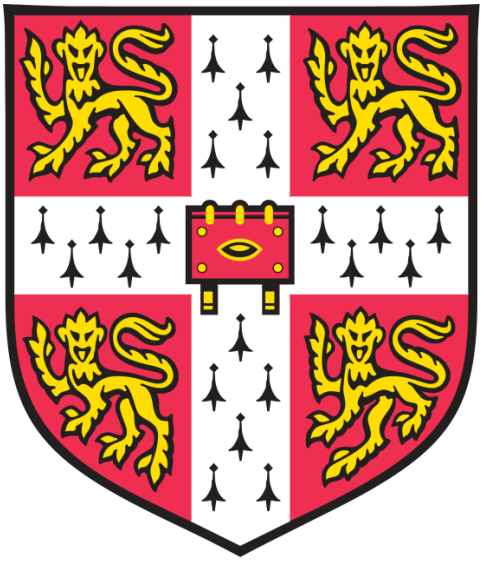
Structure of the trypanosome transferrin receptor and insights into ligand binding and therapeutic strategies



Camilla E. Trevor

Department of Biochemistry

University of Cambridge

This dissertation is submitted for the Degree of *Doctor of Philosophy*

King's College

March 2020

Declaration

This dissertation is the product of my own work and includes nothing which is the outcome of work done in collaboration, except as declared here and explicitly stated in the text. No part of this work is substantially the same as any that I have submitted, or, is being concurrently submitted for a degree, diploma, or other qualification at the University of Cambridge, at any other University, or at a similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma, or other qualification at the University of Cambridge, at any other University, or at a similar institution, except as declared here and explicitly stated in the text. Following the Degree Committee for the Faculty of Biology guidelines, this thesis does not exceed 60,000 words, excluding figures, bibliography, and appendices.

Camilla E. Trevor

March 2020

Acknowledgements

First and foremost, I would like to thank my supervisors, Mark Carrington and Andrea Gonzalez-Muñoz, for providing valuable support, guidance and mentorship over the past four years. I have enjoyed the fruitful discussions and I am particularly grateful for having the opportunity to work on this project. I would also like to thank Steven Rust, Ralph Minter and Tris Vaughan for providing help and advice during my time at AstraZeneca. I would also like to thank the BBSRC and AstraZeneca for funding the studentship.

In addition, I am grateful to all the collaborators who have contributed to the project. I would like to thank Elspeth Garman and Peter Woodcock (University of Oxford), and Gareth Rees and Alan Sandercock (AstraZeneca) for experimental data and guidance. I am also grateful to the tissue culture and biologics expression teams (AstraZeneca) for their support during the project. In particular, I would like to express my profound gratitude to Matthew Higgins (University of Oxford), for sharing his abundant expertise and providing the opportunity to work alongside him in Oxford, a privilege which I found both inspiring and rewarding.

I would like to thank past and present members of the Carrington lab (University of Cambridge) and the Ralph Minter team (AstraZeneca) for your advice and for filling the past years with joy. I would especially like to thank Olivia, Isobel, Andrea, Stephen, Sophia and Florent for the friendship, support, laughter and wonderful times. I am also grateful to King's College for the graduate student network and all the fantastic opportunities and distractions that this has provided. Finally, I am indebted to all of my family, friends and Sven for your unwavering support and for never failing to make me smile.

Abstract

African trypanosomes are extracellular protozoan pathogens that have evolved a complex and sophisticated interface with hosts to acquire nutrients and protect against the host immune system. The variant surface glycoprotein (VSG) forms a dense and dynamic protective coat, which serves to defend against the immunoglobulin-mediated host response. In addition, a unique set of receptors mediate uptake of host macromolecules and must comply with recognising ligands from a broad host range whilst also avoiding immune detection. Host transferrin (Tf) is acquired by trypanosomes through the expression of a transferrin receptor (TfR), which does not share homology with the host TfR. Multiple diverse copies of the TfR are present in the trypanosome genome but only one is expressed at a time. Several hypotheses revolve around the evolutionary advantage of multiplication and diversification of the receptor, falling into two main categories. Either the TfR repertoire serves to accommodate different Tfs from the wide vertebrate host range, or variation has arisen as part of an immune evasion strategy. To elucidate the complexities surrounding the receptor, two different TfRs were expressed as recombinant proteins to study receptor-ligand interactions at the molecular level.

Biophysical analyses of binding interactions provided no evidence to support the hypothesis that the receptor repertoire had evolved to favour host promiscuity. Instead, a single receptor was capable of binding Tf from multiple different mammals, highlighting the lack of requisite to evolve species-specific receptors. Structural determination of the trypanosome TfR in conjunction with sequence variation mapping provided further evidence. The structure revealed a predominantly conserved Tf binding site, while variation and N-linked glycans resided in the surrounding accessible regions at the apical surface of the receptor, indicating that receptor evolution was likely driven by an immune evasion strategy.

The findings presented in this thesis have served to further our understanding of the molecular architecture of the trypanosome cell surface by providing structural and biophysical data to define a complex host-parasite interaction. Therapeutic strategies have also been adapted through advances in our knowledge of the relationship between trypanosome and host. Finally, the trypanosome TfR is a remarkable example of evolutionary drift within a receptor, propagated by the advantage conferred by immune avoidance, and restricted by functional requirements of ligand uptake. Thus, the TfR repertoire has likely derived from an antigenic variation strategy to promote long-term persistence in the host.

Table of Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 African trypanosomes

Trypanosomatidae are a family of unicellular protozoa that diverged from other eukaryotes approximately 500 million years ago (Lukeš *et al*., 2014). Within the family are the African trypanosomes, a group of extracellular pathogens with an interesting set of challenges, as the parasite must hijack nutrients from the host environment whilst also avoiding immune recognition. *Trypanosoma brucei* is the most studied group of African trypanosomes and is subdivided into three subspecies: *Trypanosoma brucei (T. b.) gambiense*, *T. b. rhodesiense* and *T. b. brucei*. While all three can infect vertebrates, the two former are the only primate-infective African trypanosomes. Infection of animals causes a disease referred to as animal trypanosomiasis, or nagana, a wasting disease that can be lethal if left untreated. Nagana poses a huge economic burden in sub-Sahahran Africa due to loss of livestock. *T. b. gambiense* and *T. b. rhodesiense* are responsible for human African trypanosomiasis (HAT), commonly referred to as sleeping sickness. HAT is endemic in 36 countries in sub-Saharan Africa (Figure 1A). *T. b. gambiense* HAT accounts for 97% of HAT cases and is predominantly an anthroponosis prevalent in West Africa. On the other hand, *T. b. rhodesiense* HAT is located in East Africa and is a zoonosis, with the large animal reservoir complicating reservoir control programs (Simarro *et al*., 2008). Recent efforts focussed on eliminating HAT have shown promising results, with fewer than 3000 cases reported to the World Health Organisation in 2015 (Figure 1B). The current trend of reported cases falls below the forecasted number of cases and suggests that imminent HAT elimination may be achievable.

T. brucei is transmitted to the host via an insect vector, the tsetse fly (*Glossina* species). The parasite has adapted accordingly to proliferate in a variety of different environments such as the mammalian host bloodstream and tissues, and the insect vector midgut and salivary gland. *T. brucei* adopts a variety of different cell forms to maximise infectivity and survival in diverse settings. In its simplest form adapted for laboratory culture, the trypanosome life cycle can be compartmentalised into bloodstream-form cells that propagate in the mammalian host and procyclic-form cells that colonise the tsetse fly (Vickerman, 1985). Trypanosomes have evolved a sophisticated genome architecture to promote survival within the challenging environments.

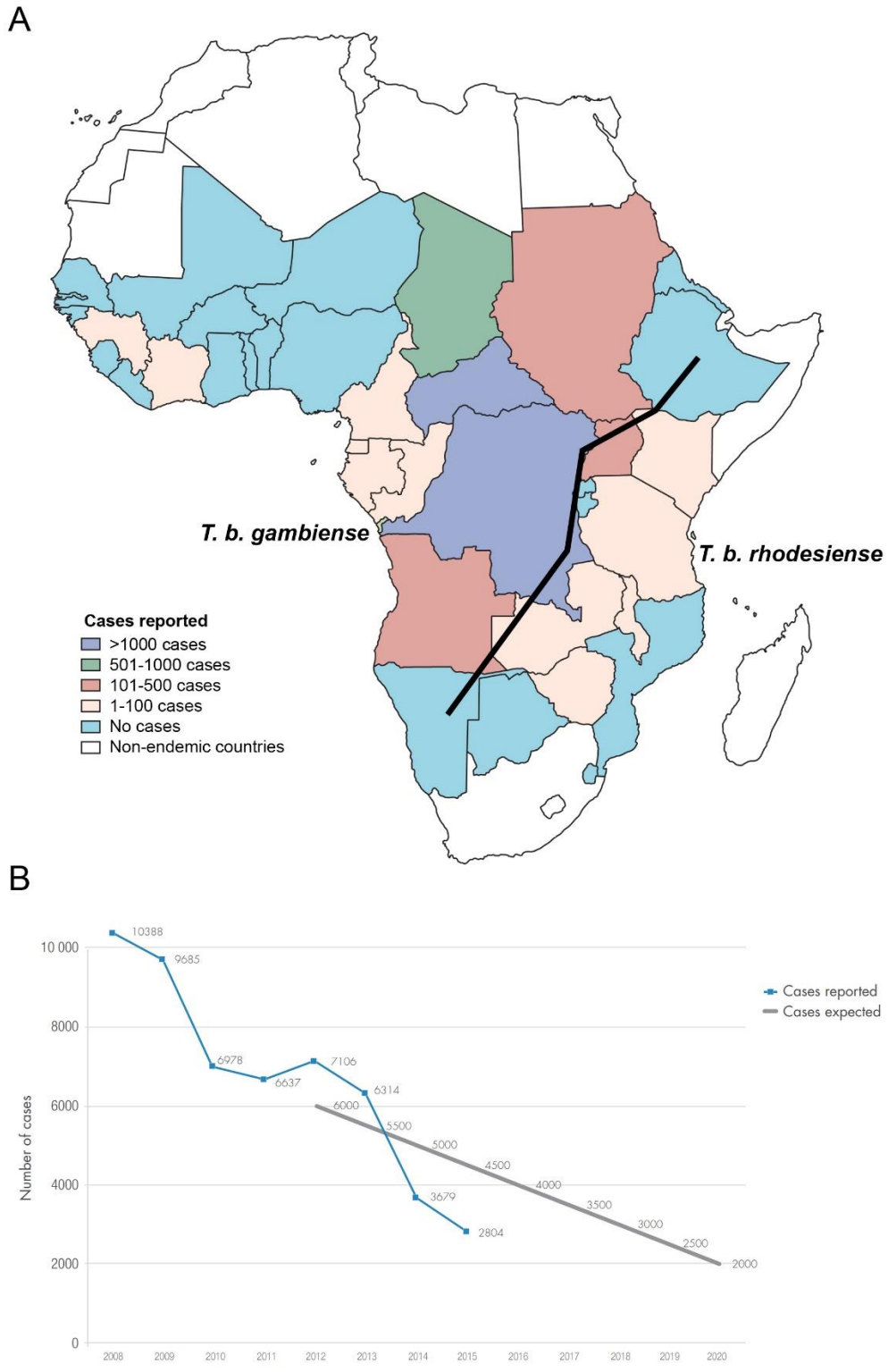


Figure 1: HAT epidemiology. A, Geographical distribution of HAT cases reported in 2009 (adapted from Kennedy, 2013). A black line depicts the separation between *T. b. gambiense* in West sub-Saharan Africa and *T. b. rhodesiense* in the East. B, Number of HAT cases reported to the World Health Organisation from 2008-2015 (from Fourth WHO report on neglected tropical diseases, 2017).

1.2 Genome architecture

The *T. brucei* nuclear genome is primarily organised into 11 megabase chromosomes, in addition to the lesser characterised ~5 intermediate chromosomes and more than 100 mini-chromosomes (Berriman *et al*., 2005). The large number of chromosomes results in over 250 telomeres, which were found to encode most of the variant surface glycoprotein (VSG) genes in sub-telomeric arrays. A small group of VSG genes are located in organised telomeric expression sites (Müller *et al*., 2018). The VSG (M_r = 50,000 – 65,000) is the major cell surface protein in trypanosomes, forming a dense and dynamic coat of $1x10^7$ VSG copies (Bartossek *et al*., 2017). The inherent instability of telomere-proximal regions has resulted in multiplication and diversification of VSGs, with identification of more than 1500 different VSG genes throughout the genome. Many reside as incomplete genes, with more than 80% of the 'VSGnome' composed of pseudogenes, and recombination is required to form a functional VSG (Cross *et al*., 2014). The large and diverse repertoire of VSGs underpins the success of antigenic variation in trypanosomes.

Telomeric expression sites, referred to as bloodstream expression sites (BESs), encode VSG and a set of expression site associated genes (ESAGs) which have also arisen through duplication and diversification (Hertz-Fowler *et al*., 2008). A typical BES contains an RNA polymerase (pol) I promoter in proximity to the ESAG7 and ESAG6 genes, an array of other ESAG1-12 genes, and a VSG encoded at the 3' end (Figure 2). DNA repeats are also present within intergenic regions, contributing to the instability of the telomeric region. Through duplication events, the set of genes has multiplied to occupy approximately 15 different BES with conserved architectures in the L427 isolate (Hertz-Fowler *et al*., 2008). Over time, the ESAG and VSG genes have diversified between expression sites. While the function of most ESAGs remain unknown, the ESAG6 and ESAG7 genes were shown to encode the two subunits of the trypanosome transferrin receptor (TfR) (Salmon *et al*., 1994) (Steverding *et al*., 1994). Only one VSG and TfR are expressed at a time, governed by tight regulation at the transcriptional level.

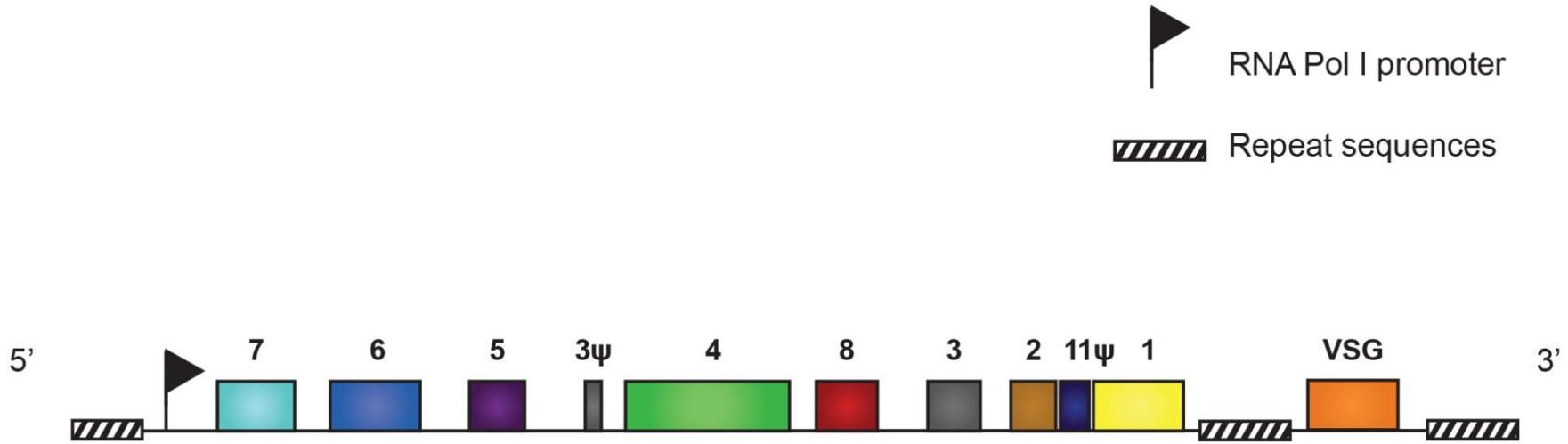


Figure 2: Organisation of a typical BES. Schematic representation of a BES encoding VSG and ESAG1-11 genes or pseudogenes, adapted from Hertz-Fowler *et al*. (2008). At the 5’ end, 50 bp repeats are often located upstream of the RNA pol I promoter, while VSG is flanked with 70 bp repeats at the 5’ end and telomere repeats at the 3’ end. Interspersed between promoter and VSG are the ESAG genes, which can range from ESAG1 to ESAG12.

1.2.1 Dynamics orchestrated at the transcriptional level

Chromosome internal protein coding genes are transcribed by RNA pol II, while the ~15 BESs are transcribed by RNA pol I (Kooter & Borst, 1984) (Günzl *et al*., 2003), a polymerase usually reserved for transcription of ribosomal DNA in eukaryotes. In trypanosomes, RNA pol I is localised in a discrete extranucleolar structure, named the expression site (ES) body (Navarro & Gull, 2001). The ES body imposes stringent restrictions on expression, with only one BES expressed at a time to maintain mono-allelic VSG expression. Recent studies have shown that under antibiotic selection, two BESs can occupy the same ES body, yet this process is reversible once the selective pressure is removed (Budzak *et al*., 2019). The factors controlling allelic exclusion of other BESs under normal conditions remain unclear. The active BES is transcribed as a polycistronic unit and subsequently *trans* spliced and polyadenylated to produce mature mRNA transcripts (reviewed by Vanhamme and Pays, 1995). VSG and ESAGs are constitutively transcribed as polycistronic units, yet expression levels of the different proteins vary dramatically, which prompted an investigation into post-transcriptional regulation in trypanosomes.

1.2.2 Dynamics orchestrated at the post-transcriptional level

The VSG accounts for more than 90% of *T. brucei* cell surface proteins (Grünfelder *et al*., 2002), yet it is transcribed at the same rate as other less represented proteins such as the TfR. This raises an interesting question of how trypanosomes achieve different levels of protein expression given that regulation is not controlled by transcription. *Cis*-acting elements that affect mRNA stability and translation have been identified in the 3' untranslated region (UTR) (Furger *et al*., 1997) (Webb *et al*., 2005). In addition, a codon bias in which certain codons favour mRNA stability has been reported (Nascimento *et al*., 2018) (Jeacock *et al*., 2018). Together, these findings explain some of the differences observed in protein expression levels, although the entire process is still yet to be resolved.

1.3 Innate immunity

1.3.1 Primate resistance to *T. b. brucei* infection

Some primates have evolved to combat *T. b. brucei* infection through the assembly of two innate immune complexes, trypanosome lytic factor 1 (TLF1) and TLF2. The complexes contain two primate-specific proteins, apolipoprotein L1 (apoL1) and haptoglobin-related protein (Hpr), associated with high density lipoprotein (HDL) particles in TLF1, and IgM in TLF2 (Raper *et al*., 1999). While initial studies suggested that Hpr may be the trypanolytic component (Smith *et al*., 1995), it was later found that Hpr mediated specificity by targeting TLF to the *T. brucei* haptoblobin-haemoglobin receptor (TbHpHbR) (Vanhollebeke *et al*., 2008). Hpr shares 91% amino acid identity with haptoglobin (Maeda, 1985) and associates with haemoglobin (Nielsen *et al*., 2006). Molecular mimicry results in uptake of TLF via TbHpHbR-mediated endocytosis (Vanhollebeke *et al*., 2008). The trypanolytic factor was identified as apoL1 (Vanhamme *et al*., 2003), a pore-forming protein owing to the lytic capacity of TLFs. It was originally suggested that insertion of ApoL1 in the endosomal membrane resulted in the formation of an anion-selective pore which triggered osmotic swelling of lysosomal compartments, followed by cell lysis (Pérez-Morga *et al*., 2005). However, more recent work has shown that the pores are cation-selective and pH-dependent (Thomson & Finkelstein, 2015). At acidic pH, apoL1 is inserted into endocytic vesicle membranes, forming a low conductance pore. As the endosome is recycled to the cell surface, neutral pH increases the conductance of the pore by 3000-fold, resulting in an influx of cations, cytoplasm swelling and cell lysis (Thomson & Finkelstein, 2015). While the route of TLF1 uptake is largely understood to rely upon TbHpHbR, loss of the receptor showed that TLF2 uptake was reduced but not abolished (Bullard *et al*., 2012). TbHpHbR-independent uptake suggested an alternative route of TLF2 entry into trypanosomes although the pathway remains unknown.

1.3.2 Evolution of a unique serum resistance associated gene in *T. b. rhodesiense*

In response to host innate immunity, *T. b. rhodesiense* has co-evolved to suppress a TLF-mediated attack through expression of a serum resistance associated (SRA) gene. A unique copy of the SRA gene is encoded in a single *T. b rhodesiense* VSG expression site (Xong *et al*., 1998). The SRA protein had a predicted GPI-anchor addition site and was shown to localise to membranes of endocytic compartments (Vanhamme *et al*., 2003). Resistance was

mediated by an interaction between N-terminus of SRA and the C-terminus of ApoL1 (Vanhamme *et al*., 2003), neutralising the pore-forming activity of ApoL1. Structural studies revealed that SRA adopts a similar fold to other trypanosome cell surface proteins, with a characteristic three-helix bundle maintaining the core fold of the protein (Zoll *et al*., 2018). In particular, the structural resemblance to the VSG N-terminus is striking and suggests that SRA has evolved from a VSG, with a loss of the multiple flexible loops that form the apical surface of the VSG, instead replaced by a single disordered loop (Zoll *et al*., 2018). The array of loops at the membrane distal surface of the VSG are involved in antigenic variation, and the absence of loops in SRA is likely due to a lack of selective pressure to generate diverse loops due to the intracellular location. Ectopic expression of the SRA gene in *T. b. brucei* conferred resistance to human serum (Xong *et al*., 1998), demonstrating that SRA was both necessary and sufficient in mediating human serum resistance.

1.3.3 Evolution of multiple resistance strategies in *T. b. gambiense*

T. b. gambiense can be separated into two genetically different groups (Balmer *et al*., 2011). Group 1 has constitutive resistance to human serum while group 2 resistance is less well-defined (Capewell *et al*., 2011). *T. b. gambiense* does not express SRA and instead has evolved several strategies to promote resistance to human serum. The strategies adopted by the well-characterised group 1 population will be discussed. First, *T. b. gambiense* was shown to express fewer HpHbR copies on the cell surface. A reported 20-fold reduction of expression compared to *T. b. brucei* was associated with a decrease in Hpr binding sites led to reduction in TLF1 uptake (Kieft *et al*., 2010). Second, a single amino acid substitution was identified in the *T. b. gambiense* HpHbR (Symula *et al*., 2012). The replacement of leucine with serine at position 210 led to a dramatic reduction of TLF1 uptake, further promoting TLF1 resistance in *T. b. gambiense* (DeJesus *et al*., 2013). Finally, a *T. b. gambiense*-specific glycoprotein (TgsGP) was identified and shown to mediate resistance to TLF2 (Uzureau *et al*., 2013) (Capewell *et al*., 2013). In contrast to the SRA gene which relied upon VSG expression site activation for SRA expression, TgsGP was constitutively expressed. The combination of these evolutionary measures has led to human serum resistance in group 1 *T. b. gambiense*.

1.4 Adaptive immunity

1.4.1 Hydrodynamic flow

The high density of VSG at the cell surface confers immunogenicity, with immunoglobulin (Ig) capable of binding the accessible N-terminal membrane distal domain of the VSG while the C-terminal domain is shielded (Schwede *et al*., 2011). At low antibody titre, Ig-VSG complexes were shown to migrate to the posterior pole of the cell (Engstler *et al*., 2007). Concomitant movement of the trypanosome and hydrodynamic forces applied to macromolecules protruding at the cell surface resulted in sorting of Ig-VSG complexes towards the flagellar pocket. As the junction between cell body and flagellum, the flagellar pocket forms a site of rapid endocytosis at the surface of trypanosomes (reviewed by Field and Carrington, 2009). Ig-VSG complexes were endocytosed, providing antibody clearance to prevent opsonisation. While this mechanism is effective during the initial stages of an emerging antibody response, hydrodynamic flow-mediated strategies cannot cope with the high antibody titres that develop with long-term infection (McLintock *et al*., 1993).

1.4.2 Antigenic variation

To promote long-term persistence in the host and address immunogenicity associated with the VSG, *T. brucei* has evolved a VSG repertoire containing more than 1500 diverse genes in the genome (Berriman *et al*., 2005). Allelic exclusion ensures that only a single VSG gene is expressed at a time within a cell (Navarro & Gull, 2001). However, the process is reversible, and a VSG switch event can occur to produce monoallelic expression of a different VSG. The switch usually occurs in two forms, often linked to the stage of infection. During early infection, a switch in expression site is predominantly observed. With the number of expression sites limited to approximately 15 BESs, antigenic variation using this method is rapidly exhausted. As the infection progresses, trypanosomes rely upon homologous recombination to insert archived VSGs from silenced genomic regions into the active expression site. The BES VSG gene is flanked by 70 base pair repeats at the 5’ end of the gene and telomere repeats at the 3’ end. The repetitive sequences facilitate DNA rearrangements through homologous recombination. The dynamics of antigenic variation within the host were studied by Mugnier *et al*. (2015). Four mice were infected with trypanosomes each expressing a different VSG, and parasites were isolated at different stages of the infection to perform RNA-seq specifically targeted to the VSG by cDNA amplification using primers targeting conserved 5’ and 3’ regions. Studies of VSG expression in mice showed that the population was rich in VSG

diversity, with expression of an average of 28 different VSGs at any time point during the first 30 days (Figure 3). Some VSG variants appear transiently and are rapidly cleared whereas others are dominant in the population. Certain VSGs were dominant in several mice, suggesting there may be a VSG expression hierarchy, whereas in other mice the dominant VSGs were merely transient, highlighting that there may be differences between humoral responses in mice (Mugnier *et al*., 2015). The findings highlighted that VSG expression was highly dynamic and suggested the presence of a VSG hierarchy. Investigation of the determinants that may be ascribed to VSG hierarchy revealed that both VSG length and genome organisation may be involved. Regarding VSG length, studies demonstrated preferential expression of shorter VSGs during infection, leading to faster growth and an energetically favourable recourse (Liu *et al*., 2018). Histone variants and genome architecture were shown to be intimately linked with VSG expression, with chromatin configurations favouring accessibility of certain expression sites which may dictate an antigenic variation hierarchy (Müller *et al*., 2018) and affect VSG switching (Povelones *et al*., 2012).

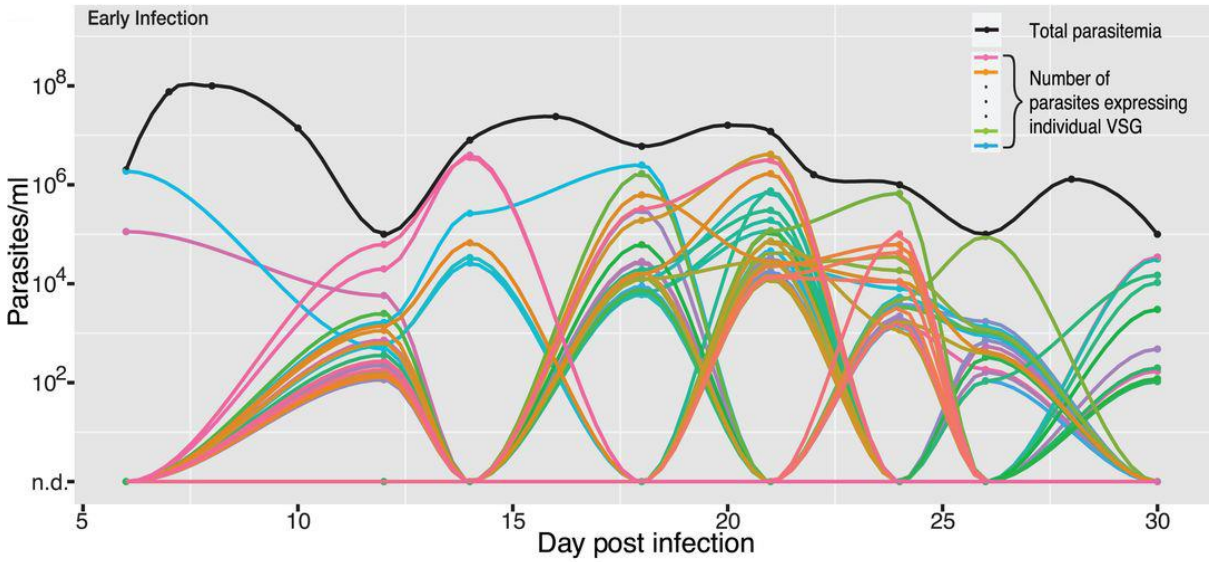


Figure 3: *In vivo* dynamics of VSG expression during early infection. Each different VSG population is represented by a coloured line, from Mugnier *et al*. (2015).

1.5 Molecular architecture of the cell surface

1.5.1 Structure of the variant surface glycoprotein

In addition to Ig-VSG endocytosis and antigenic variation, the VSG also provides protection by forming a physical barrier to shield invariant proteins. VSG molecules form homodimers and can be separated into an N-terminal and C-terminal domain. Different types of VSG have been identified and classified based on disulphide bond conservation *(*Carrington *et al*., 1991). The N-terminal domains were separated into three classes, type A, B and C, while C-terminal domains formed four classes, type I, II, III and IV. The membrane-proximal C-terminal domain has a glycosylphosphatidylinositol (GPI) anchor to mediate membrane attachment. The VSG coat was originally observed by electron microscopy in bloodstream-form trypanosomes (Vickerman, 1969), forming a dense layer with a thickness of 12 – 15 nm. Several VSGs have since been investigated by X-ray crystallography. In 1990, Freymann *et al*. reported a 2.9 Å structure of a type A VSG N-terminal domain, VSG MITat1.2 (Protein data bank (PDB): 1VSG). The core fold of the protein was formed of two long alpha helices and a third truncated alpha helix (Figure 4A). The apical surface of the structure was arranged in an array of flexible loops, which formed exposed epitopes eliciting an Ig-mediated response. The structure of a second type A VSG, ILTat1.24, was determined and revealed striking structural homology despite a lack of sequence homology (Blum *et al*., 1993) (PDB: 2VSG). The three-helix bundle formed the core fold of each monomer which showed remarkable structural conservation despite considerable sequence divergence (Figure 4A, B), with only 16% amino acid identity between the structure sequences. Thus, it was postulated that a structural motif common to VSGs was likely necessary for function and that variation of epitopes was the result of smaller structural modifications rather that overall structural rearrangements (Blum *et al*., 1993).

The structure of an intact VSG with both N- and C-terminal domains remained elusive. In 2017, Bartossek *et al.* focussed their research on elucidating the stucture of an entire VSG. First, a third N-terminal type A VSG structure was determined by X-ray crystallography, again confirming structural homology and the presence of a conserve three-helix bundle motif (MITat1.1; PDB: 5LY9) (Figure 4A). In parallel, a high-resolution structure of a VSG type II C-terminal domain was determined by Bartossek *et al*. using nuclear magnetic resonance (NMR) (PDB: 5M4T). To complement the findings, small angle X-ray scattering (SAXS) was performed with a complete VSG containing the N- and C-terminal domains. The previously solved structures were used to fit the low-resolution molecular envelope generated by SAXS.

For the first time, the structure of an intact VSG could be visualised in solution and two predominant conformations were observed for the C-terminal domain. Furthermore, the two distinct conformations were observed in lipid bilayers, confirming that the flexibility was not an artefact induced by the lack of constraints in solution. Finally, another VSG containing a type I C-terminal domain, ILTat1.24, was also studied and the same finding were observed, confirming that the flexible nature of the C-terminal domain could be extended to other VSGs. The findings were used to model intact VSG conformations within the membrane and demonstrated the nuanced density of VSG layer (Figure 5). While the VSG coat can be compact, the C-terminus confers flexibility, allowing modulable VSG cell surface coverage, which can be adapted in response to receptor accessibility requirements or immunoglobulin-mediated attack (Bartossek *et al*., 2017).

A wealth of data was available relating to VSG type A N-terminal domains, yet other VSG types were neglected. Pinger *et al*. (2018) focused their efforts on elucidating the structure of a type B VSG, MITat1.3, which revealed several striking findings. First, despite considerable sequence divergence (Figure 4B), the type B N-terminus also adopted a similar tertiary structure centred around the three-helix bundle motif (Figure 4A), suggesting that this fold was ubiquitous across VSG classes. Second, the type B VSG did not form the quaternary structure observed amongst type A VSGs. Rather than assembling as homodimers, the type B structure was monomeric in solution and the crystal packing revealed a trimer. Although the native oligomerisation status was not evident, it was likely that type B quaternary architecture had diverged from type A VSGs. Finally, an O-linked glycosylation was identified at the membrane-distal surface of the VSG. The finding represented the first discovery of an O-linked glycosylation in *T. brucei*. Furthermore, the glycosylation conferred virulence when infections of trypanosomes expressing glycosylated and mutant non-glycosylated type B VSG were compared in mice (Pinger *et al*., 2018). The findings demonstrated that antigenic variation and immune evasion were not solely governed by sequence diversity between VSGs, highlighting that post-translational modifications may also play an important role.

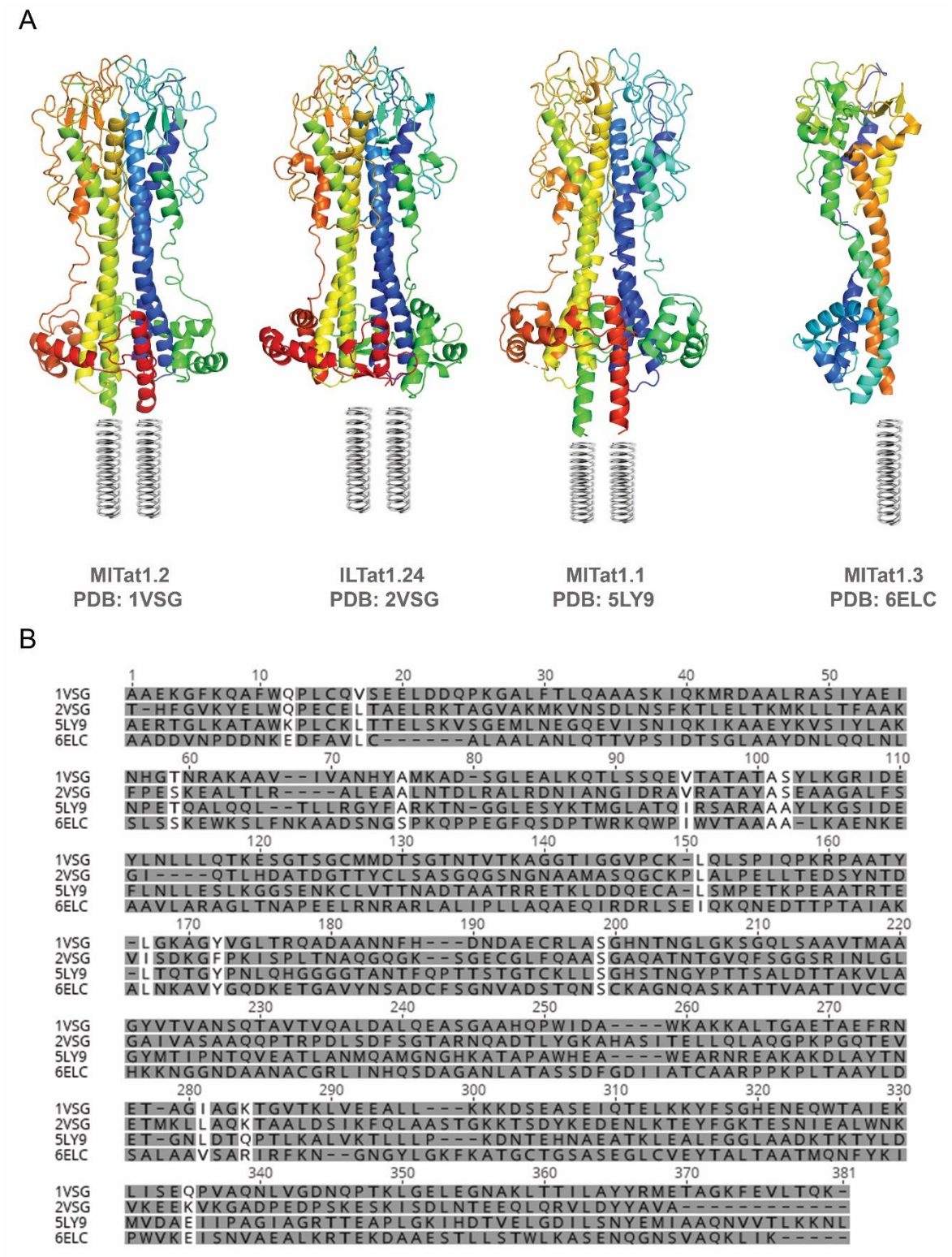


Figure 4: Comparison of the structures and sequences of VSGs. A, Structures of the N-terminal domains of three type A VSGs (MITat1.2, ILTat1.24 and MITat1.1) and one type B VSG (MITat1.3). The type A N-terminal domains are shown as homodimers while type B is represented as a monomer. C-terminal domains are illustrated as springs of various sizes to represent flexibility within the domain. B, Alignment of the protein sequences of the four VSGs, with conserved similarity shown in white and non-conserved residues coloured in grey.

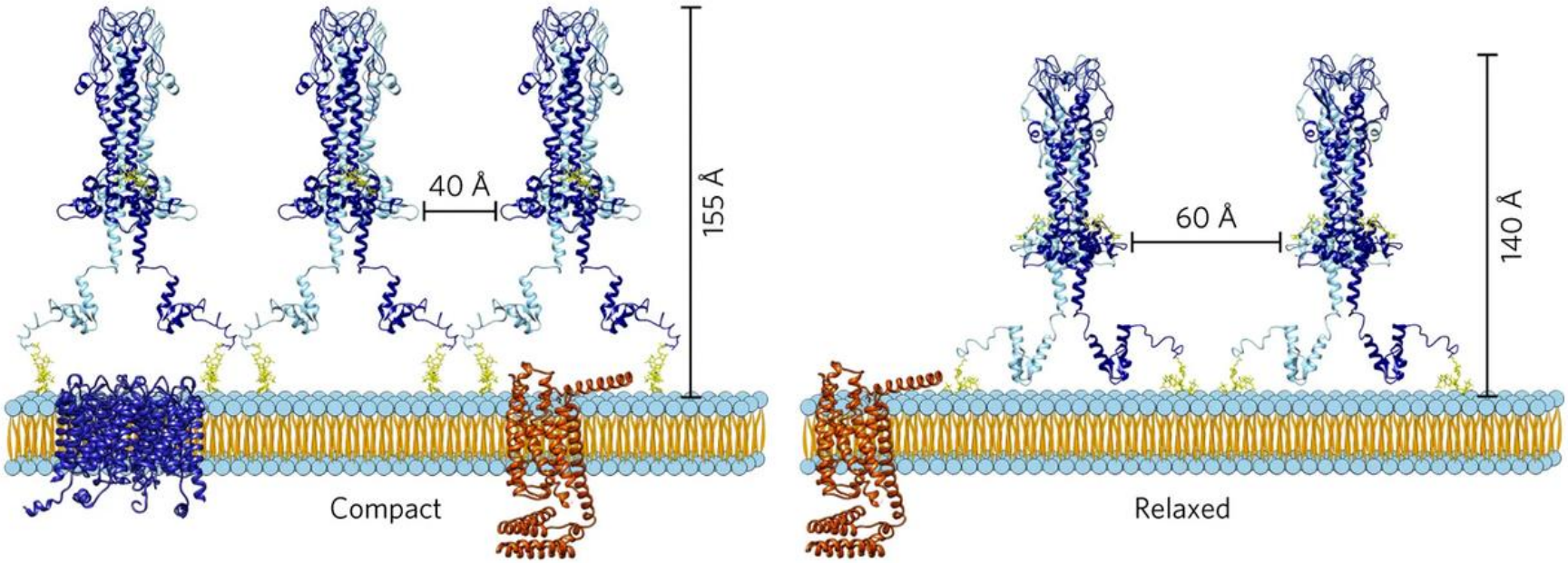


Figure 5: Nuanced density of the VSG coat. The C-terminal domain of the VSG confers flexibility and allows the VSG coat to adopt compact (left) or relaxed (right) configurations in response to other cell surface protein requirements (from Bartossek *et al*. 2017).

1.5.2 Invariant surface proteins

A less-characterised family of glycoproteins is the invariant surface glycoproteins (ISGs). Two of the most studied ISGs are ISG65 and ISG75, named to reflect their apparent molecular weight. While little is known regarding their function, a few observations have been reported. First, ISGs are present in high copy numbers, with 70,000 copies for ISG65 and 50,000 for ISG75 in *T. brucei* (Ziegelbauer & Overath, 1992). Second, ISGs rely on a transmembrane domain for membrane insertion (Carrington & Boothroyd, 1996), which distinguishes the family from GPI-anchored proteins that adopt a conserved three-helix bundle. Finally, ISG75 has been implicated in the uptake of suramin (Alsford *et al*., 2012), a compound commonly used to treat early stage HAT. Although the structures of ISGs remain elusive, conserved cysteines suggest that the extracellular domains adopt a VSG type A fold (Carrington & Boothroyd, 1996).

The recently identified factor H receptor (FHR) is an example of an invariant receptor. Factor H protects host cells against the downstream consequences of C3b deposition and is exploited by trypanosomes. By recruiting factor H to the cell surface of trypanosomes, the receptor enables inactivation of complement factor C3b through the inhibitory interactions of factor H. The FHR structure has recently been determined and the invariant receptor adopts the characteristic three-helix bundle fold (Macleod *et al*., 2020). The receptor was shown to be upregulated in stumpy bloodstream form trypanosomes which are capable of infecting tsetse flies. Furthermore, knockdown of the FHR significantly reduced the infection rate of tsetse fly midguts by more than 2.5-fold. Together, the findings provide valuable evidence to demonstrate that the *T. brucei* FHR has evolved to increase parasite transmission to tsetse flies (Macleod *et al*., 2020).

1.5.3 Evolution of the haptoglobin-haemoglobin receptor

The HpHbR is a well-characterised invariant receptor expressed in trypanosomes to fulfil haem requirements (Vanhollebeke *et al*., 2008). The structure of the HpHbR of *T. congolense*, a trypanosome that commonly infects livestock in sub-Saharan Africa, was determined by Higgins *et al*. (2013). The structure revealed a monomeric three-helix bundle (PDB: 4E40), sharing structural homology with the VSG. Subsequent structural determination of the *T. brucei* HpHbR (TbHpHbR) provided insights into the evolution of the receptor (Lane-Serff *et al*., 2014). The core TbHpHbR fold was formed of the characteristic three-helix bundle motif, similar to the *T. congolense* receptor. However, the TbHpHbR had evolved several strategies to increase

uptake of HpHb. First, the *T. brucei* receptor had evolved longer helices, allowing protrusion from the dense VSG coat (Higgins *et al.*, 2013). Second, the receptor had adopted a rigid 50° kink which would facilitate separation of VSG molecules and augment ligand accessibility through the display of a large ligand binding platform. Finally, the altered conformation promoted interaction of two TbHpHbR molecules with a dimeric HpHb ligand, producing a bivalent binding model with improved avidity. Monovalent binding of TbHpHbR to dimeric HpHb produced a dissociation constant (K_D) of 1 µM, while bivalent binding of two receptor molecules to dimeric HpHb resulted in an increased binding response, with a reported K_D of 4.5 nM (Higgins *et al*., 2013) (Lane-Serff *et al*., 2014). The accessible and invariant properties of the receptor raised an interesting question regarding immunogenicity. However, the low copy number of the receptor (~200 - 400 copies per cell) (Vanhollebeke *et al*., 2008) may reduce the likelihood of an Ig-mediated response.

While the receptors discussed thus far are invariant, the receptor responsible for transferrin (Tf) uptake presents a peculiarity as it is encoded in each of the telomeric expression sites, and multiple variants have been identified (Hertz-Fowler *et al*., 2008).

1.6 The trypanosome transferrin receptor

1.6.1 ESAG6 and ESAG7 form the transferrin receptor

The trypanosome transferrin receptor (TfR) is arguably the most highly studied yet poorly understood trypanosome receptor. The complexities began in 1991 when the discovery of a transferrin-binding protein was made (Schell et al., 1991a). The receptor was identified as a heterodimer formed of the proteins encoded by the ESAG6 and ESAG7 genes (Salmon *et al*., 1994) (Steverding *et al*., 1994). Initial reports of a receptor dimer most likely binding two molecules of transferrin (Salmon *et al*., 1994) were dispelled when it was revealed that an ESAG6 (M_r = 50,000 – 60,000) and ESAG7 (M_r = 40,000 – 42,000) heterodimer bound a single molecule of transferrin (M_r = 75,000 – 80,000) (Steverding *et al*., 1995). Studies demonstrated that ESAG6 or ESAG7 homodimers failed to support binding of transferrin, and despite their high degree of shared sequence identity, only ESAG6/7 heterodimers formed a functional transferrin-binding receptor (Salmon *et a*l., 1994). However, ESAG6 and ESAG7 homodimers can readily associate and the predisposition to form ESAG6/7 heterodimers is yet to be demonstrated.

The location of conserved cysteines revealed that ESAG6 and ESAG7 were likely to adopt a VSG N-terminus type A fold, suggesting that these genes have evolved from a common ancestor (Carrington & Boothroyd, 1996). Construction of a VSG-TfR fusion protein containing the VSG N-terminus with the ESAG6 or ESAG7 Tf-binding site produced a functional receptor with similar binding efficiency compared to the native TfR, further indicating a common origin for TfRs and VSGs (Salmon *et al*., 1997).

1.6.2 The transferrin receptor mediates iron acquisition

Vertebrates have evolved to sequester iron within the serum transporter glycoprotein transferrin. The transporter is formed of two lobes, referred to as N-lobe and C-lobe. Each lobe coordinates an iron ion in the Fe(III) state. The trypanosome TfR binds host Tf through an interaction with both the N- and C-lobes of Tf (Steverding *et al*., 2012). Studies have demonstrated that the receptor is located in the flagellar pocket (Steverding *et a*l., 1994) (Tiengwe *et al*., 2017), containing approximately 3000 receptors per cell (Steverding *et al*., 1995). The TfR-Tf complex is endoytosed via clathrin-mediated endocytosis (Coppens *et al*., 1987), but the fate of ligand and iron thereafter lacks clarity and will be further discussed in chapter 5.

To understand the requirements for Tf, several studies were performed. First, Schell *et al*. (1991) attempted to grow trypanosomes in serum depleted of Tf. Cell growth was arrested after 24 hours, and they concluded that Tf was an essential growth factor for trypanosomes. Next, studies were centred on the receptor itself. TfR knockout studies were difficult to achieve due to the multiplicity of ESAG6 and ESAG7 genes in the genome. Tiengwe *et al*. (2016) developed a method for pan-specific knockdown of TfRs by targeting common ORF sequences and 5'UTR sequences by RNAi. TfR silencing led to arrested growth in trypanosomes, and proved lethal after three days, with survival during the first 48 hours attributed to iron reserves (Taylor & Kelly, 2010). Together, the findings suggested that TfR-mediated endocytosis of Tf was essential in trypanosomes.

1.6.3 Identification of a collection of transferrin receptor sequences

The advances in DNA sequencing facilitated an analysis of the *T. brucei* genome (Berriman *et al*., 2005). However, the initial studies did not include the VSG expression sites. Efforts were

subsequently focussed on the telomeric expression sites, and isolation of these regions was achieved by transformation-associated recombination (TAR) cloning in yeast (Becker *et al*., 2004), followed by sequence determination and assembly (Hertz-Fowler *et al*., 2008). For the first time, all the telomeres of the *T. b. brucei* Lister 427 (L427) isolate could be assembled and the presence of one or multiple copies of ESAG6 and ESAG7 in each bloodstream expression sites (BES) was confirmed. The different ESAG6 and ESAG7 genes showed diversity within each gene family, yet only a single TfR is expressed at a time. Furthermore, the polymorphisms did not appear to be the result of stochastic mutations, with variation residing in hotspots referred to as hypervariable sites (Salmon *et al*., 1997). The findings prompted an interesting question, what is the function of TfR multiplication and diversification?

1.6.4 Function of the transferrin receptor repertoire

One hypothesis relating to TfR diversity stemmed from the broad host range of trypanosomes. The proposed theory was that a TfR repertoire may be required to accommodate the different Tfs from vertebrates (Bitter *et al*., 1998) (Isobe *et al*., 2003). Studies found that a single TfR had varying affinities for Tfs from different species, with differences of several orders of magnitude (Bitter *et al*., 1998) (Gerrits *et al*., 2002) (Salmon *et al*., 2005). Perhaps more convincingly, Bitter *et al*. (1998) demonstrated that trypanosomes could switch TfR depending on the host serum. In their study, trypanosomes displaying a TfR expressed from BES1 showed normal growth in bovine serum. Upon transfer of the cells to canine serum, trypanosome proliferation was reduced for almost 200 hours. Once normal proliferation had resumed, an outgrowing population of trypanosomes expressing from BES2 was identified, suggesting that the change in host Tf had induced a switch in TfR expression (Bitter *et al*., 1998). While the previous study did not isolate the specific role of Tf in switch induction, Gerrits *et al*. (2002) performed a similar study in which BES1-expressing cells grown in canine serum prevented switching when supplemented with bovine Tf. However, upon transfer of BES1-expressing trypanosomes to horse serum, the addition of bovine Tf did not prevent switching to BES2, indicating that other factors may be responsible for the observed switching events. Furthermore, it was noted that different batches of serum led to considerably variable results, with switching from BES1 to BES2 only observed in certain batches of canine serum (Gerrits *et al*., 2002). In 2003, Isobe *et al*. observed an interesting correlation between host range and TfR diversity. *T. equiperdum* is a species of trypanosome restricted to equines and had less diversity in the predicted TfR binding site compared to *T. brucei* (Isobe *et al*., 2003), providing further evidence to suggest that *T. brucei* TfR diversification was the result of adaptation to a

larger host range. In contrast, the TfR repertoire was also proposed to mediate an immune evasion strategy. Similar to the VSG repertoire, it was speculated that the TfR could alter its identity to avoid host immune detection (Borst, 1991). Due to the confusion surrounding the role of the transferrin receptor repertoire, further investigations would be performed in this work to elucidate the origin of TfR diversification.

1.7 Clinical manifestation

T. b. gambiense and *T. b. rhodesiense* are resistant to human serum and can cause a severe disease in humans, referred to as human African trypanosomiasis (HAT). The early stage infection, termed haemolymphatic, is characterised by waves of parasitemia which were historically correlated with changes in temperature and intermittent fever (Ross & Thomson, 1910). The non-specific nature of the symptoms, which include waves of fever, pruritis, headaches and lymphadenopathy, makes diagnosis reliant upon laboratory testing (Brun *et al*., 2010). Techniques for diagnosis were developed in the late 1970s and are based on antibody detection using a card-agglutination trypanosomiasis test (CATT) (Magnus *et al*., 1978). However, the low sensitivity and specificity of the test make early diagnosis a challenging feat (Bonnet *et al*., 2015). Breach of the blood brain barrier and parasite invasion of the central nervous system (CNS) marks the advancement to late stage neurological or meningoencephalitic trypanosomiasis. The ensuing deregulation of the host circadian rhythm, characterised by diurnal somnolence and nocturnal insomnia, lends itself to the name sleeping sickness. In addition, motor weaknesses, tremors, fasciculations and paralysis of limbs can occur, with aggressive or irritable behaviour and psychotic reactions also reported (Brun *et al*., 2010). *T. b. rhodesiense* produces an acute infection, with rapid progression to coma and death in under 6 months in most cases, if left untreated (Odiit *et al.*, 1997). *T. b. gambiense* is responsible for a less aggressive infection, typically described as chronic, which lasts approximately three years if untreated (Checchi *et al*., 2008).

