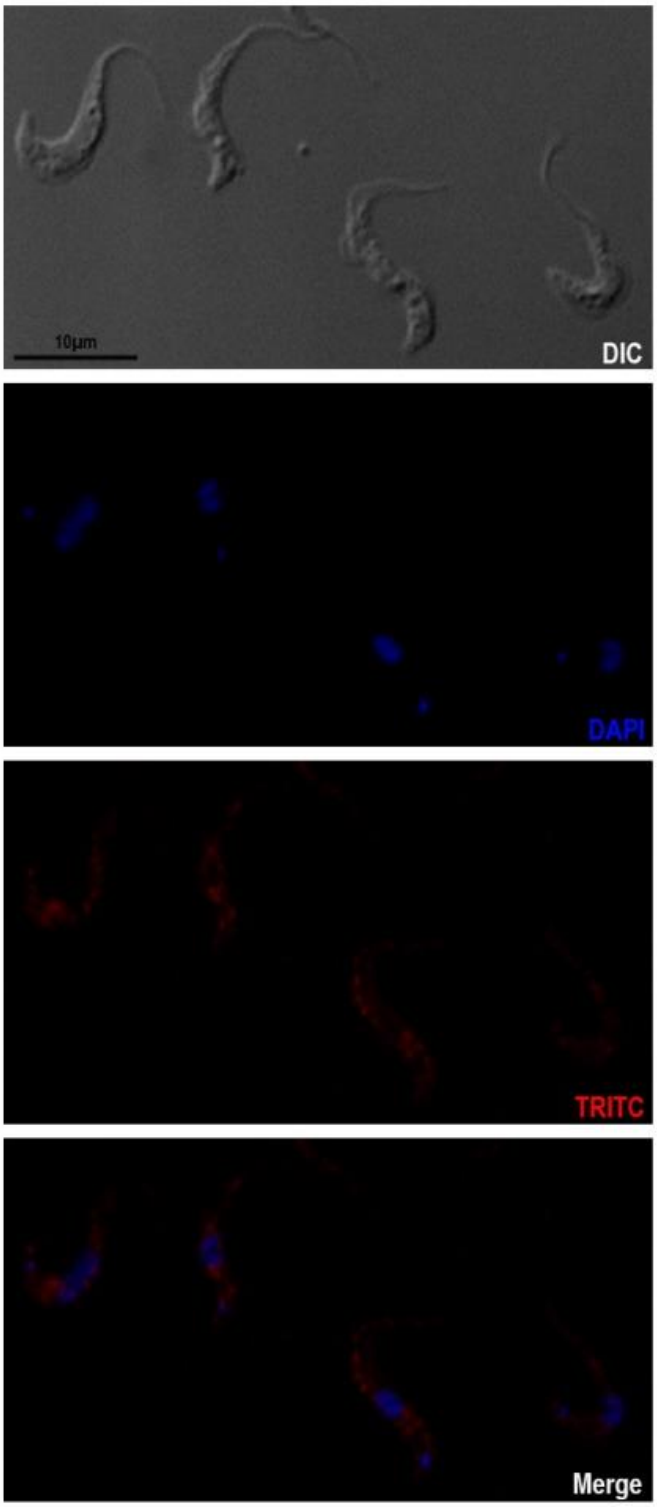
Appendix 7. Final bleed 308 1:200 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 308 to the native *T. brucei* protein FEN1, in the bloodstream 427 parasite strain. The anti-sera was used at a concentration of 1:200 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