1.8 Therapeutics

1.8.1 Current therapeutic landscape

The current treatment regimen for HAT relies predominantly on non-specific chemotherapeutic compounds associated with nocive side effects. The first-line treatment for early stage HAT is typically pentamidine and suramin for *T. b. gambiense* and *T. b. rhodesiense*, respectively (Brun *et al*., 2010). Both are delivered by injection, intramuscular for pentamidine and intravenous for suramin, requiring administration by a healthcare professional. Pentamidine can cause gastrointestinal adverse effects, while suramin can result in liver and kidney damage (Fevre *et al*., 2008). For the advanced, neurological stage of *T. b. rhodesiense*, the arsenic derivative melarsoprol can be administere, but not without risk. Encephalopathic syndrome is an often fatal adverse reaction to melarsoprol, occurring in 5 – 18% of cases (Sutherland *et al*., 2015). Late stage *T. b. gambiense* is treated with a Nifurtimox and Eflornithine combination therapy (NECT). Melarsoprol and Eflornithine are both delivered by intravenous injection, while Nifurtimox is an oral drug. The nature of the treatments, which require administration by a healthcare professional and clinical surveillance, is a drawback for those who do not have access to healthcare infrastructures. Until recently, there has been little advance in novel treatment approval for HAT, which led to its classification as a neglected tropical disease by the World Health Organisation.

1.8.2 Recent advances in therapeutic strategy

In 2018, Fexinidazole was approved for the treatment of both early and late stages of *T. b. gambiense* (Mesu *et al*., 2018). The orally administered drug, which was effective and safe for the treatment of both stages, provided a breakthrough as determination of disease progression and stage was not required and therapy could be home-based. In recent years, advances in specific delivery of toxins via antibody-based technologies have shown promising results. The remarkable endocytic capacities of trypanosomes can offer a rapid entry route for a toxin-conjugated vector. Nanobody therapeutics have gathered attention due to their small size which facilitates accessibility to invariant surface targets and favours penetration of the blood-brain barrier (reviewed by Stijlemans *et al*., 2017). However, these camelid-derived therapeutics raise the concern of immunogenicity and whether repeated injections may elicit a host immune response. Recently, MacGregor, Gonzalez-Muñoz *et al*. (2019) developed a toxin-conjugated human antibody targeting the *T. brucei* haptoglobin-haemoglobin receptor. They demonstrated that a single dose of the antibody, raised *in vitro* by phage display, was

effective in clearing a peripheral infection in mice. The promising findings raised the question of whether other trypanosome cell surface proteins could be targeted in the same manner.

1.9 Project foundations

1.9.1 Limitations in our understanding of the trypanosome transferrin receptor

As extracellular pathogens, trypanosomes have developed remarkable cell surfaces that mediate key functions such as nutrient uptake and protection against host immune effectors. Studies to elucidate the molecular architecture of the cell surface have been heavily biased towards the VSG. To date, the only well-characterised cell surface protein involved in receptor-mediated endocytosis is the HpHbR, which prompted an interest in characterisation of another trypanosome receptor. Furthermore, studies of telomeric expression sites have been centred around VSG research, while the study of other proximal genes has been overlooked in comparison, with the function of some ESAGs still unidentified. The study of two important ESAGs, ESAG6 and ESAG7, would further our understanding of the unusual genome organisation and the role of expression of multiple genes from telomeric expression sites. The project originated from several broad questions which remained poorly understood.

Why has a diverse TfR repertoire evolved?

How do two different TfRs bind Tfs from different mammals?

Can a low affinity TfR support trypanosome growth?

Previous studies of the TfR had investigated receptor-ligand interactions in the context of the trypanosome cell membrane. However, this can produce inaccuracies during determination of kinetic parameters due to variations in TfR copy numbers and detection methods. To study receptor-ligand interactions at the molecular level and obtain robust kinetic data, strategies for TfR isolation were explored. Due to low copy numbers of TfR at the trypanosome cell surface and the challenges associated with isolation of membrane proteins, expression and purification of recombinant TfR was preferential and would form the basis of the study.

1.9.2 Selection of expression system

The choice of expression system was influenced by the presence of disulphide bonds and glycans. Three disulphide bonds were present in ESAG6 and four in ESAG7, which would result in unfavourable folding conditions in the reducing environment of *E. coli* cytoplasm (Stewart *et al*., 1998). Furthermore, the trypanosome TfR is a glycoprotein with five predicted N-linked glycosylation sites in ESAG6 and three in ESAG7, based on the consensus that glycosylations can be linked to asparigines located within the N-X-S/T motif where X is any amino acid except proline (Gavel & Heijne, 1990), and confirmed by findings from Mehlert *et al*. (2012). It was not entirely clear which role glycans play and how the receptor would fold in the absence of endoplasmic reticulum (ER) processing. In previous studies, recombinant TfR had been produced in insect cells in the presence of tunicamycin to study to the involvement of glycosylation in receptor function (Maier & Steverding, 2008). Tunicamycin inhibits N-linked glycosylation and non-glycosylated TfR was produced to study ESAG6/7 heterodimerisation and Tf binding. The non-glycosylated receptor retained heterodimerisation and Tf-binding capacities, indicating that TfR glycans were not essential for receptor function (Maier & Steverding, 2008). As the role of glycans remained unknown, a eukaryotic expression system was deemed favourable. A mammalian expression system was yet to be tested for trypanosome TfR expression and could provide a lysis-free technique through the use of signal peptides to target proteins towards the cell surface.

1.10 Objectives

In contrast to other characterised receptors that bind host HpHb or Factor H and are present as a single copy in the genome, the ESAG6 and ESAG7 genes represented in multiple different copies due to their localisation in each of the ~15 BESs, yet the motives were unclear. Given the complexities that surround the receptor, this work aimed to clarify some of the ambiguities through exploration of the TfR-Tf interaction at the molecular level. By elucidating the binding mechanism, the goal was to uncover the selective pressure responsible for driving the multiplication and diversification of TfRs within the trypanosome genome.

1.10.1 Outline of the study

The overall aims of the thesis were:

1. To further investigate claims that the trypanosome TfR repertoire arose from requirements for species specificity;

2. Determine the structure of the trypanosome TfR;

3. Gain insight into iron uptake in trypanosomes;

4. Assess the viability of an anti-TfR antibody therapeutic.

First, two of the most genetically distant TfRs from different Lister 427 BESs were selected by phylogenetic analysis. The TfRs were expressed recombinantly in mammalian cells and receptor-ligand interactions were investigated at the molecular level. In addition, the growth of BES1-expressing trypanosomes in different sera was monitored, as the alleged Tf-induced switches observed in previous sera studies (Bitter *et al*., 1998) (Gerrits *et al*., 2002) and the alarming serum batch-dependant discrepancies (Gerrits *et al*., 2002) warranted further investigation. While the measurement of growth rates and the kinetic parameters of receptor-ligand binding provided valuable data, analysis of the receptor-ligand interaction at the atomic level using X-ray crystallography produced a wealth of additional information. Although the structure provided a snapshot of a single archetype from the TfR repertoire, the sequences of other TfRs could be mapped to identify hypervariable regions, resolving many profound complexities relating to TfR diversification. Next, the role of iron and the effects of pH on receptor-ligand interactions were investigated in a bid to understand the fate of the internalised complex and whether the parasite receptor mimics the native mammalian TfR. Finally, efforts were focussed towards determining whether the trypanosome TfR could be exploited as a target for a toxin-conjugated antibody therapy.

1.10.2 Study outcomes

The work presented in this thesis has provided further evidence regarding the complex interactions between trypanosome and host. Through the curation of a repertoire of receptors, trypanosomes have evolved to hijack host iron while avoiding immune recognition. A small piece of the puzzle has been solved in the quest to elucidate the molecular architecture of the

trypanosome cell surface, and the majority of the findings have been presented in a journal publication (Trevor *et al*., 2019). Finally, the viability of an antibody-drug conjugate (ADC) therapy targeting the TfR has been assessed and will provide insights into the design of future therapeutics.

Chapter 2

Materials and methods

2.1 Production of recombinant transferrin receptor

2.1.1 Transferrin receptor cloning

Recombinant transferrin receptors (TfRs) were produced by modification of ESAG6 and ESAG7 open reading frames. In the case of BES1 ESAG6 (BES1e6) and Bn2 ESAG6 (Bn2e6), the native signal peptides and glycosylphosphatidylinositol (GPI) anchor addition sites were removed. The genes (residues 20-377 for BES1 and 18-375 for Bn2) were codon optimised for mammalian expression and synthesised (GeneArt). Genes were subcloned into the pDest12 mammalian expression vector, containing a cytomegalovirus (CMV) promoter and a CD33 signal peptide for protein expression and direction towards the secretary pathway. At the C-terminus, a $(GS)_3$ linker, an AviTag and a decahistidine tag were included. For BES1 ESAG7 (BES1e7) and Bn2 ESAG7 (Bn2e7), the signal peptides were removed and the genes (residues 20-338 for BES1 and 18-337 for Bn2) were codon optimised, synthesised and subcloned into pDest12 under the CMV promoter. The CD33 signal peptide was included at the N-terminus and a $(GS)_3$ linker and StrepII-tag (Iba Life Sciences) at the C-terminus (Appendices 1 and 2).

A Bn2-TfR mutant, with all N-linked glycosylation sites mutated and tags removed, was produced for crystallisation trials. Non-glycosylated TfR was obtained by site-directed mutagenesis of predicted N-linked glycosylation sites. Asparagine was converted to aspartic acid using the QuikChange multi site-directed mutagenesis kit (Agilent Technologies, 200514), following manufacturer's protocol. For Bn2 ESAG6, five mutations (N26D, N110D, N235D, N250D, N360D) were performed and the C-terminal linker and tags were removed.

The AviTag and decahistidine tag were relocated to the N-terminus with the addition of a tobacco etch virus (TEV) protease cleavage site for tag removal. For Bn2 ESAG7, three mutations (N26D, N110D and N234D) were performed and the N-terminal linker and StrepII-tag were deleted. The genes were expressed from the CMV promoter with a CD33 signal peptide (Appendices 3 and 12).

BES1-TfR mutants were produced using the QuikChange Lightening site-directed mutagenesis kit (Agilent Technologies, 210519), as per the manufacturer's protocol. In BES1 ESAG6, glycine 139 was mutated to arginine (numbering was based on the Bn2 protein sequence to align with the crystal structure). In BES1 ESAG7, three mutations (I229V, C233R, and S246Y) were performed. The genes were expressed from the CMV promoter with a CD33 signal peptide and C-terminal tags as described for recombinant TfR production (Appendices 1 and 12).

2.1.2 Transferrin receptor expression

For individual expression of ESAG6 and ESAG7 proteins, Chinese hamster ovary (CHO) G22 cells developed by AstraZeneca (Daramola *et al*., 2014) were grown in 500 ml of serum-free CCM8 medium (Sigma-Aldrich), to a density of $2x10^6$ cells/ml. Cells were transfected with either the Bn2e6, Bn2e7, BES1e6 or BES1e7 plasmid. DNA (0.5 µg per ml of cell culture) was added to 150 mM NaCl to a final volume of 1/80 of cell culture volume. PEI max (Polyethylenimine max; Polysciences, 24765) was added to 150 mM NaCl to a final volume of 1/80 of cell culture volume, then mixed with DNA at a 1:8 mass ratio of DNA:PEI max and added to cells. Flasks were incubated overnight at 37ºC, shaking at 140 rpm with 5% CO2 and 80% humidity. Cells were supplemented with F9 and F10 media (AstraZeneca proprietary medium supplements) on days 1, 3 and 6 post-transfection. Cells were harvested on day 8 by centrifugation at 3000 rpm for 30 minutes at 4ºC. The aqueous phase containing secreted protein was recovered after centrifugation and filtered using a 0.22 µm SteriCup filter.

For BES1-TfR production, CHO G22 cells were cultured as described above and co-transfected with BES1e6 and BES1e7 plasmids in equal mass ratio. For Bn2-TfR production, Bn2e6 and Bn2e7 plasmids were co-transfected in equal mass ratio. The Bn2-TfR glycan mutant was expressed in the same way. To produce TfR chimeras, BES1e6 was co-transfected with Bn2e7 (BES1e6-Bn2e7) in equal mass ratio and Bn2e6 was co-transfected with BES1e7 (Bn2e6-BES1e7) in equal mass ratio. BES1-TfR mutants were produced in four forms. First, BES1e6 with a G139R mutation was co-transfected with BES1e7 wild type (WT)

to produce the BES1 e6 G139R mutant. Second, BES1e7 with I229V and S246Y mutations was co-transfected with BES1e6 WT to produce the BES1 e7 I229V S246Y mutant. Next, BES1e6 G139R and BES1e7 I229V and S246Y were co-transfected to form the BES1 e6 G139R e7 I229V S246Y. Finally, BES1mut was produced by co-transfection of BES1e6 G139R and BES1e7 with I229V, S446Y and C233R mutations. All proteins were expressed and recovered as described above.

To produce glycosylated TfR used for structural determination, Bn2e6 and Bn2e7 were co-transfected into CHO cells supplemented with 5 µM Kifunensine (Sigma-Aldrich, K1140), an α-mannosidase I inhibitor, to generate mannose-rich glycans (Yu *et al*., 2011). The aqueous phase containing expressed protein was recovered as described above.

2.1.3 Protein purification

ESAG6 subunits (Bn2e6 or BES1e6) were purified by nickel affinity chromatography using a His Trap EXCEL 5ml column (GE Healthcare, 17-5248-01) connected to AKTA Pure fast protein liquid chromatography (FPLC) system (GE Healthcare). The column was equilibrated with PBS (Dulbecco's Phosphate Buffered Saline, Sigma-Aldrich) and the aqueous phase was loaded onto column at 5 mL/min. A wash step was performed with five column volumes of 20 mM imidazole in PBS, pH7.4. ESAG6 was eluted with 400 mM imidazole in PBS, pH7.4. Fractions containing ESAG6 were dialysed into PBS.

ESAG7 subunits (Bn2e7 or BES1e7) were purified by streptactin affinity chromatography. The aqueous phase containing expressed ESAG7 was dialysed into PBS to remove free biotin present in the medium that would interact with streptactin. the dialysed aqueous phase was injected at 1 ml/min over a StrepTrap HP 1 ml column (GE Healthcare) using Akta Pure FPLC system (GE Healthcare). Following a wash step with five column volumes of PBS, ESAG7 was eluted with 250 µM desthiobiotin. Fractions containing ESAG7 were dialysed into PBS to remove desthiobiotin.

All TfRs including chimeras and mutants were purified by nickel affinity chromatography as stated above and fractions containing TfR (ESAG6 and ESAG7) were dialysed into PBS. Elution fractions were incubated for 10 minutes at 70°C in the presence of NuPAGE lithium dodecyl sulphate (LDS) sample buffer (Invitrogen, NP0007) and NuPAGE reducing agent (Invitrogen, NP0004). Samples were analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) using 4-12% Bis-Tris NuPAGE pre-cast gels (Invitrogen,

NP0322BOX) with NuPAGE 2-(N-morpholino)ethanesulfonic acid (MES) running buffer (Invitrogen, NP000202). Proteins were separated at 200 V for 50 minutes and stained with InstantBlue Coomassie (Expedeon, ISB1L). SDS-PAGE was performed in this manner throughout, unless stated otherwise. Pooled fractions were dialysed into PBS and concentrated for subsequent biotinylation.

2.1.4 Size exclusion chromatography

Size exclusion chromatography (SEC) was performed by Gareth Rees (AstraZeneca, Cambridge) using a high-performance liquid chromatography (HPLC) system. A Superdex 200 10/300 column (GE Healthcare) was equilibrated in PBS at 0.5 ml/min flow rate using an Agilent HP1100 HPLC. Bn2e6 at 0.9 mg/ml was injected at a volume of 10 µl with 95 µl of PBS. Bn2e7 at 0.2 mg/ml was injected at a volume of 70 µl with 35 µl of PBS. 20 µl of Bn2e6 at 0.9 mg/ml was mixed with 90 µl of Bn2e7 at 0.2 mg/ml and incubated for 24 hours prior to injection. Peaks were detected using the 280 nm signal from a diode array detector (DAD).

2.1.5 Differential scanning calorimetry

Differential scanning calorimetry (DSC) was performed by David Staunton, Department of Biochemistry, University of Oxford. Samples of purified BES1 ESAG6, BES1-TfR, Bn2 ESAG6 and Bn2-TfR were heated and the heat capacity was measured. The endothermic peak was used to determine the melting temperature (Tm).

2.2 Analysis of receptor-ligand interactions

2.2.1 Biotinylation of transferrin receptors

The AviTag provided a recognition site for BirA biotin ligase, resulting in the biotinylation of a single lysine within the AviTag sequence (Fairhead & Howarth, 2015). Biotinylation specifically at the C-terminus of ESAG6 dictated the orientation of the receptor at the surface of the streptavidin chip, mimicking attachment to the cell surface by a GPI-anchor. *E. coli* BL21 (DE3) pRIPL cells (genotype: F^- *ompT* $hsdS_B(r_B^- m_B^-)$ dcm^+ Tet^r *gal* λ(DE3) *endA* Hte [*argU proL* Cam^r] [*argU ileY leuW* Strep/$Spec^r$]) were transformed with a plasmid encoding BirA

containing an N-terminal glutathione S-transferase (GST) tag (kind gift of Matthew Higgins, University of Oxford). Cells in mid-logarithmic phase were induced with 0.02% IPTG (Isopropyl β-D-1-thiogalactopyranoside). Following overnight expression of GST-BirA at 18°C, cells were lysed using an EmulsiFlex-C5 (Avestin). Following centrifugation, the soluble fraction was incubated with Glutathione Sepharose 4B resin (GE Healthcare, 17-0756-01). A wash step was performed with PBS and GST-BirA was eluted with PBS containing 20 mM reduced glutathione. GST-BirA was dialysed into PBS and used for subsequent biotinylation reactions. Biotinylations were performed in a 1 ml reaction volume containing 30 µM purified TfR, 0.4 µM purified BirA, 0.3 mM biotin and 5 mM ATP. Reactions were performed at 25°C for 16 hours, followed by removal of GST-BirA using Glutathione Magnetic Beads (Pierce). Western blots to verify biotinylation were performed by separating proteins by SDS-PAGE. Proteins were transferred to a polyvinylidene fluoride (PVDF) membrane (Invitrogen, IB401001) using an iBlot transfer device (Thermo Fisher Scientific). The membrane was blocked with Odyssey blocking buffer (Licor, 927-40000), then incubated with IRDye 680LT Streptavidin (Licor, 926-68031) diluted 1:5000 in blocking buffer. After washing with PBS containing 0.05% Tween, the membrane was visualised at 700 nm wavelength using the Odyssey imaging system and its associated software.

2.2.2 Surface plasmon resonance using neat serum

Surface plasmon resonance (SPR) was performed using a Biacore T200 instrument (GE Healthcare) with a Series S Streptavidin (SA) chip (GE Healthcare, BR100531). Biotinylated TfR was immobilised on flow path 2, while flow path 1 remained without receptor to serve as a blank reference. Experiments were performed at 25°C in running buffer containing 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 7.4, 150 mM NaCl, 0.005 % Tween-20. Sera from cow (Sigma-Aldrich, B9433), horse (Sigma-Aldrich, H1270), mouse (Sigma-Aldrich, M5905), rabbit (Thermo Fisher Scientific, 16120099), goat (Thermo Fisher Scientific, 16210064), rat (Thermo Fisher Scientific, 10710C) and human (donor) were used. A starting concentration of neat serum diluted 1:5 into running buffer was used, and four additional two-fold serial dilutions were performed. Between each injection of diluted serum, a regeneration step was performed using 100 mM glycine pH 3. Signals observed in the blank reference flow path were subtracted to remove non-specific signals and sensorgrams were produced using the Biacore T200 evaluation software.

2.2.3 Native transferrin isolation

Native transferrins for SPR analysis were purified from serum using transferrin receptor affinity chromatography. Cow (Sigma-Aldrich, B9433), horse (Sigma-Aldrich, H1270), pig (Sigma-Aldrich, P9783), mouse (Sigma-Aldrich, M5905), rabbit (Thermo Fisher Scientific, 16120099), goat (Thermo Fisher Scientific, 16210064), rat (Thermo Fisher Scientific, 10710C) and human (donor) were used. Due to limited volumes of donor serum, human holo-transferrin (Sigma-Aldrich, T4132) was also used after no significant differences in binding kinetics were confirmed by SPR. Bn2-TfR was immobilised via amine coupling to a 1 ml HiTrap NHS-activated HP column (GE Healthcare, 17071601) following the manufacturer's protocol. Serum was passed over the column and washed with PBS. Transferrin was eluted with 100 mM citrate pH 3.5 and dialysed into PBS. Transferrins were loaded with iron using a four-fold molar excess of ammonium iron (III) sulphate in the presence of 5 mM sodium bicarbonate (Yang, Zhang, Wang, Hao, & Sun, 2012). Free iron was removed by dialysis into 10 mM HEPES pH7.4, 150 mM NaCl. Human apo-transferrin (Sigma-Aldrich, T1147) was prepared by dialysis in the presence of 10 mM ethylenediaminetetraacetic acid (EDTA) into 10 mM HEPES pH7.4, 150 mM NaCl for 15 hours to remove any traces of free iron by iron chelation.

2.2.4 Surface plasmon resonance using purified transferrin

SPR was performed using a Biacore T200 instrument (GE Healthcare) with the Biotin CAPture kit (GE Healthcare, 28920234). The dextran-coated chip surface was functionalised with single-stranded oligonucleotide deoxyribonucleic acid (DNA) which hybridised with CAPture reagent (GE Healthcare, 28920234) containing complementary DNA covalently linked to streptavidin. The methodology based on DNA hybridisation enabled reversible capture of biotinylated proteins with regeneration after each cycle using CAPture regeneration solution (GE Healthcare, 28920234). A biotin CAPture chip (GE Healthcare, 28920234) prepared with CAPture reagent was used to immobilise biotinylated TfR onto flow path 2 to a total of ~500RU. Flow path 1 was left without immobilised receptor and served as a blank reference. All experiments were performed at 25°C as described previously. Running buffer containing 10mM HEPES pH 7.4, 150mM NaCl, 0.005% Tween-20 was used for most experiments. However, the pH assay was performed using a phosphate-citrate buffer system to achieve a broad pH range that represented progression through the endocytic pathway. Solutions containing 50mM citric acid, 150mM NaCl, 0.005% Tween-20 and 50mM Na_2HPO_4, 150mM NaCl, 0.005% Tween-20 were mixed to achieve running buffers at pH 6.5, 5.5 and 4.8. Each concentration series of purified transferrin was prepared by two-fold serial dilution into running

buffer to achieve a concentration range from 1 µM to 1 nM. Each dilution was injected over flow paths 1 and 2 for 240s to measure association kinetics, followed by injection of running buffer for 500s to measure dissociation. Sensorgrams were produced following subtraction of non-specific signals as described in 2.2.2. Sensorgrams were fitted to a 1:1 interaction model and kinetic parameters were determined using the Biacore T200 evaluation software.

2.2.5 Pull-down assay

Bn2-TfR and BES1-TfR were covalently linked to cyanogen bromide-activated (CNBr-activated) sepharose 4B (Sigma-Aldrich), following manufacturer's recommendation. Receptor-coupled sepharose was incubated with bovine, goat, horse, pig and rabbit sera followed by a brief wash step with PBS. Bound ligand was eluted with 0.1 M citrate pH 3 and analysed by SDS-PAGE.

2.2.6 Biolayer interferometry

Biolayer interferometry was performed using an Octet RED384 instrument (ForteBio). Biotinylated BES1-TfR was immobilised onto streptavidin (SA) biosensors (ForteBio, 18-5021). Assay buffers were prepared at four different pH values to represent extracellular and endocytic conditions. For physiological pH, a buffer containing 10 mM HEPES pH7.4, 150 mM NaCl, 0.1% bovine serum albumin (BSA), 0.05% Tween-20 was used. For endocytic pH, a phosphate-citrate buffer system was performed by mixing 50 mM citric acid, 150 mM NaCl, 0.1% BSA, 0.05% Tween-20 and 50 mM Na_2HPO_4, 150 mM NaCl, 0.1% BSA, 0.05% Tween-20 to achieve pH 6.5, 5.5 and 4.8. Bovine holo-transferrin (Sigma-Aldrich, T1408) and bovine apo-transferrin (Sigma-Aldrich, T1428) were two-fold serially diluted into the different pH buffers to achieve concentrations ranging from 1 µM to 62.5 nM. Receptor-conjugated biosensors were dipped into each dilution of Tf for 240s, followed by dissociation in assay buffer for 500s. A regeneration step using 0.1 M citrate pH 3 and assay buffer was performed between each dilution. Sensorgrams were produced using ForteBio analysis software.

2.3 Identification of the expressed transferrin receptor in trypanosomes

2.3.1 Growth rates of BES1-expressing cells

In vitro cultures were performed using the Lister 427 isolate expressing the BES1 VSG expression site. All sera were heat-inactivated by incubation at 56°C for two hours. Cells were cultured at 37°C, 5% CO_2 in HMI-9 salts (Hirumi & Hirumi, 1989) supplemented with 10% foetal calf serum (FCS; Gibco, 10270106). Trypanosomes were passaged daily to $1x10^5$ cells/ml and initially grown for 72 hours to ensure a steady growth rate. Cells were then washed with serum-free HMI-9 and 1×10^5 cells/ml were resuspended in HMI-9 containing 10% horse (Gibco, 16050130), rabbit (Gibco, 16120099), pig (Gibco, 26250084) or FCS. Trypanosomes were cultured for an additional 309 hours in two independent flasks for each condition. Cell counts were performed manually in duplicate using a haemocytometer.

2.3.2 Reverse transcription PCR

RNA was isolated at three time points (72, 191 and 381 hours) from the cultures used to determine growth rates. A RNeasy mini kit (Qiagen, 74104) was used, following manufacturer's protocol. RNA was treated with DNase to remove residual DNA using Turbo DNA-free kit (Invitrogen). Reverse transcription of RNA was performed using random oligomers and SuperScript II reverse transcriptase kit (Invitrogen) to generate complementary DNA (cDNA). PCR was performed using the cDNA template with ESAG6 or ESAG7-specific primers (Appendix 12). Amplified PCR products corresponding to ESAG6 and ESAG7 mRNAs from cells grown in foetal calf, horse, rabbit and pig sera were sequenced using the same primers.

2.4 X-ray crystallography

2.4.1 Enzymatic glycan and tag removal

Initial deglycosylation trials were performed using peptide:N-glycosidase F (PNGase F; NEB, P0704). Glycan cleavage was preformed using 10 µg of Bn2-TfR in PBS, following the

manufacturer's protocol. After incubation for either 4 hours or 24 hours, samples were compared to an untreated control sample of Bn2-TfR by SDS-PAGE to visualise glycan removal. BES1-TfR and Bn2-TfR were also treated with endoglycosidase Hf (Endo Hf, NEB, P0703) following the manufacturer's protocol and glycan removal was monitored by SDS-PAGE. BES1-TfR and Bn2-TfR were incubated with carboxypeptidase B (Roche, 10103233001) following manufacturer's protocol. The efficiency of carboxypeptidase peptide bond hydrolysis was analysed at various time points by western blot analysis. Western blots were performed as described in 2.2.1, and probed with a 1:4000 dilution of mouse anti-polyhistidine (Sigma-Aldrich, A7058) primary antibody and a 1:15000 dilution of goat anti-mouse IRDye 680RD (Licor, 926-68070) secondary antibody.

2.4.2 Initial crystallisation trials

Bn2-TfR expressed in the presence of kifunesine was incubated with a 1.1-fold molar excess of human holo-transferrin (Sigma-Aldrich, T4132) in the presence of carboxypetidase B, Endo Hf and PNGase F (following manufacturer's recommendation) for 20 hours at 37°C. BES1-TfR expressed with kifunensine was incubated with a 1.1-fold molar excess of bovine holo-transferrin (Sigma-Aldrich, T1408) and the same enzymes for 20 hours at 37°C. Receptor-ligand complexes were isolated by size exclusion chromatography (SEC) using a Superdex 200 16/60 column (GE Healthcare) with running buffer containing 10 mM Tris pH 8, 150 mM NaCl. Peak fractions were collected and analysed by SDS-PAGE. Fractions containing Bn2-TfR in complex with human Tf were concentrated to 13.7 mg/ml, and fractions containing BES1-TfR with bovine Tf were concentrated to 14 mg/ml. Initial crystallisation trials were performed by sitting drop vapour diffusion using pre-filled SwissSci 96-well plates with four condition screens (ProPlex, Morpheus, JCSG Plus and Midas Plus). Drops were formed by mixing 100 nl of reservoir solution with 100 nl of concentrated protein complex. Plates were sealed and incubated at 18°C and 4°C. Small crystals of Bn2-TfR in complex with human Tf were fished from three conditions:

i) 10% w/v polyethylene glycol (PEG) 5000 molomethyl ether (MME), 12% v/v 1-Propanol, 0.1 M MES, pH 6.5

ii) 0.1 M sodium chloride, 25% v/v pentaerythritol propoxylate, 10% v/v dimethyl sulfoxide

iii) 20% PEG 6000, 0.1 M Tris, pH 8.5.

Crystals were washed in reservoir solution to remove residual soluble protein present in the drop, then solubilised in NuPAGE LDS loading buffer and NuPAGE reducing agent. Samples were incubated at 70°C for 10 minutes and proteins were separated by SDS-PAGE. Gel staining was performed using a SilverQuest silver stain kit (Invitrogen, LC6070).

2.4.3 Crystal optimisation

Bn2-TfR N-linked glycan mutant was incubated with a 1.1-fold molar excess of human holo-transferrin. TEV protease (Invitrogen, 12575015) was added for cleavage of N-terminal tags present on Bn2 ESAG6, and PNGase F was added to remove transferrin glycans. Following a 16 hour incubation at 37°C, the TfR-Tf complex was isolated by SEC, as described in 2.4.2, and eluted complex was concentrated to 15.5 mg/ml. Complex containing the glycan-lacking mutant and the previously isolated complex containing glycosylated Bn2-TfR expressed in kifunensine were prepared for crystallisation in optimised conditions based on the conditions which produced small crystals. Crystals were obtained in 12% (w/v) PEG 5000 MME, 12% 2-methyl-2,4-pentanediol, 0.1 M MES, pH 6.5 at 18°C for both the glycan-lacking TfR mutant in complex with human Tf and the glycosylated TfR in complex with human Tf. Crystals were fished and cryo-cooled by Matthew Higgins (Department of Biochemistry, University of Oxford).

2.4.4 Structure determination

Data were collected from the Diamond Light Source by Matthew Higgins. Crystals of the TfR glycan mutant in complex with human Tf diffracted to 2.75Å resolution. Molecular replacement using Phaser (McCoy *et al*., 2007) was initially performed using human Tf as a template. Each lobe of human Tf was interrogated separately using structure 4X1B (Wang *et al*., 2015) from the protein data bank (PDB) for molecular replacement. Next, the structure of a type A variant surface glycoprotein (VSG) N-terminus (PDB: 2VSG) (Blum *et al*., 1993) was used for molecular replacement of ESAG6 and ESAG7 based on predicted structural homology between the TfR and VSG (Salmon *et al*., 1997). Restricted regions of the VSG alpha helices were used (residues 7-112 and 239-251). Refinement was performed by Matthew Higgins using Buster (Blanc *et al*., 2004), and model building was performed alongside Matthew Higgins using Coot (Emsley *et al*., 2010).

Bn2-TfR expressed in the presence of kifunensine and treated with Endo Hf crystallised in complex with human Tf in the same conditions as the glycan-lacking TfR. Crystals diffracted to 3.6Å resolution and molecular replacement was performed with Phaser using the previously solved structure of Bn2-TfR in complex with human Tf. During sample preparation in 2.4.2, glycan removal using Endo Hf resulted in a residual N-acetylglucosamine (GlcNac). Glycans were modelled using the electron density of the residual sugars and the TfR glycan models generated by Mehlert *et al*. (2012). Sequence entropy was determined by Matthew Higgins using the Shannon Entropy-One program with all available *T. b. brucei* ESAG6/7 sequences from the Lister 427 isolate (Hertz-Fowler *et al*., 2008) (Berriman *et al*., 2002) and the EATRO 2340 isolate (Young *et al*., 2008) (Appendix 4).

2.5 Iron quantification

Iron quantification was performed by microbeam focussed particle induced X-ray emission (microPIXE). The experiment was performed by Peter Woodcock, Elspeth Garman and Matthew Higgins (Department of Biochemistry, University of Oxford). The differences in iron stoichiometries were measured over a range of pH conditions that mimicked endosomal progression. Bn2-TfR used in this assay was expressed in the presence of kifunensine and contained mannose-rich glycans which would not interfere with Tf binding as shown in the crystallography data. Bn2-TfR was incubated with human holo-Tf (Sigma Aldrich) and the receptor-ligand complex was injected over a Superdex 200 15/300 size exclusion chromatography (SEC) column (GE Healthcare) in various pH conditions to investigate the effects of pH on iron release. Human holo-Tf (Sigma-Aldrich) was studied under the same pH conditions to serve as a control and to concomitantly compare the effects of pH and receptor binding on iron retention. SEC was performed using the following buffers and, to avoid the use of chloride ions which would interfere with the microPIXE analysis, the pH was adjusted using citric acid:

(i) 50mM sodium citrate pH 4.8, 150 mM NaBr

(ii) 50mM sodium citrate pH 5.5, 150 mM NaBr

(iii) 50mM Bis-Tris pH 6.5, 150 mM NaBr

(iv) 50mM Bis-Tris pH 7.5, 150 mM NaBr.

Protein samples were subsequently concentrated to 1 mg/ml and microPIXE was performed to measure the retained iron. Measurements were performed at the Ion Beam Centre, University of Surrey. X-ray emissions from dried protein droplets (volume per droplet = 0.1 µl) were generated by a 2.5 MeV proton beam (diameter = 2.0 µm) and measured using a detector. The characteristic X-ray emission for each element allowed identification and quantification of atoms present in each sample. Sulphur was used as in internal standard as the number of sulphur atoms per molecule can be determined based on the number of methionines and cysteines present. Quantities of sulphur and iron were compared and the relative number of iron ions per protein sample was determined.

2.6 Growth rates in transferrin-depleted conditions

FCS (Gibco, 10270106) was depleted of Tf using a Bn2-TfR column. Bn2-TfR was covalently linked to a 5 ml HiTrap NHS-activated HP column (GE Healthcare, 17071701) following the manufacturer's protocol. Heat-inactivated FCS (25 ml) was passed over the column repeatedly until Tf was not detectable by western blot. After each cycle, bound Tf was removed from the column using 0.1 M citrate pH 3. Western blots were performed as described in 2.2.1, and probed using a 1:500 dilution of polyclonal rabbit anti-bovine transferrin antibody (Bethyl Laboratories, A10-114), followed by a 1:15000 dilution of goat anti-rabbit IRDye 680LT (Licor, 926-68071). A sample of 10% FCS depleted of Tf was analysed by western blot. Samples containing 10% FCS and 10% Tf-depleted FCS supplemented with 3 µM Tf were also compared. FCS was serially diluted to determine the antibody detection limit. The concentrations reported were based on the assumption that mammalian serum contained ~30 µM Tf (Dietmar Steverding, 2003).

Trypanosomes expressing BES1-TfR were cultured *in vitro* as described in 2.3.1, in HMI-9 medium containing 10% FCS or 10% Tf-depleted FCS. Cells were cultured for 200 hours and cell counts were performed manually in duplicate for two independent flasks.

2.7 Phage display

2.7.1 Media

Tryptone yeast medium (2TY) was prepared with 16 g tryptone, 10 g yeast extract and 5 g NaCl per litre of medium. *E. coli* TG1 cells (genotype: K-12 *supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5, ($r_K^- m_K^-$)* F' [*traD36 proAB*$^+$ *lacI*^{q} *lacZΔM15*]) were used throughout the phage display

process as they can read through the amber stop codon and integrate a glutamine in normal conditions but recognise the stop codon in the presence of IPTG, allowing for modulable expression of fusion proteins.

2.7.2 Soluble selections

An aliquot of $5x10^{11}$ phage from each of the bone marrow Vaughan (BMV) library (Vaughan *et al*., 1996) and the combined spleen (CS) library (Lloyd *et al*., 2009) were mixed to form an input BMV/CS library of $1x10^{12}$ phage. The BMV/CS library was then used for soluble selections against the BES1-TfR and Bn2-TfR targets. To reduce non-specific interactions, several blocking steps were performed using 3% milk powder (Marvel) in phosphate-buffered saline (PBS; Sigma, D8537), unless otherwise stated. All incubations were performed for one hour at room temperature with end-over-end rotation at 20 rpm, unless otherwise stated.

The BMV/CS phage library and magnetic streptavidin beads (Dynabeads) were blocked with milk independently, then beads were added to the library to select and remove streptavidin binders. Streptavidin-deselected phage were incubated with 100 nM biotinylated BES1-TfR or 100 nM biotinylated Bn2-TfR. Blocked streptavidin beads were added, and five wash steps were performed with PBS containing 0.1% Tween-20 (PBST) using a KingFisher magnetic particle processor (Thermo Fisher Scientific, 5400000). Beads were incubated with 10 µg/ml trypsin in 0.1 M sodium phosphate (pH 7) for 30 minutes at 37°C to mediate the release of bound phage. Recovered supernatant was added to 800 µl mid-logarithmic phase *E. coli* TG1 cells grown in 2TY medium. Phage infection was achieved by incubation for 1 hour at 37°C with slow shaking at 150 rpm. Cells were plated onto 500 cm^2 bioassay plates containing 2TY agar supplemented with 100 µg/ml ampicillin and 2% glucose (2TYAG). Diluted cells were plated onto petri dishes containing 2TYAG agar to determine colony forming units per ml (cfu/ml) for output quantification. Plates were incubated at 30°C for 16 hours and petri dish colonies were counted to determine the output library size. Colonies were scraped from bioassay plates into 2TY supplemented with 17% glycerol and stored at -80°C.

2.7.3 Phage rescue

Output library glycerol stocks were used to inoculate 25 ml of 2TYAG to achieve a starting optical density at 600 nm (OD_{600}) of 0.1. Flasks were incubated at 37°C, shaking at 280 rpm

until mid-logarithmic phase. M13KO7trp helper phage ($7.5x10^{10}$ plaque forming units; pfu) were added and cells were incubated at 37°C for 1 hour, shaking at 150 rpm. Cultures were harvested at 3200 rpm for in a Sorvall XTR centrifuge for 10 minutes and resuspended in 25 ml 2TY supplemented with 100 µg/ml ampicillin and 50 µg/ml kanamycin. The absence of glucose prevents repression of the *LacZ* promoter, enabling expression of scFv-pIII from the phagemid vector. Cultures were incubated overnight at 25°C, shaking at 280 rpm. One ml of overnight culture was centrifuged at 13000 rpm for 5 mins using a microcentrifuge. Phage input titres were calculated following *E. coli* infection and colony counts to determine number of cfu. Fifty µl of phage-rich supernatant was used to perform the next round of selection in the same way, the only difference being a decrease in antigen during each round (50 nM for round 2 and 25 nM for round 3).

2.7.4 Master plate generation

Plates from the 221-TfR and Bn2-TfR output libraries of the third round of selection were used for colony picking. Forty-four colonies from each antigen output were inoculated into wells containing 150 µl 2TYAG medium in a 96-well master plate. Colonies from TG1 cells containing a phagemid vector encoding an scFv targeting an irrelevant antigen were inoculated to provide a negative control. Following overnight culture, a small volume of culture was removed for sequencing and the master plate was supplemented with 17% glycerol and stored at -80°C. Sequencing of the CDR3 was performed following colony PCR with primers to amplify the scFv (Appendix 12). PCR products were sequenced using the amplification primers and sequence diversity of the output library was estimated.

2.7.5 Phage enzyme-linked immunosorbent assay

To determine specificity of the enriched output library, enzyme-linked immunosorbent assays (ELISAs) were performed using phage extruded from *E. coli* colonies in the master plate. The master plate was replicated by inoculating 500 µl of 2TYAG medium in a 96 deep-well plate using a metal replica plater. Following incubation of the daughter plate (37°C, 280 rpm), $1.5x10^{9}$ pfu of M13KO7 helper phage were added to each well of exponentially growing cells. Cultures were incubated at 37°C for 1 hour with 150 rpm shaking. The plate was centrifuged

at 3200 rpm for 10 minutes in a Sorvall XTR centrifuge. Supernatants were discarded and cells were resuspended in 500 µl of 2TYAK medium. Cultures were incubated overnight at 25°C, shaking at 280 rpm. Overnight cultures were incubated with 3% milk powder in PBS (MPBS) for 1 hour at room temperature, then centrifuged at 3200 rpm for 5 minutes. Supernatants containing blocked phage were transferred to 96-well assay plates coated with streptavidin and antigen (50 µl of 221-TfR or Bn2-TfR at 5 µg/ml) or streptavidin alone, that had been previously blocked with 3% MPBS for 1 hour and washed three times with PBS. Phage were incubated with antigen or streptavidin for 1 hour at room temperature, then wells were washed three times with PBST. Anti-M13-horseradish peroxidase was diluted 1:5000 in 3% MPBS and 50 µl were added per well. Following incubation for 1 hour at room temperature, plates were washed with three times with PBST. 50 µl of 3,3',5,5'-tetramethylbenzidine ELISA-substrate (Thermo Fisher Scientific, 34028) were added per well and, following colour development, the reaction was quenched using 50 µl of 0.5 M sulphuric acid. Absorbance at 450 nm wavelength was detected using an EnVision plate reader (PerkinElmer) and data were plotted using GraphPad Prism version 8 software.

2.7.6 Periplasm extraction

The scFv-pIII fusion was encoded in a lac operon inducible expression system within the phagemid vector. An amber codon separating the scFv and pIII constructs allowed for modulation of scFv-pIII fusion expression (Vaughan *et al*., 1996). IPTG induction of the lacZ promoter resulted in reduced read-through efficiency, with 90% of transcripts terminating translation at the amber codon. A leader sequence mediated delivery of soluble scFv to the periplasmic compartment, and controlled osmotic pressure resulted in rupture of the outer membrane and release of the periplasm, enabling scFv recovery as described below.

A 96 deep-well plate containing 1 ml/well of 2TY with 100 µg/ml of ampicillin and 0.1% glucose was inoculated with *E. coli* from the master plate, as described previously. Cells were incubated at 37°C with shaking at 280 rpm. Mid-logarithmic phase cells were induced with 0.02 mM IPTG and incubated overnight at 30°C, shaking at 280 rpm. Overnight cultures were centrifuged for 15 minutes at 4600 rpm in a pre-chilled (4°C) Sorvall XTR centrifuge. Supernatants were discarded and cells in each well were resuspended in osmotic shock buffer containing 200 mM Tris-HCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA) and 0.5 M sucrose, pH 8. Cultures were incubated for 30 minutes at 4°C and centrifuged for 15 minutes at 4600 rpm (4°C) to pellet cell debris and recover periplasm containing soluble scFv.

2.7.7 Homogenous time-resolved fluorescence

Homogenous time-resolved fluorescence (HTRF) assay buffer was prepared using PBS (Sigma, D8537), 0.1% BSA (Sigma, A9576), 0.4 M potassium fluoride (VWR, 26820.236). The following were combined into assay buffer to achieve final concentrations of 2 nM biotinylated antigen (BES1-TfR or Bn2-TfR), 1 ng/ml streptavidin cryptate (CisBio, 610SAKLB), 10 nM anti-Myc-XL665 (CisBio, 61MYCXLB) and 25% periplasm extract recovered previously, in a 10 µl reaction volume. Negative controls were prepared in the same way using periplasm material from scFv clones targeting an irrelevant antigen. Assays were performed in a 384-well plate and incubated for 1 hour at room temperature. Fluorescence was detected at 620 nm and 665 nm using an EnVision plate reader. Data were analysed using the following equations:

Ratio = (665nm/620nm)*10,000

% Delta F = ((Ratio Sample - Ratio Negative)/Ratio Negative))*100

The % Delta F of each sample was determined, and values were plotted using GraphPad Prism version 8 software.

2.7.8 Fab binding analysis using surface plasmon resonance

Fab conversions and purifications were performed by the biologics expression team at AstraZeneca. Surface plasmon resonance (SPR) was performed using a Biacore T100 with biotinylated BES1-TfR and Bn2-TfR in the glycosylated form immobilised on a biotin CAPture chip, as described previously. Experiments were performed at 25°C in running buffer containing 10 mM HEPES pH 7.4, 150 mM NaCl, 0.005% Tween-20. Purified Fabs were prepared in running buffer using a two-fold serial dilution from 1 uM to 62.5 nM. Data were analysed using Biacore T100 software, as described previously.

2.7.9 Fluorescence microscopy

2.7.9.1 Immunoglobulin G expression and labelling

IgG1 is typically the most abundant subclass in human serum (Vidarsson, Dekkers, & Rispens, 2014) and was the selected expression format for IgG conversion. The IgG1 framework had been engineered to contain a free cysteine at position 239 in the constant region of each heavy

chain, referred to as the IgG1-maia format. The cysteine can subsequently be used for thiol-reactive conjugation of fluorophores and toxins. V_H fragments were amplified by PCR and cloned into the IgG1-maia format, while V_L fragments were cloned into the lambda light chain framework. The use of the lambda light chain as opposed to the kappa light chain was determined by the flanking framework of the scFv sequence. Tbru108, Tbu111 and Nip228 were expressed in CHO cells by equimolar co-expression of heavy and light chains, as described previously in 2.1.2. IgGs were purified by protein A affinity chromatography using a 5 ml HiTrap Protein A HP column (GE Healthcare, 17040301) following manufacturer's protocol. AlexaFluor 488 was site-specifically conjugated to cysteine 239 of each heavy chain, by incubating purified IgG with a 40-fold molar excess of Tris-(2-carboxyethyl)-phosphine (TCEP) for three hours, followed by a 20-fold molar excess of dehydroascorbic acid (dhAA) for four hours, then a 10-fold molar excess of AlexaFluor 488 C5 maleimide (ThermoFisher Scientific, A10254) for one hour. N-acetyl-L-cysteine (NAC) was added to stop the reaction and labelled IgGs were buffer exchanged into PBS using a PD-10 column, following manufacturer's protocol.

2.7.9.2 Uptake assays

The Lister 427 isolate expressing from BES1 was cultured as described in 2.3.1. Trypanosomes were washed in serum-free HMI-9 and resuspended in serum-free HMI-9 with 0.5 mg/ml bovine serum albumin (BSA). A lysosome protease inhibitor FMK024 (morpholinourea-phenylalanine-homophenylalanine-fluoromethyl ketone) was added to a final concentration of 2 µM and cells were incubated with 500 nM of Tbr108, Tbu111 or Nip228 at 37°C for 1.5 hours, based on previous experimental conditions defined by MacGregor *et al*. (2019). As a control, 100 nM of bovine Tf, labelled with AlexaFluor 488 (ThermoFisher Scientific, A10235) following manufacturer's recommendation, was added to cells in the same conditions. Cells were washed with HMI-9 and fixed with 1% formaldehyde. Cells were mounted on microscope slides and visualised at 100x magnification using identical settings for all samples with a Nikon Eclipse Ti widefield microscope.

Chapter 3

Investigation of transferrin receptor species specificity

3.1 Introduction

3.1.1 The transferrin receptor: a functional heterodimer

The trypanosome transferrin receptor (TfR) is encoded by two genes, ESAG6 and ESAG7, which are homologous and have evolved from a gene duplication event. ESAG6 and ESAG7 protein sequences from the Bn2 expression site are represented in Figure 6A. The primary divergence between ESAG6 and ESAG7 is the C-terminal truncation of ESAG7 and absence of a GPI-anchor addition site. Membrane insertion of ESAG6 and ESAG7 heterodimers is mediated by a single GPI anchor at the ESAG6 C-terminus. Excluding the ESAG6 C-terminus, the Bn2 ESAG6 and ESAG7 proteins have 82% sequence identity. ESAG6 or ESAG7 homodimers can form but heterodimerisation is necessary for Tf binding (Salmon *et al*., 1994) (Chaudhri *et al*., 1994). Understanding why both subunits are required for Tf binding and how trypanosomes orchestrate the assembly of heterodimers over homodimers are key questions that will be investigated in this chapter.

3.1.2 The TfR: a variant receptor

DNA sequencing of the telomeric expression sites of the T. b. brucei L427 isolate enabled a comprehensive analysis of the VSG and ESAG genes (Hertz-Fowler *et al*., 2008). ESAG6 and ESAG7 genes were present in one or multiple copies in each BES. Furthermore, these genes had diversified, and the sequence variation could be investigated. Alignments of the L427 ESAG6 (Figure 7) and ESAG7 (Figure 8) protein sequences demonstrated that polymorphisms were not randomly distributed and clustered in hotspots. Thus, variation was restricted to a mere tenth of the amino acid sequence with more than 90% sequence identity between BESs. These regions of variation were common between all ESAG6 or all ESAG7 sequences and have been described as hypervariable sites (Salmon *et al*., 1997). Solved VSG structures from Blum *et al*. (1993) provided clues on the location of these sites, as the TfR was predicted to share structural homology with the N-terminal domain of the VSG. Modelling of the TfR onto the VSG N-terminus revealed a membrane distal positioning within loops at the hypothetical ligand binding pocket of the receptor. Diversification of these residues was shown to produce differences in Tf affinity between variants of the TfR repertoire (Salmon *et al*., 1997).

3.1.3 Species specificity hypothesis

In 1998, Bitter *et al*. noted a marked difference in trypanosome proliferation depending on the species of serum in which the cells were cultured. In their example, trypanosomes expressing the 221-TfR (BES1-TfR) showed a normal growth rate in bovine serum, with rare switching events. When transferred to canine serum, a reduction in proliferation was observed for 200 hours. After 200 hours, a normal growth rate was observed and the outgrowing population of cells had switched expression site to express the VO2-TfR (BES2-TfR) (Bitter *et al*., 1998). The proposed hypothesis was that a change in mammalian serum led to a switch in transferrin receptor to accommodate the different transferrin. Over the years, this hypothesis was either supported (Gerrits *et al*., 2002) (Isobe *et al*., 2003) (Van Luenen et al., 2005) (Cordon-Obras *et al*., 2013) or contested (Salmon *et al*., 2005), with some arguing that the requirement for a diverse TfR repertoire may be driven by immune avoidance (Borst, 1991), and others claiming that both host adaptation and immune components may play a role in driving receptor switching (Gerrits *et al*., 2002).

3.1.4 A comparative study between two TfRs

To further disentangle the complexities surrounding the role of TfR multiplication and diversification, two different TfRs were chosen for comparative studies. Recombinant TfRs were expressed and differences in affinity for Tf from a range of mammals was investigated. A mammalian expression system was selected for recombinant TfR expression, due to the presence of glycans and disulphide bonds in both ESAG6 and ESAG7. Surface plasmon resonance (SPR) was be performed to measure the kinetic binding parameters. To compare Tf binding between the two receptors, a panel of mammalian species was selected to represent diverse families and differences in transferrin affinities were measured.

3.2 Aims

To understand how polymorphisms in the TfR repertoire may affect the affinity for Tf, my aim was to conduct a comparative study to identify the extent of species specificity by comparing kinetic measurements for Tf binding to two of the most different TfRs. SPR was performed to accurately determine the differences in receptor affinity for transferrin of different species and several questions would be addressed:

Do ESAG6 and ESAG7 subunits preferentially heterodimerise?

To what extent have TfRs evolved to bind Tf in a species-specific manner?

Is a high affinity receptor necessary to support growth?

Does BES switching occur if a low affinity receptor is expressed?

3.3 Results

3.3.1 Selection of TfRs and design of recombinant constructs

Phylogenetic trees were assembled containing all ESAG6 or ESAG7 amino acid sequences from expression sites in L427 genome to inform the choice of receptor (Figure 6B). As a starting point, the TfR encoded in BES1 was selected (also commonly referred to as the 221 expression site) based on reports that switching had been observed in cells expressing BES1-TfR (Bitter *et al*., 1998) (Gerrits *et al*., 2002). The Bn2 expression site encoded a TfR distant from BES1-TfR in both the ESAG6 and ESAG7 protein sequences and was chosen for further investigation. The Bn2 expression site was studied during characterisation of trypanosome expression sites (Berriman *et al*., 2002) and contains an ESAG6 sequence identical to the BES17 ESAG6, and an ESAG7 with a difference of ten amino acids to BES17 ESAG7. While the BES nomenclature (Hertz-Fowler *et al*., 2008) was used for naming BES1-TfR, Bn2-TfR cannot be assigned to the BES nomenclature due to the fact that the Bn2 expression site was not identified in subsequent sequencing efforts, which may be the result of gene conversion or sequencing ambiguities.

The ESAG6 and ESAG7 genes from BES1 (referred to as BES1e6 and BES1e7) and the ESAG6 and ESAG7 genes from Bn2 (Bn2e6 and Bn2e7) were codon optimised and modified to facilitate expression in a mammalian expression system. The native trypanosome signal peptide was removed and replaced with mammalian CD33 signal peptide to allow for secretion in a mammalian cell expression system. The GPI-anchor addition site on ESAG6 was removed (Figure 9A) and an AviTag for biotinylation and deca-His-tag for purification were added at the carboxy (C) terminus (Figure 9B). For the ESAG7 constructs, a strepII tag was added at the C-terminus. Genes were expressed under the cytomegalovirus (CMV) promoter in CHO cells.

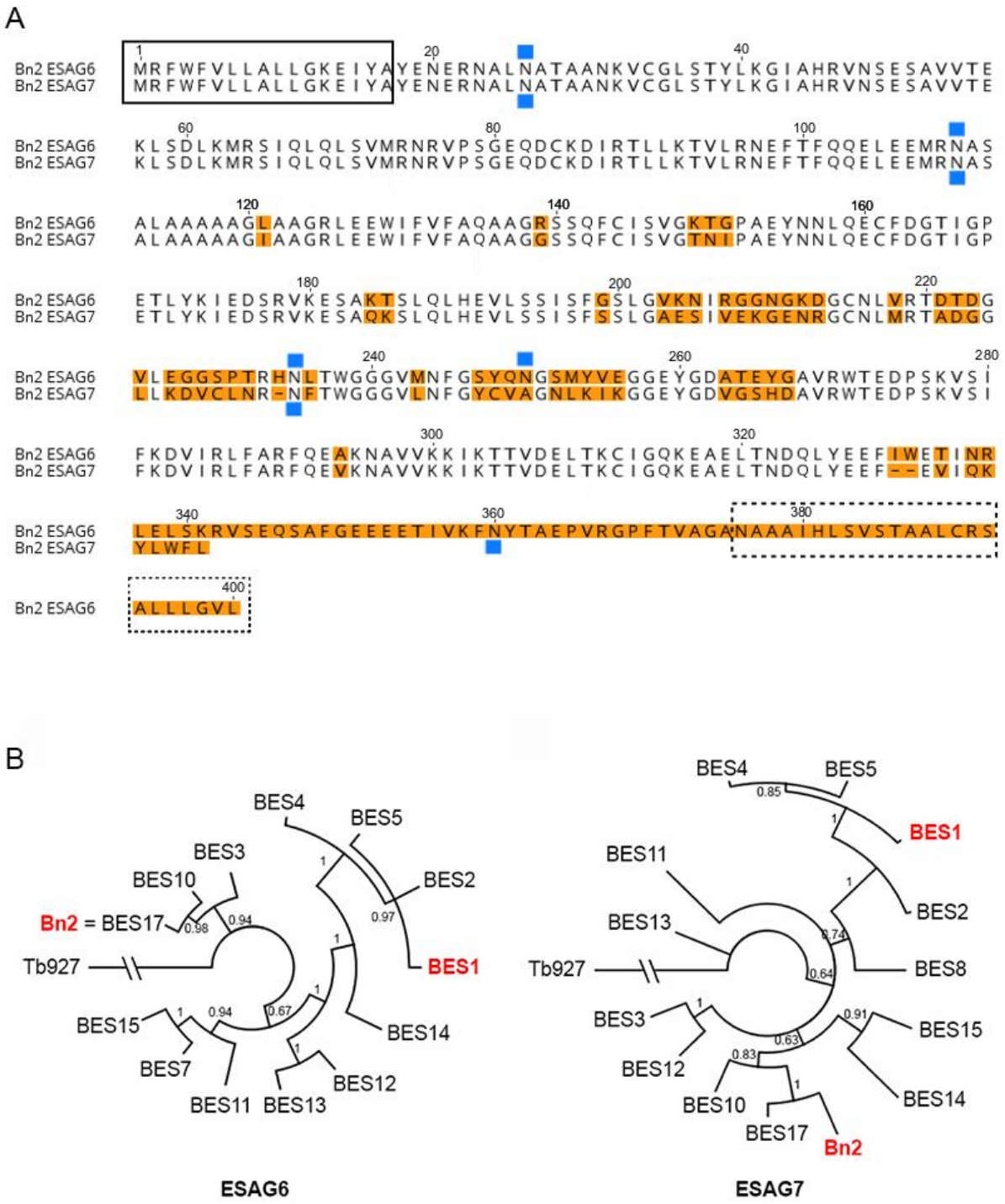


Figure 6: Alignment of ESAG6 and ESAG7 sequences and phylogenetic analysis. A, Protein sequence alignment of the Bn2-TfR ESAG6 and ESAG7 sequences. A solid box indicates the signal peptide, a dashed box indicates the GPI-anchor addition site and variation is highlighted in orange. B, Protein sequences of the L427 ESAG6 and ESAG7 from the expression sites of the L427 strain are used to construct phylogenetic trees and inform the selection of evolutionarily distant TfRs indicated in red.

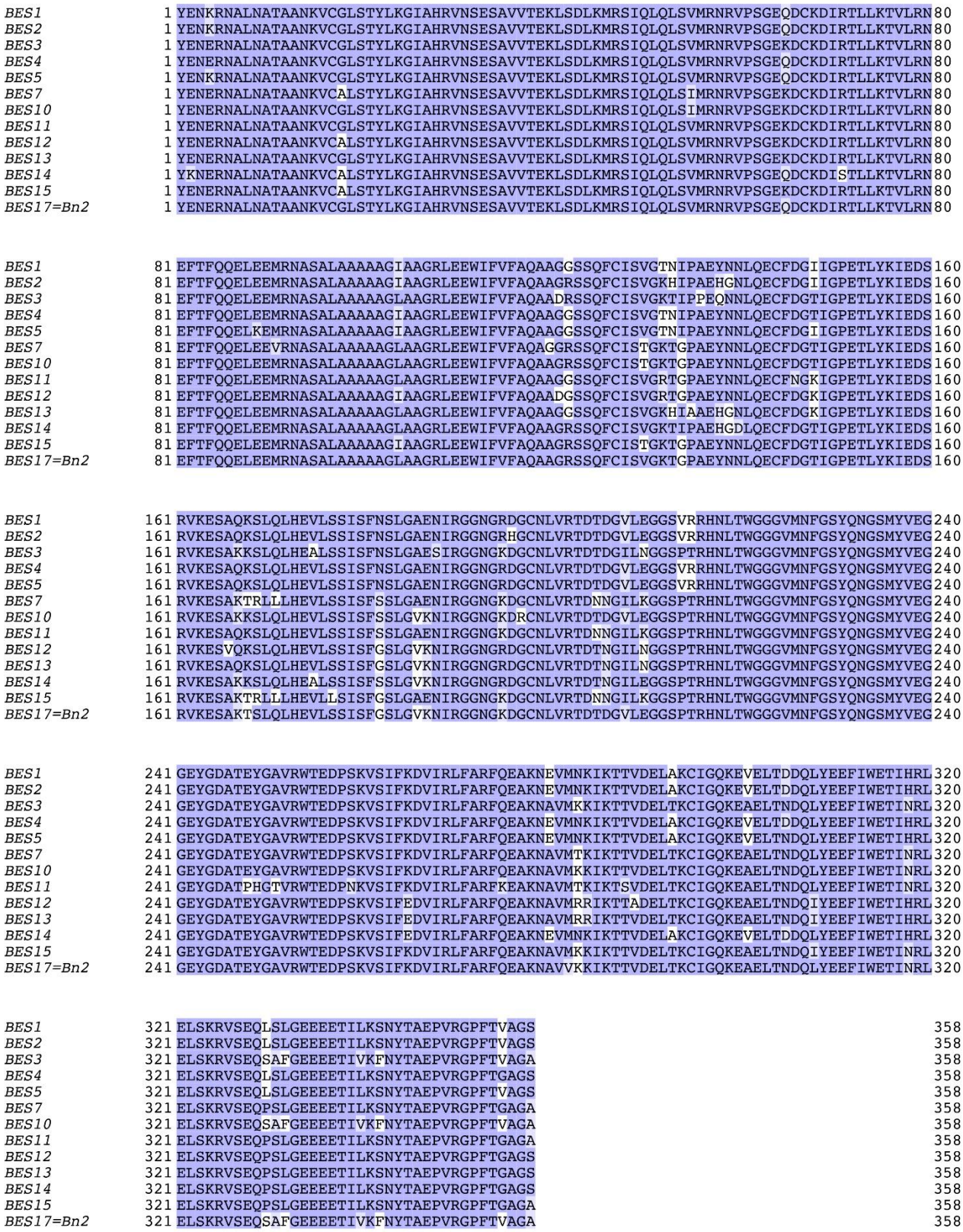


Figure 7: Alignment of the L427 ESAG6 sequences. ESAG6 amino acid sequences were aligned with diversity highlighted in white and conserved residues indicated in blue.

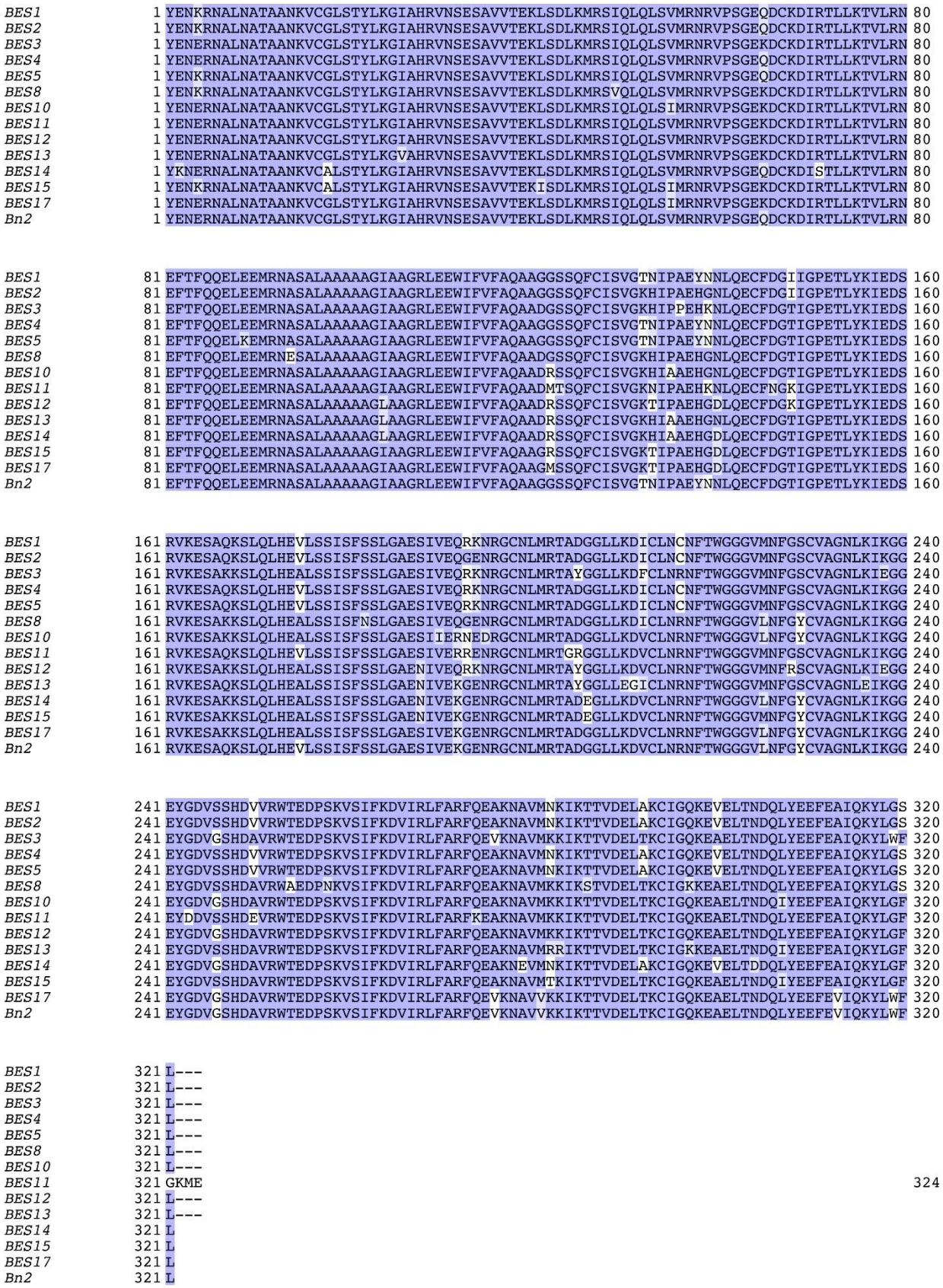


Figure 8: Alignment of the L427 ESAG7 sequences. ESAG7 amino acid sequences were aligned with diversity highlighted in white and conserved residues indicated in blue.

3.3.2 Recombinant ESAG6 and ESAG7 expression

In the pilot expression test, BES1e6, BES1e7, Bn2e6 and Bn2e7 were expressed individually in CHO cells. BES1e6 and Bn2e6 containing His-tags were purified by nickel affinity chromatography. BES1e7 and Bn2e7 with strepII tags were purified by streptactin resin (Figure 9A, B). Each purified ESAG6 was subsequently incubated with the corresponding ESAG7 from the same expression site. The assembly of a heterodimeric receptor was assessed by size exclusion chromatography and showed no evidence of the presence of heterodimers (Appendix 5). The ESAGs remained as homodimers and did not reassemble as heterodimers. Co-expression of ESAG6 and ESAG7 from the same expression site resulted in the formation of heterodimers that could be isolated by nickel affinity chromatography (Figure 9C), resulting in the formation of BES1-TfR (Figure 9D) and Bn2-TfR (Figure 9E). Recombinant ESAG6 (M_r = 44,000) had an additional five predicted glycan modifications, producing a higher molecular weight separation by SDS-PAGE. Recombinant ESAG7 (M_r = 37,000) had three predicted N-linked glycosylation sites. Although ESAG7 did not play a role in affinity chromatography, equimolar amounts of ESAG6 and ESAG7 appear to co-elute. Thus, the propensity for the receptor to preferentially heterodimerise was investigated.

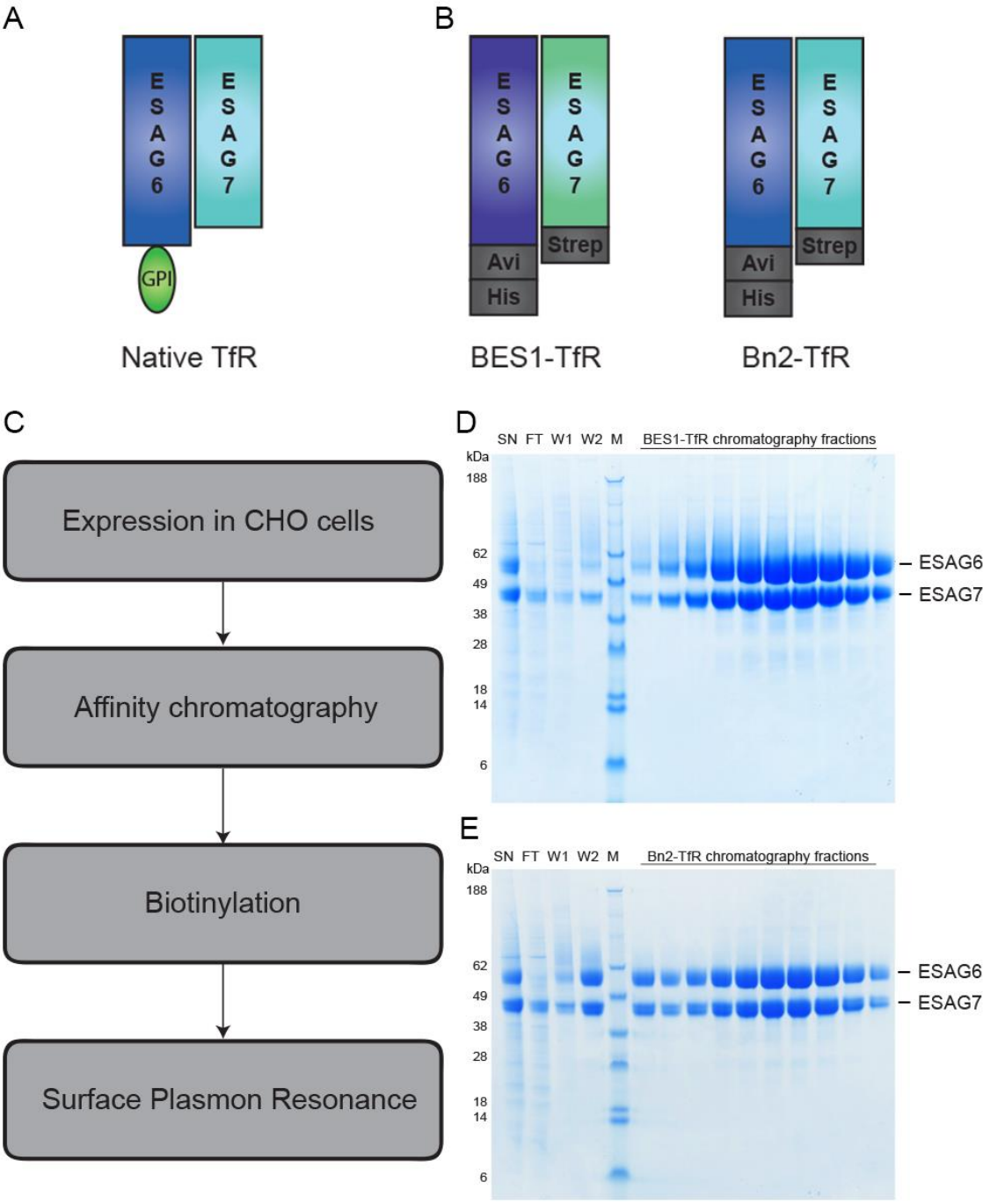


Figure 9: Recombinant TfR design and purification. A, Native TfR schematic depicting an ESAG6 and ESAG7 heterodimer. The GPI-anchor addition site in the ESAG6 C-terminus mediates GPI anchor attachment for membrane anchorage. B, Recombinant TfR constructs were made by removing the ESAG6 GPI-anchor addition site and adding a C-terminal Avitag for site-specific biotinylation and a His-tag for affinity chromatography. A StrepII-tag was added to the ESAG7 C-terminus for purification. C, Workflow for TfR characterisation by SPR. Recombinant TfR constructs are expressed in CHO cells, purified by nickel affinity chromatography and biotinylated by BirA biotin ligase in preparation for SPR. D, BES1-TfR nickel affinity chromatography analysis by SDS-PAGE. E, Bn2-TfR nickel affinity chromatography. SN = supernatant; FT = flowthrough; W1 and W2 = washes; M = protein marker MW ladder.

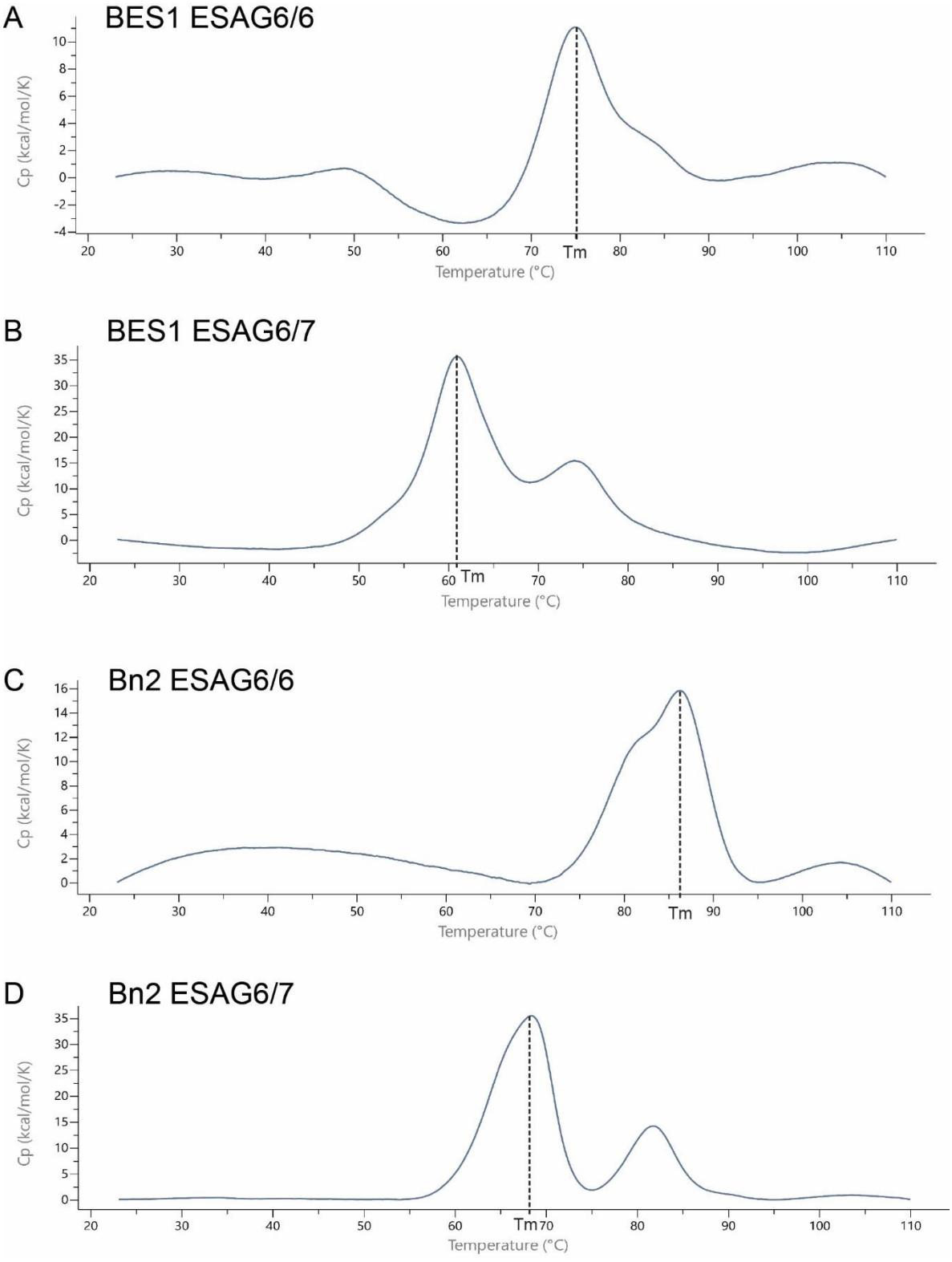


Figure 10: Differential scanning calorimetry (DSC) to compare melting temperatures (Tm) of ESAG6 homodimers and ESAG6/7 heterodimers. A, ESAG6 homodimers from BES1 were tested by DSC and a Tm of 75°C was measured. B, BES1 ESAG6/7 heterodimers recorded a Tm of 61°C. C, Bn2 ESAG6 homodimers were melted at 86°C. D, A Tm of 68°C was measured for Bn2 ESAG6/7 heterodimers. Tm = melting temperature; Cp = heat capacity.

3.3.3 Investigation of preferential heterodimerisation

The hypothesis was that the formation of heterodimers would be more energetically favourable than the formation of homodimers, thereby making heterodimers more thermostable. Differential scanning calorimetry (DSC) was performed to compare the thermostability of different dimer configurations. A lower melting temperature (Tm) would indicate a less stable structure while a high Tm correlates with an energetically favourable and stable conformation. Melting temperatures of ESAG6/6 homodimers from the BES1 and Bn2 expression sites were compared to ESAG6/7 heterodimers from the same expression sites. The BES1 ESAG6 homodimers denatured with a Tm of 75°C (Figure 10A) while ESAG6/7 heterodimers forming the BES1-TfR denatured at the lower temperature of 61°C (Figure 10B). ESAG6 homodimers from the Bn2 expression site produced a higher Tm of 86°C (Figure 10C) compared to its ESAG6/7 counterparts which denatured at 68°C (Figure 10D). ESAG6 homodimers showed increased thermostability for both expression sites compared to the ESAG6/7 heterodimers, providing the unexpected conclusion that homodimers have increased stability and require more energy to denature than heterodimers. The preferential formation of ESAG6/7 heterodimers in a co-expression system, and indeed most likely within the parasite, remains unsolved.

3.3.4 Surface plasmon resonance to investigate species specificity hypothesis

3.3.4.1 Site-specific biotinylation

TfRs were prepared for surface plasmon resonance (SPR) by site-specific biotinylation of the ESAG6 C-terminus. Recombinant BirA containing an N-terminal glutathione S-transferase (GST) tag was expressed in *E. coli* and purified by glutathione sepharose (Figure 11A). GST-BirA was incubated with BES1-TfR or Bn2-TfR in the presence of biotin and ATP to mediate site-specific biotinylation within the AviTag motif. A western blot probed with fluorescently-labelled streptavidin indicated biotinylation of the ESAG6 subunit in BES1-TfR and Bn2-TfR (Figure 11B). The corresponding Coomassie-stained SDS-PAGE confirmed the presence of heterodimers (Figure 11C), while the western blot demonstrated the absence of biotinylation prior to BirA addition, and ESAG6-specific biotinylation following BirA incubation (Figure 11B).

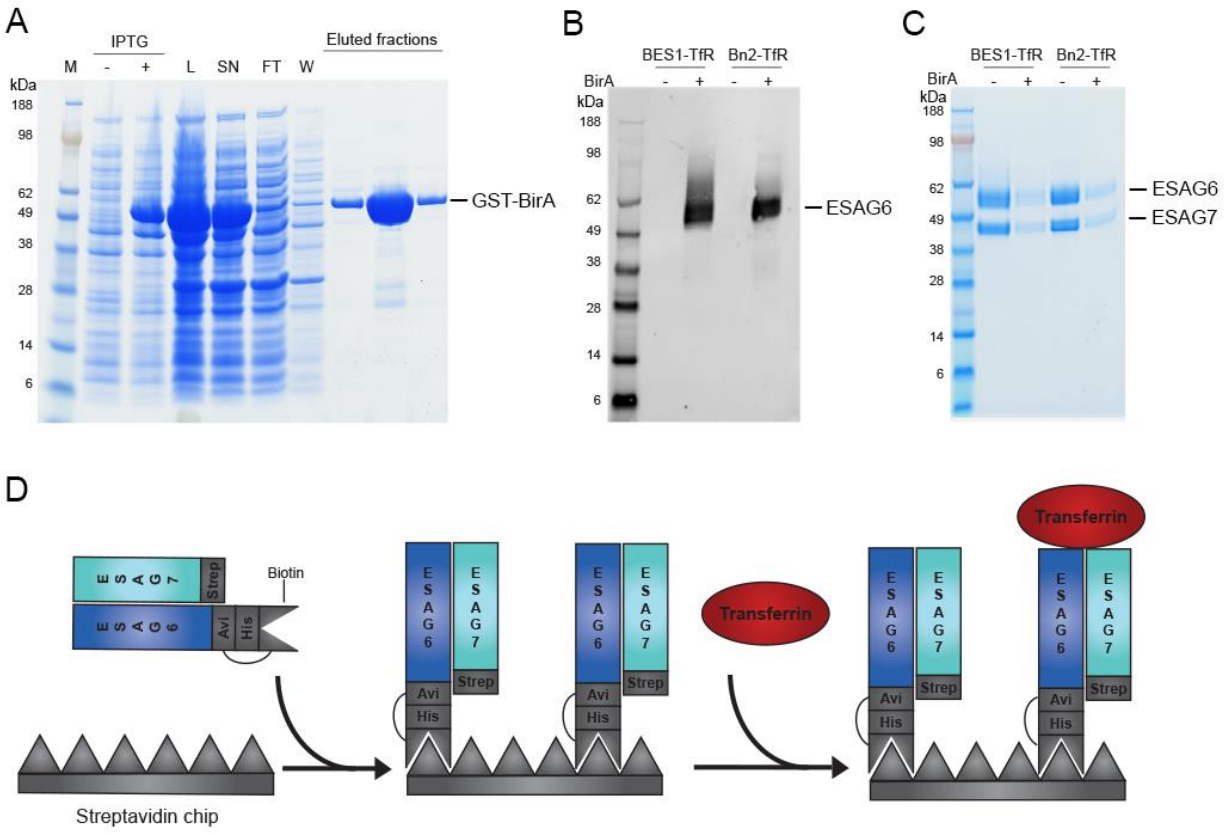


Figure 11: TfR biotinylation for surface plasmon resonance (SPR). A, GST-BirA expression by IPTG induction and purification by glutathione affinity chromatography. B, Analysis of TfR biotinylation by western blot with fluorescently-labelled streptavidin. C, Coomassie-stained SDS-PAGE containing samples analysed by western blot in B. D, SPR is performed using a streptavidin chip for biotinylated TfR immobilisation. Serum or purified transferrin is flowed over the chip and Tf binding is measured.

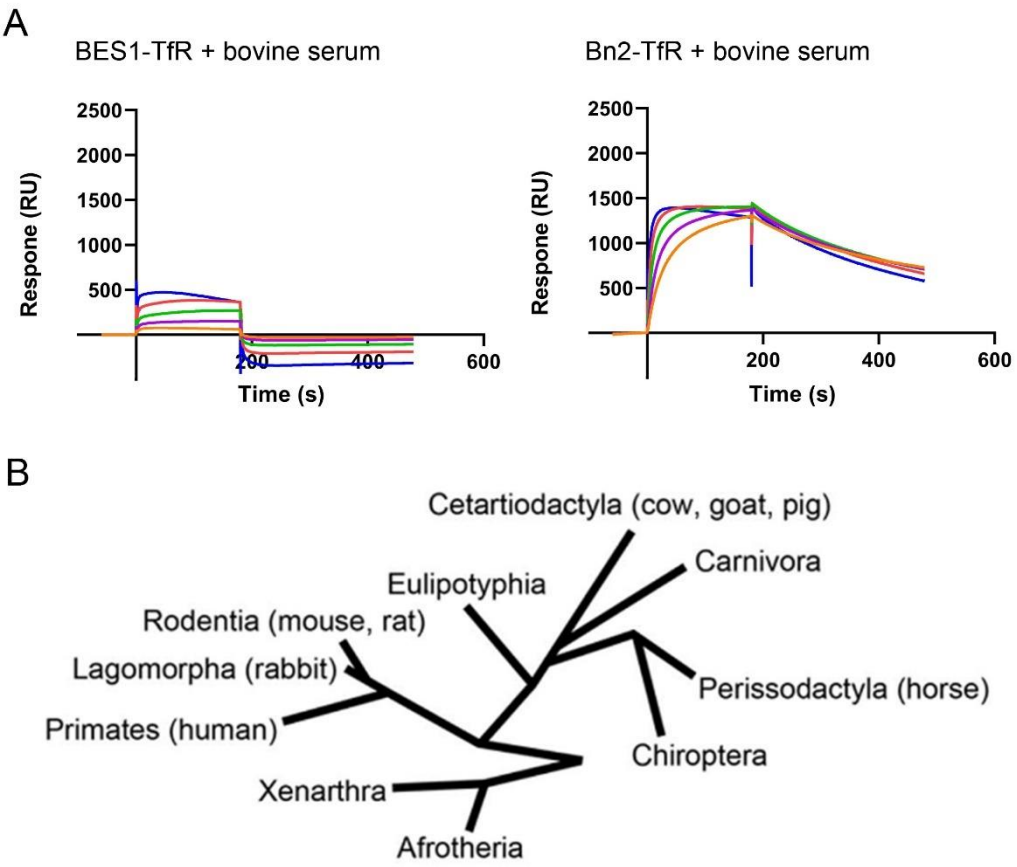


Figure 12: Recombinant TfR function was assessed by SPR. A, Measurement of bovine serum response to TfR variants by SPR. Biotinylated BES1-TfR and Bn2-TfR were immobilised and the binding response of bovine Tf present in serum was recorded. Serum was injected in the following dilution range: 1:10 (blue sensorgram); 1:20 (red); 1:40 (green); 1:80 (violet); 1:160 (orange). B, Evolutionary tree of mammals, adapted from Graphodatsky *et al*., 2011. Species indicated in brackets were selected for further investigation to represent the various evolutionary diversifications within the broad mammalian host range.

3.3.4.2 Testing transferrin binding to recombinant TfR

To verify recombinant TfR functionality, biotinylated receptors were prepared for SPR by immobilisation onto an SPR streptavidin (SA) chip and bovine serum dilutions were flowed over the SPR chip to screen for Tf binding (Figure 11D). A response was observed for both BES1-TfR and Bn2-TfR and, following subtraction of the non-specific response produced by serum interactions with the chip surface in the absence of receptor, the specific binding response was reported as a sensorgram (Figure 12A). BES1-TfR produced a lower binding response to bovine Tf compared to Bn2-TfR, indicating differences in affinity between receptor variants. An additional six mammalian species were chosen to represent the broad mammalian host range, informed by the mammalian phylogenetic tree (Figure 12B).

3.3.4.3 SPR with mammalian sera

Sera from goat, horse, human, rat, mouse and rabbit were tested to compare binding of Tf from different species to TfR variants. Bn2-TfR displayed evidence of high affinity binding to transferrins from all species tested, while BES1-TfR showed a weaker binding response to all sera tested (Figure 13). Crude serum produced high levels of background signals due to non-specific binding of serum components to the chip. Following subtraction of the non-specific response, a negative signal was observed for BES1-TfR with human and horse sera, masking any TfR-specific binding. Furthermore, goat and rabbit sera produced atypical binding responses to BES1-TfR that are not representative of a physiological receptor-ligand interaction, while the horse serum response to Bn2-TfR negatively correlates with serum concentration. Finally, determination of receptor-ligand dissociation constants was not possible due to unknown concentrations of Tf in each serum. I concluded that isolation of Tf from serum was necessary to further investigate the kinetic binding parameters of the receptor-ligand interaction.

3.3.4.4 Transferrin isolation from serum

To isolate transferrin from serum, affinity chromatography was performed using a TfR column. As a promiscuous TfR that exhibited high affinity binding to all sera tested, Bn2-TfR was selected to purify transferrin. After covalently linking Bn2-TfR to an N-hydroxysuccinimide (NHS)-activated sepharose column, serum was loaded and bound Tf ligand was eluted at low pH (Figure 14A). The purity of the transferrin samples obtained was testament to the high degree of specificity of the receptor (Figure 14B). The decrease in pH may have resulted in a loss of iron, with Tf transitioning from the holo-Tf iron-bound form to the apo-Tf form devoid of iron. Purified apo-Tf was subsequently loaded with Fe(III) ions to obtain holo-Tf.

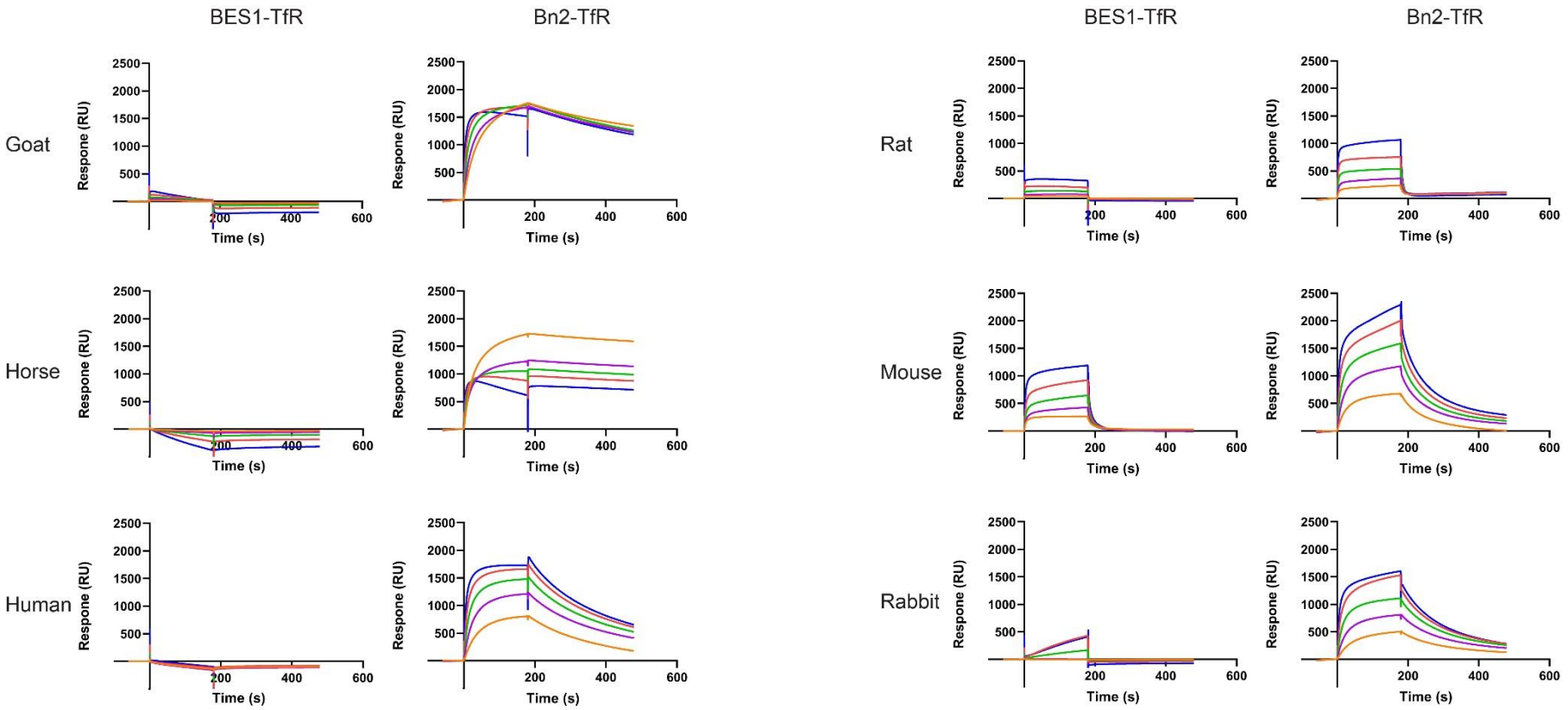


Figure 13: Measurement of serum response to TfR by SPR. Biotinylated BES1-TfR and Bn2-TfR were immobilised and whole serum was injected in the following dilution range: 1:10 (blue sensorgram); 1:20 (red); 1:40 (green); 1:80 (violet); 1:160 (orange).

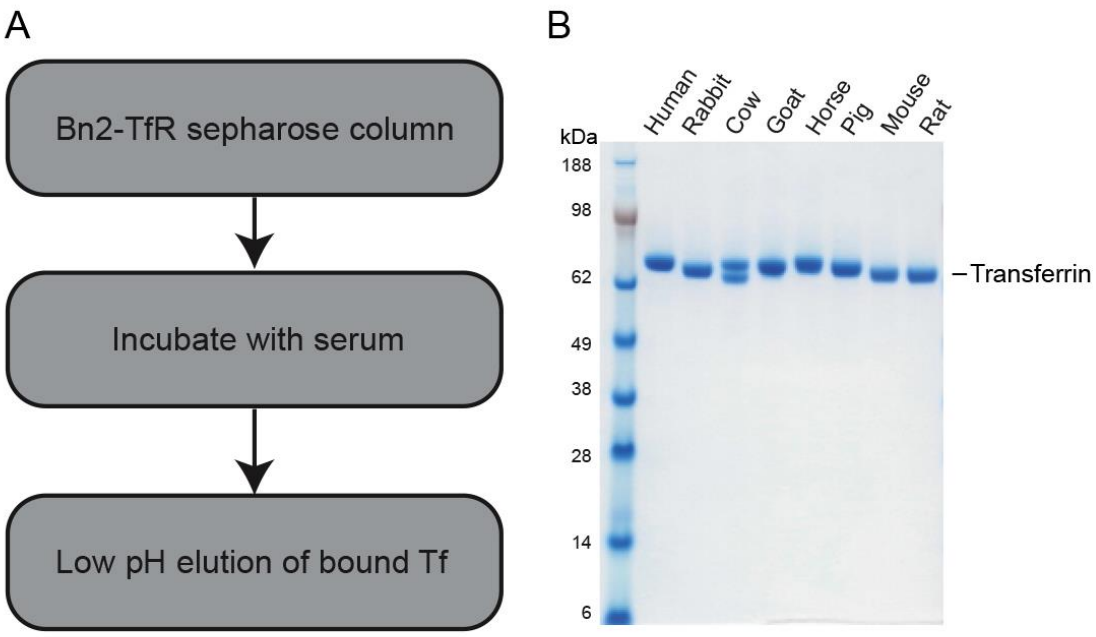


Figure 14: Isolation of native Tf from serum. A, Workflow for Tf purification from serum using a trypanosome TfR column. B, Tf from different species eluted from Bn2-TfR column. Bovine Tf separated into two distinct polypeptides due to cleavage of a peptide fragment (M_r = 6,000) (Tsuji *et al*., 1984).

3.3.4.5 SPR with purified transferrin

The use of affinity-purified native transferrin immensely improved the quality of the SPR data and provided a concentration range to allow accurate determination of the dissociation constants of transferrin from each mammal. Whilst Bn2-TfR is a broadly promiscuous receptor producing a strong binding response to transferrin from every species tested, BES1-TfR shows moderate to low affinity for its ligand (Figure 15). Analysis of the binding kinetics showed that Bn2-TfR has low to medium nanomolar affinity for each transferrin, with the range spanning from 1 to 500 nM (Table 1). Rat transferrin had the weakest affinity for Bn2-TfR (K_D = 500 nM) while transferrins from the other species tested produced low nanomolar affinities. In contrast, affinities for the BES1 receptor were lower, with K_D values ranging from 80 nM to > 3.6 µM. For most species, the affinity was predominantly in the micromolar range, with only cow, goat and rat exhibiting nanomolar affinities. Horse, human and pig transferrins produced a low signal that could not support K_D measurement within the Tf concentration range used. The K_D of horse, human and pig was therefore approximated to a value greater than the lowest detectable K_D of 3.6 µM. While differences in binding were observed between receptor variants, the data highlight that there was no emergence of species specificity, as one receptor had an indiscriminate high affinity for all Tfs while the other receptor had unselective moderate to poor binding to all. There was not any evidence of species-specific evolution in which each receptor had diversified to generate a high affinity for a subset of Tfs. BES1-TfR did not have a greater affinity than Bn2-TfR for any of the Tfs tested, suggesting that the selective pressure is not likely to be driven by species-specific Tf affinity.

To ensure the absence of Bn2 receptor bias for Tfs isolated during the purification method, pull-downs using native sera were performed. BES1-TfR and Bn2-TfR beads were incubated with sera from cow, goat, horse, pig and rabbit (Figure 16A). Bound ligand was eluted at low pH and visualised by SDS-PAGE (Figure 16B), corroborating the results observed by SPR using Tf purified by Bn2-TfR (Figure 15). While cow and goat Tfs produced an SPR binding response to BES1-TfR and were present in the BES1-TfR pull-down elution, horse and pig Tfs did not produce detectable binding by SPR, which was reflected in the absence of Tf in the horse and pig sera elution fractions for BES1-TfR. Furthermore, rabbit Tf produced a weak yet detectable SPR response to BES1-TfR, which was reproduced with native serum in the pull-down confirming the translatability between Tf purified by Bn2-TfR and Tf present in native serum conditions. Bn2-TfR served as a positive control for transferrin binding and elution in TfR pull-downs.

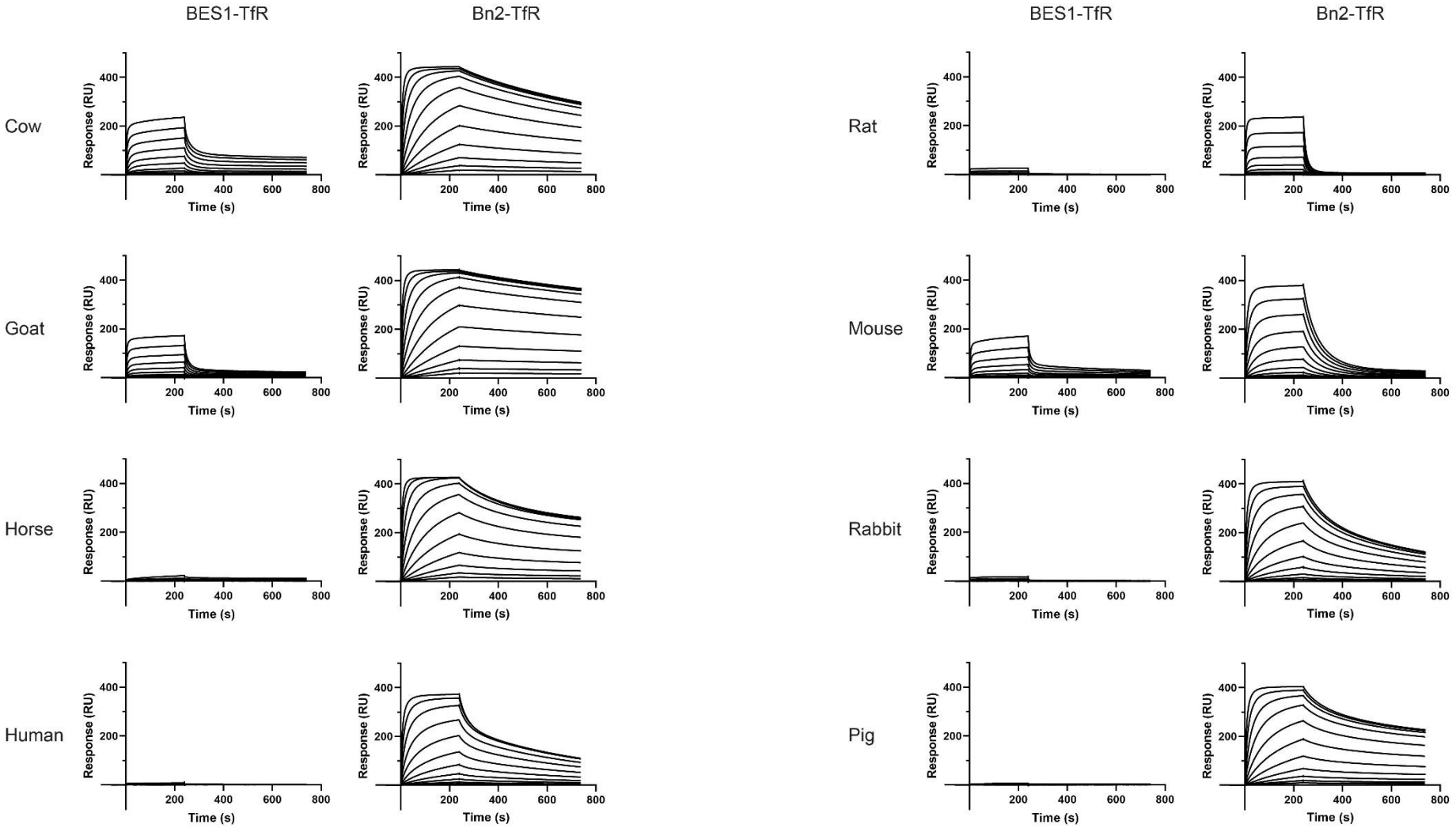


Figure 15: Measurement of purified Tf binding to TfR by SPR. Biotinylated BES1-TfR and Bn2-TfR were immobilised and purified transferrins were injected in a concentration series ranging from 1 µM to 1 nM.

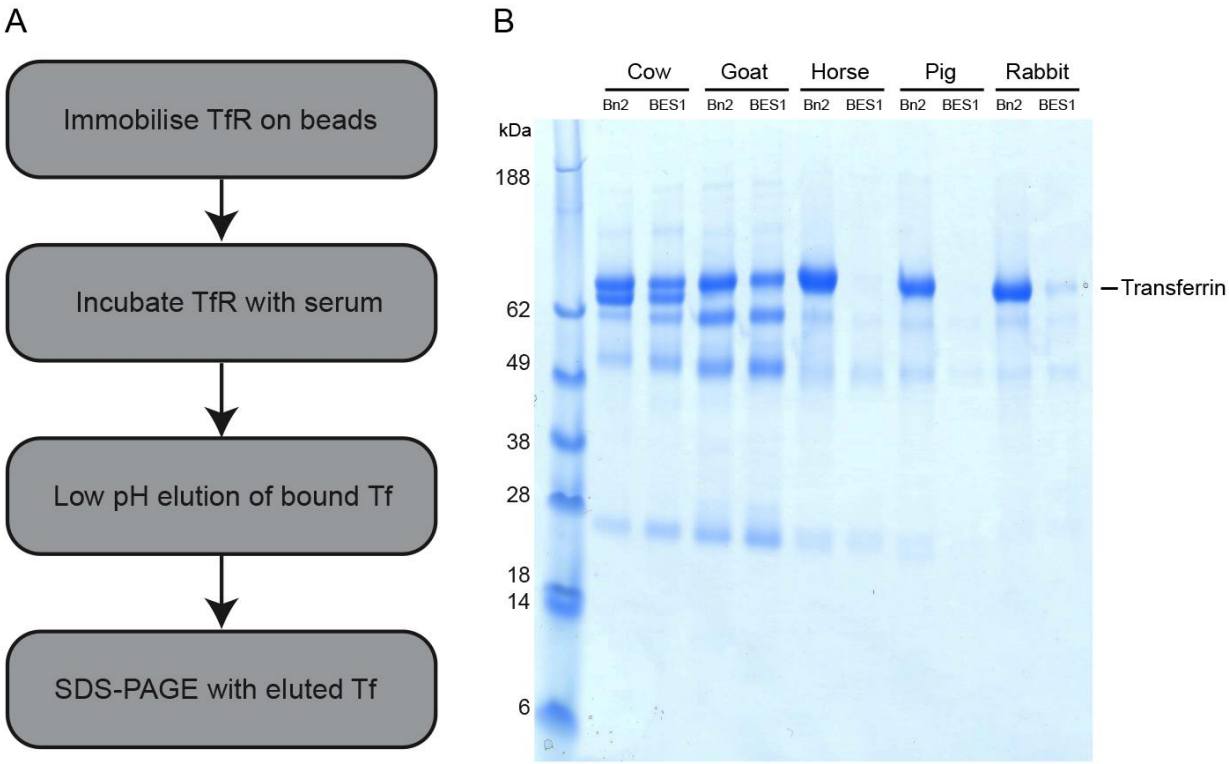


Figure 16: Pull-down assay between TfR variants and mammalian sera. A, Workflow depicting the pull-down strategy using CNBr-activated beads to mediate BES1-TfR or Bn2-TfR coupling. B, Following serum incubation and washing, ligands eluted at low pH were separated by SDS-PAGE.