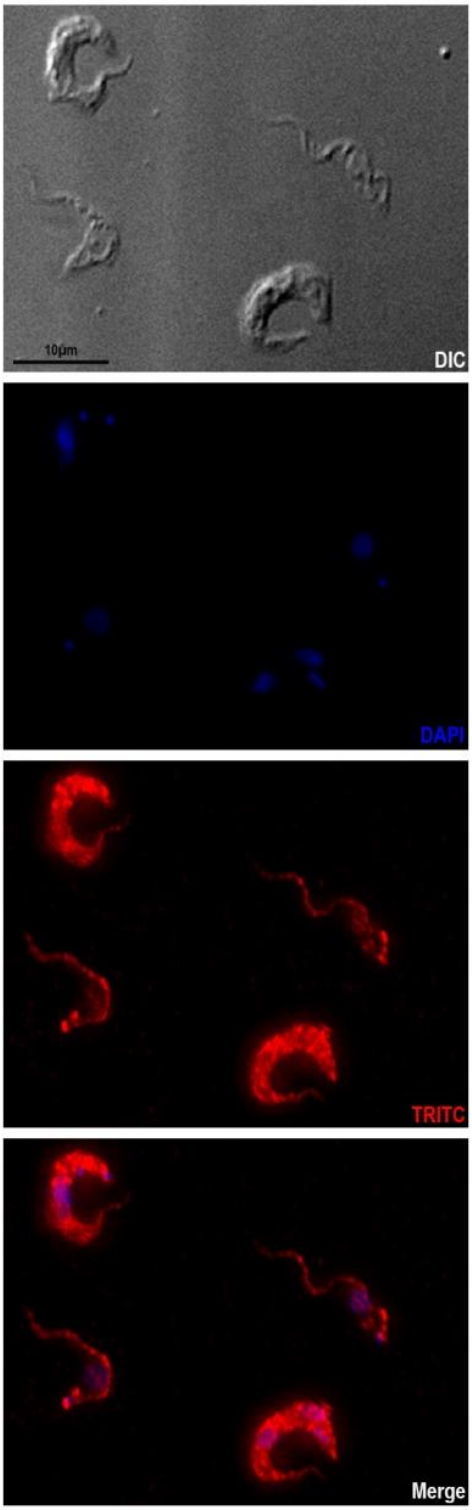
Appendix 8. Final bleed 308 1:1,000 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy images showing the affinity of the final bleed anti-sera of rabbit 308 to the native *T. brucei* protein FEN1 during the different stages of the bloodstream 427 parasite cell cycle. The anti-sera was used at a concentration of 1:1,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



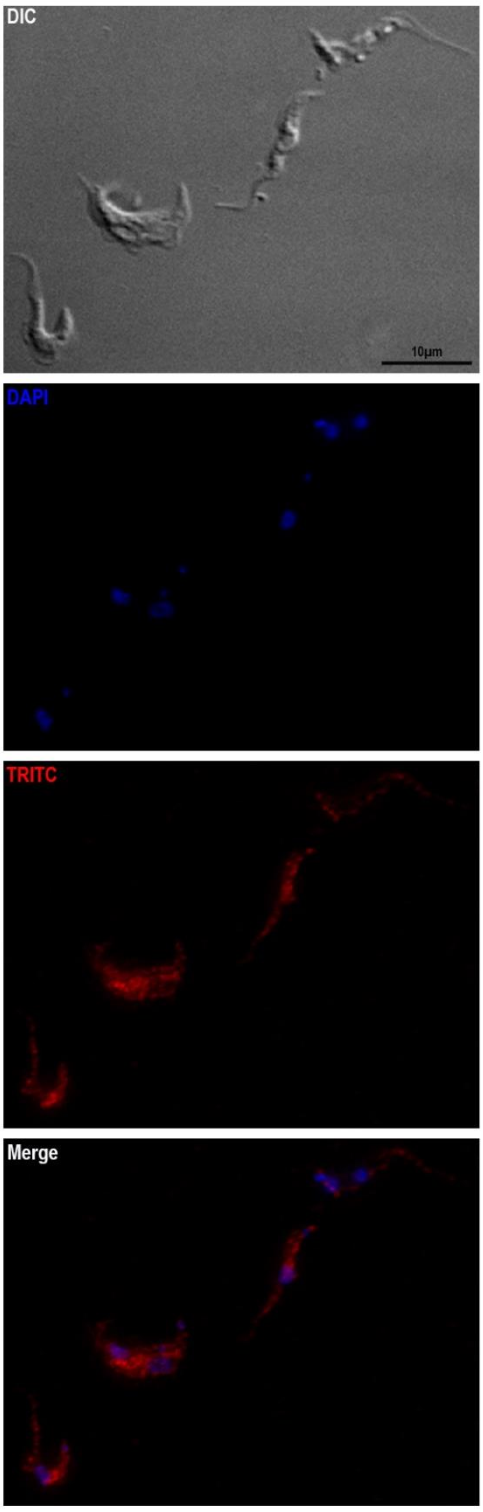
Appendix 9. Final bleed 308 1:5,000 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy images showing the affinity of the final bleed anti-sera of rabbit 308 to the native *T. brucei* protein FEN1 during the different stages of the bloodstream 427 parasite cell cycle. The anti-sera was used at a concentration of 1:5,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