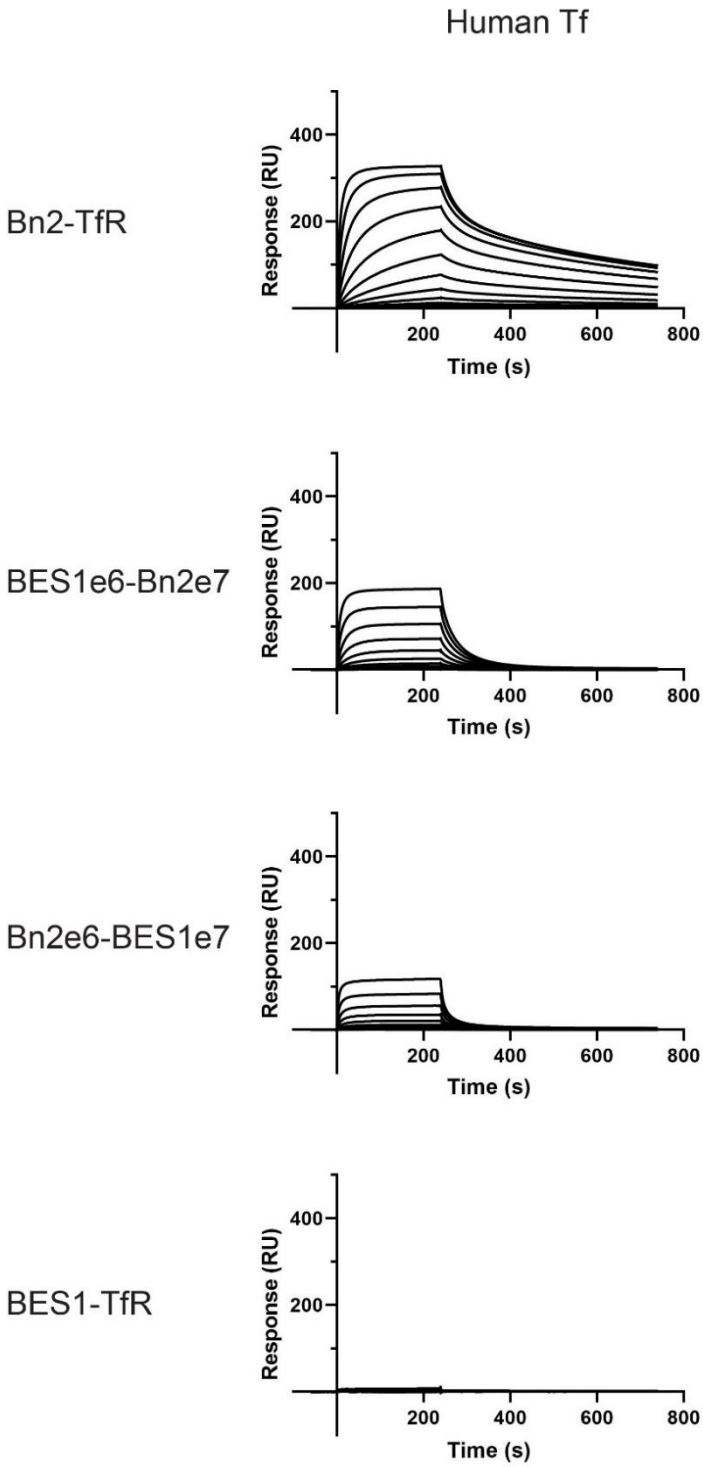


Figure 17: Measurement of human Tf binding to TfR chimeras by SPR. Biotinylated Bn2-TfR, BES1-TfR and TfR chimeras (BES1e6-Bn2e7 and Bn2e6-BES1e7) were immobilised and human Tf was injected in a concentration series ranging from 1 µM to 1 nM.

Table 1: SPR measurement of binding parameters. Kinetic values for interactions between immobilised BES1-TfR, Bn2-TfR or TfR chimeras and purified transferrin. Kinetic data for low binding responses were not determined (ND).

TfR	Transferrin	K_D(nM)	k_{on}($M^{-1}s^{-1}$)	k_{off}(s^{-1})
BES1	Human	ND	ND	ND
BES1	Rabbit	3600	$5.3x10^{4}$	$1.9x10^{-1}$
BES1	Mouse	350	$2.6x10^{5}$	$9.0x10^{-2}$
BES1	Rat	2900	$1.2x10^{5}$	$3.6x10^{-1}$
BES1	Cow	80	$4.5x10^{5}$	$3.6x10^{-2}$
BES1	Goat	240	$2.8x10^{5}$	$6.7x10^{-2}$
BES1	Horse	ND	ND	ND
BES1	Pig	ND	ND	ND
Bn2	Human	13	$1.6x10^{5}$	$2.0x10^{-3}$
Bn2	Rabbit	15	$1.8x10^{5}$	$2.7x10^{-3}$
Bn2	Mouse	75	$3.6x10^{5}$	$2.8x10^{-2}$
Bn2	Rat	500	$1.3x10^{5}$	$6.3x10^{-2}$
Bn2	Cow	2.8	$2.7x10^{5}$	$7.5x10^{-4}$
Bn2	Goat	1.4	$2.5x10^{5}$	$3.5x10^{-4}$
Bn2	Horse	3.7	$2.4x10^{5}$	$8.9x10^{-4}$
Bn2	Pig	6.2	$1.5x10^{5}$	$9.1x10^{-4}$
BES1e6-Bn2e7	Human	240	$1.2x10^{5}$	$2.8x10^{-2}$
Bn2e6-BES1e7	Human	350	$8.9x10^{4}$	$3.1x10^{-2}$

3.3.4.6 TfR chimeras

To further investigate the differences in Tf binding between variants, I wanted to pinpoint whether a subunit may be responsible for mediating a low affinity or high affinity response. BES1-Bn2 chimeras containing an ESAG from each receptor were produced by co-expressing BES1-TfR ESAG6 with Bn2-TfR ESAG7 (BES1e6-Bn2e7), and Bn2-TfR ESAG6 with BES1-TfR ESAG7 (Bn2e6-BES1e7). SPR was performed using human Tf which previously showed weak binding (K_D > 3.6 µM) to BES1-TfR (Figure 15). The addition of a Bn2-TfR subunit improved BES1-TfR binding to human Tf in both chimeras, demonstrating that ESAG6 and ESAG7 subunits both play a role in Tf binding (Figure 17). Bn2-TfR ESAG7 complementation produced the greatest increase in affinity, with a K_D of 240 nM compared to 350 nM for ESAG6 switching (Table 1), suggesting that ESAG7 has a larger Tf interaction interface. While it is possible to improve BES1-TfR affinity through ESAG switching, my next aim was to determine whether the low affinity of native BES1-TfR for certain species of Tf could suffice to support growth.

3.3.5 Investigation of TfR affinity and cell growth

3.3.5.1 Growth rates of cells expressing a low affinity TfR

Concentrations of transferrin in serum and tissues are high, with human transferrin concentrations of 30 - 40 µM in serum (Shih *et al*., 1990) and 15 µM in tissues (Steverding, 2003). These values are considerably higher than the determined Tf dissociation constants, allowing us to speculate that sera from species exhibiting weak TfR binding would support growth. To test this hypothesis, cells expressing BES1-TfR were grown in horse, pig or rabbit sera which all displayed a weak binding response by SPR (K_D > 3.6 µM) (Table 1). Growth rates were compared to that of BES1-TfR cells in foetal calf serum (FCS) as cow transferrin displayed the highest affinity for BES1-TfR (K_D = 80 nM) and would serve as a positive control. Adult bovine and canine sera were not used, despite being tested in previous studies (Bitter *et al*., 1998) (Gerrits *et al*., 2002) (Van Luenen *et al*., 2005), as I was unable to culture trypanosomes in these sera, indicating the presence of unidentified serum factors that compromised cell growth. Cells were initially cultured for 72 hours in FCS to ensure that the growth rate was steady prior to passaging cells into media containing each of the four different sera (Figure 18). In all four sera, growth rates remained stable and there was no

evidence of arrested proliferation that resembled previous observations (Bitter *et al*., 1998). While growth rates were marginally reduced for horse, rabbit and pig compared to cow, these differences did not correlate with Tf affinities as cells grew faster in horse serum than rabbit serum despite having a weaker affinity for horse Tf compared to rabbit Tf. Furthermore, population doubling times were closely aligned, with 5.9 hours in FCS compared to 6.6 hours in horse serum, despite differences in affinity of more than 40-fold between cow and horse Tf (Table 1). Thus, even in 10% serum from species that expressed Tf with a low affinity for 221-TfR, growth is supported at a steady rate. Furthermore, mammalian serum has a high Tf concentration of ~30 μM (Steverding, 2003), which is far greater than the dissociation constants reported. Together, these findings suggested that expression of a receptor with high Tf affinity is not an important factor controlling trypanosome growth.

3.3.5.2 Identification of the expressed TfR

While continuous cell growth after transfer to a different serum was a good indication of the absence of an orchestrated switching event, verification of the predominantly expressed TfR was performed. Following mRNA extraction and reverse transcription, primers specific to ESAG6 and ESAG7 were used to amplify all transcripts encoding the TfR. ESAG7 sequencing reads were used to identify switching as a discrete single nucleotide polymorphism (SNP) distinguishes BES1 from other expression sites (Figure 19A). Sequencing data revealed that the expressed TfR remained largely unchanged with predominant BES1 ESAG7 expression (Figure 19B). Although a unique SNP does not differentiate ESAG6 residing in BES1 from other expression sites, comparison of nucleotides in several locations also revealed predominant expression of ESAG6 from BES1. A low rate of switching was observed over time although this was comparable to the rate of switching observed in the bovine positive control and confirmed that the assay was performed for a sufficient duration to allow for switching. It is entirely plausible that a basal level of switching could occur, and our data suggest that this is independent of transferrin affinity for the predominant TfR.

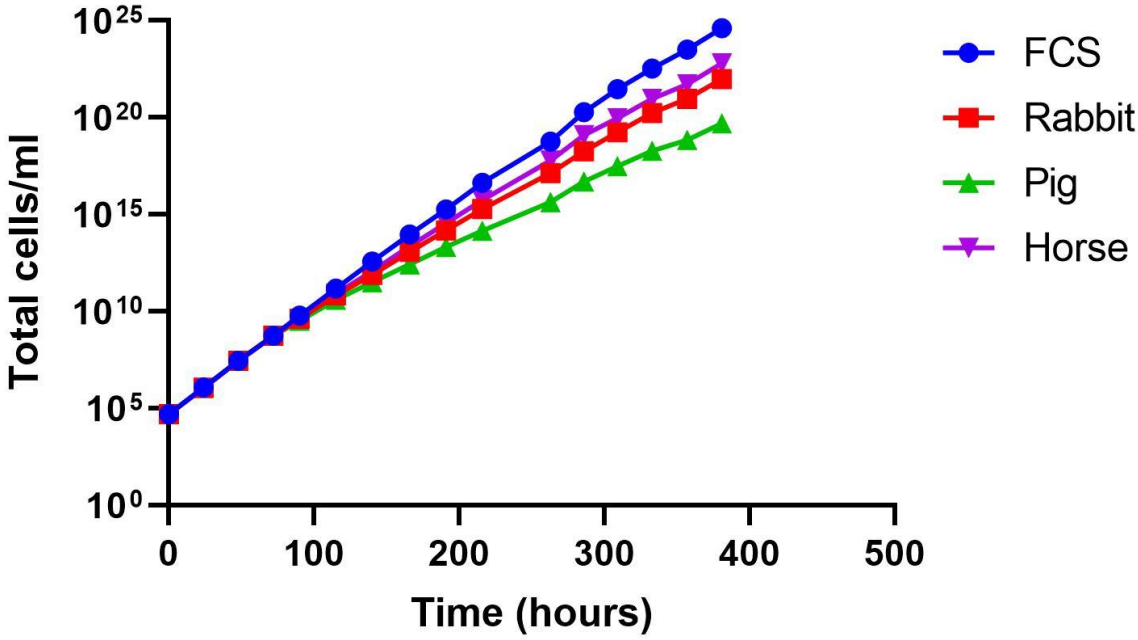


Figure 18: Growth of trypanosomes expressing the BES1-TfR in different sera. After culturing BSF cells in FCS for 72 hours to ensure a steady growth rate, cells were grown in FCS, horse, rabbit and pig sera. Error bars were too slight for visualisation on the graph.

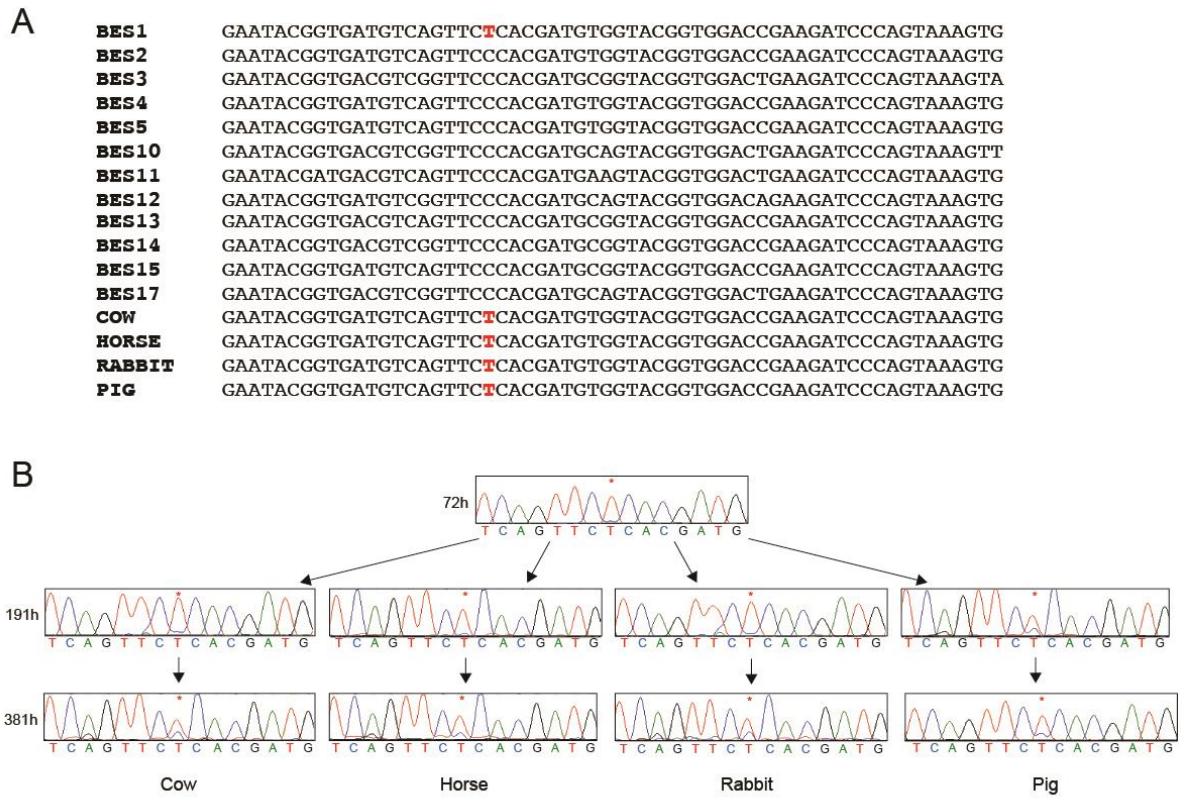


Figure 19: RT-PCR sequencing reveals a lack of TfR switching. A, The ESAG7 present in BES1 has a single nucleotide polymorphism (SNP) allowing BES1 TfR expression to be distinguished from TfRs in other expression sites. B, Sequencing chromatograms of ESAG7 RT-PCR transcripts following growth of BES1-expressing trypanosomes in various sera. Trypanosomes were initially grown in FCS for 72 hours then split into different sera. RNA transcripts from cells were analysed by RT-PCR at 72 hours, 191 hours and 381 hours to evaluate the extent of ESAG7 switching.

3.4 Discussion

3.4.1 A propensity to heterodimerise by an unknown mechanism

It has previously been demonstrated that ESAG6 and ESAG7 homodimers can form (Salmon *et al*., 1994), and our data confirmed these findings. However, I have found that heterodimers preferentially associated when both ESAGs were co-expressed and that their association must occur at an early stage following translation as homodimers did not readily dissociate to form heterodimers. The preferential heterodimerisation in a co-expression system was apparent following ESAG6 isolation by nickel affinity chromatography, which co-eluted in apparent equimolar ratios with ESAG7. The evidence that ESAG6 and ESAG7 may have an inherent propensity to form heterodimers would be logical from an evolutionary perspective due to the energetic burden that would occur if homodimers were formed. ESAG7 homodimers lacking GPI anchor addition sites would not be trafficked to the cell surface and instead would be destined for degradation. ESAG6 homodimers, while capable of cell surface anchorage, are refractory to transferrin binding and would not form a functional transferrin receptor. While it is apparent that preferential heterodimerisation occurs, the molecular features that may favour heterodimerisation remain unknown.

3.4.2 A lack of evidence to support emergence of species specificity

Alignments of ESAG6 and ESAG7 sequences have shown that polymorphisms are not the result of stochastic mutations but rather diversification is clustered in hypervariable hotspots suggesting active selection. The selective pressure has been proposed as the requirement to bind a variety of different Tfs to accommodate the broad mammalian host range. However, our data do not support this hypothesis based on several findings. First, comparison of Tf binding to two TfR variants by SPR did not demonstrate the emergence of discriminate species specificity. Second, growth of trypanosomes expressing BES1-TfR is supported in the presence of diluted sera containing low affinity Tf. Finally, BES1-TfR expression persists during growth in sera

containing low affinity Tf, suggesting that expression of a high affinity TfR is not necessary to support growth. In conclusion, active selection of sequence variation does not appear to be driven by Tf affinity, thus indicating that a different form of evolutionary pressure may exist.

As extracellular parasites, trypanosomes have evolved a specialised cell surface in response to host immune system exposure. While VSG switching is a well-characterised strategy to combat immune onslaught (Mugnier *et al*., 2015), variation of trypanosome receptors for immune evasion is less studied. I have shown that two antigenically different TfRs can survive in sera from different species without deleterious effects to cell growth and survival. Thus, it is conceivable that the primary selection factor for TfR variation may be immune avoidance as opposed to species specificity.

Chapter 4

Structure of the trypanosome transferrin receptor

4.1 Introduction

4.1.1 Previous insights into the trypanosome transferrin receptor structure

Predictive studies were previously performed to provide insights into the trypanosome transferrin receptor (TfR) structure. Initial investigations were focused on mapping the ligand binding site (Salmon *et al*., 1997). Sequence homology between ESAG6, ESAG7 and some VSGs including those with known structures (Freymann *et al*., 1990) (Blum *et al*., 1993) allowed identification of four blocks which varied between ESAG6 and ESAG7 sequences. These blocks aligned with the four most exposed surface loops of solved VSGs that formed the variant epitopes. The four variable loops in ESAG6 and ESAG7 were predicted to form the ligand binding site due to a potentially exposed localisation that would be accessible. Furthermore, sequence variation in the ligand binding site elucidated differences observed in Tf binding between receptor variants (Steverding *et al*., 1995). In more recent studies in 2012, Mehlert *et al*. generated a TfR homology model based on the structure of VSG MITat1.2 (Freymann *et al*., 1990). As a structure of the trypanosome TfR was yet to be obtained, efforts here were focussed towards determining the structure. To visualise the TfR and its

interactions with Tf at atomic resolution, we performed structural studies with the receptor-ligand complex using X-ray crystallography.

4.1.2 X-ray crystallography

X-ray crystallography is a technique that allows visualisation of protein structures at atomic resolution. The amalgamation of X-rays and crystals stemmed from size restrictions imposed by molecular structures. First, atomic bonds have an average length of 1.5 Å (0.15 nm) which is far smaller than visible light wavelengths of 400 – 800 nm. X-rays have a typical wavelength of 0.01 – 10 nm, thus suitable for providing diffraction patterns with atomic resolution. Second, the diffraction pattern produced by X-rays scattered from a single molecule would be very weak. Hence, crystals were required to provide a homogenous array of ordered molecules to amplify the diffraction signal.

The transition of soluble proteins to a crystalline solid state required precipitation whilst avoiding denaturation of molecules. Thus, precipitation must be slow and controlled, achieved by the addition of precipitant to the aqueous protein solution at sub-precipitation concentrations. The ensuing water evaporation resulted in increased precipitant and protein concentrations to favour gradual protein precipitation and crystal nucleation. The requirement of a sweet spot in precipitant concentration without additional evaporation which would not be conducive to crystal growth led to the development of the hanging or sitting drop vapour diffusion method. Within a sealed environment, water could diffuse between the drop containing protein and precipitant, and the reservoir solution containing precipitant. Crystal nucleation occurs once equilibrium is achieved, with a closed environment maintaining ideal nucleation and growth conditions.

4.1.3 A quest for protein homogeneity

Obtaining suitable crystals is not trivial and often entails a multifactorial optimisation. A homogenous sample of target protein favours crystal nucleation. Heterogeneity is commonly introduced by the presence of contaminants, hence the necessity for sample purity. Furthermore, glycosylations can be heterogeneous, resulting in different

glycosylation profiles between molecules. Finally, flexibility within glycans and disordered regions can introduce heterogeneity as the orientation of molecular features may vary. The ESAG6 and ESAG7 constructs contained C-terminal linkers and tags that may adopt varied orientations. Carboxypeptidase treatment can be performed to remove the flexible C-terminus.

To avoid glycan heterogeneity, removal is usually best and can be performed by several types of glycosidase which either cleave the entire glycan or leave a sugar moiety. Peptide:N-glycosidase F (PNGase F) is a promiscuous endoglycosidase that cleaves the glycoside bond between asparagine and N-acetylglucosamine (GlcNAc) through a hydrolysis reaction that transforms aspargine into aspartic acid. Endoglycosidase Hf (Endo Hf) is another type of endoglycosidase that selectively cleaves the β(1-4) glycoside bond between GlcNAc sugars of mannose-rich carbohydrates, requiring production of mannose-rich glycans during protein expression. Endo Hf can achieve more efficient deglycosylation due to accessibility as the base of the glycan does not have to be reached.

4.2 Aims

The multiplication and diversification of TfRs has long been a focus of discussion. It is either the result of a requirement for a repertoire of different binding sites to accommodate the range of different mammalian transferrins (Bitter *et al*., 1998) (Gerrits *et al*., 2002) (Isobe *et al*., 2003); or diversity has arisen from an immune evasion strategy (Borst, 1991). While work in the Chapter 3 suggested that the repertoire had not evolved to achieve a broader host range, the trypanosome TfR variation remained shrouded in complexities, requiring structural determination to answer key questions:

Does the structure adhere to the same three-helix fold adopted by other trypanosome surface proteins?

How does the receptor bind its ligand?

Why do ESAG6 and ESAG7 preferentially heterodimerise?

Can the location of polymorphisms and glycans provide insights into the origin of a diverse TfR repertoire?

4.3 Results

4.3.1 Crystallisation trials

4.3.1.1 Sample homogeneity

BES1- and Bn2-TfRs were produced as previously described in Chapter 3, and initials trials to remove TfR glycans were performed with PNGase F. Complete removal of glycans was not successful and heterogenous glycan cleavage was apparent by SDS-PAGE (Figure 20A). A smear pattern on the gel indicated incomplete hydrolysis, producing a protein sample with molecules of varying sizes depending on the cleavage efficiency. Complete deglycosylation was only achieved after TfR denaturation, indicating that unfolding favoured glycan removal and thus accessibility may be the factor compromising PNGase F activity in folded TfR. Endo Hf provided greater glycan trimming efficiency compared to PNGase F, yet complete deglycosylation was not achieved when compared to the denatured and fully deglycosylated TfR control (Figure 20B). Attempts were made to remove C-terminal flexible structures by carboxypeptidase treatment. Enzyme efficiency was monitored by western blot probed with anti-polyhistidine antibody to detect a C-terminal His-tag in ESAG6 (Figure 20C). Longer incubations with carboxypeptidase led to increased C-terminal cleavage efficiency, with an overnight incubation of 20 hours drastically increasing His-tag trimming compared to incubations of 4 hours or less. A greater concentration of carboxypeptidase also led to improved cleavage for shorter incubations of 4 hours, although an overnight incubation produced the most favourable results in the presence of lower concentrations of enzyme. SDS-PAGE of the samples used for western blot illustrated that equal amounts of TfR were used throughout the carboxypeptidase trial (Figure 20D).

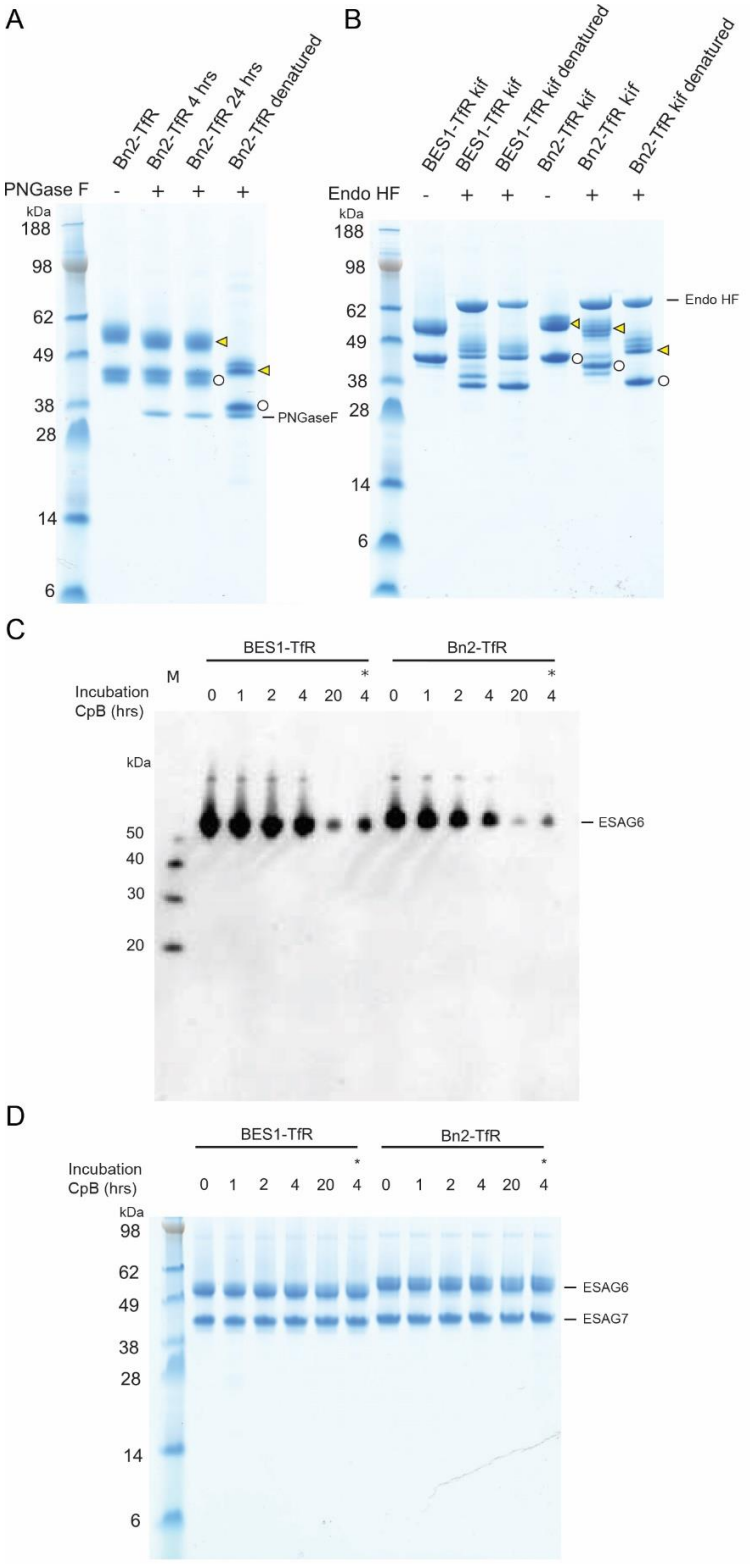


Figure 20: Preparation of homogenous TfR for crystallisation. A, Pilot test of Bn2-TfR glycan removal using PNGase F. B, Trial removal of BES1-TfR and Bn2-TfR glycans using Endo Hf. ESAG6 subunits are highlighted with a yellow triangle; ESAG7 subunits are indicated with a white circle. C, Carboxypeptidase B treatment of BES1-TfR and Bn2-TfR, followed by western blot probed with anti-polyhistidine. D, Coomassie-stained SDS-PAGE of western blot samples shown in C.

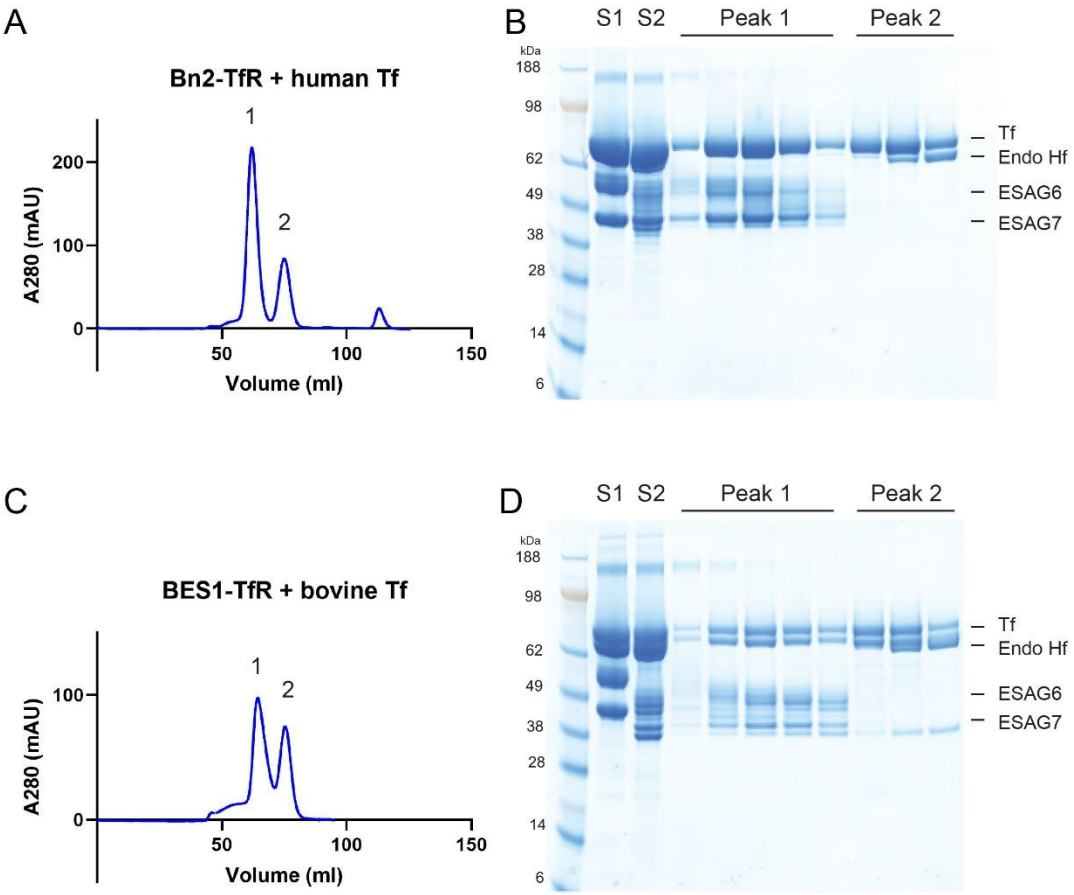


Figure 21: Isolation of TfR-Tf complexes for crystallisation. A, Size exclusion chromatography (SEC) was performed to isolate the Bn2-TfR and human Tf complex. The absorbance at 280 nm was measured and represented as a chromatogram. B, Coomassie-stained SDS-PAGE depicting Bn2-TfR with human Tf samples before and after enzymatic treatment and SEC. Bn2-TfR with human Tf prior to enzymatic treatment is shown in S1. Bn2-TfR with human Tf following Endo Hf, PNGase F and Carboxypeptidase B treatment is shown in S2. Sample S2 was subsequently separated by SEC to isolate the TfR-Tf complex (Peak 1 samples) and remove free Tf ligand (Peak 2 samples). C, Chromatogram depicting isolation of the BES1-TfR and bovine Tf complex by SEC. D, Coomassie-stained SDS-PAGE of BES1-TfR and bovine Tf samples as described in B. AU = absorbance units.

4.3.1.2 Isolation of the TfR-Tf complex

Purified Bn2-TfR was incubated with human transferrin in the presence of Endo Hf for removal of mannose-rich glycans present on the TfR, PNGase F for cleavage of glycans present on Tf and carboxypeptidase for removal of TfR C-terminal tags. Purified BES1-TfR was incubated with bovine Tf (K_D = 80 nM) as the affinity for human Tf was weak (K_D > 3.6 µM), in the presence of the same cocktail of enzymes. Enzyme-treated receptor-ligand complexes were isolated by size exclusion chromatography. Absorbance at 280 nm was monitored and fractions from peak 1 (Figure 21A, C) contained the TfR-Tf complex (Figure 21B, D) while peak 2 contained unbound ligand. In Figure 21D, bovine Tf separated into two distinguishable polypeptides due to peptide cleavage that can occur (Tsuji *et al*., 1984). Samples were set up for crystallisation in a variety of conditions using the sitting drop vapour diffusion method.

4.3.1.3 Initial crystals

Following sample preparation in crystallisation plates, small crystalline formations appeared for Bn2-TfR in complex with human Tf. BES1-TfR in complex with bovine Tf did not produce any crystals. The presence of a cleavage product in bovine Tf may have introduced heterogeneity that hindered crystallisation.

To verify crystallisation of the complex, crystals were solubilised and analysed by SDS-PAGE. Following silver staining, the presence of human Tf, ESAG6 and ESAG7 within the crystals was confirmed (Figure 22). However, the size and quality of the crystals were not sufficient for X-ray diffraction and further optimisation was required.

4.3.1.4 Optimisation of the TfR-Tf crystals

The optimisation approach was two-fold and involved adjusting the crystallisation conditions and improving protein homogeneity. In previous samples, deglycosylation of human Tf produced a single product (Figure 21B), in contrast to Bn2-TfR in which ESAG6 and ESAG7 resolved as a smeared pattern suggesting heterogenous deglycosylation. To circumvent TfR glycan heterogeneity and achieve complete absence of N-linked oligosaccharides, site-directed mutagenesis of the expression constructs was performed. The five predicted N-linked sites in Bn2 ESAG6 and three

sites in Bn2 ESAG7 were mutated to aspartic acid to abolish glycosylation. To address heterogeneity introduced by flexible linkers, tags and variable carboxypeptidase cleavage efficiency, the ESAG6 and ESAG7 constructs were further modified. C-terminal linkers and tags were removed, and N-terminal tags were cloned upstream of ESAG6 with an interposed TEV protease site for subsequent tag cleavage following purification. The cleavage site was inserted at the N-terminus to leave a residual glycine after cleavage. All tags were removed from ESAG7 as they were surplus to requirement, with heterodimerisation sufficient for co-elution of ESAG6 and ESAG7. The Bn2-TfR glycan mutant was incubated with human Tf and the receptor-ligand complex was isolated by size exclusion chromatography, with ESAG6 and ESAG7 showing less heterogeneity by SDS-PAGE (Figure 23). Samples were prepared for crystallisation using optimised conditions based on the conditions in which small crystals were obtained (Figure 22).

Crystals of the non-glycosylated Bn2-TfR mutant in complex with human Tf were obtained in 12% (w/v) PEG 5000 MME, 12% 2-methyl-2,4-pentanediol, 0.1 M MES, pH 6.5 at 18°C. Crystals diffracted to 2.75 Å resolution and molecular replacement was first performed using available structural data for human Tf (PDB: 4X1B) (Wang *et al*., 2015). Next, the VSG (PDB: 2VSG) (Blum *et al*., 1993) was used as a homologous model for Bn2-TfR molecular replacement, followed by model building and refinement. In addition, crystals were obtained in the same conditions with Endo Hf-treated Bn2-TfR in complex with human Tf. Crystals diffracted to 3.6Å resolution and were solved by molecular replacement using the non-glycosylated TfR structure (Appendix 6).

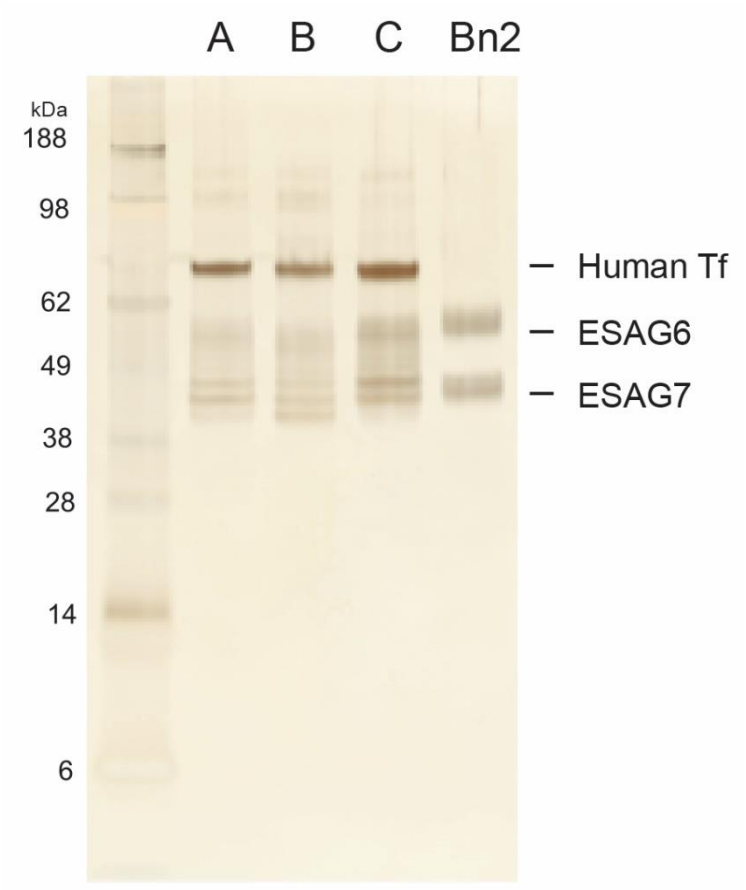


Figure 22: Visualisation of crystal protein composition by SDS-PAGE. Crystals were solubilised and proteins were separated by SDS-PAGE and silver stained. Conditions for sample A: 10% w/v polyethylene glycom (PEG) 5000 monomethyl ether (MME), 12% v/v 1-Propanol, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES), pH 6.5; sample B: 0.1 M sodium chloride, 25% v/v pentaerythritol propoxylate, 10% v/v dimethyl sulfoxide; and sample C: 20% PEG 6000, 0.1 M Tris, pH 8.5.

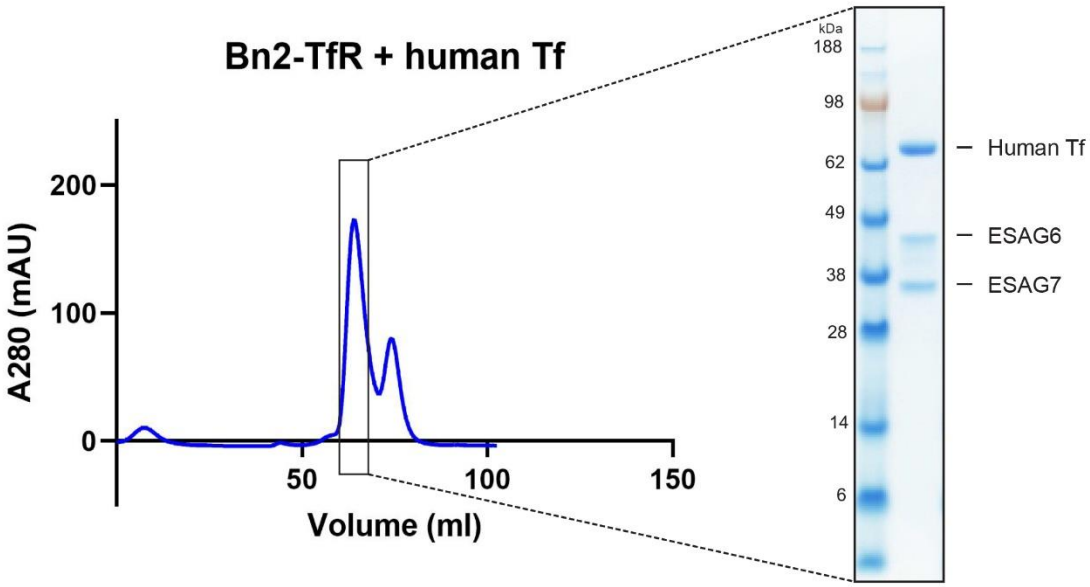


Figure 23: Isolation of Bn2-TfR mutant in complex with human Tf. Size exclusion chromatogram produced by the absorbance at 280 nm, with visualisation by SDS-PAGE of the isolated sample. AU = absorbance units.

4.3.2 The structure revealed a VSG-like fold

The VSG family contains a remarkable set of proteins which have undergone considerable sequence diversification whilst maintaining a conserved structure (Figure 4) (Freymann *et al*., 1990) (Blum *et al*., 1993) (Bartossek *et al*., 2017) (Pinger *et al*., 2018). The solved TfR structure revealed that the receptor also adopted a similar fold to other characterised trypanosome cell surface proteins, with an elongated three-helix bundle that formed the core of each subunit (Figure 24 and Figure 25A). Sequence alignments of TfR with other structurally characterised VSGs highlighted the lack of conservation (Figure 25B). However, superimposition of the TfR structure on solved structures of the VSG N-terminal domain showed close alignment and structural conservation (Figure 25A). The two N-terminal alpha-helices aligned with the VSG helices, while the third alpha-helix, partially absent in the VSG, was full-length in the TfR to reinforce the core fold. At the membrane proximal side of each ESAG, a short helix provided a wedge between the two subunits of the heterodimer (Figure 24). This likely dictated adequate membrane proximal spacing between molecules within the dense and dynamic VSG coat (Bartossek *et al*., 2017). A network of intramolecular disulphide bridges maintained the core fold of the individual receptor subunits, whilst heterodimer pairing was reliant on extensive hydrogen bonding. Thirty-three residues at the C-terminus of ESAG6 were not resolved due to poor electron density, suggesting flexibility. A C-terminal GPI-anchor would mediate membrane integration. The apical surface of the receptor was arranged in a complex array of intertwined loops that formed the ligand binding surface, with both ESAG6 and ESAG7 subunits interacting with a single molecule of human transferrin. The receptor-ligand interface was restricted to the membrane distal tip of the receptor, which differed from the large ligand binding surfaces adopted by other structurally characterised *T. b. brucei* cell surface receptors. In the case of the haptoglobin-haemoglobin receptor (HpHbR), the three-helix motif adopted a rigid 50° kink to displace the compact VSG layer and augment the ligand binding site (Lane-Serff *et al*., 2014). The factor H receptor consists of a similar motif to HpHbR and has evolved a membrane distal binding pocket, with two of the three helices splayed open to accommodate factor H (Macleod *et al*., 2020). In a bid to maximise epitope occlusion in the VSG coat, the transferrin receptor has likely evolved an exclusively apical binding pocket allowing lateral surfaces of the receptor to remain embedded in the VSG coat.

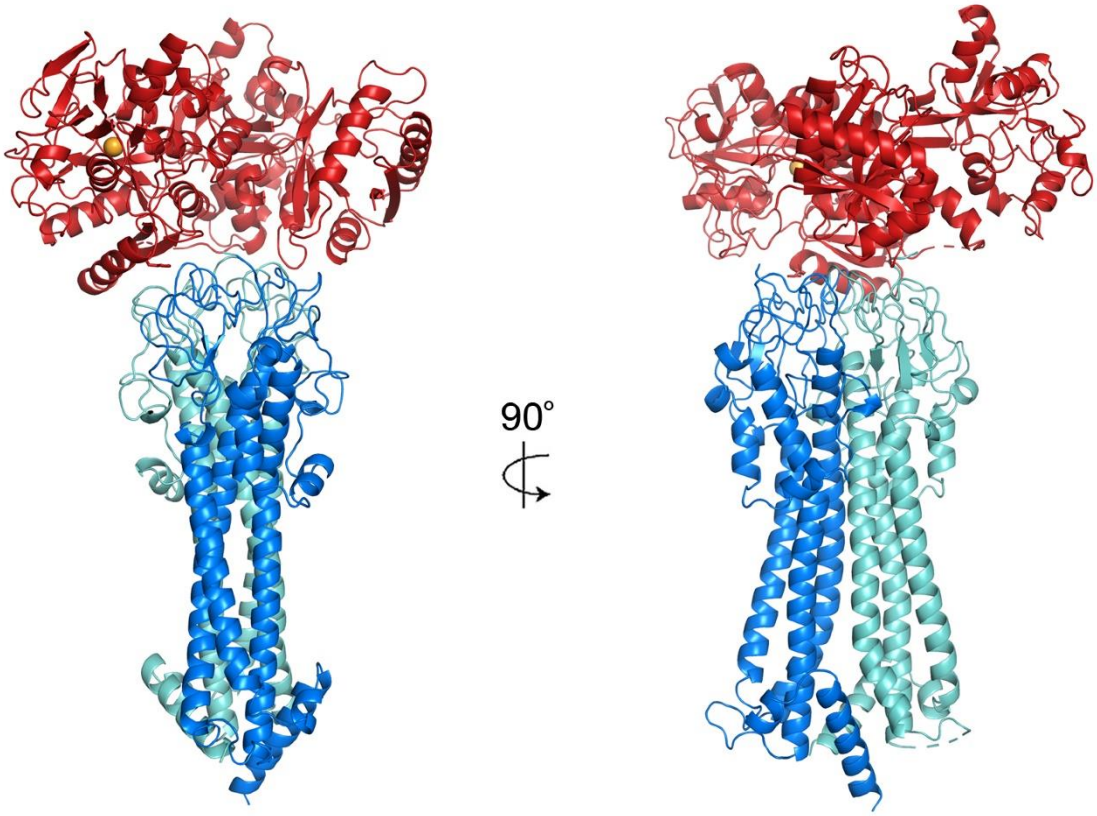


Figure 24: Structure of the trypanosome TfR in complex with human Tf. Bn2-TfR is represented in dark blue (ESAG6) and cyan (ESAG7). Human Tf is shown in red and Fe(III) is represented as an orange sphere.

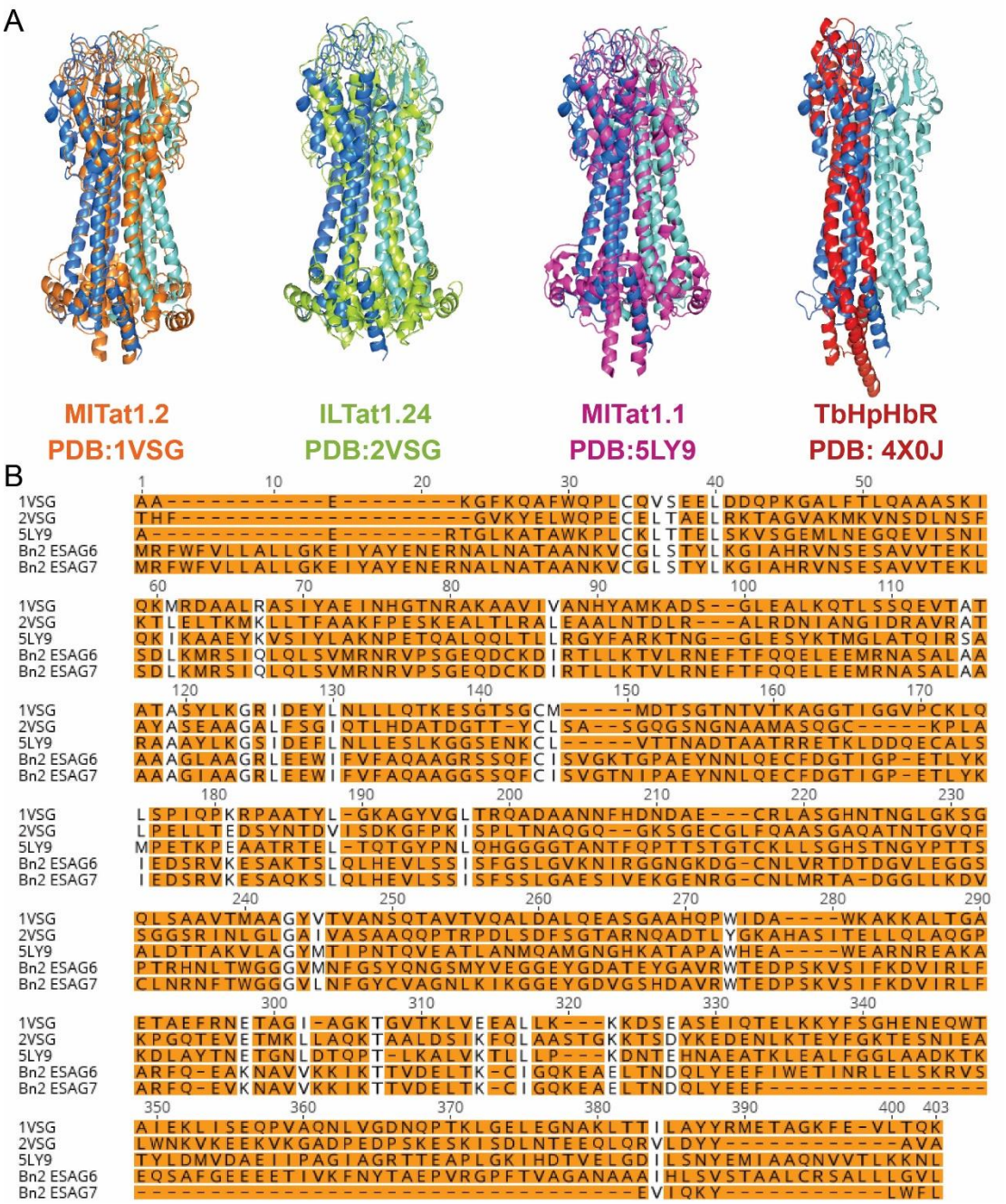


Figure 25: Comparison of TfR with structures and sequences of other trypanosome cell surface proteins. A, Superimposition of TfR structure with *T. brucei* VSG structures (MITat1.2, ILTat1.24 and MITat1.1) and the *T. brucei* haptoglobin-haemaglobin receptor (TbHpHbR). The PDB codes of the structures used for the overlay are listed. B, Alignment of the protein sequences used for structural comparison, with sequence variation highlighted in orange. The TbHpHbR sequence was not aligned due to greater sequence diversity.

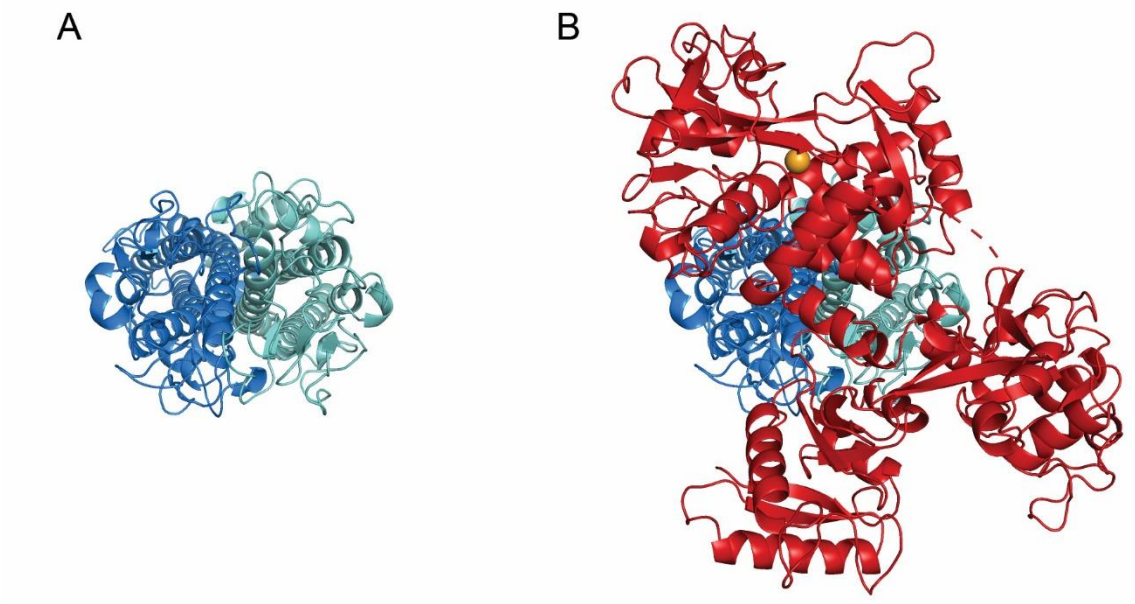


Figure 26: Tf binds both subunits of TfR. A, Top view of the membrane distal tip of the TfR. ESAG6 is represented in dark blue and ESAG7 is shown in cyan. B, Human Tf, shown in red, bound to the membrane distal tip of the TfR.

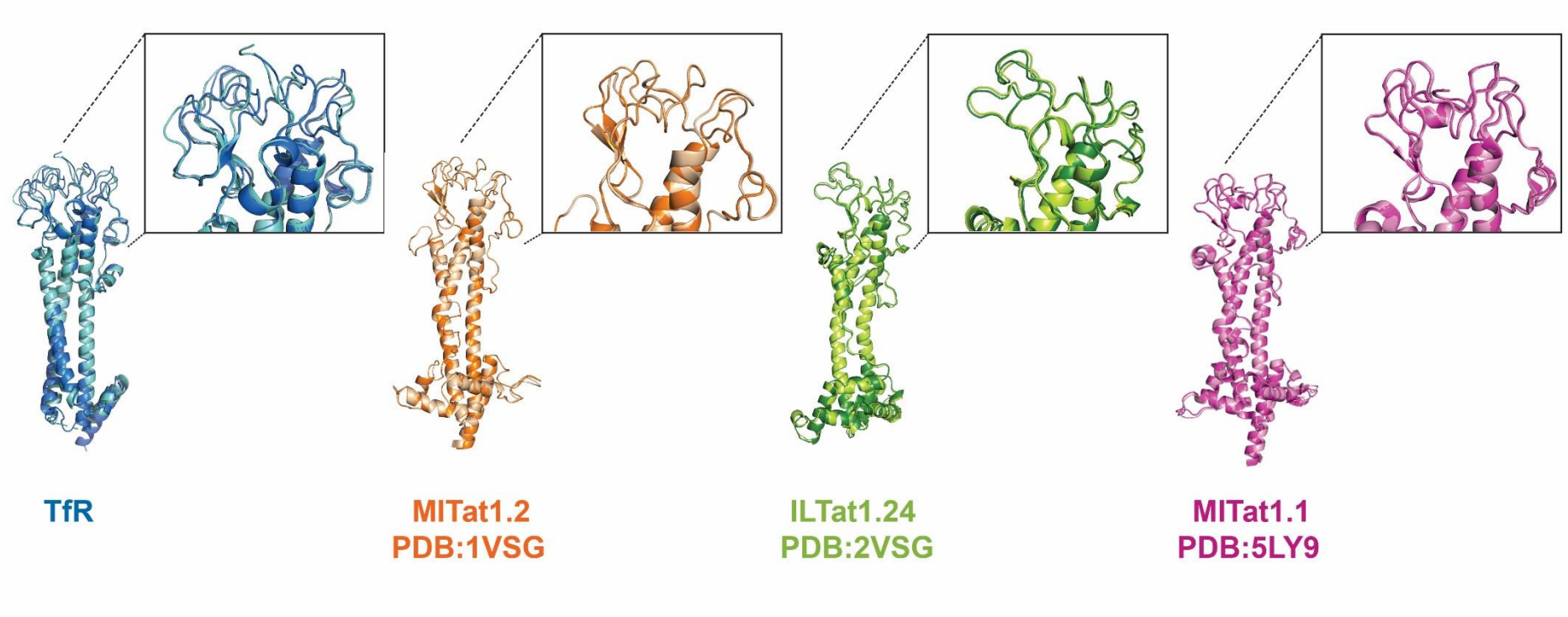


Figure 27: Overlay of TfR subunits and VSG monomers. Structural divergence and asymmetry within the loops of ESAG6 and ESAG7 is compared to an overlay of VSG monomers. Structures of VSG homodimers MITat1.2, ILTat1.24 and MITat1.1 were used, with PDB codes listed.

4.3.3 The receptor-ligand interface elucidates heterodimerisation requirement

ESAG6 and ESAG7 displayed remarkable structural similarity (Figure 24) yet they are not interchangeable, with the presence of both subunits required for functional ligand binding (Salmon *et al*., 1994) (Chaudhri *et al*., 1994). Alignment of the sequences showed 80% identity within the structurally represented sequences (Figure 6). Initial observations of the receptor-ligand interface provided clues for the requirement of a heterodimer, illustrating that Tf binding spanned across both ESAG6 and ESAG7 in an asymmetric manner, with an apparent larger Tf footprint spanning ESAG7 (Figure 26A, B). A greater contribution of ESAG7 in Tf binding corroborated previous SPR data produced using TfR chimeras (Figure 17). Superimposition of the ESAG6 and ESAG7 structures revealed a perfectly aligned core with structural divergence observed solely at the apical surface (Figure 27). On the other hand, an overlay of VSG monomers illustrated that the flexible loops of each molecule were aligned, suggesting that structural divergence at the apical surface of TfR was caused by sequence variation rather than flexibility and movement within the loops during crystallisation, as similar differences in loop structure were not observed for VSG monomers. The subtle structural variation is the reason both ESAGs are required for transferrin binding. Transferrin itself is formed of two lobes, termed N-lobe and C-lobe. Structural differences within transferrin lobes further elucidated the requirement for heterodimerisation. The receptor must accommodate the bilobal heterogeneity of the ligand by providing an asymmetric binding surface.

4.3.4 Receptor-ligand interface and diversification

A single molecule of human transferrin binds at the surface of Bn2-TfR, predominantly through the formation of a network of hydrogen bonds. Investigation of the binding interactions at the receptor-ligand interface demonstrated that the majority of the interactions were mediated by ESAG7 (Table 2), further corroborating the SPR data with TfR chimeras (Figure 17). While ESAG6 predominantly contacts the C-lobe, ESAG7 binds both N- and C-lobes (Table 2).

Alignments of the protein sequences of Bn2-TfR and BES1-TfR demonstrated the high degree of conservation between receptor variants (Figure 28). The ESAG6 sequences shared 91% sequence identity while the ESAG7 sequences were 92% identical, despite being two of the most different TfRs in the L427 genome. Comparison of the Tf-binding residues with the

residues that vary between Bn2-TfR and BES1-TfR revealed that interacting amino acids were conserved. Only one variable residue within the interaction site was identified, an arginine/glycine polymorphism located at residue 139 in ESAG6, numbered according to the Bn2-TfR sequence to match numbering of the crystal structure. Site-directed mutagenesis was performed to investigate the variable residue involved in Tf binding. Glycine 139 was mutated in BES1-TfR ESAG6 to the corresponding residue in Bn2-TfR by converting to arginine. To investigate ESAG7, three variable residues in proximity to the Tf binding site were selected for site-directed mutagenesis (Figure 29A, B). In BES1-TfR, residues 229, 233 and 246 of ESAG7 were mutated to the corresponding residues in Bn2-TfR. SPR was performed to study the effects of mutations on Tf affinity. Human and rat Tf were initially tested for binding as they previously produced the lowest affinities for BES1-TfR and Bn2-TfR respectively. Human Tf showed negligible binding to BES1-TfR (K_D > 3.6 µM), while rat Tf produced the weakest binding response to Bn2-TfR (K_D = 500 nM) (Table 1). Each of the mutants tested improved binding for human and rat Tf (Figure 30). The greatest improvement in affinity was observed for human Tf and was produced by combining all four mutations, with a previous K_D > 3.6 µM improving more than 10-fold to 300 nM (Table 3). Thus, the quadruple mutant, referred to as BES1mut, was tested for binding to the mammalian Tf panel. The mutations improved the affinity of BES1-TfR for all transferrins tested (Figure 31), with at least a two-fold increase in binding observed for rat, mouse and rabbit. Horse affinity improved more than 10-fold; while cow, goat and rabbit affinities improved more than 25-fold (Table 3 and Table 1). In conclusion, mutation of a small number of residues in a low affinity TfR can greatly increase Tf binding, indicating that Tf affinity may not drive active selection of variation.

Table 2: List of interactions between trypanosome TfR and human Tf. Chain A represents EAG6, chain B is ESAG7 and chain C corresponds to human Tf.

Transferrin receptor			Transferrin				
Chain	**Residue**	**Group**	**Chain**	**Lobe**	**Residue**	**Group**	**Type of interaction**
A	R139	Side chain	C	N	A54	Backbone CO	Hydrogen bond
A	D221	Side chain	C	C	H349	Side chain	Hydrogen bond
A	T222	Side chain	C	C	H349	Side chain	Hydrogen bond
A	T222	Side chain	C	C	E372	Side chain	Hydrogen bond
A	G228	Backbone CO	C	C	R352	Side chain	Hydrogen bond
A	G228	Backbone CO	C	C	S370	Backbone NH	Hydrogen bond
A	Y248	Side chain	C	C	H349	Side chain	Aromatic stacking
A	Y266	Side chain	C	C	H349	Side chain	Hydrogen bond
B	Y18	Side chain	C	N	T330	Side chain	Hydrogen bond
B	Y18	Side chain	C	N	R324	Side chain	Cation-π
B	Y18	Side chain	C	N	N325	Side Chain	Hydrogen bond
B	E19	Side chain	C	N	Y71	Side chain	Hydrogen bond
B	E19	Side chain	C	N	R324	Side chain	Hydrogen bond
B	E19	Side chain	C	N	K312	Side chain	Hydrogen bond
B	N20	Side chain	C	C	N383	Side chain	Hydrogen bond
B	G138	Backbone CO	C	C	R352	Side chain	Hydrogen bond
B	S140	Backbone CO	C	C	R352	Side chain	Hydrogen bond
B	S140	Side chain	C	C	S359	Side chain	Hydrogen bond
B	S141	Side chain	C	C	C368	Backbone NH	Hydrogen bond
B	Q142	Side chain	C	C	R352	Side chain	Hydrogen bond
B	Q142	Side chain	C	C	C368	Backbone CO	Hydrogen bond
B	N150	Side chain	C	C	S359	Side chain	Hydrogen bond
B	N150	Side chain	C	C	E367	Side chain	Hydrogen bond
B	R213	Side chain	C	C	E367	Side chain	Hydrogen bond
B	D222	Side chain	C	N	N76	Side chain	Hydrogen bond
B	H265	Side chain	C	N	P74	Side chain	Stacking

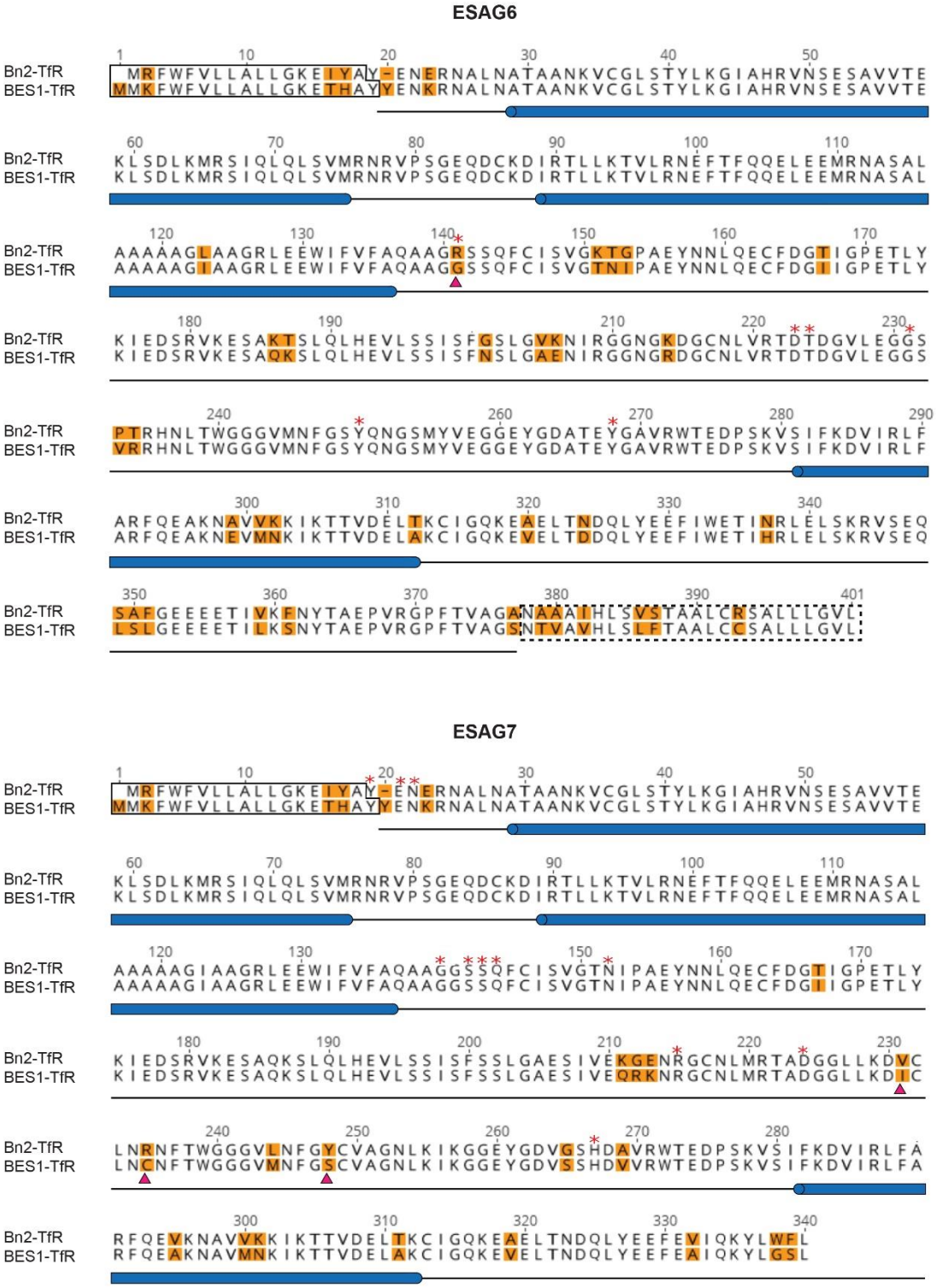


Figure 28: Protein sequence alignments of Bn2-TfR and BES1-TfR. Residues that vary are highlighted in orange. Bn2-TfR residues that bind human Tf are indicated with a red asterisk and residues in BES1-TfR selected for mutagenesis are indicated with a pink triangle. Helices are represented by a blue cylinder, signal peptides are indicated by a black box and the GPI anchor addition site is shown with a dashed box.

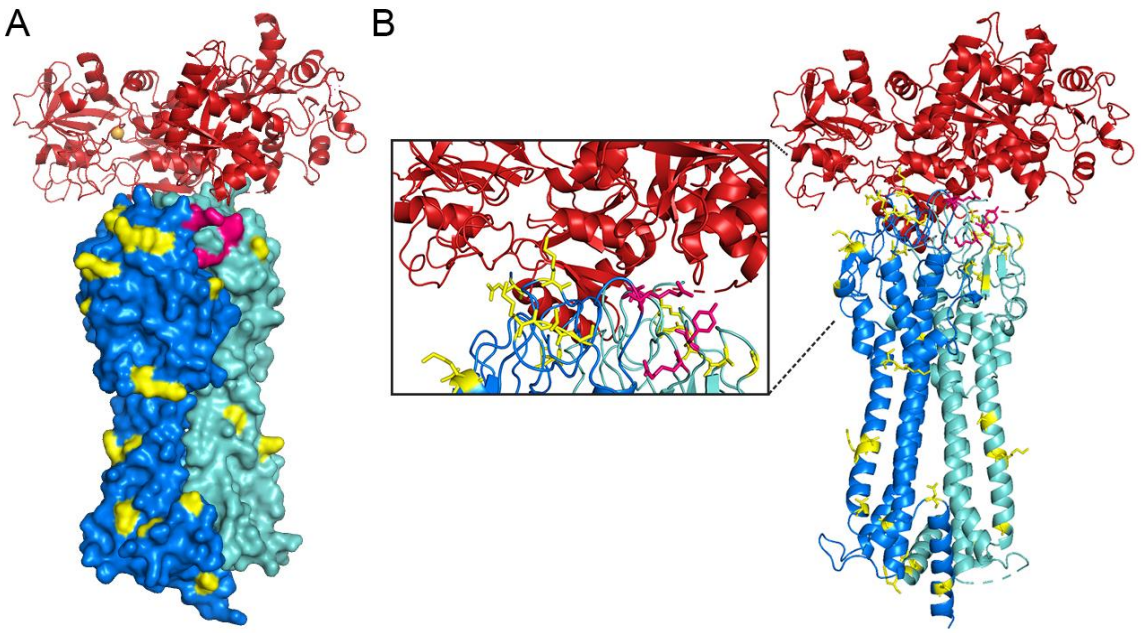


Figure 29: Sequence variation between Bn2-TfR and BES1-TfR. A, Surface representation of Bn2-TfR with ESAG6 shown is dark blue and ESAG7 represented in cyan. Residues that vary between Bn2-TfR and BES1-TfR are highlighted in yellow. Pink indicates residues selected for mutagenesis, and human Tf is shown in red. B, Cartoon representation of Bn2-TfR bound to human Tf with an identical colour scheme as described in A. Variable residues are shown in stick representation.

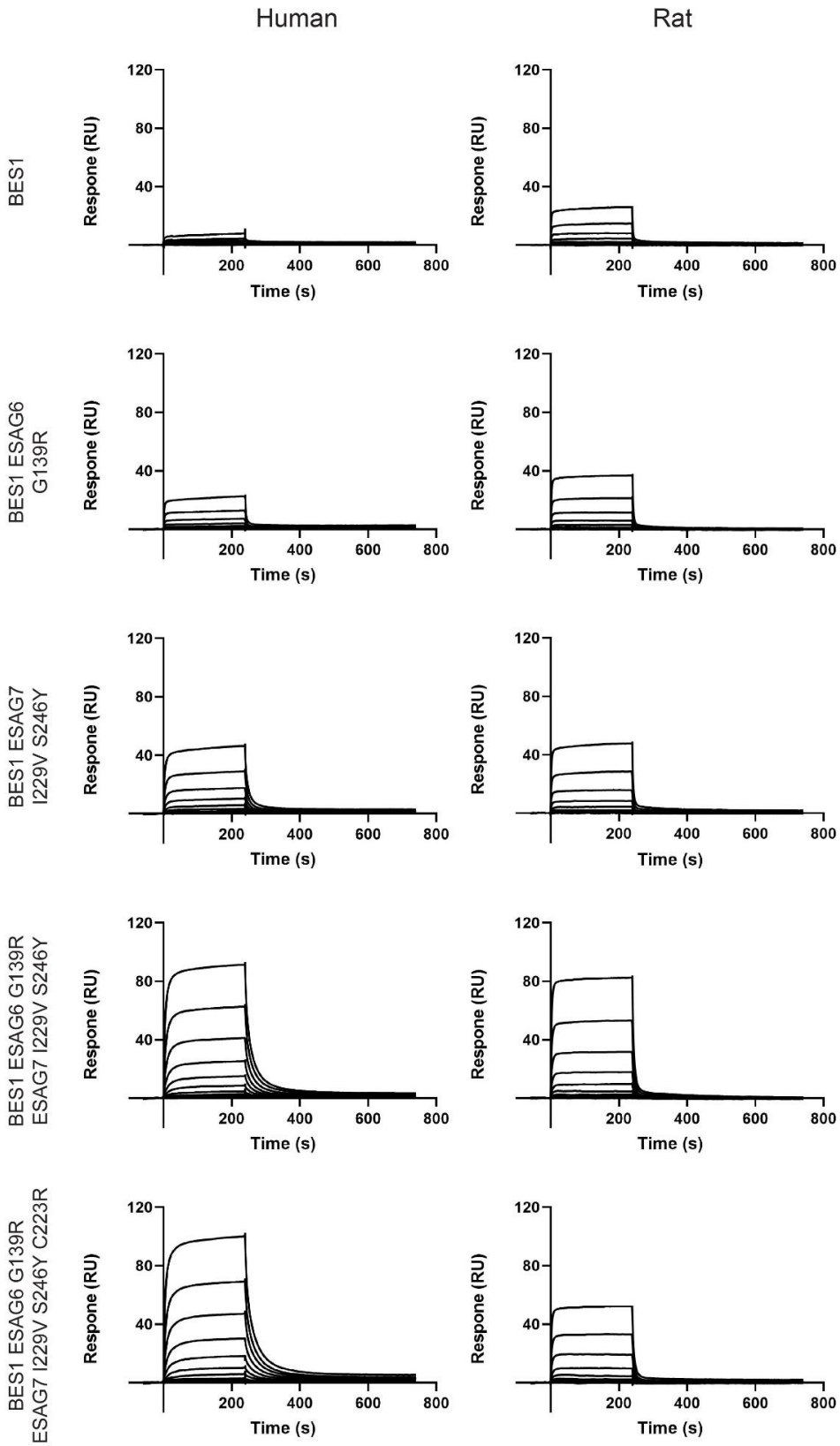
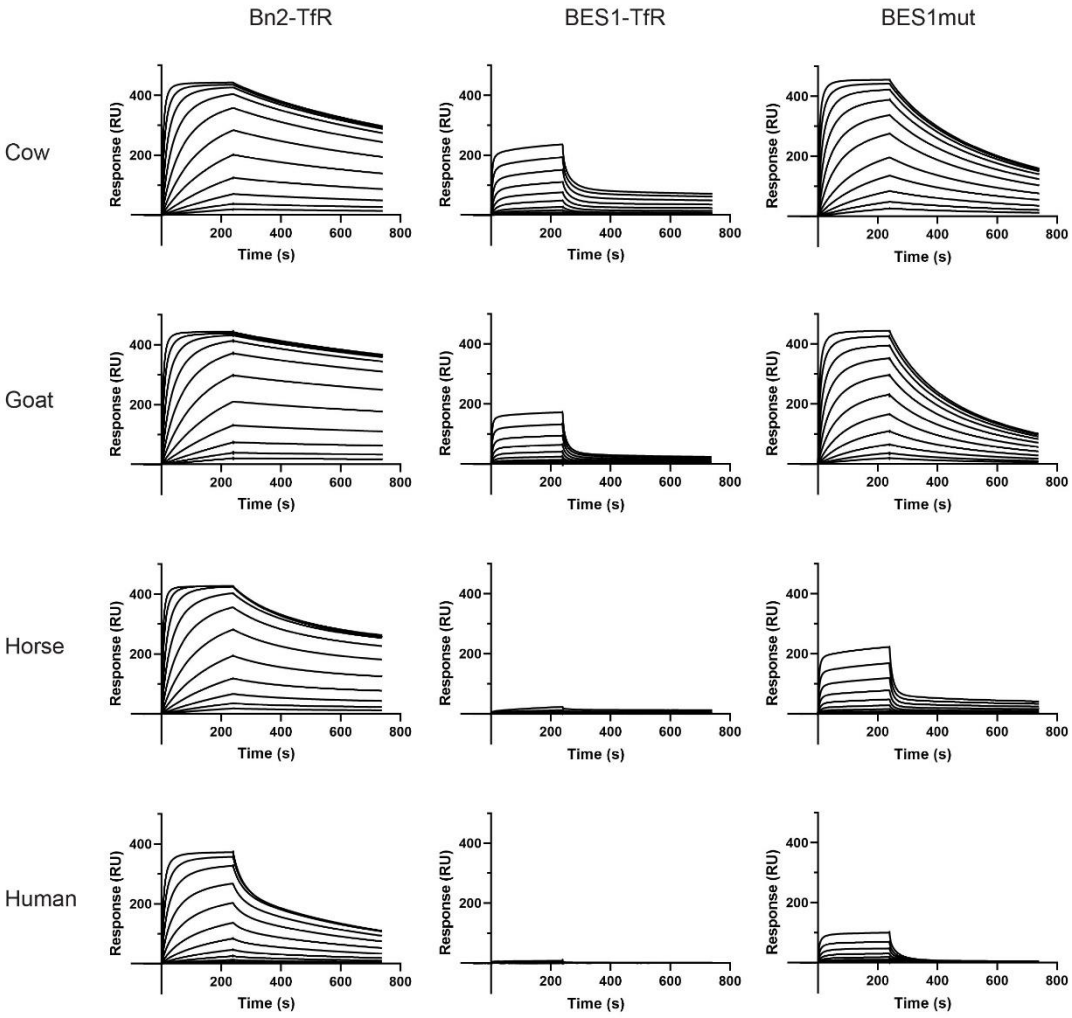


Figure 30: Measurement of human and rat Tf binding to BES1-TfR mutants by SPR. Biotinylated BES1-TfR mutants were immobilised and purified Tf was injected in a concentration series ranging from 1 µM to 1 nM.


Bn2-TfR
BES1-TfR
BES1mut
Cow
Goat
Horse
Human
Response (RU)
Time (s)
200
400
600
800

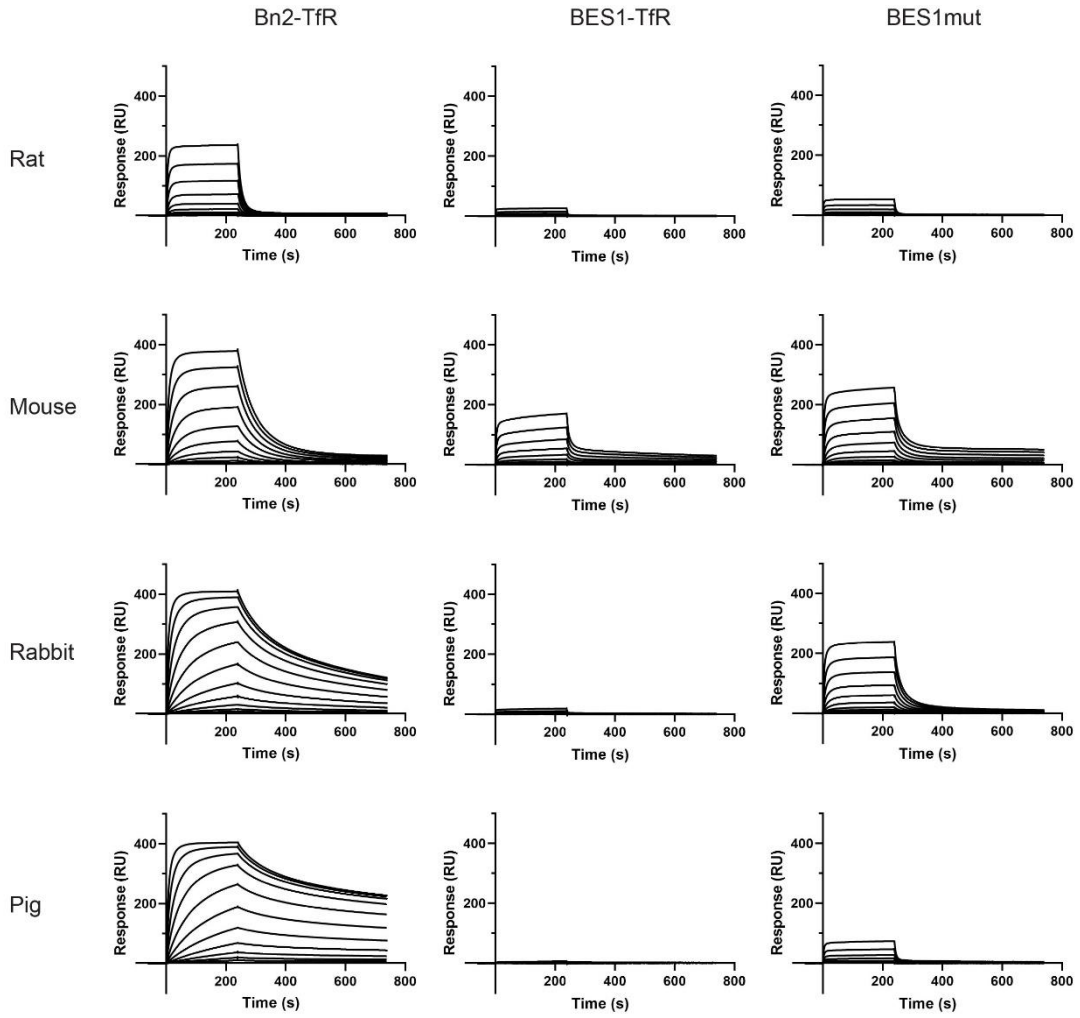


Figure 31: Measurement of Tf binding to Bn2-TfR, BES1-TfR and BES1mut by SPR. Biotinylated TfRs were immobilised and purified Tf was injected in a concentration series ranging from 1 µM to 1 nM.

Table 3: SPR measurement of binding parameters. Kinetic values for interactions between immobilised BES1-TfR mutants and purified transferrin. BES1mut contains four mutations: G139R in ESAG6 (e6); and I229V, S246Y and C233R in ESAG7 (e7). Kinetic data for low binding responses were not determined (ND).

TfR	Transferrin	K_D(nM)	k_{on}($M^{-1}s^{-1}$)	k_{off}(s^{-1})
BES1	Human	ND	ND	ND
BES1	Rat	2900	$1.2x10^5$	$3.6x10^{-1}$
BES1 e6 G139R	Human	2600	$7.2x10^4$	$1.9x10^{-1}$
BES1 e6 G139R	Rat	2800	$1.1x10^5$	$3.1x10^{-1}$
BES1 e7 I229V S246Y	Human	1200	$7.3x10^4$	$8.3x10^{-2}$
BES1 e7 I229V S246Y	Rat	2100	$1.3x10^5$	$2.7x10^{-1}$
BES1 e6 G139R e7 I229V S246Y	Human	410	$7.2x10^4$	$2.9x10^{-2}$
BES1 e6 G139R e7 I229V S246Y	Rat	1100	$1.4x10^5$	$1.6x10^{-1}$
BES1mut	Human	300	$6.6x10^4$	$2.0x10^{-2}$
BES1mut	Rat	1100	$1.3x10^5$	$1.4x10^{-1}$
BES1mut	Mouse	150	$2.9x10^5$	$4.2x10^{-2}$
BES1mut	Rabbit	99	$2.1x10^5$	$3.9x10^{-3}$
BES1mut	Cow	3.2	$6.5x10^5$	$2.1x10^{-3}$
BES1mut	Goat	7.6	$4.1x10^5$	$3.1x10^{-3}$
BES1mut	Horse	250	$3.1x10^5$	$8.0x10^{-2}$
BES1mut	Pig	1300	$9.6x10^4$	$1.2x10^{-1}$

4.3.5 Mapping of TfR polymorphisms

To study the locations in which polymorphisms are clustered, *T. b. brucei* ESAG6 and ESAG7 sequences from the Lister 427 and EATRO 2340 isolates were analysed. The frequency of variation of each residue was calculated and the entropy was mapped onto a surface representation of the structure using the Shannon entropy scale. Polymorphic residues were predominantly located at the apical tip of the receptor (Figure 32A), indicating that sequence variation was a result of active selection. While active selection within the TfR apex could have arisen through a necessity to either increase Tf affinity or avoid immune recognition, SPR data with BES1mut (Figure 31) suggested that the former was not a viable hypothesis as a small number of point mutations considerably improved the affinity of BES1-TfR for all Tfs tested. Furthermore, surface representations of the structure were used to compare polymorphic residues (Figure 32B, C) with residues that formed the Tf-binding site (Figure 32D). The residues interacting with transferrin (Figure 32D) aligned with conserved regions at the surface of the receptor (Figure 32C). The binding footprint of Tf appeared to form a negative image of the variable sites, illustrating that Tf binding and diversification were not aligned. Mapping of sequence variation provided evidence that receptor evolution was not the result of a requirement for species specific TfRs, as diversification did not reside in the Tf binding site.

4.3.6 Glycan positioning

Endo Hf cleaved N-linked glycans at the β(1-4) glycoside bond between two GlcNAc sugars, leaving a residual Glc-NAc attached to the protein. The residual sugar served as a guide for positioning and orienting glycosylations. Previously published predicted chemical structures of trypanosome N-linked glycans were used for modelling (Mehlert *et al*., 2012). One glycan in ESAG6 was not modelled as it was located in the unresolved C-terminus close to the GPI anchor. Resolved glycans were found in two different locations on the receptor. A single glycan was located in the core three-helix bundle of each ESAG within the narrowest region of the receptor (Figure 33A). The second glycosylated region was found at the apical surface of the receptor. Three glycans from ESAG6 and two glycans from ESAG7 were arranged in a ring formation located beneath the ligand binding site (Figure 33A, C). The structure indicated that the glycan ring was not directly involved in ligand binding, corroborating previous findings (Maier & Steverding, 2008) (Mehlert *et al*., 2012). In addition, comparison of glycan positioning with polymorphisms revealed precise positioning of glycans in less diverse regions of the structure (Figure 33B). Glycans were located between variable patches and appeared to serve as a shield for conserved epitopes.

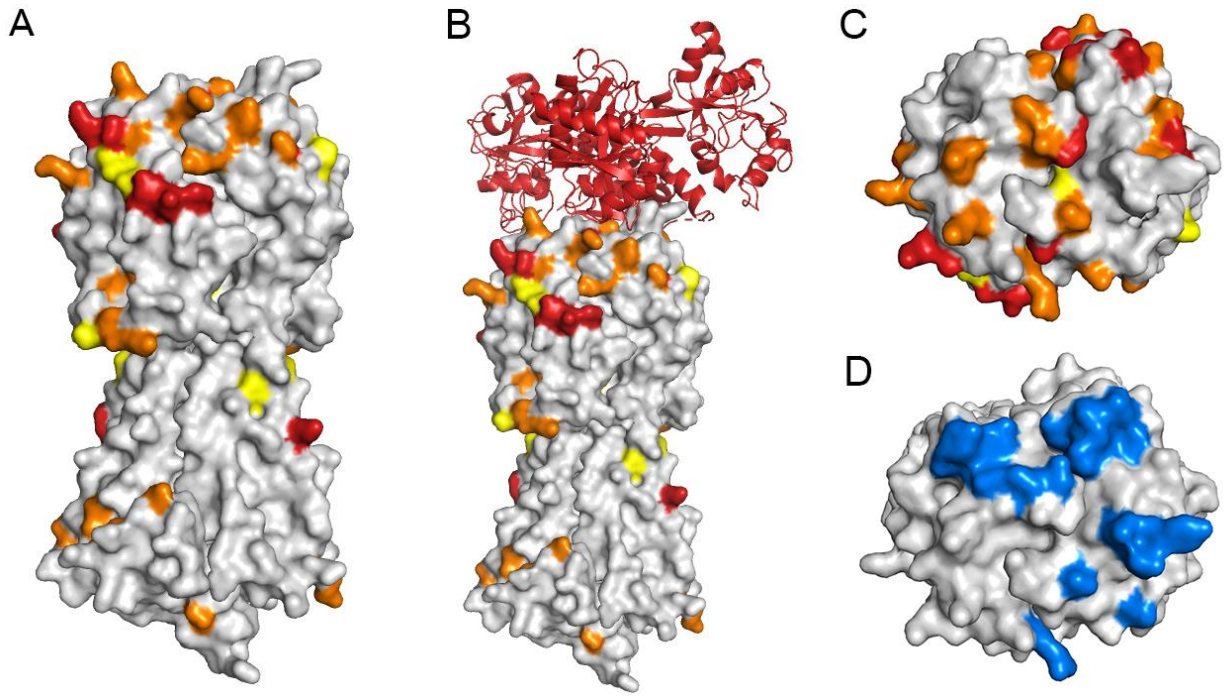


Figure 32: Mapping of TfR polymorphisms. A, Surface representation of Bn2-TfR depicted in grey, with variable residues highlighted in yellow (Shannon sequence entropy 0.3-0.5), orange (0.5-0.8) and red (>0.8). B, Bn2-TfR is shown in grey, with sequence variation concentrated towards the apical surface that surrounds the Tf-binding site. Human Tf is shown as a cartoon in red. C, Top view of the TfR apical surface with polymorphisms coloured as described in A. D, Top view of the TfR with residues that bind Tf highlighted in blue.

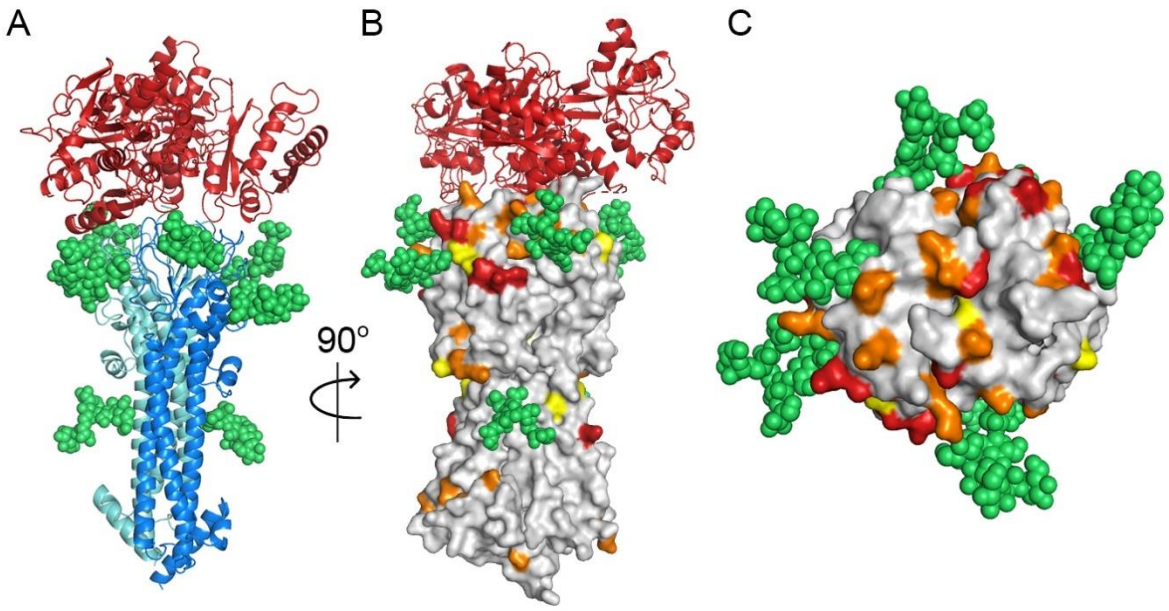


Figure 33: Localisation of TfR glycans. A, Trypanosome TfR structure shown in blue (ESAG6) and cyan (ESAG7). Modelled TfR glycans are represented in green and human Tf is shown in red. B, Surface representation of the TfR in grey, with polymorphisms highlighted in yellow, orange and red. Glycans and human Tf are coloured as described in A. C, Top view of the TfR apical surface with glycans shown in green and variable regions highlighted in yellow, orange and red.

4.4 Discussion

Structural characterisation of the TfR revealed that the receptor adopted a structure similar to other trypanosome cell surface proteins, with a three-helix bundle motif maintaining the core fold of the receptor. In particular, the TfR showed striking similarity to the N-terminal domain of type A VSGs. Despite sequence divergence between surface proteins, structural homology is conserved.

The heterodimerisation interface of the receptor extends along the alpha-helix. Initial hopes that structural data would provide clues to solve the preferential heterodimerisation conundrum were not fulfilled. There were no obvious regions in which the assembly of a homodimer would produce an electrostatic clash or steric hinderance. Due to the concentration of structural differences at the apical surface of the receptor, it is likely that discrete variations within the folding of loops in this region would favour the formation of heterodimers over homodimers. Ultimately, the receptor acquired an ESAG6/7 asymmetric binding site to accommodate the bilobal heterogeneity of the transferrin ligand.

Several discreet mutations provided an increase in human Tf binding affinity for BES1-TfR, suggesting that Tf affinity was not the driving factor for TfR variation. In addition, mapping of TfR sequence diversity and glycans delivered further evidence indicating that host promiscuity was not the origin of receptor multiplication and diversification. The location of polymorphisms at the surface of the receptor within exposed epitopes that were not protected by bound ligand and could be susceptible to immunoglobulin attack strongly suggested that variation was driven by an immune avoidance strategy. Furthermore, N-linked glycans were placed to shield conserved epitopes. Concomitantly, polymorphic clusters and glycan barriers supported the hypothesis that evolutionary divergence of TfRs originated from an advantage conferred by avoidance of the host adaptive immune system, thus promoting long term persistence.

Chapter 5

Iron acquisition

5.1 Introduction

5.1.1 Iron acquisition in mammals

Iron is an essential nutrient and one of the most abundant elements on Earth. The redox potential of iron underpins its requirement in cells, with transitions between Fe(III) and Fe(II) constituting an important role in electron transfer reactions. Due to the necessity of iron, transport and uptake of iron has shaped evolution. Vertebrates sequester Fe(III) within the iron transporter protein transferrin (Tf) to avoid conversion to Fe(II) and the associated toxicity. Thus, a cellular transferrin receptor (TfR) has evolved across vertebrates to specifically bind circulating Tf. Furthermore, iron sequestration serves to limit iron scavenging from invading pathogens, which in turn has led to pathogen adaptation in the form of sophisticated iron acquisition mechanisms to combat host defence strategies. Bloodstream-form trypanosomes have evolved a TfR which is not homologous to the human TfR but can efficiently hijack host Tf via receptor-mediated endocytosis (Coppens *et al*., 1987).

Iron acquisition has been extensively characterised in humans. The human TfR exists in two forms, TfR1 and TfR2. The latter is predominantly expressed in the liver, while the TfR1 homologue is ubiquitously expressed and will be referred to in this chapter. The homodimeric transmembrane receptor binds two molecules of transferrin and relies on pH changes during cellular trafficking to efficiently release iron and transferrin. The human TfR has a high affinity for iron-bound holo-Tf (K_D = 1 to 10 nM) (Giannetti *et al*., 2003). Upon binding, the receptor-ligand complex is internalised via clathrin-mediated endocytosis (Pearse & Robinson, 1990). As the complex transits through the endocytic system, it experiences a progressive decrease

in pH. Iron is released from the complex at between pH 6.5 and pH 5.5 (Morgan, 1979). Studies have shown that the receptor potentiates the release of iron at a slightly higher pH when compared to iron dissociation from Tf alone (Giannetti *et al*., 2005) (Eckenroth *et al*., 2011). The human TfR retains a high affinity for iron-free apo-Tf at low pH. However, the receptor has a low affinity for apo-Tf at neutral pH (Morgan, 1983). Upon recycling of the receptor-ligand complex to the cell surface, apo-Tf is released during exposure to neutral pH. The receptor can now resume availability and selectively bind holo-Tf.

5.1.2 Iron acquisition in trypanosomes

Procyclic-form trypanosomes that proliferate in the tsetse fly midgut are dependent upon iron for many cellular processes. Procyclic-form cells rely on oxidative phosphorylation for energy production, in which iron-sulphur clusters play a critical role. Iron acquisition during the insect stage remains relatively unknown, although it has been shown that the process may involve uptake and reduction of ferric complexes (Mach *et al*., 2013). In contrast, bloodstream-form trypanosomes, which have evolved to express a TfR, required little iron due to the absence of an active respiratory chain (Hannaert *et al*., 2003). Previous calculations revealed that bloodstream-form trypanosomes required 40,000 Fe(III) atoms per generation but had the capacity to internalise 85 000 iron atoms (Steverding, 1998). During mammalian infection, it is believed that Fe(III) is required to form Fe-dependent superoxide dismutases involved in converting superoxide radicals to protect against oxidative stress (Kabiri & Steverding, 2001). Haem containing iron may also play a role in antioxidant defence in bloodstream-form trypanosomes (Kořený *et al*., 2010) and can be acquired by uptake of haptoglobin-haemoglobin mediated by the haptoglobin-haemoglobin receptor (HpHbR) (Vanhollebeke *et al*., 2008). Whether iron can be extracted from haem is uncertain and seemingly implausible, as an orthologue of haem oxygenase has not been identified in trypanosomes (Taylor & Kelly, 2010). Furthermore, various studies suggested that iron transported by Tf is essential for trypanosomes, based on evidence which demonstrated that Tf is required for growth (Schell *et al*., 1991b) and knockdown of TfR variants by RNAi led to trypanosome growth arrest (Tiengwe *et al*., 2017).

There is a paucity of data relating to Tf uptake and iron release in trypanosomes. The TfR-Tf complex is endocytosed at the flagellar pocket (Coppens *et al*., 1987), which prompted speculation that Tf and iron release may operate in a similar manner to the human pathway. However, the fate of the receptor-ligand complex remained unclear, with one study showing

that the TfR-apoTf complex is stable from pH 8 to pH 5 (Steverding *et al*., 1995), while another study demonstrated dissociation of apo-Tf from the receptor at pH 5 (Maier & Steverding, 1996). In this chapter, affinity measurements would be performed to clarify whether binding of Tf to the trypanosome TfR was altered depending on the pH and iron-bound state of the ligand.

5.2 Aims

To investigate the implication of iron and the pH-dependency of receptor-ligand interactions, surface plasmon resonance (SPR) was performed using holo- and apo-Tf and a pH range to represent progression through the endocytic pathway. Several key questions would be addressed:

1. Does the trypanosome TfR mimic the human TfR regarding iron and ligand release?

2. How do pH and the presence of iron affect ligand binding?

3. At which stage of the endocytic pathway does iron release from the receptor-ligand complex occur?

5.3 Results

5.3.1 A conserved binding interface on transferrin

Unlike the human TfR, previous studies have shown that the trypanosome TfR does not discriminate between iron-bound holo-Tf and iron-free apo-Tf (Steverding *et al*., 1995). However, a relation in ligand binding between the human and trypanosome receptor was observed. An overlay of the structure of human TfR-Tf complex with the trypanosome TfR-Tf complex revealed that both receptors contacted a similar surface on human Tf (Figure 34). Trypanosomes have likely evolved to bind the same ligand surface to prevent the possibility of Tf escape mutations. If escape mutants arose to avoid binding to the trypanosome receptor, a loss in binding to the native human TfR would also occur. Furthermore, the principle of minimalising escape mutants can be applied to other mammals as the TfR-Tf interface is almost certainly conserved in mammals. The evidence for this is: First, mammalian TfRs are homologous, with TfRs from each species used in this study exhibiting more than 75% amino

acid identity with the human TfR, and an alignment demonstrated that mammalian TfRs were conserved (Appendix 7). Second, an alignment of mammalian Tfs also showed homology (Figure 35), thus suggesting that the TfR-Tf binding site is conserved across mammalian species. Finally, an alignment of mammalian Tfs illustrated that the majority of human Tf residues interacting with the trypanosome TfR were either identical or similar in other species (Figure 35), demonstrating that trypanosomes would bind a similar surface on other mammalian Tfs. Thus, the binding strategy had evolved to limit the propagation of Tf escape mutants and could be extended to other mammalian hosts.

5.3.2 The structure revealed monoferric transferrin bound to the trypanosome receptor at pH 6.5

The crystal structure confirmed that the trypanosome TfR bound a single Tf molecule. Tf can be subdivided into two lobes, termed N-lobe (residues 1-330) and C-lobe (residues 340-679), with a hinge region separating the two lobes. The structure of the trypanosome TfR-Tf complex was obtained at pH 6.5 and electron density corresponding to an iron atom was only present in the C-lobe (Figure 36). Each lobe can coordinate an Fe(III) atom via a conserved aspartic acid, histidine and two tyrosines within each lobe (Figure 36, inset). A synergistic anion, typically carbonate, completed the coordination (Lambert *et al*., 2005). A lack of electron density at the site of iron coordination within the N-lobe indicated that Fe(III) was absent.

Previously solved structures of human apo-transferrin lacking iron (Wally *et al*., 2006) and human holo-transferrin with iron bound in both lobes (Noinaj *et al*., 2012) were superimposed onto the structure of human transferrin bound to trypanosome TfR (Figure 37) described in the previous chapter. Indeed, the N-lobe aligned with the apo-transferrin structure adopting an open conformation of the iron binding cleft, permissive to iron release (Figure 37A). The C-lobe aligned with holo-transferrin, showcasing a closed fold of the cleft to retain iron (Figure 37B). Adoption of the closed conformation in the C-lobe and an open conformation in the N-lobe provided further evidence that the structure had a single iron atom bound in the C-lobe. Although the structure gave us a snapshot of the iron state of the complex at pH 6.5, understanding transferrin binding and iron presence at different pH values of the endocytic pathway was essential.

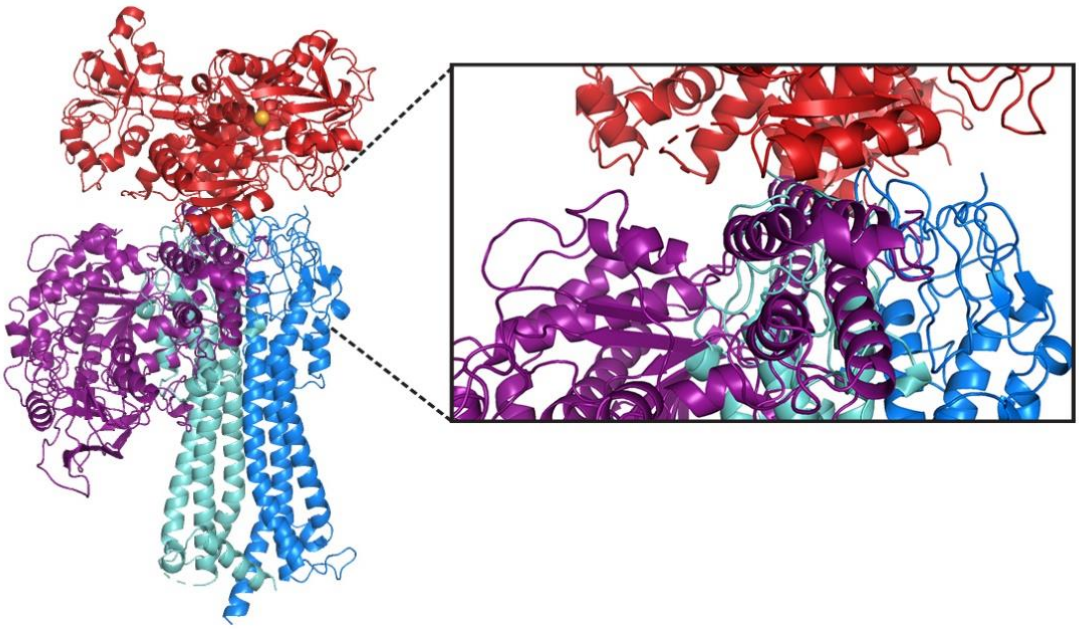


Figure 34: Human TfR and trypanosome TfR bind a similar surface on Tf. Overlay of the structures of human TfR-Tf complex and the trypanosome TfR-Tf complex. Human TfR is represented in purple, ESAG6 in blue, ESAG7 in cyan and human Tf in red.

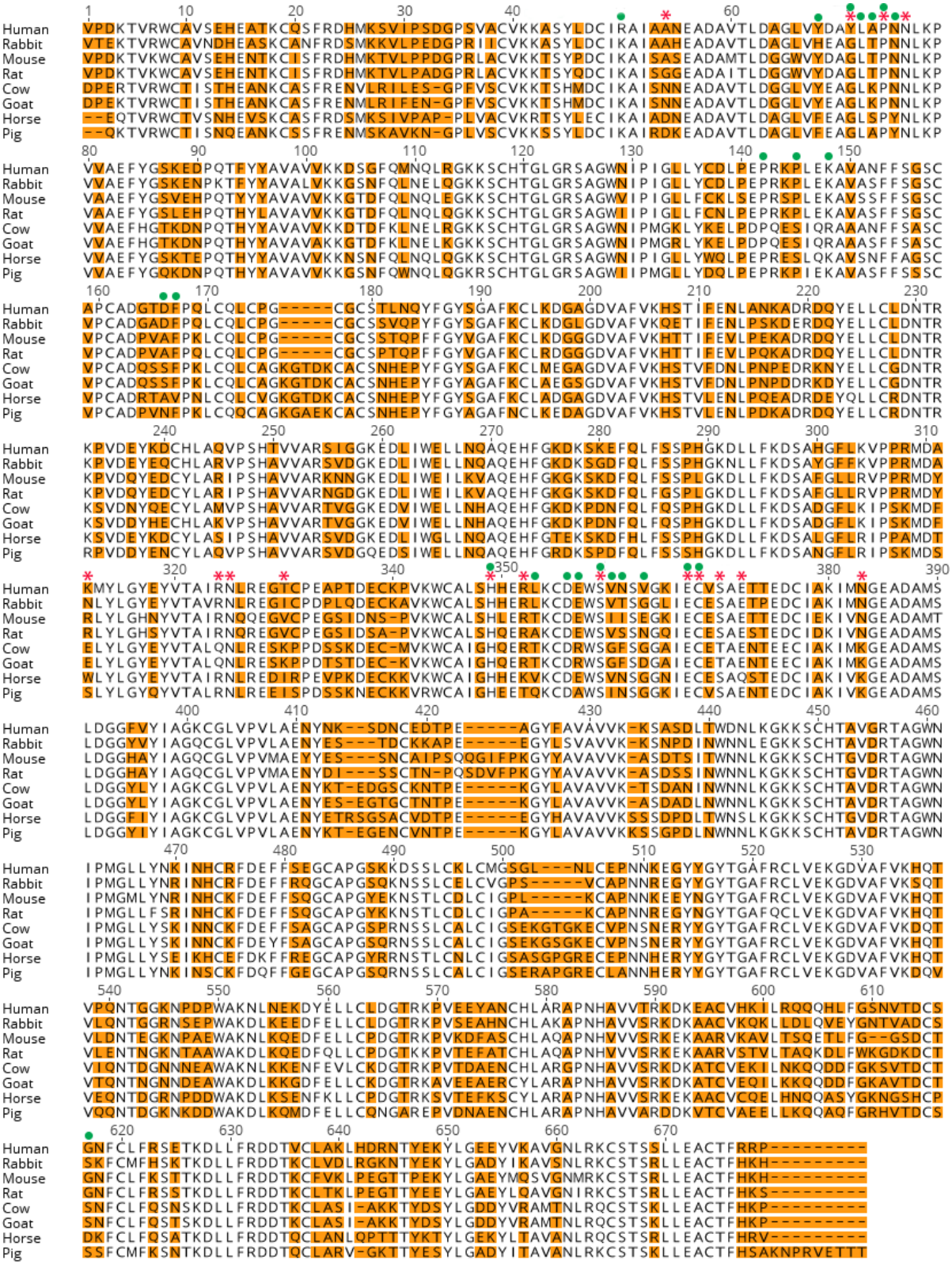


Figure 35: Alignment of mammalian transferrins. Protein sequences of mammalian Tfs used in binding studies were aligned and variation was highlighted in orange. Residues in human Tf that interact with the trypanosome TfR are indicated with a red asterisk and those that interact with the human TfR are indicated with a green circle.