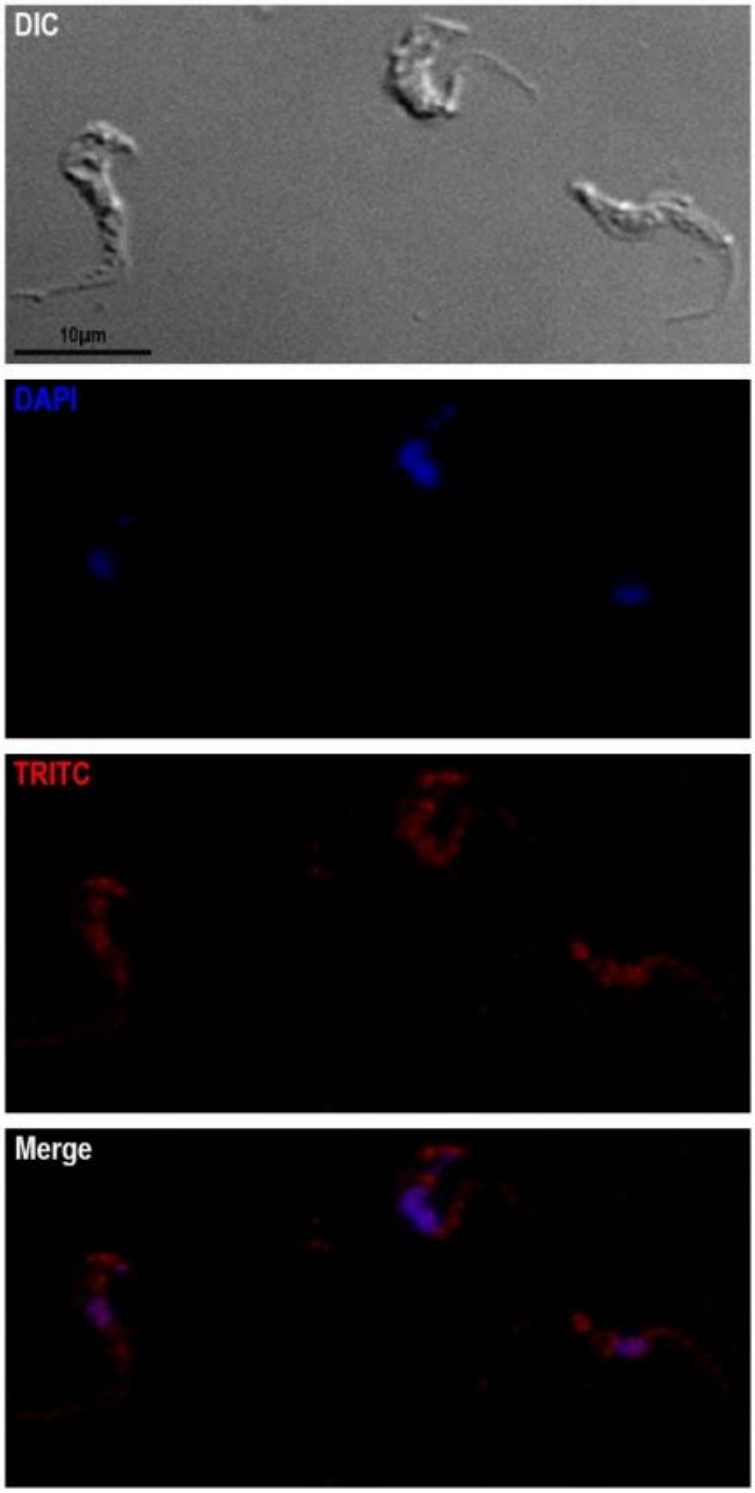
Appendix 10. Pre-immune 309 1:1,000 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy panels showing the affinity of the pre-immune bleed anti-sera of rabbit 309 to the native *T. brucei* protein FEN1, in the bloodstream 427 parasite strain. The anti-sera was used at a concentration of 1:1,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



Appendix 11. Final bleed 309 1:200 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 309 to the native *T. brucei* protein FEN1, in the bloodstream 427 parasite strain. The anti-sera was used at a concentration of 1:200 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



Appendix 12. Final bleed 309 1:1,000 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 309 to the native *T. brucei* protein FEN1, in the bloodstream 427 parasite strain. The anti-sera was used at a concentration of 1:1,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.



Appendix 13. Final bleed 309 1:5,000 anti-sera against *T. brucei* bloodstream 427.
Fluorescent microscopy panels showing the affinity of the final bleed anti-sera of rabbit 309 to the native *T. brucei* protein FEN1, in the bloodstream 427 parasite strain. The anti-sera was used at a concentration of 1:5,000 and was detected using the anti-rabbit 568 secondary to interact with TRITC excitation and emission wavelengths. Images were taken using the Deltavision Elite.