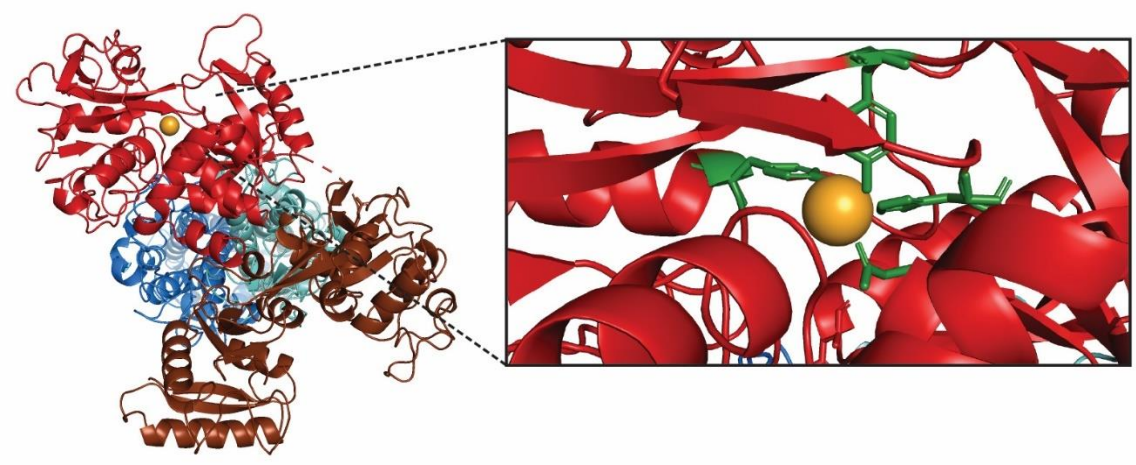


Figure 36: Structure of human Tf N- and C-lobes bound to trypanosome TfR. The N-lobe of human Tf is shown brown, the iron-bound C-lobe is coloured in red and Fe(III) is depicted as an orange sphere. The four Tf residues that coordinate iron are shown as a stick representation in green in the inset.

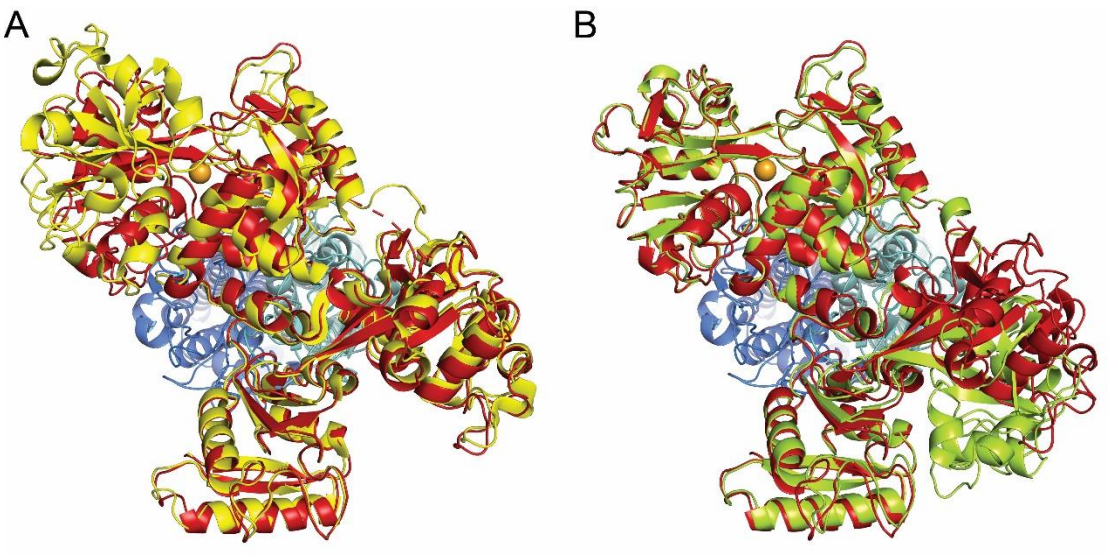


Figure 37: Comparison of apo- and holo-Tf structures with Tf bound to the trypanosome TfR. A, Structure of human apo-Tf (PDB: 2HAU) shown in yellow was superimposed onto the structure of human Tf (red) bound to the trypanosome TfR. B, An overlay of diferric human Tf (PDB: 3V83) in green and human Tf (red) bound to the trypanosome TfR.

5.3.3 Effects of pH on ligand binding

In light of the similarities in Tf-binding site between the trypanosome TfR and the native TfR, a hypothesis was formulated that molecular mimicry may occur. Following iron release from the endocytosed human TfR-Tf complex, apo-Tf remains bound at acidic pH. Recycling of the complex to the cell surface leads to Tf dissociation as apo-Tf has a low affinity for the human TfR at neutral pH (Morgan, 1983). Thus, binding of apo- and holo-Tf to trypanosome TfR was investigated by SPR using a pH range that represented the changes in environment during endocytosis. Previous studies had shown that the lysosomal pH encountered in bloodstream-form trypanosomes is 4.8 (McCann *et al*., 2008), and this increase in acidity was reflected in the pH spectrum selected. The sensorgrams showed binding of iron-bound and iron-free transferrin to the trypanosome TfR at every pH tested (Figure 38), with dissociation constants (K_D) remaining in the low nanomolar range (Table 4). At pH 7.4, a three-fold difference in affinity was observed between holo- and apo-Tf. While holo-Tf bound Bn2-TfR with a K_D of 9.1 nM, apo-Tf showed weaker binding with a K_D of 31 nM (Table 4). While a reduction in binding was observed, it was clear that the loss of iron did not abolish Tf binding to the trypanosome TfR at neutral pH, indicating that the parasite receptor did not operate in the same way as the human receptor. Contrary to previous findings (Maier & Steverding, 1996), low pH did not disrupt the TfR-Tf complex. One possibility is that a lower affinity TfR, such as BES1-TfR, may facilitate the release of Tf at acidic pH. To test this, a similar experiment was performed by biolayer interferometry (BLI) to study the effect of pH on the lower affinity interaction between BES1-TfR and bovine Tf. The sensorgrams showed that Tf binding was conserved at low pH (Appendix 8). A difference in binding response was observed between two iron states at pH 7.4, with a reduction in binding observed for apo-Tf compared to holo-Tf. The decrease in binding was consistent with the differences in binding responses between Bn2-TfR and human Tf in two iron states, measured by SPR at pH 7.4 (Figure 38). The reduction in binding was attributed to the differences in Tf conformation depending on the iron state (Figure 37), suggesting that these changes were ubiquitous in mammalian Tfs. The open conformation of apo-Tf was likely responsible for the weaker affinity observed. At lower pH, binding of bovine Tf to BES1-TfR was not disrupted (Appendix 8). While it appeared that holo-Tf showed weaker binding compared to apo-Tf as the acidity increased, it was difficult to draw robust conclusions as there were discrepancies between the initial baseline signal and kinetic parameters could not be determined. However, the data suggest that Tf would remain bound to its receptor throughout the endocytic pathway, regardless of the host species or the TfR actively expressed.

Although we had determined that the trypanosome receptor-transferrin complex does not respond to pH in the same way as the native human receptor, the fate of iron remained elusive. A decrease in affinity was observed in human holo-Tf binding to Bn2-TfR between pH 7.4 (K_D = 9.1 nM) and pH 6.5 (K_D = 31 nM) (Table 4). The weaker binding of holo-Tf at pH 6.5 closely resembled the affinity of apo-Tf at pH 6.5 (K_D = 35 nM), suggesting that there may be iron release from the TfR-Tf complex between pH 7.4 and pH 6.5. Additional studies would be conducted to further elucidate iron release during endosomal progression.

5.3.4 Investigation of iron release by microPIXE

Particle-induced X-ray emission with a microfocused beam (microPIXE) is a technique pioneered by Elspeth Garman and Geoffrey Grime (Garman & Grime, 2005), which allowed investigation of iron release in atomic detail (Appendix 9). An identical pH range was selected to reflect endocytic conditions and provide a comparison with SPR data. Iron quantification was performed to measure the number of iron atoms present in unbound human Tf and human Tf bound to Bn2-TfR (Figure 39), providing several observations. First, it is worth noting that the starting iron saturation at physiological pH for both free Tf and receptor-bound Tf is ~0.6 atoms per Tf, which correlated with the reported 30-60% iron saturation of circulating serum Tf (Williams & Moreton, 1980) (Eckenroth *et al*., 2011). Second, in the absence of receptor, Tf displayed a dramatic release of iron following a decrease by one pH unit, with a loss of ~80% of bound iron. By pH 4.8, almost all iron has been released from Tf, with 0.055 atoms of iron remaining per molecule of Tf. In contrast, the fate of iron for receptor-bound Tf was different with a reduced amount of iron loss during acidification compared to Tf alone. Approximately 50% of bound iron was released from the complex between physiological pH and pH 6.5, which correlated with previous SPR data suggesting a loss of iron between pH 7.4 and pH 6.5 (Figure 38). Finally, binding of the trypanosome TfR to human Tf appeared to stall the release of iron during transit of the complex through the endosomal lumen. Thus, iron could be retained until the complex reached the lysosome.

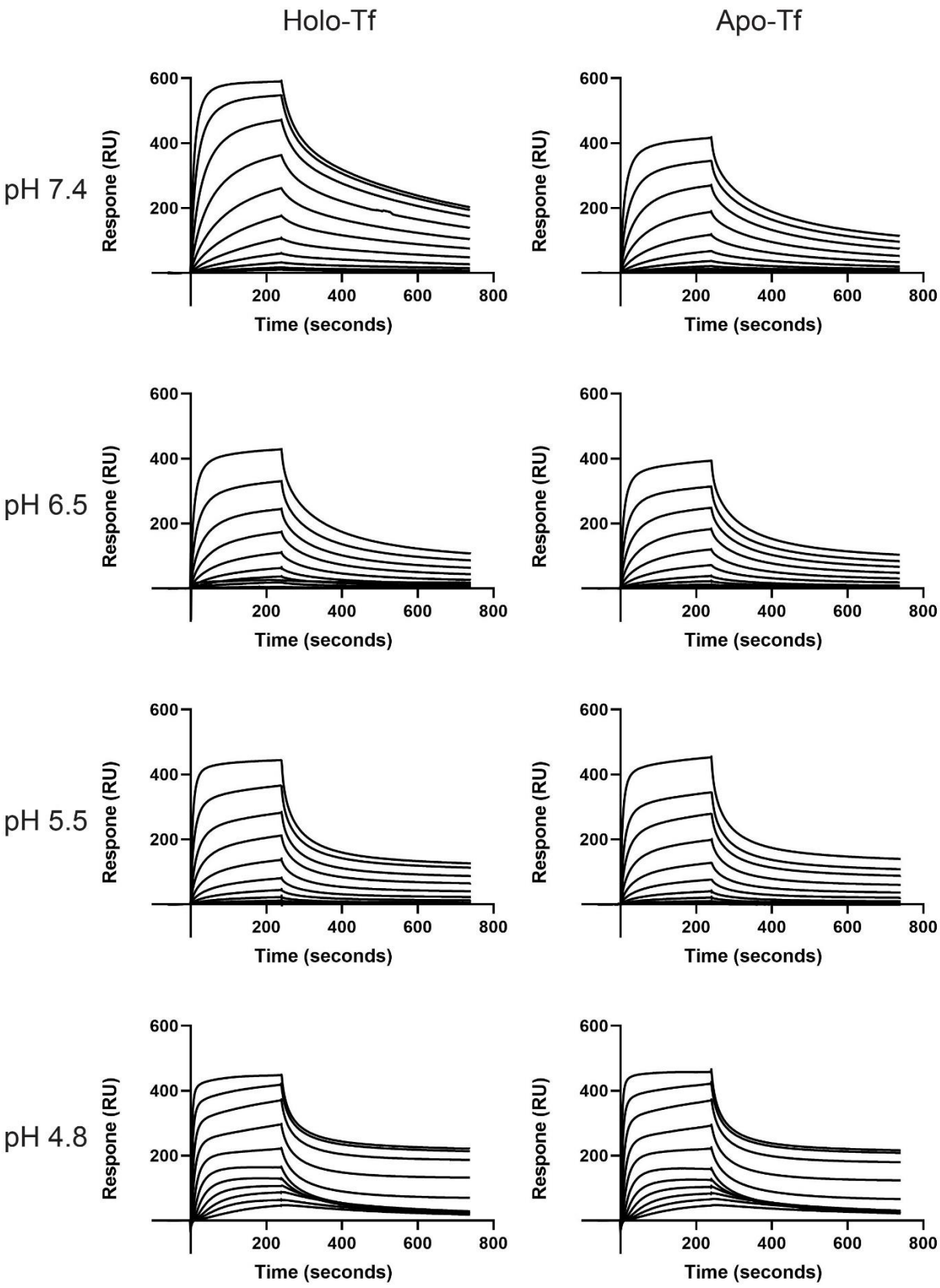


Figure 38: Measurement of human holo- and apo-Tf binding to Bn2-TfR at different pH by SPR. Biotinylated TfRs were immobilised and purified Tf was injected in a concentration series ranging from 1 µM to 1 nM at pH 7.4, 6.5, 5.5 and 4.8.

Table 4: SPR measurement of binding parameters. Kinetic values for interactions between immobilised Bn2-TfR and humo holo- and apo-transferrin. Kinetic data for pH 5.5 and pH 4.8 were not reported due to inaccuracies caused by the low pH buffer.

TfR	pH	Transferrin	K_D(nM)	k_{on}($M^{-1}s^{-1}$)	k_{off}(s^{-1})
Bn2	7.4	Human holo-Tf	9.1	2.4×10^5	2.2×10^{-3}
Bn2	7.4	Human apo-Tf	31	1.3×10^5	4.1×10^{-3}
Bn2	6.5	Human holo-Tf	41	2.4×10^5	9.8×10^{-3}
Bn2	6.5	Human apo-Tf	35	2.1×10^5	3.7×10^{-2}

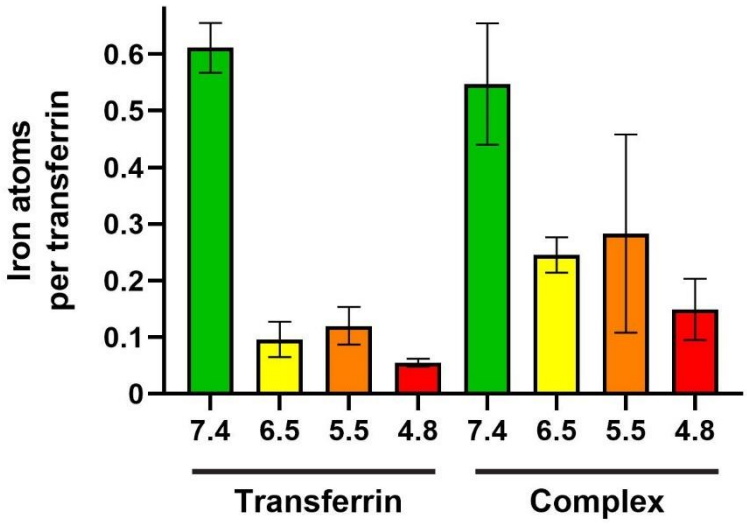


Figure 39: Quantification of bound iron by microPIXE. The number of iron atoms bound per Tf was measured at each pH value (7.4, 6.5, 5.5 and 4.8) for free human Tf (Transferrin) and human Tf bound to the trypanosome TfR (Complex). Mean values of three technical replicates were plotted while error bars represent standard deviation.

5.3.5 How essential is transferrin?

Previous studies in chapter 3 and this chapter showed that a low affinity TfR can support growth and that the trypanosome TfR does not discriminate between holo- and apo-Tf, which prompted questioning of the essentiality attributed to Tf uptake in trypanosomes. To test this, bovine Tf was selectively depleted from foetal calf serum (FCS) by affinity chromatography using a Bn2-TfR column. Tf was not detected by western blot probed with anti-bovine Tf antibody in Tf-depleted FCS (Figure 40). A titration was performed to investigate the sensitivity of anti-bovine Tf antibody and low (4.5 nM) concentrations of Tf were detectable (Appendix 10), indicating that Tf-depleted serum was likely to have sub-nanomolar concentrations of bovine Tf. BES1-TfR was shown to bind bovine Tf with a K_D of 80 nM (Table 1), which was far greater than the concentration of bovine Tf in depleted FCS. Trypanosomes expressing the BES1-TfR were cultured in medium containing Tf-depleted FCS and compared to a control culture in medium containing FCS. In both groups, cells proliferated without any marked differences in growth rates (Figure 40). Cells were cultured for 200 hours, a sufficient amount of time to exhaust iron reserves, and the low Tf concentrations (< 1 nM following dilution in medium) did not negatively impact cell proliferation.

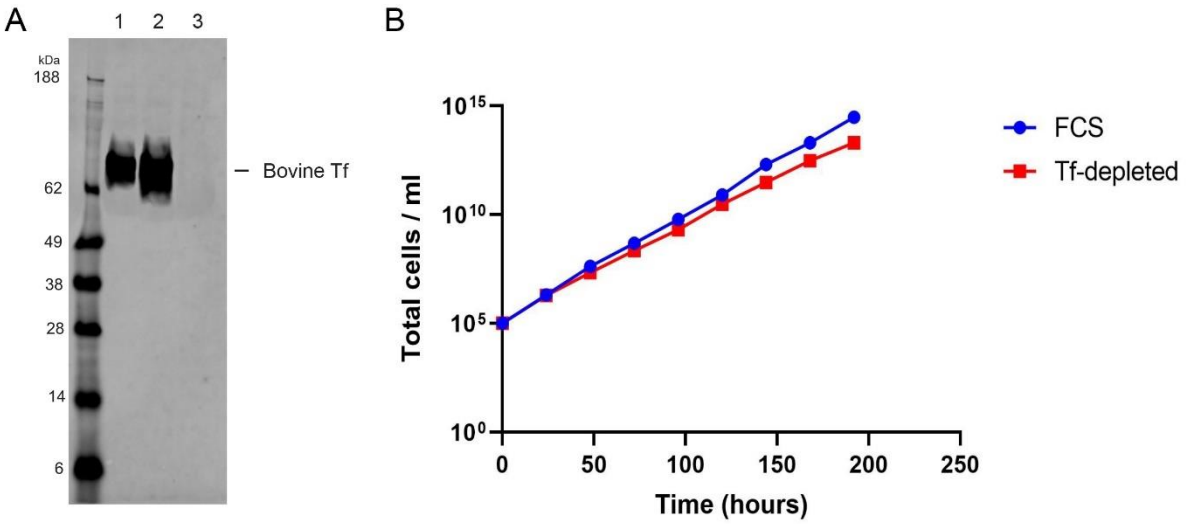


Figure 40: Comparison of growth rates of trypanosomes in FCS and Tf-depleted serum. A, Samples containing 10% FCS (1), 10% FCS depleted of Tf and repleted with 3 µM bovine Tf (2) and 10% FCS depleted of Tf (3) were analysed by western blot probed with anti-bovine Tf antibody. B, Bloodstream form trypanosomes expressing BES1-TfR were grown in medium containing FCS (blue) or FCS depleted of Tf (red).

5.4 Discussion

In this study, pH appeared to have little effect on transferrin binding, indicating that receptor-ligand interactions are likely to be maintained through the endocytic pathway. Furthermore, binding of Tf to the trypanosome receptor stalled iron release compared to unbound Tf, with some iron shown to remain bound at lysosomal pH. The delayed release of iron observed in receptor-bound Tf could indicate an evolutionary advantage conferred by iron retention until the lysosome. Iron sequestration until the lysosome may have physiological significance, as previous studies have shown that *T. brucei* mucolipin-like protein (*Tb*MLP), involved in iron transport to the cytosol, is located in proximity to the lysosome (Taylor *et al*., 2013). However, release of the remaining iron and the fate of the TfR-Tf complex remains unknown, prompting speculation of several possible scenarios. First, the entire complex could be degraded allowing iron release, although this is inconsistent with previous reports that the TfR is recycled (Kabiri & Steverding, 2001). Second, the entire complex could be recycled to the cell surface, but this does not explain how the remaining sequestered iron is released. Third, the receptor and ligand could be differentially sorted and embark on separate routes, resulting in Tf proteolysis and TfR recycling as suggested in previous studies (Pal *et al*., 2003). Previous investigations have shown that the TfR may have a half-life between 1.5 hours (Schwartz *et al*., 2005) (Tiengwe *et al*., 2017) and 7 hours (Kabiri & Steverding, 2000), with a cycling time of 11 minutes (Kabiri & Steverding, 2000), suggesting that the receptor is recycled. However, whether the ligand remains bound or the factors involved in receptor-ligand dissociation remain to be elucidated and further investigation is required.

Previous studies have shown that iron is essential for bloodstream form trypanosomes, and depletion of iron using the iron chelator deferoxamine resulted in growth inhibition (Breidbach *et al*., 2002). It would be useful to investigate iron metabolism through titration of deferoxamine to modulate iron starvation and further understand iron requirements in bloodstream form trypanosomes. Tf-depletion studies in section 5.3.5 have shown that uptake of Tf may not be essential as proliferation was not affected, which is in contrast to previous findings (Schell *et al*., 1991b). In the study conducted by Schell *et al*., cell growth was arrested in Tf-depleted FCS at 24 hours, with a decline in trypanosome population attributed to cell death observed between 24 and 48 hours. The discrepancies were not likely to be due to differences in the expressed TfR as both studies used trypanosomes expressing from the BES1 (221) site. However, Tf-depletion methods did vary. While Schell *et al*. immunised rabbits with bovine Tf to generate polyclonal antibodies for Tf depletion from FCS, the approach adopted in this study (section 5.3.5) relied upon the selectivity of Bn2-TfR for Tf-depletion from FCS. FCS is likely to contain lactoferrin, another iron transporter protein which is homologous to Tf.

Concentrations of lactoferrin detected in serum were 10-fold lower than Tf in 3-month old calves (Tóthová *et al*., 2016), and may remain at low levels in adults. An alignment of bovine Tf and bovine lactoferrin showed that the iron transporter proteins shared 58% amino acid identity (Appendix 11). The use of a polyclonal antibody may have resulted in depletion of both Tf and lactoferrin as the alignment suggested a likelihood for the presence of shared epitopes. It is unclear whether the polyclonal antibody used in 5.3.5 for Tf detection would cross-react and recognise lactoferrin, thus making it difficult to comment on whether lactoferrin would be present in Tf-depleted serum. If Bn2-TfR selectively depleted Tf and lactoferrin remained in the serum, two possibilities could be formulated. Either BES1-TfR does not discriminate between the two iron transporter proteins and can bind and endocytose lactoferrin, or trypanosomes have evolved an alternative strategy for lactoferrin uptake. If lactoferrin was not present in the Tf-depleted serum, it may be possible that trypanosomes can rely on an alternative iron source, whether that be through uptake of HpHb in conjunction with an uncharacterised pathway for iron extraction from haem, or bloodstream-form trypanosomes may have evolved an alternative route for iron uptake.

Chapter 6

Development of TfR-specific monoclonal antibodies using phage display

6.1 Introduction

Trypanosome populations have inherent capacities to avoid host immune detection by switching expression of cell surface proteins. Antibodies that successfully bind prior to switching can, at low to medium immunoglobulin titres, be cleared by endocytosis. Trypanosomes have remarkable endocytic capacities, a feat that is further enhanced by hydrodynamic flow forces that propel large, globular molecules such as antibodies towards the flagellar pocket (Engstler *et al*., 2007). Internalisation of surface bound immunoglobulins led to the formulation of the hypothesis that trypanosomes could be specifically targeted and killed if a toxin were conjugated to the IgG. This concept was investigated and validated during a collaboration involving our laboratory and AstraZeneca, led by MacGregor and Gonzalez-Munoz. The study was an elegant demonstration of a novel therapeutic strategy that can draw upon the body of research conducted in cancer ADC therapeutics by using a related ADC scaffold, with the addition of modified specificity to target trypanosomes (MacGregor, Gonzalez-Munoz *et al.*, 2019).

6.1.1 Trypanosome receptors as a target for ADC-based therapeutics

The *T. b. brucei* haptoglobin-haemoglobin receptor was selected as the target for an ADC therapeutic for several reasons. First, the receptor is encoded by a single copy gene within the genome which will remain expressed during dynamic VSG switching. Second, HpHbR is capable of receptor-mediated endocytosis of macromolecules including haptoglobin-haemoglobin (M_r = 150,000) (Lane-Serff *et al*., 2014) and the host immune complex TLF1 (M_r = 500,000) (Raper *et al*., 1999). Third, the *T. b. brucei* HpHbR structure revealed a three-helix bundle featuring a rigid 50° kink forming an extensive ligand interaction surface (Lane-Serff *et al*., 2014), which is likely accessible to bivalent IgG binding. Finally, a knockout HpHbR-/- cell line was available, as a control, to further validate the specificity of the proof-of-principle study. Specificity of the antibody drug conjugate (ADC) was achieved by performing phage display with a diverse scFv library. The choice of receptor proved fruitful as the phage display screen produced several antibodies that were endocytosed following HpHbR binding. Two antibodies were toxin-conjugated and exhibited excellent trypanosome-specific lethality with minimal toxicity to human cells. The IC50 concentrations of the ADCs tested in *T. b. brucei* were ~ 1 pM while human cell lines tolerated a dose 50,000 times greater, providing a remarkably wide therapeutic window (MacGregor, Gonzalez-Munoz *et al.*, 2019). However, there were a few drawbacks in the choice of receptor: (i) A low copy number of 200 – 400 molecules per cell (Vanhollebeke *et al*., 2008); (ii) A polymorphism in the human infective *T. b. gambiense* HpHbR leads to reduced efficiency of the ADCs in their current format (personal communication); (iii) Knockout studies show that the receptor is not essential and resistance may arise. Together, these formed the premise for an investigation of additional trypanosome cell surface receptors as targets for ADCs. The trypanosome transferrin receptor is an attractive target as it has a higher copy number of ~ 3,000 receptors/cell (Steverding *et al*., 1994) and has, in previous studies, been shown to be essential (Tiengwe *et al*., 2016). Although the TfR is variable, the different sequences are highly conserved and common epitopes may be exploitable. Following the previous success of a phage display selection screen generating trypanosome-specific antibodies, phage display would be employed to raise antibodies against the trypanosome transferrin receptor. BES1-TfR and Bn2-TfR are two of the most divergent TfRs in the Lister 427 *T. b. brucei* strain and would serve as a substrate for developing a pan-specific TfR antibody.

6.1.2 Origins of phage display

In addition to the endocytic potential of an ADC-based therapy to target trypanosomes, antibodies provide an attractive therapeutic strategy due to their compatibility with the host immune system. Phage display offers a high-throughput, *in vitro* approach for raising specific antibodies against a target. Two properties of bacteriophage were harnessed to develop 'phage display': their capacity to express surface proteins and their ability to replicate in bacteria. This concept of coupling genotype to phenotype within a virion and achieving clonal propagation from *E. coli* was originally pioneered in the 1990s (Scott & Smith, 1990) (McCafferty *et al*., 1990). Filamentous M13 phage infect *E. coli* expressing the F-pilus and are commonly used in phage display. The variable regions from the light and heavy chains of an antibody (Figure 41A) can be engineered in various configurations to form bivalent antigen-binding fragments ($F(ab')_2$) (Figure 41B) or monovalent antigen-binding fragments (Fabs) (Figure 41C). Additionally, the variable heavy (V_H) and variable light (V_L) chains can be fused with a flexible linker to form a single chain variable fragment (scFv) (Figure 41D). The scFv format was initially selected during the origins of phage display for two reasons. First, the binding kinetics of an scFv were shown to be equivalent to antigen binding by the parental antibody (Ward *et al*., 1989). Second, scFv can be inserted into a filamentous fd phage vector at the N-terminus of the gene III adsorption protein, and expressed to achieve scFv – protein III (pIII) surface display (McCafferty *et al*., 1990) (Figure 42). Four copies of pIII are typically expressed at the surface of each virion and mediate bacterial adhesion. McCafferty and colleagues demonstrated that a one million-fold enrichment of anti-lysozyme scFv-expressing phage could be achieved after two rounds of purification over a lysozyme-Sepharose column (McCafferty *et al*., 1990). This principle was adapted to soluble and panning selections in which antigens are immobilised on beads or a plate surface respectively. Two to four rounds of selection are typically performed with library output quantification and diversity monitoring to determine the success of the screen.

6.1.3 Filamentous phage properties

Filamentous phage are used in phage display due to their parasitic, as opposed to lytic, infection of bacteria. This results in phage extrusion from *E. coli* without killing the cells (Marvin and Hohn, 1969). The phage single stranded genome is typically replaced with a phagemid vector encoding a selectable marker and the scFv-pIII fusion protein under the *lacZ* promoter. The remainder of the genetic material required for virion production is supplied by helper phage. M13KO7trp helper phage have been designed to preferentially package phagemid

vector DNA over phage DNA, ensuring effective assembly and extrusion of scFv-phages rather than helper phage production. Any helper phage DNA that is packaged and extruded encodes a trypsin-sensitive motif within the pIII protein that will render the phage non-infective following trypsin treatment.

6.1.4 Library generation

ScFv phage display concomitantly holds a phenotype and genotype by providing an immune component display system coupled to its DNA cargo. The success of the technology can be underpinned by the size and diversity of the scFv-expressing phage library. The diversity of scFv-expressing phage was achieved by building an immune naïve library that harnesses the inherent variation of antibodies expressed from lymphocytes. Peripheral blood lymphocytes, tonsil B-cells and bone marrow B-cells were collected from 43 non-immunised donors (Vaughan *et al*., 1996). Messenger RNA was isolated and complementary DNA was transcribed using random hexamer primers. PCR was performed with specific primers to amplify V_H and V_L chains. PCR fragments were inserted into a phagemid vector with a $(Gly_4Ser)_3$ linker to connect the two chains. The resulting bone marrow Vaughan (BMV) phage library achieved $1.4x10^{10}$ individual scFv recombinants. The combined spleen (CS) library was created in the same way using splenic and foetal liver B-cells from 20 non-immunised donors (Lloyd *et al*., 2009). BMV and CS phage libraries were pooled to produce a library size of $9.6x10^{10}$ scFv clones.

6.1.5 Phage display principle

Phage display is performed by incubation of the target receptor with the BMV/CS scFv-expressing phage library (Figure 43). Following removal of unbound phage, trypsin was added to fulfil several functions. First, trypsin-mediated proteolysis released bound phage through cleavage of the trypsin-sensitive Myc-tag adjacent to the scFv. Second, removal of the scFv fragment revealed pIII, rendering the phage *E. coli* infective. While library phage retained the ability to infect *E. coli*, they lacked the capacity to replicate the phagemid vector and egress from cells. Library phage were supplemented with helper phage to provide the genetic material to complete a productive infection and release phage particles. The resulting output population was enriched with phage displaying scFv specific to the target receptor.

6.1.6 Triage of output phage library

6.1.6.1 Phage ELISA

Phage-based enzyme-linked immunosorbent assays (ELISA) provided a high-throughput screening method to visualise specificity and triage the enriched output population to remove weak or non-specific binders. Each well of a 96-well master plate had a single TG1 colony that contained the phagemid vector encoding scFv-pIII. The pIII leader sequence localised pIII to the periplasmic space. Insertion of pIII in the outer membrane was refractory to subsequent phage infection, preventing superinfection of multiple phages in TG1 cells. To enable helper phage superinfection, glucose was added to the medium to mediate lacZ repression and prevent scFv-pIII expression. Upon addition of M13KO7trp helper phage, extrusion of scFv-phage was achieved.

The phage output raised against either BES1-TfR or Bn2-TfR was subsequently used in a phage ELISA to assess interactions with each antigen (Figure 44). A 96-well streptavidin assay plate was coated with BES1-TfR or Bn2-TfR. A third streptavidin plate without immobilised antigen was used to assess the prevalence of streptavidin binders. Clonally-extruded scFv-phage from each well of the master plate were added to each well of the antigen-coated assay plate. Following a wash step, bound scFv-phage were detected by anti-M13. HRP-conjugated anti-M13 reacted with TMB to produce a chromogenic substrate. The reaction was quenched by the addition of sulfuric acid and absorbance of the chromogen could be visualised at 450 nm. The absorbance signal was measured in absorbance units (AU) and positively correlated with phage binding.

6.1.6.2 Homogenous time resolved fluorescence

While phage ELISA served as a first-line indication of the success of a phage display screen, its reliability as a quantitative assay was debatable as several shortcomings exist. First, expression of phage and display proteins can vary between clones. Second, anti-M13-HRP binds pVIII, the most abundant phage coat protein. Phage express ~ 2,700 pVIII copies per virion (Newman *et al*., 1977), resulting in significant signal amplification and difficulties in discriminating between weak and strong binders. Homogenous time resolved fluorescence (HTRF) can be considered a more reliable screening technique to scrutinise binding as *E. coli*-

derived soluble scFv were used to eliminate the complications related to the presence of phage.

HTRF can be used to measure protein-protein interactions based on a combination of the fluorescence resonance energy transfer (FRET) principle and time-resolved measurement. An energy donor, such as cryptate, was coupled to streptavidin in order to bind the biotinylated target receptor (Figure 45). The acceptor fluorophore, XL665, was coupled to an anti-myc antibody for specific targeting to the myc-tagged scFv. Upon excitation from a light source, emission at 620nm occurred in the absence of proximity between the two fluorophores. However, if the scFv binds the target receptor, energy transfer from the donor to the acceptor fluorophore resulted in emissions at 665nm. The time resolved concept was derived from the occurrence of background short-lived fluorescence following excitation. To avoid interference of this signal in the analysis, measurements were taken 50-150 µs after excitation to allow for decay of the non-specific signal. Emissions at 665 nm and 620 nm wavelengths were measured following the time delay. The signal ratio of 665:620 nm positively correlated with the binding response between the scFv and the target receptor.

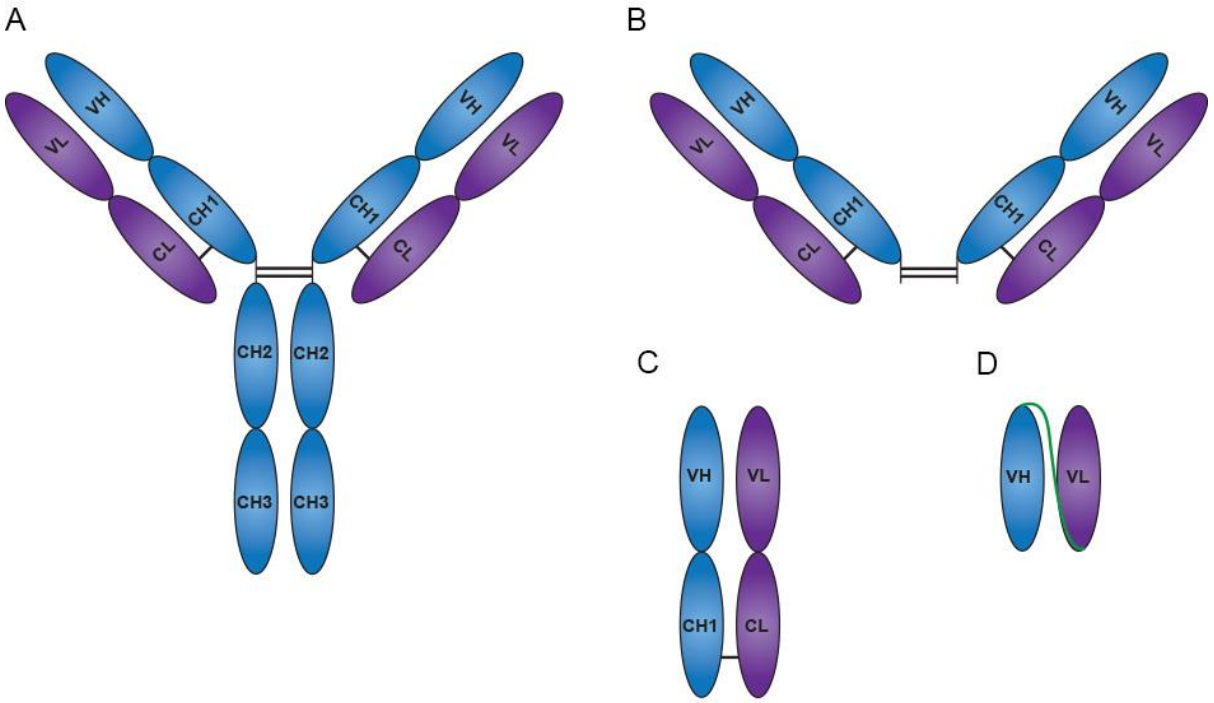


Figure 41: Immunoglobulin engineering schematic. A, Schematic representation of an IgG, with heavy chains shown in blue and light chains in violet. B, Representation of a F(ab')$_2$ fragment containing the two disulphide bridges in the hinge region to link the heavy chains. C, Fab fragment containing a disulphide bond connecting the heavy and light constant regions. D, Schematic of an scFv, composed of the variable domains of the light and heavy chain, fused by a linker peptide as they do not contain thiols for disulphide linkage.

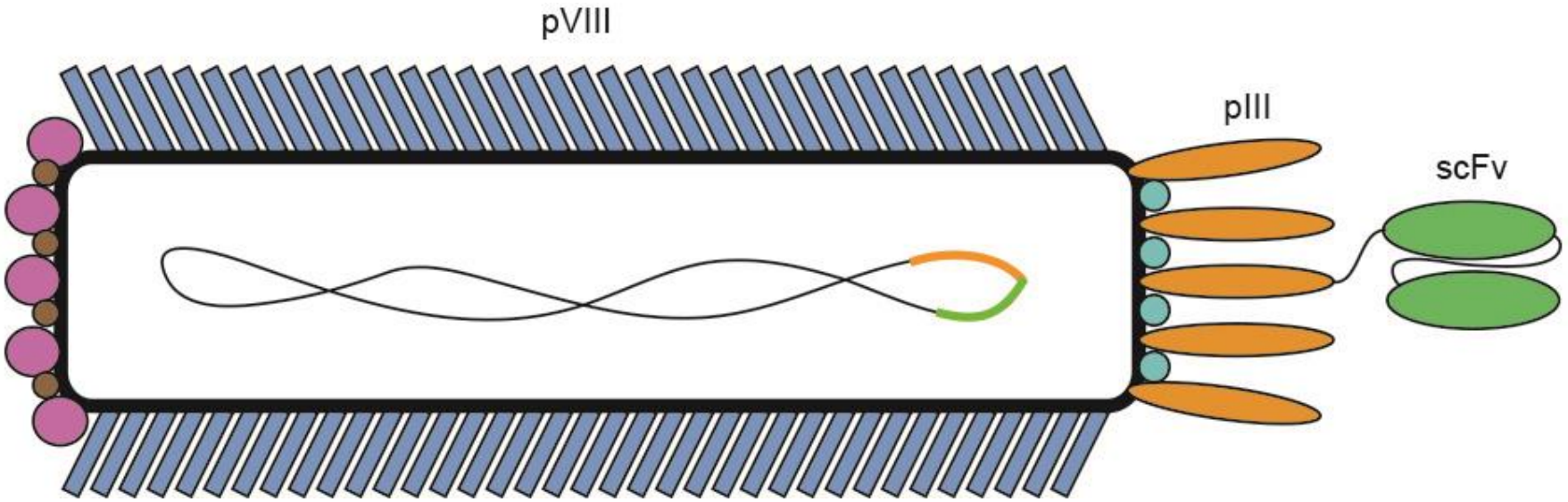


Figure 42: Schematic of the engineered filamentous M13 phage. The abundant pVIII protein forms the coat of the bacteriophage. Four to five copies of pIII are located at the anterior tip and mediate *E. coli* infection. A phagemid vector encodes scFv at the N-terminus of pIII to achieve scFv display.

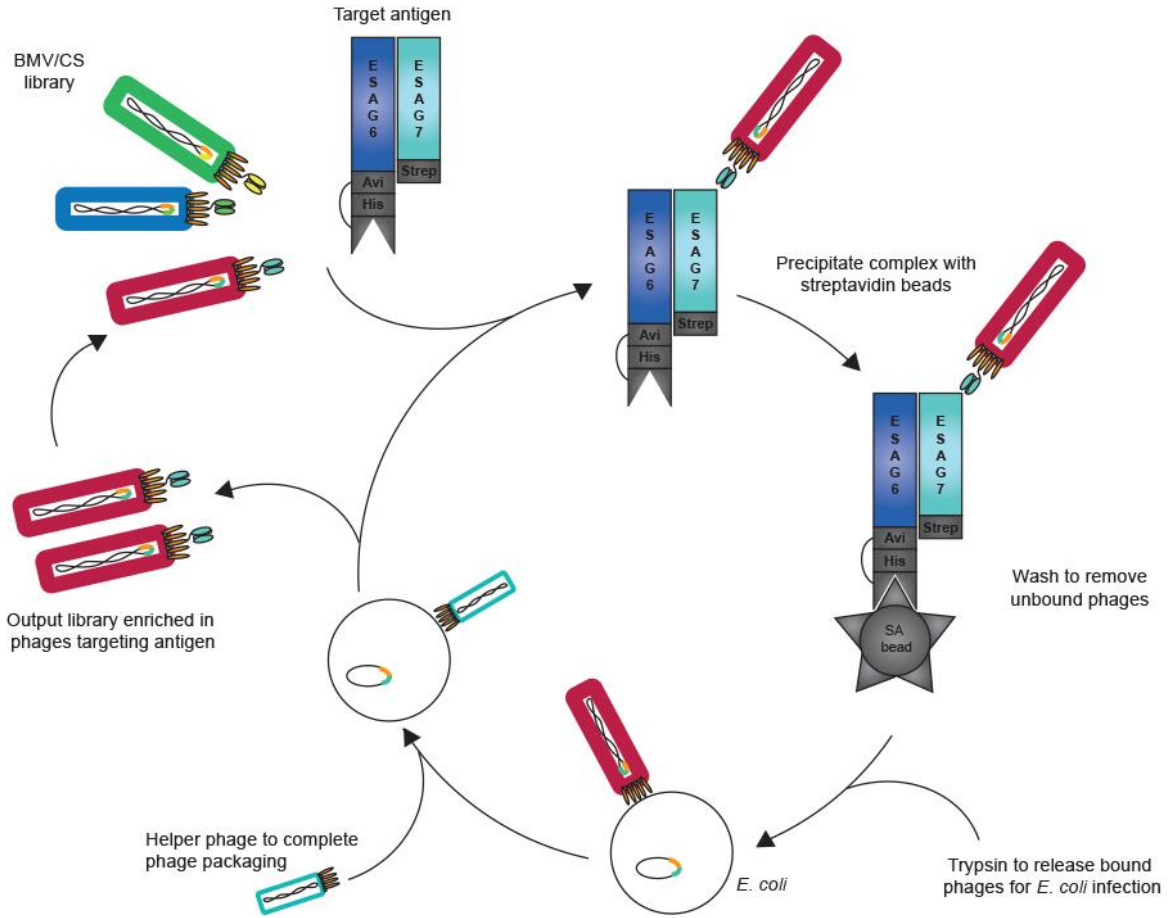


Figure 43: Schematic of the phage display soluble selection principle. Biotinylated target antigen is incubated with the scFv-phage library and binders are isolated by streptavidin bead precipitation. Trypsin digest fulfils the dual function of releasing bound phage and revealing pIII for *E. coli* infection. Helper phage are added to provide the lacking genetic material for virion packaging and extrusion. The resulting output population is enriched in phage expressing scFv that bind the target antigen.

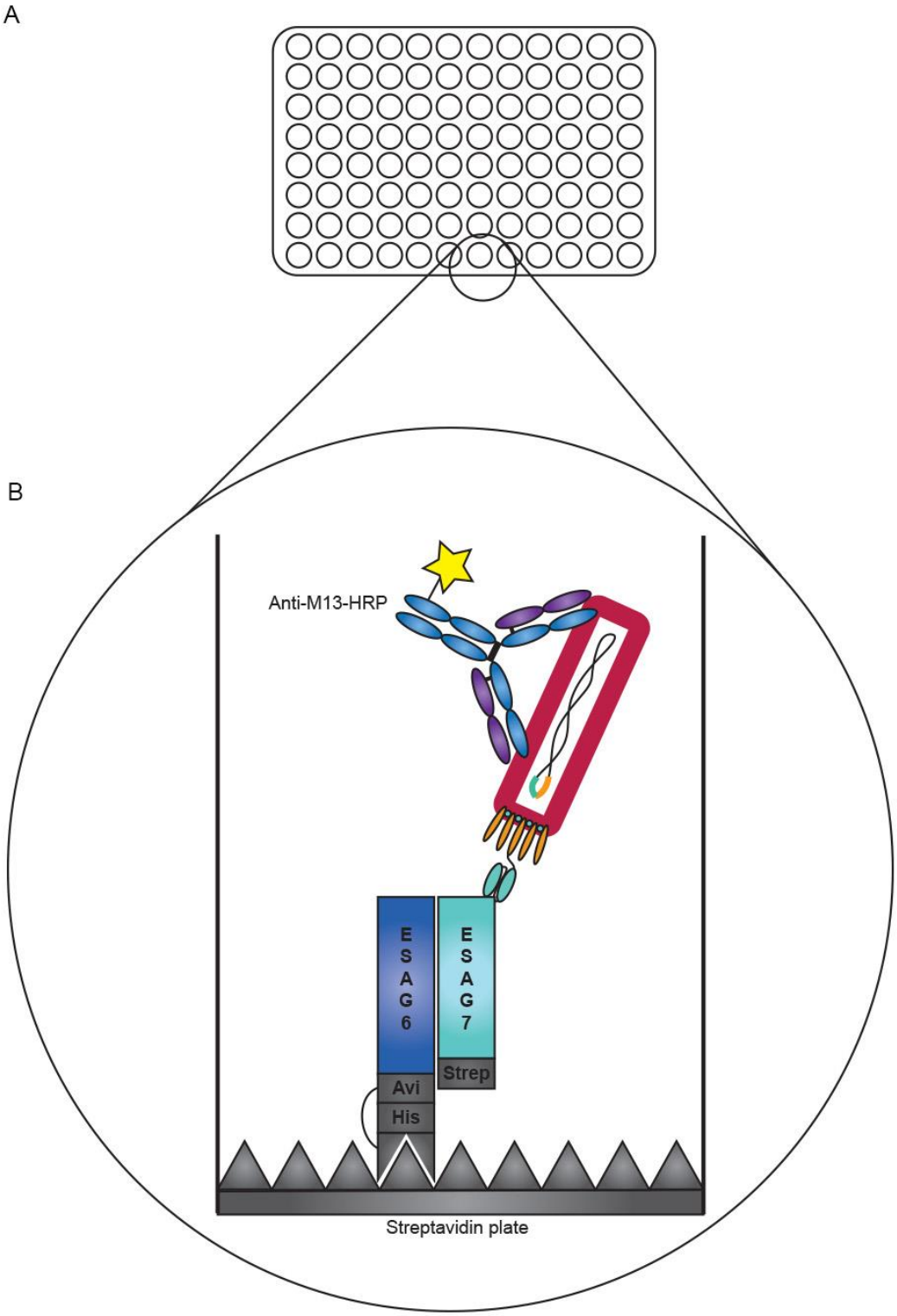


Figure 44: Schematic of the phage ELISA principle. A, 96-well plates are coated with biotinylated target antigen BES1-TfR or Bn2-TfR. B, A single well of a streptavidin-coated 96-well plate is represented. Biotinylated target antigen is immobilised on the streptavidin surface. Phage expressing scFv from clonal *E. coli* expansion are added and wells containing bound scFv-phage are visualised by anti-M13-HRP.

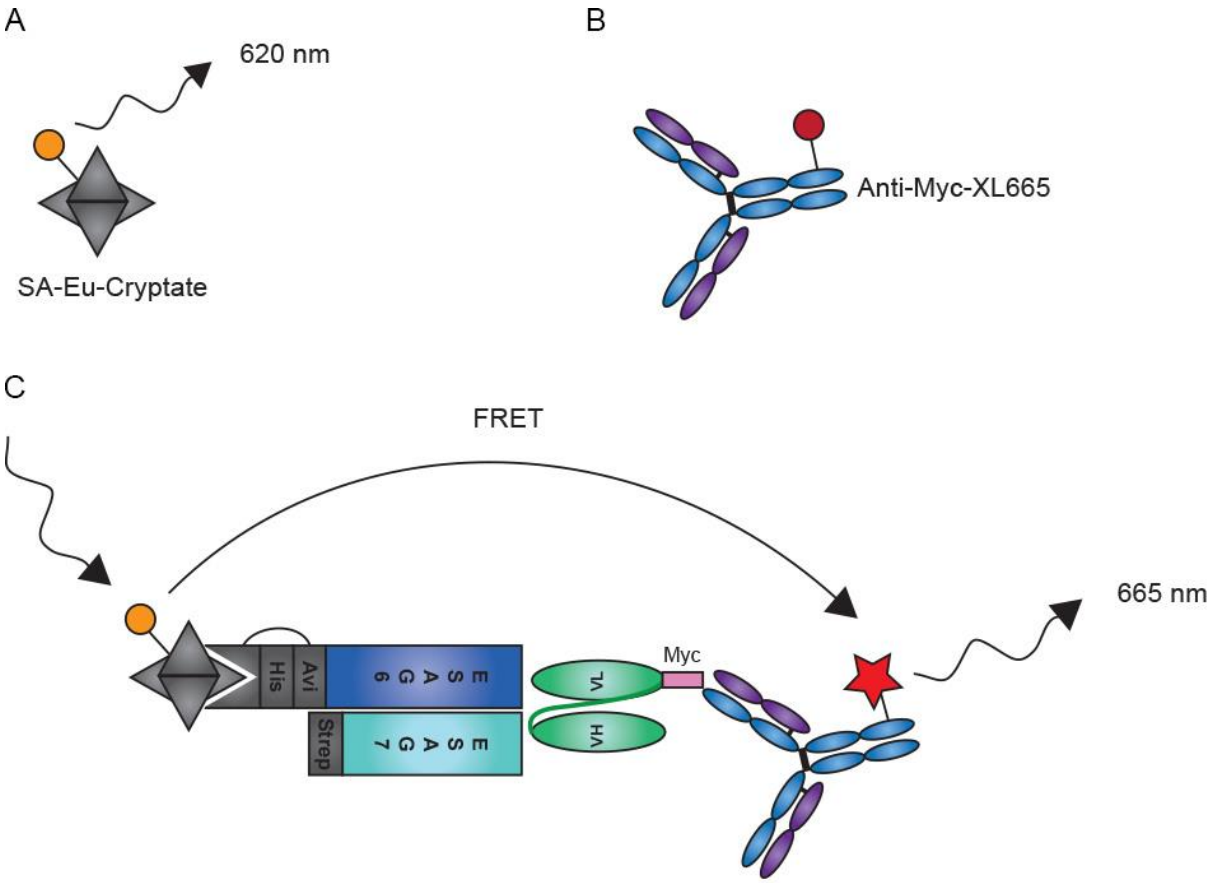


Figure 45: HTRF schematic. A, Europium-cryptate, a fluorescent donor that emits at 620 nm, is conjugated to streptavidin. B, XL665, an acceptor fluorophore, is conjugated to an anti-Myc-tag antibody. C, Anti-Myc-XL665 binds scFv via the Myc-tag while SA-Eu-cryptate is complexed with biotinylated target antigen. Proximity-induced energy transfer between Eu-cryptate and XL-665 is achieved upon scFv binding to target antigen, resulting in fluorescence emission at 665 nm.

6.2 Aims

The aim of the phage display screen was to isolate antibodies targeting the trypanosome transferrin receptor (TfR), and answer several questions surrounding the development of TfR monoclonal antibodies (mAbs):

1. Can a human-derived scFv library be enriched for trypanosome TfR binders?

2. Can a glycan-rich protein be targeted?

3. Would *in vitro* binders function in the context of a dense VSG coat?

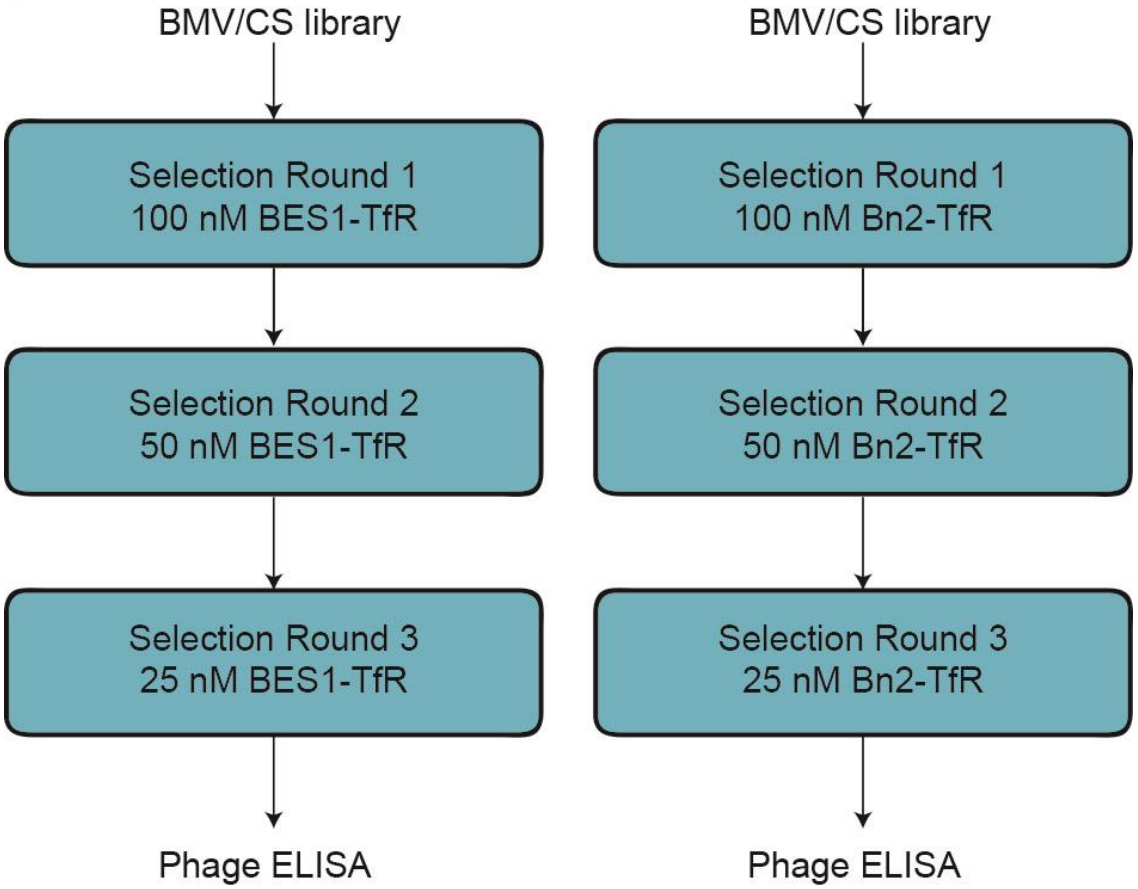
To answer these questions, phage display against BES1-TfR and Bn2-TfR was performed and scFv-binders were further triaged and characterised by homogenous time-resolved fluorescence (HTRF), SPR and trypanosome uptake assays.

6.3 Results

6.3.1 Enrichment of BMV/CS library for TfR binders

Three serial rounds of soluble selections were performed by addition of the BMV/CS library to biotinylated BES1-TfR or biotinylated Bn2-TfR. Phage binders were captured with streptavidin beads and, following phage infection of *E. coli* TG1 cells, the library output size could be estimated based on the number of TG1 colony forming units (cfu) while the diversity was assessed via sequencing of the complimentary determining region 3 (CDR3). The diversity typically begins to fall as the output increases, indicating enrichment of a subset of high affinity binders. The antigen concentration is decreased during each round of phage display to drive the selection of high affinity binders (Figure 46A).

A



B

	BES1-TfR output (cfu/ml)	Bn2-TfR output (cfu/ml)
Round 1	$6x10^3$	$1.4x10^3$
Round 2	$2.2x10^7$	$4.6x10^7$
Round 3	$8.4x10^7$	$8x10^7$

C

	BES1-TfR output diversity	Bn2-TfR output diversity
Round 3	44%	53%

Figure 46: Phage display soluble selection outputs and diversity. A, Workflow highlighting the selection strategy for each round of selection. B, Table indicating the output library size following each round, determined by quantification of *E. coli* colony forming units. C, Diversity of the output library after three rounds of selection.

The output figures revealed a low round 1 output with an unusually large enrichment after round 2 (Figure 46B). The low round 1 output could be related to the presence of glycans on the receptors, resulting in a reduced interaction interface for the scFv-phage due to steric hinderance and impaired accessibility. Nevertheless, glycan binders may exist, but the origin of the library must be considered. As the library was generated from human lymphocytes, many glycan binders could potentially be screened and removed during the B-cell education process to avoid self-reactivity. The BES1-TfR and Bn2-TfR were expressed in CHO cells which display similar glycans to humans. It is therefore unlikely that a human-derived library would be rich in mammalian glycan binders.

A considerable enrichment was observed after round 2 which, although usual, may indicate a plummet in diversity and propagation of dominant scFv-phage within the population. Indeed, analysis of the CDR3 sequences after round 3 indicated a low diversity, with 53% of the Bn2-TfR output population representing unique scFvs, and less than half of the population for BES1-TfR (Figure 46C). The library outputs show marginal increases between round 2 and round 3 for both TfR variants (Figure 46B). Stabilisation of the library output size and the concomitant reduction in diversity could be indicative of a saturated effort in library enrichment and thus a fourth round of selection was not performed.

6.3.2 Phage ELISA reveals variant specificity and a lack of cross-reactivity

Phage enriched for BES1-TfR binders, referred to as the BES1-TfR output, were screened by phage ELISA for binding interactions against the three immobilised targets: BES1-TfR, streptavidin and Bn2-TfR. The Bn2-TfR output would be analysed in an identical screen.

To ascertain the number of hits in each screen, the hit threshold was determined. The average signal obtained from the negative control containing phage targeting an irrelevant antigen, carcinoembryonic antigen (CEA), was used as a starting point. A hit threshold was determined as three-fold the negative control signal. ELISA response signals above the hit threshold were considered as positive hits. The BES1-TfR output appeared to be selective for BES1-TfR binding. All phage clones gave a positive binding response to immobilised BES1-TfR and showed negligible binding to the streptavidin control plate (Figure 47A). Furthermore, the BES1-TfR output showed little interaction with immobilised Bn2-TfR, with only two scFv-phage clones producing a response above the hit threshold (Figure 47C). On the other hand, the

Bn2-TfR output appears to have produced more cross-reactive binders (Figure 47D) in addition to specific binding to the enrichment target (Figure 47B). The larger number of cross-reactive hits generated from the Bn2-TfR library may correlate with the greater diversity of the enriched output, as the diverse scFvs may bind a variety of epitopes, increasing the chances of targeting shared epitopes.

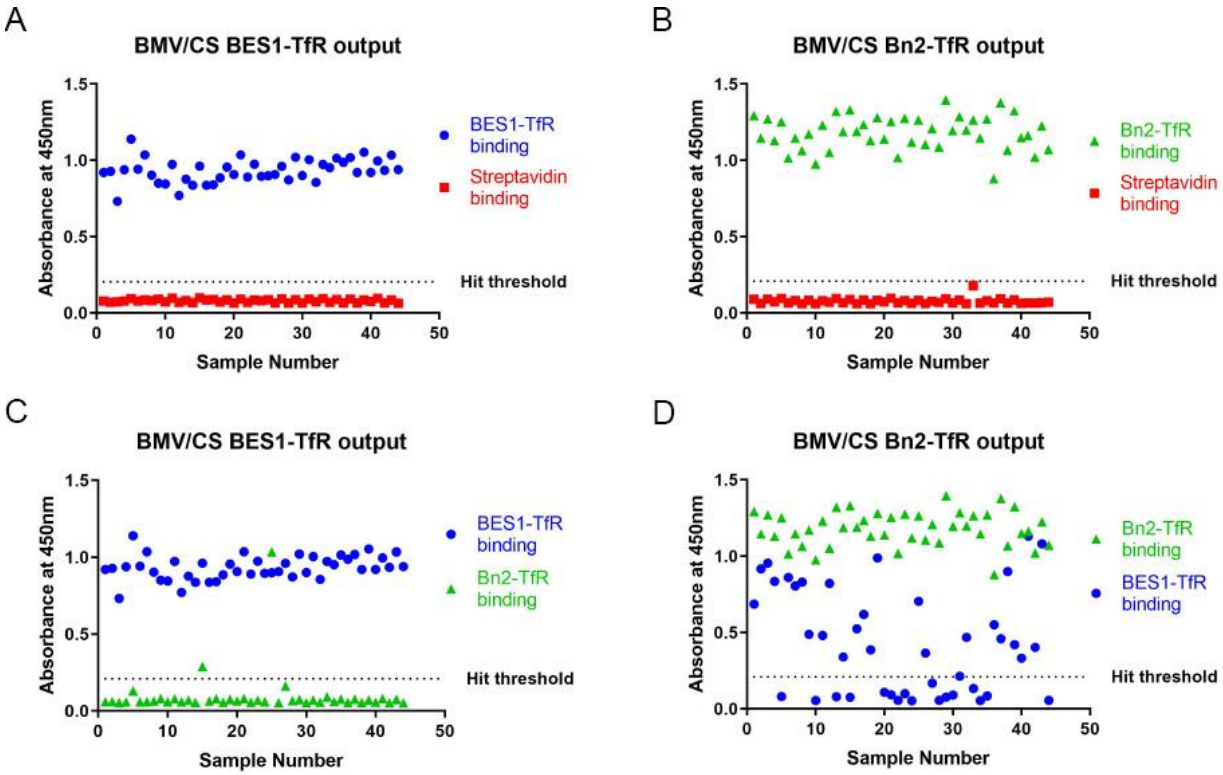


Figure 47: Visualisation of binding by phage ELISA reveals variant specificity and infrequent cross-reactivity. A, The BMV/CS library enriched for BES1-TfR binders was screened for binding against immobilised BES1-TfR and streptavidin. B, The BMV/CS library output selected against Bn2-TfR was tested for Bn2-TfR and streptavidin binding. C, Cross-reactivity was investigated by testing binding of the BES1-TfR output library to Bn2-TfR. D, The Bn2-TfR output library was screened for cross-reactive binding to BES1-TfR.

6.3.3 Output triage by HTRF

Target-specific binding was determined by the following equations:

$$Ratio = \frac{Emission\ at\ 665\ nm}{Emission\ at\ 620\ nm}\ \text{x}\ 10^4$$

$$\%\Delta F = \frac{Ratio_{sample} - Ratio_{background}}{Ratio_{background}}\ \text{x}100$$

The background ratio is determined by averaging the ratios produced by a negative control anti-CEA scFv against the target of interest. Thus, ΔF is a measure of specific binding obtained after the removal of non-specific noise.

The specificity of the phage selection screen was apparent from the segregation of binders depending on the library origin (Figure 48). Samples 1 - 44 represent the scFv library targeting BES1-TfR while samples 45 - 88 contain scFv raised against Bn2-TfR. The results corroborate previous data from the phage ELISA, revealing a high degree of specificity for each target and a lack of cross-reactivity. ELISA data showed a higher number of cross-reactive hits derived from the Bn2-TfR library which, coupled with the greater diversity of phage, suggested a variety of scFv targeting conserved epitopes. However, the cross-binding hits showed a weaker HTRF signal which appears to be sub-detection in HTRF as fewer hits are obtained, highlighting the extent of signal amplification in phage ELISA. The increased number of cross-reactive hits from the Bn2-TfR library is still visible although less marked.

Although HTRF provides a high-throughput screening method, quantitation can be unreliable due to variation in *E. coli* density, scFv expression and periplasm extraction between samples. The results provided a platform for sample triage to select for several types: scFvs that bound both receptors and showed promising cross-reactivity, and scFvs that produced a unique binding response to BES1-TfR or Bn2-TfR alone. Based on these criteria, six clones were further investigated by SPR using immobilised receptor to directly measure the affinities of variable antibody fragments.

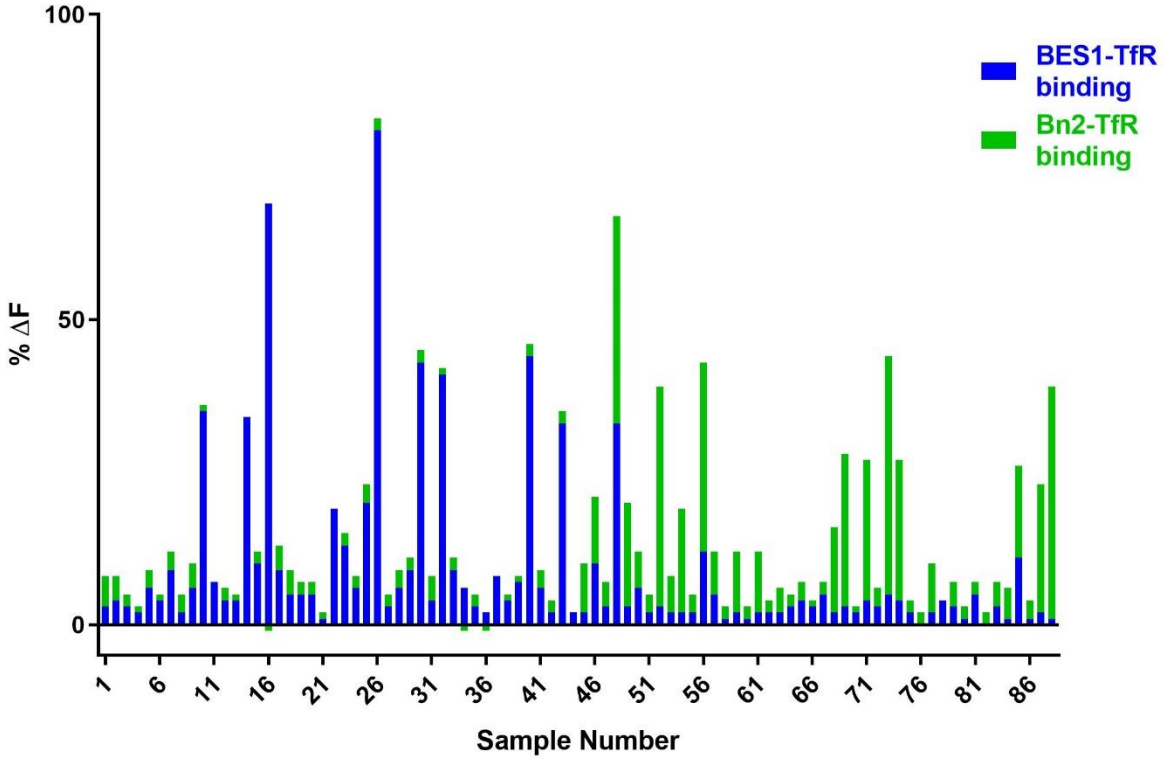


Figure 48: Detection of scFv binding to target antigen by HTRF confirms low cross-reactivity. The %ΔF indicating the binding response of each scFv with BES1-TfR and Bn2-TfR is represented as a stacked bar chart. Samples 1 – 44 inclusive contain scFv raised against BES1-TfR while samples 45 – 88 were raised against Bn2-TfR.

6.3.4 Kinetic analysis

6.3.4.1 Fab conversion

The choice of fragment framework was informed by several factors. First, the requirement of a single juxtaposed V_H and V_L domain as opposed to the bivalent format offered by IgG and $F(ab')_2$ (Figure 1A and B). Bivalency promotes avidity, thus making accurate kinetic measurements between a single variable domain and an epitope challenging. Fab fragments (Figure 41C) and scFv (Figure 41D) offer single domain formats which would enable affinity determination. Previous observations from fellow lab members revealed that scFv have the propensity to aggregate by forming intermolecular interactions in addition to intramolecular interactions, resulting in oligomeric configurations. On the other hand, Fab fragments retain a conserved domain from each chain and disulphide bonding maintains pairing between the light and heavy chain. Variable domains are cloned into a Fab framework, expressed in CHO cells and purified for SPR.

6.3.4.2 Kinetic measurements by SPR

The binding of Fabs to immobilised BES1-TfR and Bn2-TfR was measured by SPR (Figure 49). The majority of the purified Fabs tested showed little to no binding to BES1-TfR. Tbru104, 109 and 110 did not produce a binding response, while Tbru10 showed a negligible response. Two Fabs, Tbru108 and Tbru111, produced a significant response indicative of binding to BES1-TfR. However, the determined kinetic values are low due to the rapid off rates, with an estimated K_D of 1 µM for Tbru108 and 1.4 µM for Tbru111 (Table 5). The Fabs showed a greater binding to Bn2-TfR, with the exception of Tbru104 and Tbru105 which did not produce a binding response. Tbru108, Tbru109, Tbru110 and Tbru111 Fabs bound Bn2-TfR. The highest affinities were observed with Tbru108 and Tbru111 (Table 5), suggesting the potential for a pan-specific TfR binding. Nip228 is a negative control antibody commonly used at AstraZeneca, targeting 4-hydroxy-3-iodo-5-nitrophenylacetic acid (Haqqani *et al*., 2018). Fab fragments of Nip228 did not interact with BES1-TfR although low levels of negligible binding were observed with Bn2-TfR. Human holo-Tf was used as a positive control and produced the previously observed low affinity binding to BES1-TfR and high affinity binding to Bn2-TfR. A positive binding response from Tbru108 and Tbru111 to both TfR variants warranted further investigation by conversion into fluorescently-labelled IgGs to visualise binding and uptake in trypanosomes.

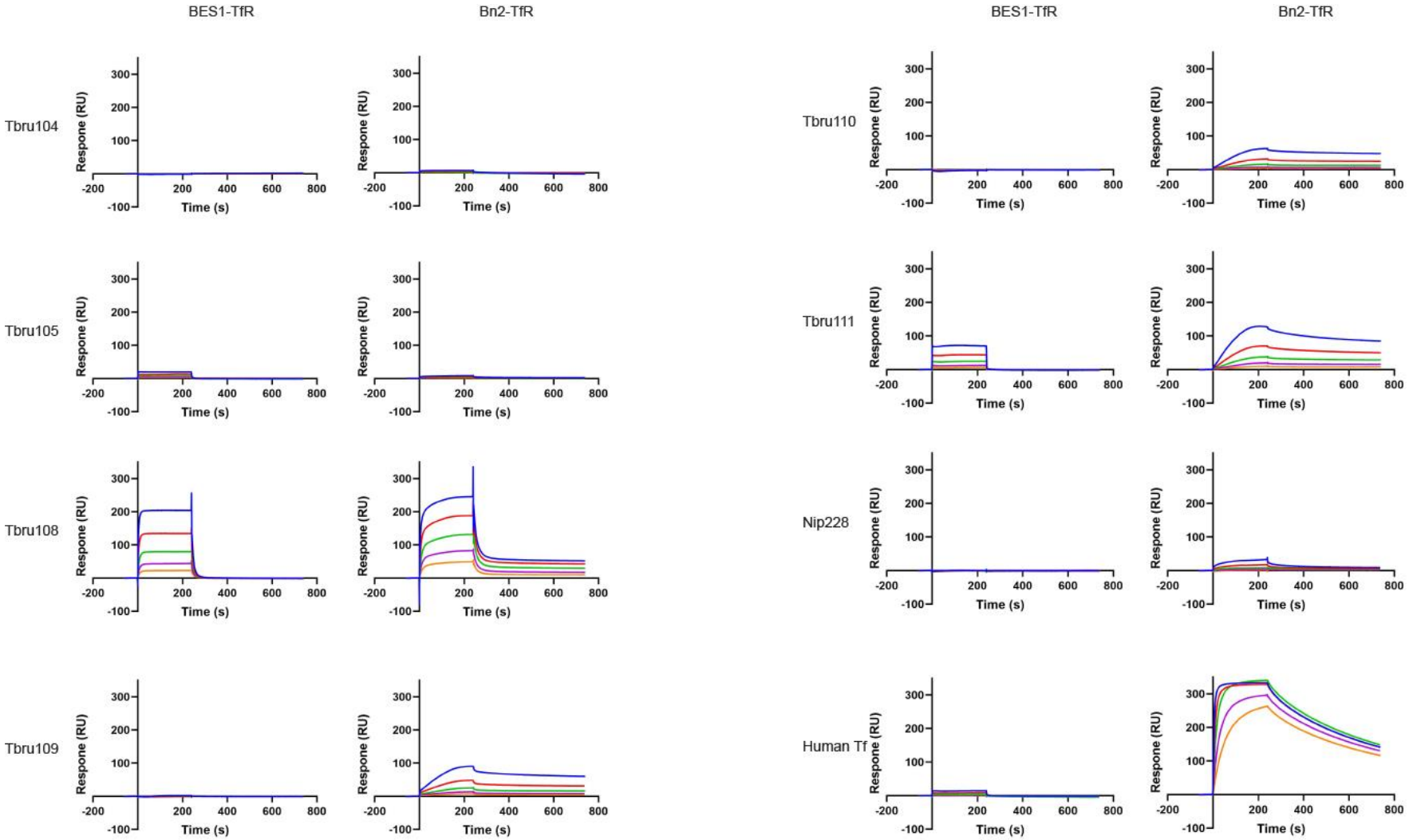


Figure 49: Measurement of Fab binding to target antigen by SPR. Biotinylated BES1-TfR and Bn2-TfR were immobilised and purified Fabs were injected in a concentration series ranging from 1000 nM to 62.5 nM.

Table 5: SPR analysis of binding kinetics**.** Kinetic values for interactions between soluble Fabs and immobilised BES1-TfR or Bn2-TfR. Kinetic data from Fabs that produced a binding response were reported; data for low binding responses were not able to be determined.

TfR	Transferrin	K_D(nM)	k_{on}($M^{-1}s^{-1}$)	k_{off}(s^{-1})
BES1	Tbru108	1000	2.2×10^{5}	2.2×10^{-1}
BES1	Tbru111	1400	7.4×10^{5}	1.0
Bn2	Tbru108	110	1.7×10^{4}	1.8×10^{-3}
Bn2	Tbru109	610	8.2×10^{2}	5.0×10^{-4}
Bn2	Tbru110	800	4.8×10^{2}	3.8×10^{-4}
Bn2	Tbru111	230	3.1×10^{3}	7.2×10^{-4}

6.3.5 Testing internalisation of anti-TfR mAbs in trypanosomes

To investigate IgG binding to TfR in live trypanosomes, antibodies were prepared and used in uptake assays. Variable regions were reformatted into an IgG1 framework and fluorescently-labelled (AlexaFluor 488). Antibodies were tested for BES1-TfR binding and uptake by adding mAbs to trypanosomes expressing from BES1. Assays were performed in serum-free conditions to avoid competition of serum Tf. The addition of FMK, a lysosomal protease inhibitor, allowed accumulation of endocytosed protein. Internalisation of Tbru108 and Tbru111 was compared to fluorescently-labelled bovine Tf uptake and the negative control antibody Nip228 (Figure 50). While Tf was internalised, forming a bright spot in the lysosome, mAb binding and uptake was not apparent. Neither Tbru108 nor Tbru111 showed accumulation within the cell and Nip228 appeared to produce a greater signal inside the cell under the same conditions.

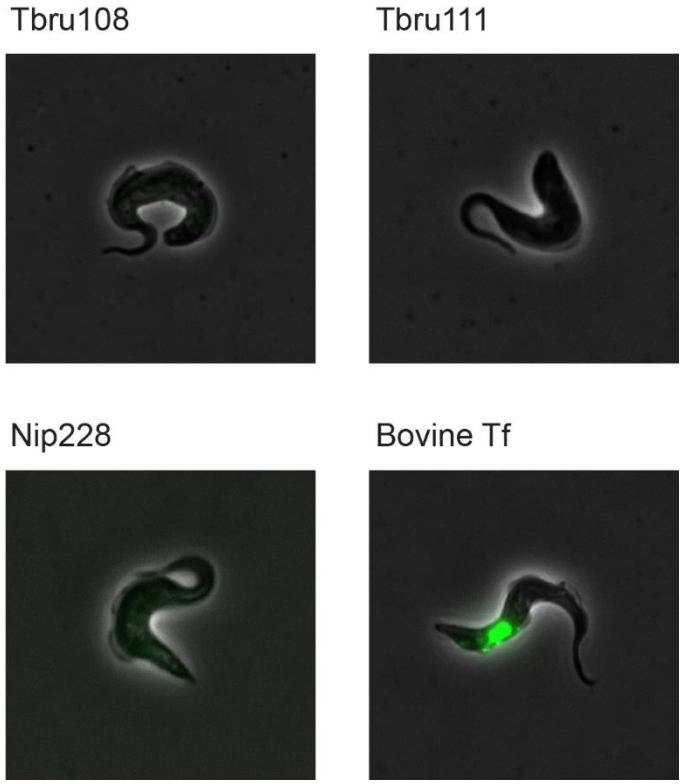


Figure 50: Testing internalisation of anti-TfR mAbs. Fluorescently-labelled (AlexaFluor 488) mAbs (Tbru108 and Tbru111) were added to BES1 trypanosomes in serum-free medium. Uptake of a negative control antibody (Nip228) and bovine Tf was also visualised.

6.4 Discussion

6.4.1 TfR as a therapeutic target

The screen has provided further insights into the feasibility of TfR targeting as a successful therapeutic strategy. Considering the high concentration of mAb tested (500 nM), it was evident that mAbs generated here did not bind BES1-TfR in their current format. It is possible that the mAbs may bind a TfR from another expression site and further testing would be required. Low output yields obtained in the first round of selection are likely a product of heavily and heterogeneously glycosylated antigens which may also account for the low binding response of Fabs observed by SPR. The lack of binding and uptake in trypanosomes may be due to minor differences in glycosylation profile and composition between CHO cells and trypanosomes. In addition, while an IgG format increases avidity through bivalency, the larger structure of an IgG may encounter accessibility constraints depending on the location of the epitope.

In light of the high degree of sequence conservation amongst trypanosome TfRs, the lack of cross-reactive BES1-TfR and Bn2-TfR binders was the most unexpected outcome. The majority of enriched scFv were receptor-specific and only a discrete population was cross-reactive. Our previous data have shown that conservation is predominantly located in the membrane-proximal region of the receptor. Following scFv-phage binding, the immobilisation and stringent washing of C-terminally biotinylated receptors on streptavidin beads may have resulted in loss of phage bound to the C-terminus due to steric hinderance caused by crowding at the surface of the beads. Thus, binders located towards the tip of the receptor would be enriched. This coincides with the most polymorphic region of the receptor in which BES1-TfR and Bn2-TfR show the greatest disparity.

Furthermore, the TfR is rich in glycans, with eight predicted N-linked glycosylation sites per heterodimer. Due to the inherently variable nature of glycans, it is possible that differences in glycosylation profile and composition between BES1-TfR and Bn2-TfR may lead to differences in scFv binding. Moreover, while the structure of BES1-TfR is yet to be solved, it is plausible that subtle structural differences between variants may confer a loss of cross-reactivity. Finally, the shortcomings of the phage display screen provided further evidence to support the hypothesis of TfR diversification driven by an immune avoidance strategy. The lack of exposed conserved epitopes, coupled with glycan shielding and C-terminal receptor packing to protect less diverse regions, are all likely to contribute to immunoglobulin inaccessibility.

6.4.2 Future perspectives

Several strategies could be explored to improve the selection output library, cross-reactivity and affinities. First, removal of glycans prior to phage display selections could be considered. The TbHpHbR was expressed in *E. coli* resulting in the absence of glycosylations during the phage display trials, without impinging on IgG binding to native, glycosylated HpHbR (MacGregor, Gonzalez-Munoz *et al.*, 2019). A crossover of the output libraries during the selection process to raise an output concomitantly enriched against both antigens would further improve cross-reactivity of the selection library. *In vitro* affinity maturation could be performed to increase the affinities and cross-reactivity of the output library. A third-round output enriched for antigen binders can be randomly mutated in the V_H and V_L using error-prone PCR to drive the production of higher affinity binders. Finally, a library formatted on a smaller framework such as a small fragment peptide library could be tested to achieve greater accessibility to conserved epitopes. However, the structure of the receptor has since revealed that the TfR is a complex target. The lateral surfaces are protected by the VSG coat and glycan shielding, likely rendering the conserved core inaccessible to immunoglobulins (Figure 51). In addition, ligand occlusion of the receptor by apo- and holo-Tf, coupled with diversification of the regions surrounding the Tf-binding site, will contribute to difficulties in the production of a viable pan-specific antibody therapeutic targeting the TfR.

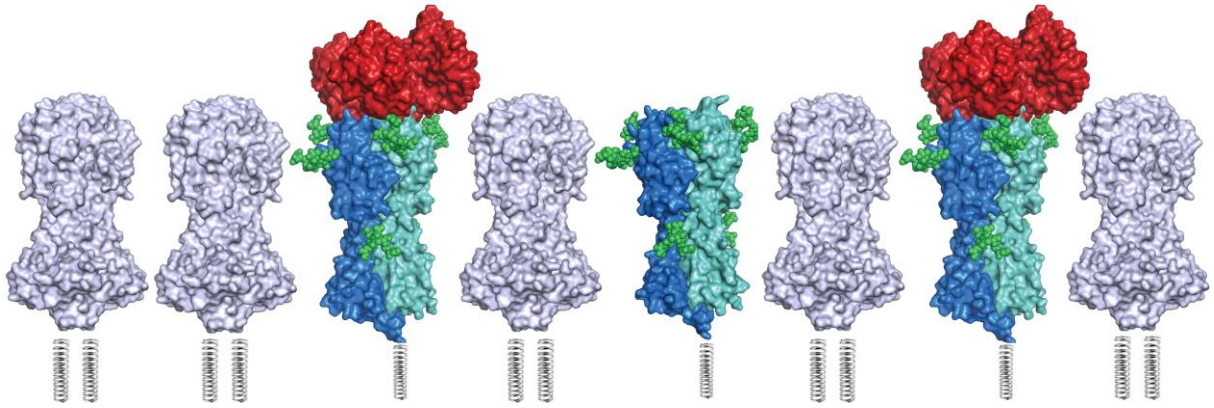


Figure 51: Model of the *T. brucei* surface coat. Surface representation of MITat1.2 (PDB = 1VSG) homodimers (grey) with TfR heterodimers in blue and cyan. C-terminal domains and GPI-anchors are represented as springs to represent flexibility. Glycans are modelled in green and Tf is represented in red. The VSG coat and glycans shield lateral surfaces of the TfR, while the residues at the tip of the receptor are engaged in ligand binding or have diversified, thus complicating the development of a pan-specific TfR therapeutic.

Chapter 7

General discussion

More than a century since the first reports on what we now know to be antigenic variation in trypanosomes (Franke, 1905), progress is still being made to further our understanding of the strategies deployed to avoid host detection whilst responding to nutrient requirements. In this work, the findings can be summarised and further explored through the following questions.

7.1 How do trypanosomes acquire iron whilst avoiding host recognition?

The trypanosome is an elegant example of an extracellular pathogen that has developed multiple strategies to avoid host immune recognition. In contrast to other kinetoplastids (*T. cruzi* and *Leishmania* spp.) which have a coat rich in glycoconjugates, the *T. brucei* coat is predominantly protein (reviewed by Ferguson, 1997). As the coat assumes the role of the host-parasite interface, evolutionary adaptations in the coat to avoid host immune detection have shaped the genome. While the VSG is central to antigenic variation, the findings presented in this thesis suggest that TfR variation may play a supporting role by forming an additional degree of variation.

Trypanosomes were considered to rely primarily on single copy receptors to bind host macromolecules, such as the haptoglobin-haemoglobin receptor (HpHbR) to fulfil haem requirements (Vanhollebeke *et al*., 2008) and the factor H receptor (FHR) to mediate protection from the host innate immune response and promote insect vector transmission (Macleod *et*

al., 2020). This raises the interesting question of whether single copy receptors elicit an immune response. In the case of FHR, expression is low on proliferating bloodstream-form cells (< 200 copies per cell) (Macleod *et al*., 2020), which may confer protection. HpHbR is also expressed at low levels (200 to 400 copies per cell), yet the receptor has elicited an innate immune response through targeting by trypanosome lytic factor (TLF) (Vanhollebeke *et al*., 2008) and can be targeted by an immunoglobulin-based therapy (MacGregor *et al.*, 2019), suggesting immunogenicity. Likely in a bid to combat immunogenicity, the trypanosome TfR is peculiar as it has evolved a family of receptors for iron acquisition. Understanding how trypanosomes orchestrate immune avoidance whilst maintaining a functional Tf-binding site was a key focus of the research in this thesis.

Based on the structure of the receptor and mapping of sequence diversification, it was evident that variation was selective due to the preponderance of apical polymorphisms. Further experiments produced a number of salient findings relating to the origin of selective variation. In the first instance, biophysical analyses complemented with trypanosome growth experiments confirmed that high affinity binding of Tf was not necessary for cell proliferation. Moreover, the structure of the receptor-ligand complex and investigation of the polymorphic regions revealed that ligand binding may not be the determinant for selective variation as the binding site was predominantly conserved. In addition, point mutation of one to four residues in BES1-TfR produced a considerable increase in Tf binding affinity, suggesting that Tf affinity was not the factor driving receptor evolution. Finally, the positioning of glycans to shield conserved epitopes suggested that the TfR may be a target of immunoglobulin-mediated attack. Thus, a picture emerged in which variation was likely derived from antigenic drift and restricted by a requirement for maintenance of Tf binding.

Investigation of the receptor-ligand complex in different pH conditions revealed Tf remained bound to the receptor at lysosomal pH. Moreover, iron was likely to be sequestered in the complex during endosomal progression to the lysosome. This finding is compatible with reports that a trypanosome iron transporter protein is located in proximity to the lysosome (Taylor *et al*., 2013). The TfR has been shown to undergo recycling, with conflicting reports on TfR half-life ranging from 1.5 hours (Schwartz *et al*., 2005) (Tiengwe *et al*., 2017) to 7 hours (Kabiri & Steverding, 2000). Further work is required to fully understand how iron is released from the complex and the fate of Tf and TfR thereafter.

7.2 Did the VSG evolve from a TfR?

Structural determination of the trypanosome TfR revealed striking similarities between the receptor and type A VSGs. While it is highly likely that the VSG and TfR evolved from a common ancestor, an interesting question in precedence arose. Evolutionary analyses suggest that an ancestral procyclic form trypanosome residing in the insect vector predates the bloodstream-form trypanosome (Lake *et al*., 1988). Iron requirements for procyclic form cells exceed those of bloodstream-form cells (Stijlemans *et al*., 2015). Thus, it is plausible that the ancestral parasite had greater requirements for iron during the early stages of adaptation to the mammalian host. Over time, bloodstream-form cells have evolved to require less iron whilst developing more advanced immune evasion strategies in the form of the VSG coat.

How have the TfR and VSG evolved since divergence? Much like the VSG, evolution of the TfR is centred around diversification of immunogenic surfaces. For the TfR, variation stringency is imposed by functional requirements to maintain Tf binding. This is largely reminiscent of VSG selection, in which diversification whilst preserving the core structure to maintain VSG coat function has shaped evolution of the VSG repertoire. However, one interesting distinction was observed. Diversity in the core three-helix bundle is less pronounced in the TfR compared to the VSG, which prompted the question of why the TfR core was less variable. Conservation mediated by core structural constraints seemed implausible as sequence diversity is compatible with three-helix bundle conservation (Blum *et al*., 1993) (Carrington & Boothroyd, 1996), and studies have shown that an N-terminal VSG core fusion with a TfR binding site supports Tf binding (Salmon *et al*., 1997). A conceivable explanation is that the conserved core may be testament to the shielding capacities of ESAG6 and ESAG7 glycans. The ring of five sugars at the periphery of the Tf-binding site favours occlusion of the three-helix core, which may have reduced the requirement for variation. In contrast, type A VSGs were shown to have one or two glycans per monomer in most cases, with a location predominantly towards the membrane proximal base of the VSG (Mehlert *et al*., 2002). Thus, diversity within the core fold of the VSG may be required to protect against an adaptive immune response in the absence of glycan shielding.

7.3 Is TfR-mediated endocytosis of host Tf essential?

Investigation of TfR-Tf interactions revealed two interesting findings. First, a high affinity receptor is not essential for trypanosome growth. Second, TfR can bind both apo- and holo-Tf at extracellular pH. If Tf were critical for fulfilment of iron requirements, the expectation is that evolution of the TfR repertoire would be centred around the production of high affinity receptors or, at the very least, receptors that discriminate between apo- and holo-Tf for selective uptake of iron-laden Tf. The evidence produced here suggested that TfR-Tf interactions may not be essential and the hypothesis was tested. Trypanosomes cultured in Tf-depleted medium did not exhibit a growth phenotype, and the surprising findings warrant further investigation to address several questions:

1. Do trypanosomes rely on uptake of lactoferrin or another iron transport protein to fulfil iron requirements in the absence of Tf?

2. Why is TfR knockdown lethal and does the TfR have an additional function?

3. Does an alternative iron uptake pathway exist?

4. Are TfRs or any other proteins upregulated in response to Tf depletion?

The lack of discrimination between apo- and holo-Tf may be explained by the advantage of TfR occupancy, as ligand-bound receptor would provide protection against an adaptive immune response. Numerous findings in this thesis suggest that immune avoidance is paramount for host persistence and surpasses trypanosome requirements for iron during TfR evolutionary selection.

In conclusion, the work described in this thesis has elucidated one of the complex interactions that occurs between host and parasite during infection, offering several key findings. First, the future of an antibody therapy targeting the TfR is uncertain and trypanosome efficiency in immunoglobulin avoidance may explain the challenges in TfR targeting. Second, trypanosomes have a broad host range, yet host promiscuity does not appear to be the selective pressure for diversity within the TfR repertoire. Finally, evolution of the TfR is likely selected by an advantage conferred in immune avoidance to promote long-term persistence in the host. These findings will provide valuable insights for the design of future therapeutics

and will further our understanding of the molecular architecture of the trypanosome cell surface.

Appendices

Appendix 1: List of GeneArt gene constructs. Amino acid sequences of the expression constructs, containing the CD33 leader sequence. **Linkers**, AviTag and **His-tag** are present in ESAG6, and residues mutated in BES1 are shown in **pink**. ESAG7 contains **linker** and **StrepII-tag**, with location of primers used for RT-PCR marked in BES1.

BES1 ESAG6

MPLLLLLPLLWAGALAAAYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVGT
NIPAEYNNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDTDG
VLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEVM
NKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVAG
S**GSGSGSASG**GLNDIFEAQKIEWHE**GG**HHHHHHHHHH

BES1 ESAG7

MPLLLLLPLLWAGALAAAYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVGT
NIPAEYNNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQRKNRGCNLMRTADGG
LLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMN
KIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGS**GSGSGS**WSHPQFEK

Bn2 ESAG6

MPLLLLLPLLWAGALAAAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISVGK
TGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTSLQLHEVLSSISFGSLGVKNIRGGNGKDGCNLVRTDTDG
VLEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVV
KKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAG
A**GSGSGSASG**GLNDIFEAQKIEWHE**GG**HHHHHHHHHH

Bn2 ESAG7

MPLLLLLPLLWAGALAAAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVGT
NIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEKGENRGCNLMRTADGG
LLKDVCLNRNFTWGGGVLNFGYCVAGNLKIKGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEVKNAVVK
KIKTTVDELTKCIGQKEAELTNDQLYEEFEVIQKYLWFL**GSGSGS**WSHPQFEK

Appendix 2: List of GeneArt gene constructs. DNA sequences of the expression constructs used in this work.

BES1 ESAG6

ATGCCTCTGCTGCTGCTGCTGCCTCTGCTGTGGGCTGGCGCTCTGGCGGCCGCATACGAGAACAAGCGGAACGCC
CTGAATGCCACCGCCGCCAACAAAGTGTGCGGCCTGAGCACCTACCTGAAGGGAATCGCCCACAGAGTGAACAGC
GAGAGCGCCGTCGTGACCGAGAAGCTGAGCGACCTGAAGATGCGGAGCATCCAGCTGCAGCTGAGCGTGATGCGG
AACAGAGTGCCTAGCGGCGAGCAGGACTGCAAGGACATCCGGACCCTGCTGAAAACCGTGCTGAGAAACGAGTTC
ACCTTCCAGCAGGAACTGGAAGAGATGAGAAACGCCAGCGCCCTGGCCGCTGCCGCTGCTGGAATTGCTGCTGGC
AGACTGGAAGAATGGATCTTCGTGTTCGCCCAGGCCGCAGGCGGCAGCTCCCAGTTCTGTATCAGCGTGGGCACC
AACATCCCCGCCGAGTACAACAACCTGCAGGAATGCTTCGACGGCATCATCGGCCCCGAGACACTGTACAAGATC
GAGGACAGCAGAGTGAAAGAGTCCGCCCAGAAGTCCCTGCAGCTGCACGAGGTGCTGAGCAGCATCAGCTTCAAC
AGCCTGGGCGCCGAGAACATCAGAGGCGGCAATGGCAGAGATGGCTGCAACCTCGTGCGGACCGATACCGACGGC
GTGCTGGAAGGCGGATCTGTGCGGAGACACAACCTGACATGGGGCGGAGGCGTGATGAACTTCGGCAGCTACCAG
AACGGCTCTATGTACGTGGAAGGGGGCGAGTACGGCGACGCCACAGAATATGGCGCTGTGCGGTGGACCGAGGAC
CCCAGCAAGGTGTCCATCTTCAAGGACGTGATCCGGCTGTTCGCCCGGTTCCAGGAAGCCAAGAACGAAGTGATG
AACAAGATCAAGACCACCGTGGACGAGCTGGCCAAGTGCATCGGCCAGAAAGAGGTGGAACTGACCGACGACCAG
CTGTACGAGGAATTCATCTGGGAGACAATCCACCGGCTGGAACTGAGCAAGCGGGTGTCCGAGCAGCTGTCCCTG
GGCGAGGAAGAGGAAACCATCCTGAAGTCCAACTACACCGCCGAGCCCGTGCGGGGACCTTTTACAGTGGCCGGC
TCTGGAAGCGGCTCTGGCTCTGCTAGCGGCGGACTGAACGACATCTTCGAGGCCCAGAAAATCGAGTGGCACGAG
GGCGGACACCACCACCATCACCATCATCACCACCATTGA

BES1 ESAG7

ATGCCTCTGCTGCTGCTGCTGCCTCTGCTGTGGGCTGGCGCTCTGGCGGCCGCATACGAGAACAAGCGGAACGCC
CTGAATGCCACCGCCGCCAACAAAGTGTGCGGCCTGAGCACCTACCTGAAGGGAATCGCCCACAGAGTGAACAGC
GAGAGCGCCGTCGTGACCGAGAAGCTGAGCGACCTGAAGATGCGGAGCATCCAGCTGCAGCTGAGCGTGATGCGG
AACAGAGTGCCTAGCGGCGAGCAGGACTGCAAGGACATCCGGACCCTGCTGAAAACCGTGCTGAGAAACGAGTTC
ACCTTCCAGCAGGAACTGGAAGAGATGAGAAACGCCAGCGCCCTGGCCGCTGCCGCTGCTGGAATTGCTGCTGGC
AGACTGGAAGAATGGATCTTCGTGTTCGCCCAGGCCGCAGGCGGCAGCTCCCAGTTCTGTATCAGCGTGGGCACC
AACATCCCCGCCGAGTACAACAACCTGCAGGAATGCTTCGACGGCATCATCGGCCCCGAGACACTGTACAAGATC
GAGGACAGCAGAGTGAAAGAGTCCGCCCAGAAGTCCCTGCAGCTGCACGAGGTGCTGAGCAGCATCAGCTTCAGC
AGCCTGGGCGCCGAGAGCATCGTGGAACAGCGGAAGAACCGGGGCTGCAACCTGATGAGAACAGCCGATGGCGGC
CTGCTGAAGGACATCTGCCTGAACTGCAACTTCACCTGGGGCGGAGGCGTGATGAACTTCGGCTCTTGCGTGGCC
GGCAATCTGAAGATCAAGGGCGGCGAGTACGGCGACGTGTCCAGCCATGATGTCGTGCGGTGGACCGAGGACCCC
AGCAAGGTGTCCATCTTCAAGGACGTGATCCGGCTGTTCGCCCGGTTCCAGGAAGCCAAGAACGCTGTGATGAAC
AAGATCAAGACCACCGTGGACGAGCTGGCCAAGTGCATCGGCCAGAAAGAGGTGGAACTGACCAACGACCAGCTG
TACGAGGAATTCGAGGCCATCCAGAAGTACCTGGGCAGCGGCTCTGGCTCCGGATCTTGGAGCCACCCCCAGTTC
GAGAAATGA

Bn2 ESAG6

ATGCCTCTGCTGCTGCTGCTGCCTCTGCTGTGGGCTGGCGCTCTGGCGGCCGCATACGAGAACGAGAGAAACGCC
CTGAATGCCACCGCCGCCAACAAAGTGTGCGGCCTGAGCACCTACCTGAAGGGAATCGCCCACAGAGTGAACAGC
GAGAGCGCCGTCGTGACCGAGAAGCTGAGCGACCTGAAGATGCGGAGCATCCAGCTGCAGCTGAGCGTGATGCGG
AACAGAGTGCCTAGCGGCGAGCAGGACTGCAAGGACATCCGGACCCTGCTGAAAACCGTGCTGAGAAACGAGTTC
ACCTTCCAGCAGGAACTGGAAGAGATGAGGAACGCCTCTGCCCTGGCCGCTGCTGCTGCAGGACTGGCTGCTGGC
AGACTGGAAGAATGGATCTTCGTGTTCGCCCAGGCCGCTGGCAGAAGCTCCCAGTTCTGTATCAGCGTGGGCAAG
ACCGGCCCTGCCGAGTACAACAACCTGCAGGAATGCTTCGACGGCACCATCGGCCCCGAGACACTGTACAAGATC
GAGGACAGCAGAGTGAAAGAGTCCGCCAAGACCTCCCTGCAGCTGCACGAGGTGCTGAGCAGCATCAGCTTTGGC
AGCCTGGGCGTGAAGAACATCAGAGGCGGCAACGGCAAGGACGGCTGCAATCTCGTGCGGACCGATACCGACGGC
GTGCTGGAAGGCGGCAGCCCCACAAGACACAACCTGACATGGGGCGGAGGCGTGATGAACTTCGGCAGCTACCAG
AACGGCTCTATGTACGTGGAAGGGGGCGAGTACGGCGACGCCACAGAATATGGCGCTGTGCGGTGGACCGAGGAC
CCCAGCAAGGTGTCCATCTTCAAGGACGTGATCCGGCTGTTCGCCCGGTTCCAGGAAGCCAAGAACGCCGTCGTG
AAGAAAATCAAGACCACCGTGGACGAGCTGACCAAGTGCATCGGCCAGAAAGAGGCCGAGCTGACAAACGACCAG
CTGTACGAGGAATTCATCTGGGAGACAATCAACCGGCTGGAACTGAGCAAGCGGGTGTCCGAGCAGAGCGCCTTC
GGCGAAGAGGAAGAGACAATCGTGAAGTTCAACTACACCGCCGAGCCCGTGCGGGGACCTTTTACAGTGGCTGGC
GCCGGATCTGGCAGCGGCTCTGCTAGCGGCGGACTGAACGACATCTTCGAGGCCCAGAAAATCGAGTGGCACGAG
GGCGGACACCACCACCATCACCATCATCACCACCATTGA

Bn2 ESAG7

ATGCCTCTGCTGCTGCTGCTGCCTCTGCTGTGGGCTGGCGCTCTGGCGGCCGCATACGAGAACGAGAGAAACGCC
CTGAATGCCACCGCCGCCAACAAAGTGTGCGGCCTGAGCACCTACCTGAAGGGAATCGCCCACAGAGTGAACAGC
GAGAGCGCCGTCGTGACCGAGAAGCTGAGCGACCTGAAGATGCGGAGCATCCAGCTGCAGCTGAGCGTGATGCGG
AACAGAGTGCCTAGCGGCGAGCAGGACTGCAAGGACATCCGGACCCTGCTGAAAACCGTGCTGAGAAACGAGTTC
ACCTTCCAGCAGGAACTGGAAGAGATGAGGAACGCCTCTGCCCTGGCCGCTGCCGCTGCTGGAATTGCTGCTGGC
AGACTGGAAGAATGGATCTTCGTGTTCGCCCAGGCCGCTGGCGGCAGCTCCCAGTTCTGTATCAGCGTGGGCACC
AACATCCCCGCCGAGTACAACAACCTGCAGGAATGCTTCGACGGCACCATCGGCCCCGAGACACTGTACAAGATC
GAGGACAGCAGAGTGAAAGAGTCCGCCCAGAAGTCCCTGCAGCTGCACGAGGTGCTGAGCAGCATCAGCTTCAGC
AGCCTGGGAGCCGAGAGCATCGTGGAAAAGGGCGAGAACCGGGGCTGCAACCTGATGAGAACAGCCGATGGCGGC
CTGCTGAAGGACGTGTGCCTGAACCGGAACTTCACCTGGGGAGGCGGAGTGCTGAACTTCGGCTATTGCGTGGCC
GGCAATCTGAAGATCAAGGGCGGCGAGTATGGCGACGTGGGCTCTCATGATGCCGTGCGGTGGACCGAGGACCCC
AGCAAGGTGTCCATCTTCAAGGACGTGATCCGGCTGTTCGCCCGGTTCCAGGAAGTGAAGAACGCCGTCGTGAAG
AAAATCAAGACCACCGTGGACGAGCTGACCAAGTGCATCGGCCAGAAAGAGGCCGAGCTGACAAACGACCAGCTG
TACGAGGAATTCGAAGTGATCCAGAAATACCTGTGGTTCCTGGGCAGCGGCTCCGGCTCTTGGAGCCACCCTCAG
TTCGAGAAATGA

Appendix 3: Sequence of the Bn2-TfR mutant. Amino acid and DNA sequences of N-linked glycan mutant used for crystallography. The **linkers**, AviTag and **His-tag** are cleavable at the TEV-protease cut site. ESAG7 does not contain affinity tags and N-linked glycosylation sites were mutated to aspartic acid.

Bn2 ESAG6

MPLLLLLPLLWAGALAAA**HHHHHHHHHH**GG**GLNDIFEAQKIEWHE**ASENLYFQGSGYENERNALDATAANKVCGL
STYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRDA
SALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISVGKTGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTS
LQLHEVLSSISFGSLGVKNIRGGNGKDGCNLVRTDTDGVLEGGSPTRHDLTWGGGVMNFGSYQDGSMYVEGGEYG
DATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVVKKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINR
LELSKRVSEQSAFGEEEETIVKFDYTAEPVRGPFTVAGA

ATGCCTCTGCTGCTGCTGCTGCCTCTGCTGTGGGCTGGCGCTCTGGCGGCCGCTCACCACCACCATCACCATCAT
CACCACCATGGCGGCGGACTGAACGACATCTTCGAGGCCCAGAAAATCGAGTGGCACGAGGCTAGCGAGAACCTC
TATTTCCAAGGGTCTGGCTACGAGAACGAGAGAAACGCCCTGAATGCCACCGCCGCCAACAAAGTGTGCGGCCTG
AGCACCTACCTGAAGGGAATCGCCCACAGAGTGAACAGCGAGAGCGCCGTCGTGACCGAGAAGCTGAGCGACCTG
AAGATGCGGAGCATCCAGCTGCAGCTGAGCGTGATGCGGAACAGAGTGCCTAGCGGCGAGCAGGACTGCAAGGAC
ATCCGGACCCTGCTGAAAACCGTGCTGAGAAACGAGTTCACCTTCCAGCAGGAACTGGAAGAGATGAGGAACGCC
TCTGCCCTGGCCGCTGCTGCTGCAGGACTGGCTGCTGGCAGACTGGAAGAATGGATCTTCGTGTTCGCCCAGGCC
GCTGGCAGAAGCTCCCAGTTCTGTATCAGCGTGGGCAAGACCGGCCCTGCCGAGTACAACAACCTGCAGGAATGC
TTCGACGGCACCATCGGCCCCGAGACACTGTACAAGATCGAGGACAGCAGAGTGAAAGAGTCCGCCAAGACCTCC
CTGCAGCTGCACGAGGTGCTGAGCAGCATCAGCTTTGGCAGCCTGGGCGTGAAGAACATCAGAGGCGGCAACGGC
AAGGACGGCTGCAATCTCGTGCGGACCGATACCGACGGCGTGCTGGAAGGCGGCAGCCCCACAAGACACAACCTG
ACATGGGGCGGAGGCGTGATGAACTTCGGCAGCTACCAGAACGGCTCTATGTACGTGGAAGGGGGCGAGTACGGC
GACGCCACAGAATATGGCGCTGTGCGGTGGACCGAGGACCCCAGCAAGGTGTCCATCTTCAAGGACGTGATCCGG
CTGTTCGCCCGGTTCCAGGAAGCCAAGAACGCCGTCGTGAAGAAAATCAAGACCACCGTGGACGAGCTGACCAAG
TGCATCGGCCAGAAAGAGGCCGAGCTGACAAACGACCAGCTGTACGAGGAATTCATCTGGGAGACAATCAACCGG
CTGGAACTGAGCAAGCGGGTGTCCGAGCAGAGCGCCTTCGGCGAAGAGGAAGAGACAATCGTGAAGTTCAACTAC
ACCGCCGAGCCCGTGCGGGGACCTTTTACAGTGGCTGGCGCCTGA

Bn2 ESAG7

MPLLLLLPLLWAGALAAAYENERNALDATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRDASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVGT
NIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEKGENRGCNLMRTADGG
LLKDVCLNRDFTWGGGVLNFGYCVAGNLKIKGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEVKNAVVK
KIKTTVDELTKCIGQKEAELTNDQLYEEFEVIQKYLWFLGS

ATGCCTCTGCTGCTGCTGCTGCCTCTGCTGTGGGCTGGCGCTCTGGCGGCCGCTTACGAGAACGAGAGAAACGCC
CTGAATGCCACCGCCGCCAACAAAGTGTGCGGCCTGAGCACCTACCTGAAGGGAATCGCCCACAGAGTGAACAGC
GAGAGCGCCGTCGTGACCGAGAAGCTGAGCGACCTGAAGATGCGGAGCATCCAGCTGCAGCTGAGCGTGATGCGG
AACAGAGTGCCTAGCGGCGAGCAGGACTGCAAGGACATCCGGACCCTGCTGAAAACCGTGCTGAGAAACGAGTTC
ACCTTCCAGCAGGAACTGGAAGAGATGAGGAACGCCTCTGCCCTGGCCGCTGCCGCTGCTGGAATTGCTGCTGGC
AGACTGGAAGAATGGATCTTCGTGTTCGCCCAGGCCGCTGGCGGCAGCTCCCAGTTCTGTATCAGCGTGGGCACC
AACATCCCCGCCGAGTACAACAACCTGCAGGAATGCTTCGACGGCACCATCGGCCCCGAGACACTGTACAAGATC
GAGGACAGCAGAGTGAAAGAGTCCGCCCAGAAGTCCCTGCAGCTGCACGAGGTGCTGAGCAGCATCAGCTTCAGC
AGCCTGGGAGCCGAGAGCATCGTGGAAAAGGGCGAGAACCGGGGCTGCAACCTGATGAGAACAGCCGATGGCGGC
CTGCTGAAGGACGTGTGCCTGAACCGGAACTTCACCTGGGGAGGCGGAGTGCTGAACTTCGGCTATTGCGTGGCC
GGCAATCTGAAGATCAAGGGCGGCGAGTATGGCGACGTGGGCTCTCATGATGCCGTGCGGTGGACCGAGGACCCC
AGCAAGGTGTCCATCTTCAAGGACGTGATCCGGCTGTTCGCCCGGTTCCAGGAAGTGAAGAACGCCGTCGTGAAG
AAAATCAAGACCACCGTGGACGAGCTGACCAAGTGCATCGGCCAGAAAGAGGCCGAGCTGACAAACGACCAGCTG
TACGAGGAATTCGAAGTGATCCAGAAATACCTGTGGTTCCTGGGCAGCTGA

Appendix 4: List of sequences used for sequence entropy calculations

ESAG6 sequences:

>ACH41858
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRHGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNAAAVHLSVSTAALCFSVLLLGVL

>ACH41863
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNAAAVHLSVSTAALCFSVLLLGVL

>ACH41864
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGHFTVA
GSNAAAVHLSVSTAALCFSVLLLGVL

>ACH41855
MRFLFVLLALLGKKTHAYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAADGSSQFCISVGT
NIPAEHNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGVKNIRGGNGRDGCNLVRTDTNG
ILNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFEDLIRLFARFQEAKNAVM
KKIKTTVDELTKCIGQKEAELTNDQIYEEFIWETINRLELSKRMSEQSAFGEEEETILKSNYTAEPVRGPFTGAG
SNTVAVHLSFSTAALCCSVLLLGVL

>ACH41854
MRFLFVLLALLGKKTHAYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADGSSQFCISVGT
NIPAEHNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGVKNIRGGNGRDGCNLVRTDTNG
ILNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
RRIKTTVDELTKCIGQKEAELTNDQIYEEFIWETINRLELSKRMSEQPTLGEEEETILKSNYTAEPVRGPFTGAG
ANTVALHLSVSTAALCCSVLLLGVL

>ACH41849
MRFWFVLLALLGKETYAYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNTSALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVGK
TGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTSLQLHEVLSSISFGSLGVKNIRGGNGKDGCNLVRTDTDG
VLEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTDDPSKVSIFKDVIRLFARFQEAKNAVM
KKIKTTVDELTKCVGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAG
ANAAAIHLSVSTAALCRSALLLGVL

>ACH41867
MRFWFVLLAVLGKETYAYYENERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISTGK
TGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTRLLLHEVLSSISFGSLGAENIRGGNGKDGCNLVRTDNNG
ILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
KKIKTTVDELTKCIGQKEAELTNDQIYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
ANTVAVQSSVFTAALCCSALLLGVL

>ACH41861
MRFLFVLLALLGKKTHAYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSIMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVGK
TGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEVLSSISFGSLGVKNIRGGNGKDRCNLVRTDTDG
VLEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
KKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAG
ANAAAIHLSVSTAALCRSALLLGVL

>ACH41852
MMRFLFVLLALLRKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSV
MRNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGGSSQFCISV
GKNIPAEHKNLQECFDGKIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAENIRGGNGKDGCNLVRTDN
NGILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATPHGTVRWTEDPNKVSIFKDVIRLFARFKEAKNA
VMTKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTG
AGANTVAVQSSVSTAALCCSVLLFGVL

>ACH41856
MRFLFVLLALLGKETYAYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNTSALAAAAAGLAAGRLEEWIFVFAQAADRSSQFCISVGK
HIAAEHGNLQECFDGTIGPETLYKIEDSRVKESAKTSLQLHEVLSSISFGSLGVKNIRGGNGRDGCNLVRTDTDG
VLEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
TKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAG
ANAAAIHLSVSTAALCRSALLLGVL

>ACH41865
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAGGRSSQFCISTGKT
GPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTSLQLHEVLSSISFGSLGVKNIRGGNGRDGCNLVRTDTDGV
LEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMT
KIKTTVDELTKCIGHKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

>ACH41859
MRFLFVLLAVLGKETYAYYENERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADGSSQFCISVGK
TGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTRLLLHEVLSSISFGSLGAENIRGGNGKDGCNLVRTDNNG
ILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFKEAKNAVM
TKIKTTVDELTKCIGQKEAELTNDQIYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
ANTVAVQSSVSTAALCCSVLLLGVL

>ACH41866
MRFLFLLLVLLGKKTHANYYENKRNALNATAANKVCGLSTYLKGIAHRVNGESAVVTEKLSDLKIRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNTSALAAAAAGIAAGRLEEWIFVFAQAADRSSQFCISVG
KTIPPEHNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFGSLGVKNIRGGNGKDGCNLVRTDTD
GILNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAV
MKKIKTTVDELTKCTGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETILKSNYTAEPVRGPFTVA
GSNAVAIHLSVSTAALCRSALLLGVL

>ACH41853
MRFWLVLLALLGKETYANYYENKRNALNATAANKVCRLSTYLKGIAHRVSSESAVVTEKLSDLKMKSIQLQLSIL
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGMTSKFCISVG
GSRPAVHDKLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEALSSISFGSLGVKNIRGGNGMDGCNLVRTDTD
GILAGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTKDPSKVSIFKDVIRLFARFQEAKNAV
MNKIKTTVDELTKCIGHKEAELTDYQLYEEFIWETIHRLELSKRVSEQPSLGEQEETILKSNYTAEPVRGPFTGA
GADAAAIQSSVSTAALCCLALLLGVM

>ACH41847
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADRSSQFCISVGKT
IPPEQNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFNSLGAESIRGGNGKDGCNLVRTDTDGI
LNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMK
KIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

>ACH41848
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
TNIPAEYNNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNTVAVHLSVSTAALCFSVLLLGVL

>ACH41851
MRFLFVLLALLGKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGGSSQFCISVG
RTGPAEYNNLQECFDGKIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAENIRGGNGKDGCNLVRTDNN
GILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATPHGTVRWTEDPNKVSIFKDVIRLFARFKEAKNAV
MTKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGA
GANTVAVQSSVSTAALCCSVLLFGVL

>ACH41857
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRHGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNAAAVHLSVSTAALCFSVLLLGVL

>ACH41860
MRFLFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSIMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISTGKT
GPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEVLSSISFSSLGVKNIRGGNGKDRCNLVRTDTDGV
LEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMK
KIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

>ACH41862
MRFLFVLLALLGKKTHAYYKNERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDISTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISVGK
TIPAEHGDLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGVKNIRGGNGRDGCNLVRTDTNG
ILEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFEDVIRLFARFQEAKNEVM
NKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
SNTVAVHLSVSTAALCRSALLLGVL

>ACH41868
MMRFWFVLLALLGKETYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGGSSQFCISVGK
HIAAEHGNLQECFDGKIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFGSLGVKNIRGGNGRDGCNLVRTDTNG
ILNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFEDVIRLFARFQEAKNAVM
RRIKTTVDELTKCIGQKEAELTNDQIYEEFIWETIHRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
SNTVAVHLSVSTAALCCLALLLGVL

>CAQ57275 **(BES1)**
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
TNIPAEYNNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNTVAVHLSLFTAALCCSALLLGVL

>CAQ57298
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRHGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNAAAVHLSVSTAALCFSVLLLGVL

>CAQ57310
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADRSSQFCISVGKT
IPPEQNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFNSLGAESIRGGNGKDGCNLVRTDTDGI
LNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMK
KIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

>CAQ57333
MMKFWFVLLALLGKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSV
MRNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISV
GTNIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDT
DGVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNE
VMNKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTG
AGSNKVAVHLSVSTAALCFSVLLLGVL

>CAQ57361
MMRFLFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELKEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
TNIPAEYNNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFNSLGAENIRGGNGRDGCNLVRTDTD
GVLEGGSVRRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNEV
MNKIKTTVDELAKCIGQKEVELTNDQLYEEFIWETIHRLELSKRVSEQLSLGEEEETILKSNYTAEPVRGPFTVA
GSNTVAVHLSVSTAALCFSALLLGVL

>CAQ57373
MRFLFVLLALLGKETHANYYENERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSIM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEVRNASALAAAAAGLAAGRLEEWIFVFAQAGGRSSQFCISTG
KTGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTRLLLHEVLSSISFSSLGAENIRGGNGKDGCNLVRTDNN
GILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAV
MTKIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGA
GANTVAVQSSVSTAALCCSALLLGVL

>CAQ57403
MRFLFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSIMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISTGKT
GPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEVLSSISFSSLGVKNIRGGNGKDRCNLVRTDTDGV
LEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMK
KIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

>CAQ57408
MRFLFVLLALLGKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGGSSQFCISVG
RTGPAEYNNLQECFNGKIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAENIRGGNGKDGCNLVRTDNN

GILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATPHGTVRWTEDPNKVSIFKDVIRLFARFKEAKNAV
MTKIKTSVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGA
GANTVAVQSSVSTAALCCSVLLFGVL

>CAQ57419
MMRFWFVLLALLGKETYAYYENERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAADGSSQFCISVG
RTGPAEYNNLQECFDGKIGPETLYKIEDSRVKESVQKSLQLHEVLSSISFGSLGVKNIRGGNGRDGCNLVRTDTN
GILNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFEDVIRLFARFQEAKNAV
MRRIKTTADELTKCIGQKEAELTNDQIYEEFIWETIHRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGA
GSNTVAVHLSVSTAALCCSVLLLGVL

>CAQ57428
MMRFWFVLLALLGKETYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGGSSQFCISVGK
HIAAEHGNLQECFDGKIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFGSLGVKNIRGGNGRDGCNLVRTDTNG
ILNGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFEDVIRLFARFQEAKNAVM
RRIKTTVDELTKCIGQKEAELTNDQIYEEFIWETIHRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
SNTVAVHLSVSTAALCCLALLLGVL

>CAQ57442
MRFLFVLLALLGKKTHAYYKNERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDISTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISVGK
TIPAEHGDLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGVKNIRGGNGRDGCNLVRTDTNG
ILEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFEDVIRLFARFQEAKNEVM
NKIKTTVDELAKCIGQKEVELTDDQLYEEFIWETIHRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
SNTVAVHLSVSTAALCRSALLLGVL

>CAQ57453
MRFWFVLLAVLGKETYAYYENERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISTGK
TGPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTRLLLHEVLLSISFGSLGAENIRGGNGKDGCNLVRTDNNG
ILKGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
KKIKTTVDELTKCIGQKEAELTNDQIYEEFIWETINRLELSKRVSEQPSLGEEEETILKSNYTAEPVRGPFTGAG
ANTVAVQSSVFTAALCCSALLLGVL

>CAQ57480
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISVGKT
GPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTSLQLHEVLSSISFGSLGVKNIRGGNGKDGCNLVRTDTDGV
LEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVVK
KIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

>CAD21457 **(Bn2)**
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAAGRSSQFCISVGKT
GPAEYNNLQECFDGTIGPETLYKIEDSRVKESAKTSLQLHEVLSSISFGSLGVKNIRGGNGKDGCNLVRTDTDGV
LEGGSPTRHNLTWGGGVMNFGSYQNGSMYVEGGEYGDATEYGAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVVK
KIKTTVDELTKCIGQKEAELTNDQLYEEFIWETINRLELSKRVSEQSAFGEEEETIVKFNYTAEPVRGPFTVAGA
NAAAIHLSVSTAALCRSALLLGVL

ESAG7 sequences:

>CAQ57274 **(BES1)**
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
TNIPAEYNNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQRKNRGCNLMRTADG
GLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
NKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57296
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQGENRGCNLMRTADG
GLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
NKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57297
MMKFWFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
KHIPAEHGNLQECFDGIIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQGENRGCNLMRTADG
GLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
NKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57309
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAADGSSQFCISVGKH
IPPEHKNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGAESIVEQRKNRGCNLMRTAYGGL
LKDFCLNRNFTWGGGVMNFGSCVAGNLKIEGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEVKNAVMKK
IKTTVDELTKCIGQKEAELTNDQLYEEFEAIQKYLWFL

>CAQ57332
MMKFWFVLLALLGKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSV
MRNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISV
GTNIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQRKNRGCNLMRTAD
GGLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAV
MNKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57345
MMKFWFVLLALLGKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSV
MRNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISV
GTNIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQRKNRGCNLMRTAD
GGLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFEEAKNAV
MNKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57346
MMKFWFVLLALLGKETHANYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSV
MRNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISV
GTNIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQRKNRGCNLMRTAD
GGLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAV
MNKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57360
MMRFLFVLLALLGKETHAYYENKRNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEQDCKDIRTLLKTVLRNEFTFQQELKEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVG
TNIPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEQRKNRGCNLMRTADG
GLLKDICLNCNFTWGGGVMNFGSCVAGNLKIKGGEYGDVSSHDVVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
NKIKTTVDELAKCIGQKEVELTNDQLYEEFEAIQKYLGSL

>CAQ57402
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSIMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAADRSSQFCISVGKH
IAAEHGNLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGAESIIERNEDRGCNLMRTADGGL
LKDVCLNRNFTWGGGVLNFGYCVAGNLKIKGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMKK
IKTTVDELTKCIGQKEAELTNDQIYEEFEAIQKYLGFL

>CAQ57407
MMRFWFVLLALLGKETYAYYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAADMTSQFCISVG
KNIPAEHKNLQECFNGKIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVERRENRGCNLMRTGRG
GLLKDVCLNRNFTWGGGVMNFGSCVAGNLKIKGGEYDDVSSHDEVRWTEDPSKVSIFKDVIRLFARFKEAKNAVM
KKIKTTVDELTKCIGQKEAELTNDQLYEEFEAIQKYLGFGKME

>CAQ57420
MMRFWFVLLALLGKETYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADRSSQFCISVGK
TIPAEHGDLQECFDGKIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGAENIVEQRKNRGCNLMRTAYGG
LLKDVCLNRNFTWGGGVMNFRSCVAGNLKIEGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVMK
KIKTTVDELTKCIGQKEAELTNDQLYEEFEAIQKYLGFL

>CAQ57429
MMRFWFVLLALLGKKTHAYYENERNALNATAANKVCGLSTYLKGVAHRVNSESAVVTEKLSDLKMRSIQLQLSVM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADRSSQFCISVG
KHIAAEHGNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEALSSISFSSLGAENIVEKGENRGCNLMRTAYG
GLLEGICLNRNFTWGGGVMNFGSCVAGNLEIKGGEYGDVSSHDAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
RRIKTTVDELTKCIGKKEAELTNDQIYEEFEAIQKYLGFL

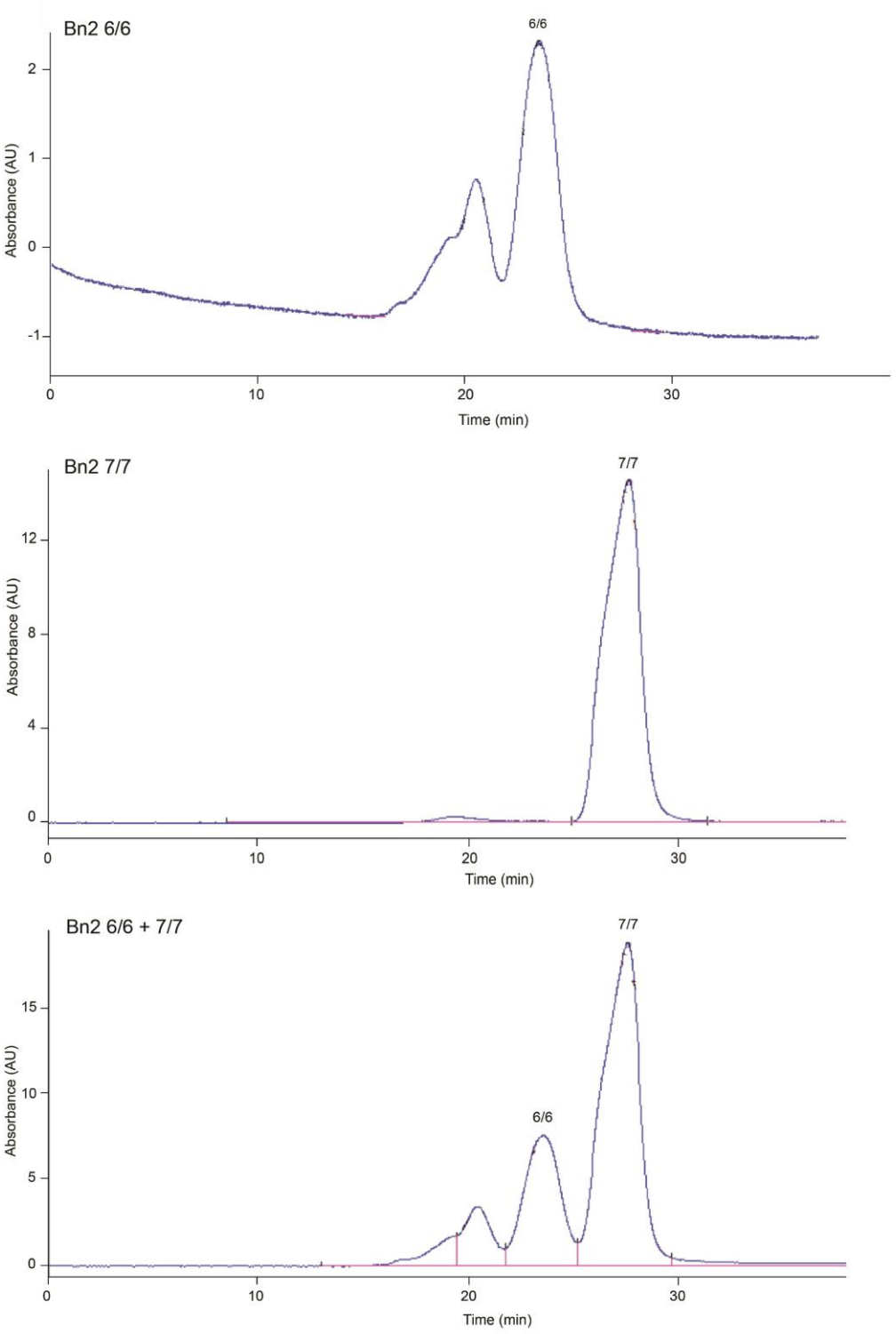
>CAQ57441
MRFLFVLLALLGKKTHAYYKNERNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMR
NRVPSGEQDCKDISTLLKTVLRNEFTFQQELEEMRNASALAAAAAGLAAGRLEEWIFVFAQAADRSSQFCISVGK
HIAAEHGDLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGAENIVEKGENRGCNLMRTADEG
LLKDVCLNRNFTWGGGVLNFGYCVAGNLKIKGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEAKNEVMN
KIKTTVDELAKCIGQKEVELTDDQLYEEFEAIQKYLGFL

>CAQ57452
MRFWFVLLALLGKETHAYYYENKRNALNATAANKVCALSTYLKGIAHRVNSESAVVTEKISDLKMRSIQLQLSIM
RNRVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGRSSQFCISVG
KTIPAEHGDLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGAENIVEKGENRGCNLMRTADE
GLLKDVCLNRNFTWGGGVMNFGYCVAGNLKIKGGEYGDVSSHDAVRWTEDPSKVSIFKDVIRLFARFQEAKNAVM
TKIKTTVDELTKCIGQKEAELTNDQIYEEFEAIQKYLGFL

>CAQ57467
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSIMRN
RVPSGEKDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGMSSQFCISVGKT
IPAEHGDLQECFDGTIGPETLYKIEDSRVKESAKKSLQLHEALSSISFSSLGAESIVEKGENRGCNLMRTADGGL
LKDVCLNRNFTWGGGVLNFGYCVAGNLKIKGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEVKNAVVKK
IKTTVDELTKCIGQKEAELTNDQLYEEFEVIQKYLWFL

>CAD21456 **(Bn2)**
MRFWFVLLALLGKEIYAYENERNALNATAANKVCGLSTYLKGIAHRVNSESAVVTEKLSDLKMRSIQLQLSVMRN
RVPSGEQDCKDIRTLLKTVLRNEFTFQQELEEMRNASALAAAAAGIAAGRLEEWIFVFAQAAGGSSQFCISVGTN
IPAEYNNLQECFDGTIGPETLYKIEDSRVKESAQKSLQLHEVLSSISFSSLGAESIVEKGENRGCNLMRTADGGL
LKDVCLNRNFTWGGGVLNFGYCVAGNLKIKGGEYGDVGSHDAVRWTEDPSKVSIFKDVIRLFARFQEVKNAVVKK
IKTTVDELTKCIGQKEAELTNDQLYEEFEVIQKYLWFL

Appendix 5: Size exclusion high performance liquid chromatography. Elution profiles of Bn2 ESAG6/6 homodimers (Bn2 6/6) (M_r = 87,000) and ESAG7/7 (M_r = 74,000) homodimers (Bn2 7/7) were investigated separately. Incubation of ESAG6 with ESAG7 (Bn2 6/6 + 7/7) (M_r = 81,000) did not result in the formation of a novel peak indicating the formation of heterodimers.



Appendix 6: Data collection and refinement statistics for crystal structures.

Data collection	**Non-glycosylated TfR**	**Glycosylated TfR**
Space group	*C*121	*C*121
Cell dimensions		
a, *b*, *c* (Å)	163.49, 108.11, 115.00	128.18, 117.87,134.55
α, β, γ (°)	90.0, 128.74, 90.0	90.0, 111.5, 90.0
Wavelength (Å)	1.000	1.000
Resolution (Å)	46.3 – 2.75 (2.80 - 2.75)	39.6 – 34.2 (3.48 – 3.42)
Total Observations	134161 (7123)	83799 (4432)
Total Unique	40238 (2045)	25075 (1291)
R_{pim} (%)	6.0 (62.8)	3.3 (67.3)
R_{merge} (%)	9.3 (99.9)	4.9 (116.3)
R_{meas} (%)	11.1 (118.3)	5.9 (126.2)
$CC_{1/2}$	0.991 (0.553)	0.988 (0.692)
$I/\sigma(I)$	9.1 (1.2)	11.6 (1.1)
Completeness (%)	98.9 (99.8)	99.0 (99.5
Multiplicity	3.3 (3.5)	3.3 (3.4)
Wilson B factor (Å^2)	59.95	152.06
Refinement		
Reflections	38216	25066
R_{work} / R_{free}	18.3/23.3	21.1/24.1
Number of residues		
Protein	1288	1288
Sugar	0	12
Fe(III)	1	1
water	38	0
R.M.S. deviations		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	1.21	1.22
Ramachandran plot		
Favoured (%)	93.6%	94.4%
Allowed (%)	6.3%	5.5%
Outliers (%)	0.1%	0.1%

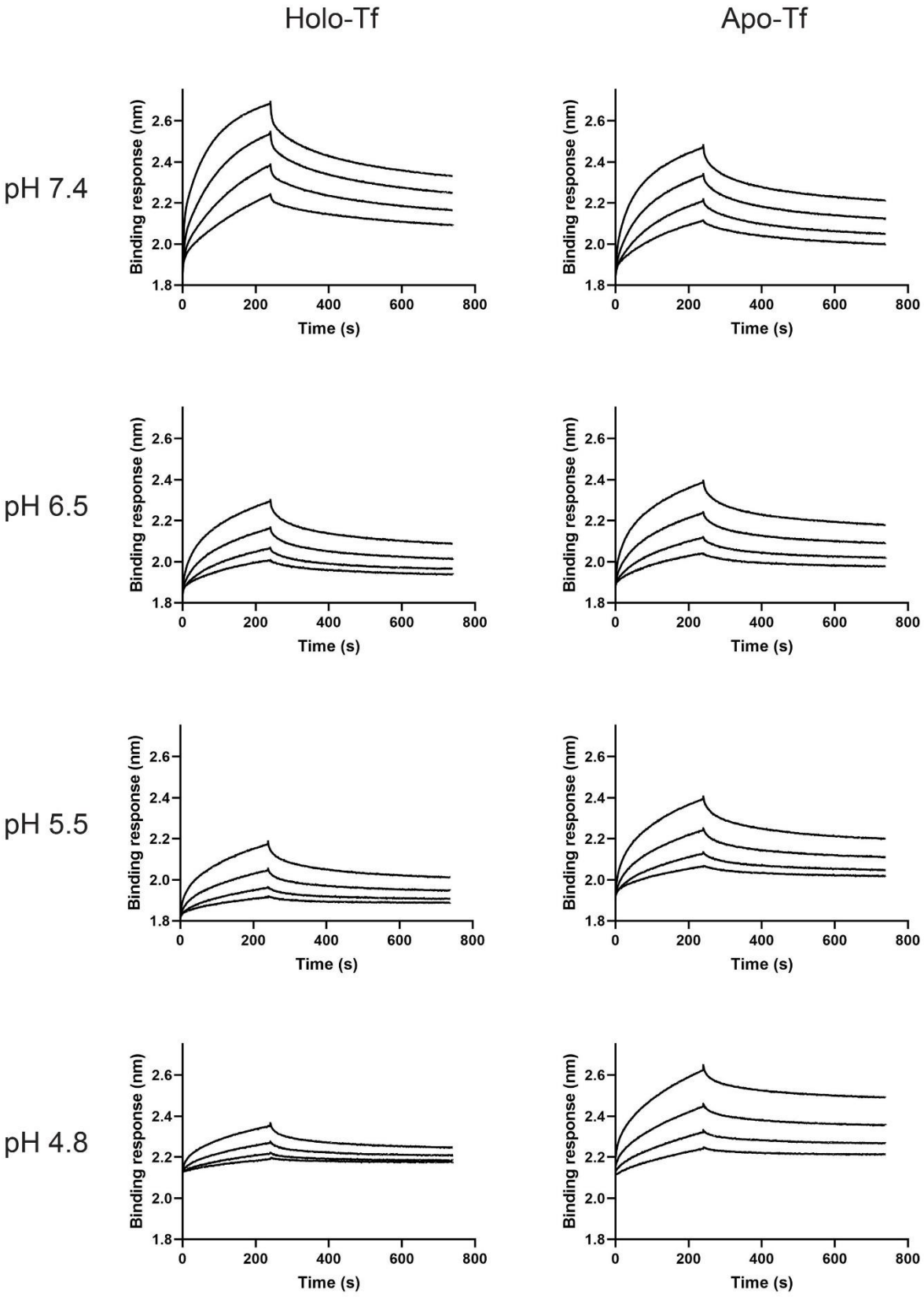
Appendix 7: Alignment of eight mammalian TfRs.

```
mouse       MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENADNNMKAS---VR    57
rat         MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENADSNMKAS---VR    57
pig         MMDQARSAFSSLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENVDSNTRSNHIGVA    60
cow         MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENVDSNMRGNQTSIA    60
goat        MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENVDSNMRGNQTSIT    60
horse       -MDQARSAFSNLFGGAPLSYTRFSLARQVDGDNSHVEMKLAVDEEENVDNNVRSNHASLT    59
human       MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAVDEEENADNNTKAN---VT    57
rabbit      MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAVDEEENVENNVRAN---VA    57
             *********.**** ************************.*****.:.* :..   :

mouse       KPKRFNGRLCFAAIALVIFFLIGFMSGYLGYCKRVEQKEECVKLAETEETDKSE--TMET    115
rat         KPKRFNGRLCFATIAVVIFFLIGFMIGYLGYCKRVEQKEECVRLAEAEEADKSE--NDET    115
pig         KPKRLNGYVCYGIIAVITFFLIGFMIGYLAYCKRVESKTDCKTLVPTEPSETEETETFEA    120
cow         KPKRLNGYVCYGIIAVIVFFLIGFMIGYLGYCRRVES-QDCGKEAGTQPSCPEETETFES    119
goat        KPKRLNGYICYGIIAVIIFFLIGFMIGYLGYCKRVEPQGDCGKEAGTQPSCPEETEAFEP    120
horse       KPKRFNGSFCYAVIAVIIFFLIGFMIGYLGYCKRVEPKSECGRSGDSKEIE--GTEPPET    117
human       KPKRCSGSICYGTIAVIVFFLIGFMIGYLGYCKGVEPKTECERLAGT-ESP--VR--EEP    112
rabbit      KPKRCNGTLCYAMVAIIIFFLIGFMIGYLGNCKRVEQKAECESVAGT-EVE--QSENAEP    114
            **** .* .*:. :*:: ******* ***. *: **   :*     :           *

mouse       -EDVPT-SSRLYWADLKTLLSEKLNSIEFADTIKQLSQNTYTPREAGSQKDESLAYYIEN    173
rat         -EYVPK-SSRLFWADLKTLLSEKLNSIEFTDIIKQLSQNTYTPREAGSQKDENLAYYIEN    173
pig         E-NFPQ-TPRLFWADLKILLSKGLDTTDFTRTIKMLNE-DYAPREAGSQKDESLGFFIEN    177
cow         EEQLPG-VPRIFWADLKSTLSGKLDAVDFARAIKMLNENSYVPREAGSEKDTSLAFFIEN    178
goat        EEQLPG-VPRIFWADLKSTLSDRLDAVDFTRAIKMLNENSYVPREAGSQKDSSLAFYIES    179
horse       EEYFPETPSRLLWTDLRTMLSERLTATEFTNTIKRLNGNSYVPREAGSQKDESLAFFIEN    177
human       GEDFPA-ARRLYWDDLKRKLSEKLDSTDFTGTIKLLNENSYVPREAGSQKDENLALYVEN    171
rabbit      EEDLPVIPPRLYWEDLRTMLSEKLDTTDFTSTIRQLNENSYIPREAGSQNDENLASFVEN    174
               .*    *: * **:  **  * : :*:  *: *.   * ******::* .*. ::*.

mouse       QFHEFKFSKVWRDEHYVKIQVKSSIGQNMVTIVQS--NGNLDPVESPEGYVAFSKPTEVS    231
rat         LFHDFKFSKVWRDEHYVKIQVKNSVSQNLVTI-NS--GSNIDPVEAPEGYVAFSKAGEVT    230
pig         QFREFKLSKVWHDEHFVKIQVKGSNAENSVTLVNTDSNSLVYPVESPEGYVAYSKATTVT    237
cow         QLQDCKLGKVWHDEHFIKIQVKGSSQNSVSIVSTSGNGSQAYPVESPEGYVAYSKAATVT    238
goat        QLRDYKLDKVWHDEHFVKIQVKGSSQNSVSIVSTSGDGSQAYSVENPEGYVAYSKAATVT    239
horse       QFREFKLNKVWRDEHFVKIQVKGSNAQSSVTVVN-GSGDMISLVENPTGYVAYSKATTVT    236
human       QFREFKLSKVWRDQHFVKIQVKDSA-QNSVIIVD-KNGRLVYLVENPGGYVAYSKAATVT    229
rabbit      KFRDFKLSKVWHDEHFVKIQVKDSA-QNSVIIVN-KNSEAPDLVENPMGYVAYSKATTVT    232
             ::: *:.***:*:*::*****.*  :.   :     .     ** * ****:**   *:
```

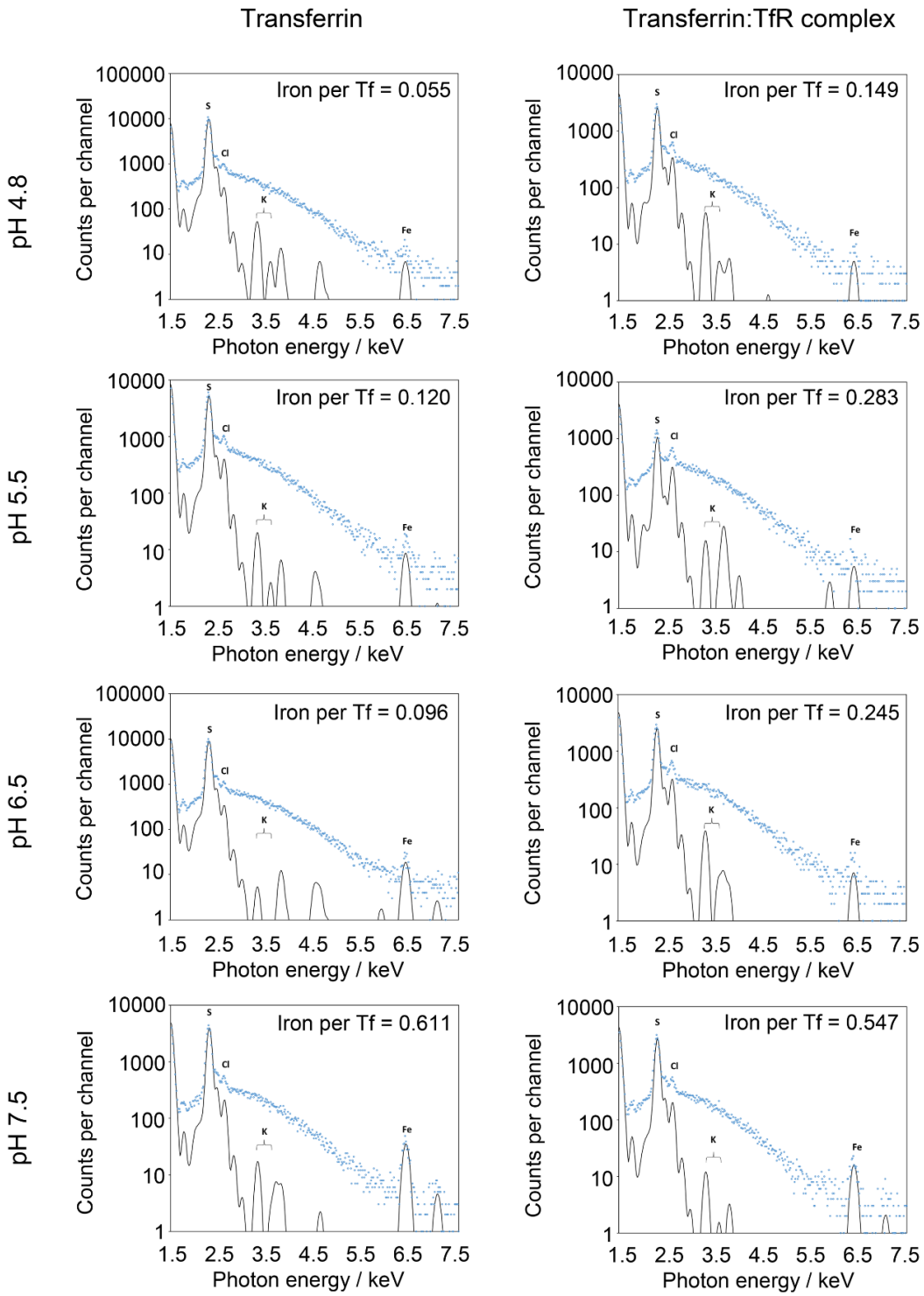
```
mouse       GKLVHANFGTKKDFEELSYSVNGSLVIVRAGEITFAEKVANAQSFNAIGVLIYMDKNKFP    291
rat         GKLVHANFGTKKDFEELNYSVNGSLVIVRAGKITFAEKVANAQSFNAIGVLIYMDRNTFP    290
pig         GKLIFANFGTKKDFEDLKMPVNGSLVIVRAGKITFAEKVANAQSLDAIGVLIYMDRANFP    297
cow         GKLVHANFGTKQDFEDLNMPVNGSLVIVRAGKISFAEKVANAESLNAIGVLIYMDYSKYP    298
goat        GKLVHANFGTKQDFEDLNMPVNGSLVIVRAGKITFAEKVANAERLNAIGVLIYMDHSKFP    299
horse       GKLVHANFGTKEDYEALSYPVNGSLVIVRAGEITFAQKVANAESLNAVGVLIYMDQAKFP    296
human       GKLVHANFGTKKDFEDLYTPVNGSIVIVRAGKITFAEKVANAESLNAIGVLIYMDQTKFP    289
rabbit      GKLIHANFGTKKDFEDLNSPVNGSIVIVRAGEITFAEKVANAESLNAIGMLIYMDQMKFP    292
            ***:.******:*:* *   ****:******:*:**:*****: ::*:*:*****  .:*

mouse       VVEADLALFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISRAAAEKLFGKM    351
rat         VVEADLQFFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPSIPVQTISRAAAEKLFKNM    350
pig         IINADVPVFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISRAGAEKLFGNM    357
cow         IVNANLPVFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTITRAGAEKLFQNM    358
goat        IVNANLPVFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISRAGAERLFQNM    359
horse       IVNANLPVFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISRAAAEALFANM    356
human       IVNAELSFFGHAHLGTGDPYTPGFPSFNHTQFPPSRSSGLPNIPVQTISRAAAEKLFGNM    349
rabbit      IVKADLPLFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISREAAEKLFGNM    352
            :::*:: .****************************:*****.******:* .** ** :*

mouse       EGSCPARWNIDSSCKLELSQNQNVKLIVKNVLKERRILNIFGVIKGYEEPDRYVVVGAQR    411
rat         EGNCPPSWNIDSSCKLELSQNQNVKLTVNNVLKETRILNIFGVIKGYEEPDRYIVVGAQR    410
pig         EQDCPLTWRTDFPCKLVSSPSKNVKLTVNNVLKEIKILNIFGVIKGFEEPDRYVIVGAQR    417
cow         EGDCPRIWGTDSSCKLVSSQDKNVKLSVNNVLKEIRILNVFGVIKGFEEPDRYVIVGAQR    418
goat        QGDCPSTWGTDSSCKLVSSQDKNVKLSVNNVLKEIRIFNVFGVIKGFEELDRYIIVGAQR    419
horse       KGDCPSSWKTDSSCRLEFPGDKNVKLTVNNELKEIRIFNVFGVIKGFEEPDRYVVIGAQR    416
human       EGDCPSDWKTDSTCRMVTSESKNVKLTVSNVLKEIKILNIFGVIKGFVEPDHYVVVGAQR    409
rabbit      EGACPSEWQTDSTCRLQSSQNKNVKLTVNNVLKEARILNVFGVIKGFEEPDRYIVVGAQR    412
            :  **  *  *  *::    .:**** *.* *** :*:*:******: * *:*:::****

mouse       DALGAGVAAKSSVGTGLLLKLAQVFSDMISKDGFRPSRSIIFASWTAGDFGAVGATEWLE    471
rat         DAWGPGVA-KSSVGTGLLLKLAQVFSDMISKDGFRPSRSIIFASWTAGDYGAVGATEWLE    469
pig         DAWGPGA-AKSSVGTSLLLNLAQILSDMVIKGQFKPSRSIVFASWSAGDFGAIGATEWLE    476
cow         DAWGPGA-AKSSVGTSLLLTLARILSDMVLKGQFKPSRSIVFASWSGGDFGAVGATEWLE    477
goat        DAWGPGA-AKSSVGTSLLLTLARILSDMVLKGQFKPSRSIVFASWSGGDFGAVGATEWLE    478
horse       DAWGPGA-AKSSVGTALLLELARIFSDMVSKGGFKPSRSIVFASWGAGDFGAIGATEWLE    475
human       DAWGPGA-AKSGVGTALLLKLAQMFSDMVLKDGFQPSRSIIFASWSAGDFGSVGATEWLE    468
rabbit      DAWGPGA-AKSSVGTALLLELARMFSDMVLKGGFKPSRSIIFASWSAGDFGAIGATEWLE    471
            ** * *.  **.***.*** **:::***: *. *:*****:**** .**:*::*******

mouse       GYLSSLHLKAFTYINLDKVVLGTSNFKVSASPLLYTLMGKIMQDVKHPVDGKSLYRDSNW    531
rat         GYLSSLHLKAFTYINLDKVVLGTSNFKVSASPLLYTLMGKIMQDVKHPIDGKYLYRDSNW    529
pig         GYLSSLHLKAFTYINLDKAVLGTSNFKVSASPLLYSLIEKMMQDVKNPVTGQSLYRDSNW    536
cow         GYLSSLHLKAFTYINLDKAVVGTTNFKVSASPLLYSLIEKIMKDVKHPLNGLSLYRDSNW    537
goat        GYLSSLHLKAFTYINLDKAVVGTTNFKVSASPLLYSLIEKIMKDVKHPLTGLSLYRDSNW    538
```

```
horse       GYLSSLHLKAFTYINLDKAVLGAKNFKVSASPLLYSLIEKTMQEVKHPVTGLSLYRDSNW    535
human       GYLSSLHLKAFTYINLDKAVLGTSNFKVSASPLLYTLIEKTMQNVKHPVTGQFLYQDSNW    528
rabbit      GYLSSLHLKAFTYINLDKAVLGTTNFKVSASPLLYKLIEKTMQDVKHPVTGMSLYQDSNW    531
            ******************.*:*:.***********.*: * *::**:*: *  **:****

mouse       ISKVEKLSFDNAAYPFLAYSGIPAVSFCFCEDADYPYLGTRLDTYEALTQKVPQLNQMVR    591
rat         ISKIEELSLDNAAFPFLAYSGIPAVSFCFCEDEDYPYLGTKLDTYEILIQKVPQLNQMVR    589
pig         INKVEKLSFDDAAFPFLAYSGIPAVSFCFCEDTDYPYLGTTMDTYDVLSKRVPQLNRMAR    596
cow         ISKVQKLSLDNAAFPFLAYSGIPAVSFCFCEDTDYPYLGTPMDTYETLNREVPQLNRVAR    597
goat        ISKVKKLSLDNAAFPFLAYSGIPAVSFCFCEDTDYPYLDTPMDTYETLNGQVPQLNRVAR    598
horse       INKVEKLSFDNAAFPFLAYSGIPALSFCFCEDTEYPYLGTTMDTYEVLSQNVPELSRLTR    595
human       ASKVEKLTLDNAAFPFLAYSGIPAVSFCFCEDTDYPYLGTTMDTYKELIERIPELNKVAR    588
rabbit      VNKVEKLSFDNAAFPFLADSGIPSLSFWFCEDTDYPYLGTTFDTYGMLIQKIPQLNKMAR    591
             .*:::*::*:**:**** ****::** **** :****.* :***  *  .:*:*.::.*

mouse       TAAEVAGQLIIKLTHDVELNLDYEMYNSKLLSFMKDLNQFKTDIRDMGLSLQWLYSARGD    651
rat         TAAEVAGQFIIKLTHDIELTLDYEMYNSKLLSFMKDLNQFKADIKDMGLSLQWLYSARGD    649
pig         AAAEVAGHLVIKLTIDFELNLNYEMYNDKILSFVREMNQFRVDIREMGLSLQWLYSARGD    656
cow         AAAEVAGQLVIKLTHGVELNLNYEMYNDEILRFVKEMNLFRADIRDMGLNMQWLYSARGD    657
goat        AAAEVAGQLVMKLTHGVELDLNYEMYNDEILRFVKEMNLFRADIRDMGLNMQWLYSARGD    658
horse       AAAEVAGQLLIKLSYDVELNLNYDMYNDKILSFVKDMNQFRADIKEMGLNLQWLYSARGD    655
human       AAAEVAGQFVIKLTHDVELNLDYERYNSQLLSFVRDLNQYRADIKEMGLSLQWLYSARGD    648
rabbit      AAAEVAGQFMIKLTHDAELNLDYEMYNNKILSFVRELNQYRAYIKELGLNLQWLYSARGD    651
            :******::::**: . ** *:*: **.::* *:::** ::. *:::**.:*********

mouse       YFRATSRLTTDFHNAEKTNRFVMREINDRIMKVEYHFLSPYVSPRESPFRHIFWGSGSHT    711
rat         YFRATSRLTTDFHNAEKTNRFVMREINDRIMKVEYHFLSPYVSPRESPFRHIFWGSGSHT    709
pig         FFRATSRLTSDYRNVETRDKFVMREINDRIMKVEYHFLSPYVSPRESPFRHIFWGSGSHT    716
cow         FFRATSRLTTDYKNAEKTDRSVMREINDRIMKVEYHLLSPYVSPREFPFRHIFWGSGSHT    717
goat        FFRATSRLTTDYKNAEKTDRSVMREINDRIMKVEYHLLSPYVSPREFPFRHIFWGSGSHT    718
horse       FFRATSRLTTDYKNAERANRVVMREINDRIMKVEYHFLSPYVSPRESPFRHIFWGSGSHT    715
human       FFRATSRLTTDFGNAEKTDRFVMKKLNDRVMRVEYHFLSPYVSPKESPFRHVFWGSGSHT    708
rabbit      FFRATSRLTTDFKNAEITNRFVMREINDRIMKVEYHFLSPYVSPRESPFRHIFWGSGSHT    711
            :********:*: *.*  :: **:::***:*:****:*******:* ****:********

mouse       LSALVENLKLRQKNITAFNETLFRNQLALATWTIQGVANALSGDIWNIDNEF   763
rat         LSALVENLRLRQKNITAFNETLFRNQLALATWTIQGVANALSGDIWNIDNEF   761
pig         LSALVEHLKLRQKNSSAFNQTLLKNQLALATWTIQGAANALSGDIWDIDNEF   768
cow         LSALLEHLKLRKKNNGAFNQTLLENQLALATWTIQGAANALSGDIWDIDNEF   769
goat        LSALLEHLKLRKKLNGAFNQTLLENQLALATWTIQGAANALSGDIWDIDNEF   770
horse       LSALLEHLKLRQKNSGAFNETLLRNQLALATWTIQGAANALSGDIWDIDNEF   767
human       LPALLENLKLRKQNNGAFNETLFRNQLALATWTIQGAANALSGDVWDIDNEF   760
rabbit      LSALLENLNLRQKNNGAFNETLLRNQLALATWTIQGAANALSGDIWDIDNEF   763
            * **:*:*.**::   ***:**:.************.*******:*:*****
```

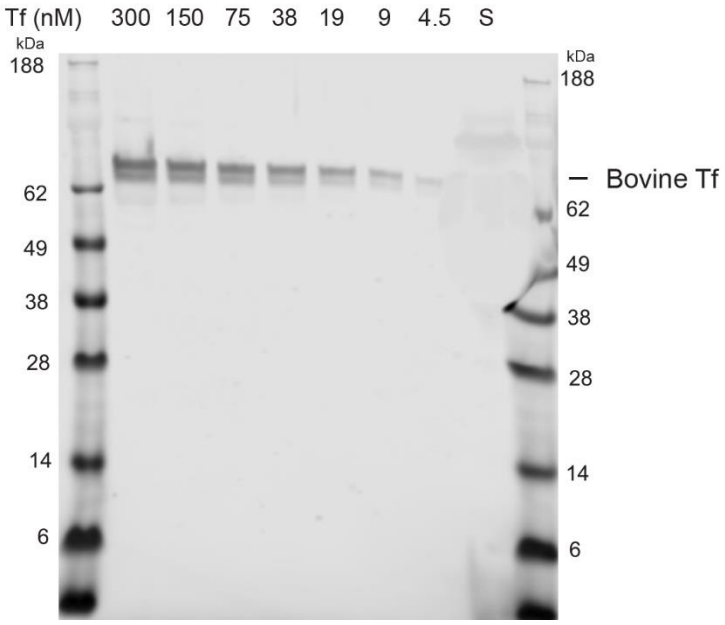
Appendix 8: Biolayer interferometry measurement of binding between BES1-TfR and bovine Tf. Biotinylated BES1-TfR was immobilised on a streptavidin biosensor and binding of bovine Tf was measured for different iron states and at different pH values.



Appendix 9: Quantification of elements present in Tf and TfR-Tf complex at different pH by microPIXE. The photon emissions associated with each element (sulphur, chloride, potassium and iron) were measured and counts per channel were plotted. Sulphur measurements were used in conjunction with the number of sulphurs present in the protein sequence to extrapolate the quantity of iron.



Appendix 10: Measurement of bovine anti-Tf antibody sensitivity. Serial dilution of FCS was performed to test the detection limit of the antibody by western blot, with estimated concentrations reported above each lane in nM. The Tf-depleted serum sample (S) was also analysed by western blot.



Appendix 11: Alignment of bovine lactoferrin and transferrin protein sequences.

```
Identity:     403/693 (58.2%)
Similarity:   515/693 (74.3%)

lactoferrin         1 --------CTISQPEWFKCRRWQWRMKKL--GAPSITCVRRAFALECIRA     40
                            ||||..|..||..::..:.::  ..|.::||::...::||:|
transferrin         1 DPERTVRWCTISTHEANKCASFRENVLRILESGPFVSCVKKTSHMDCIKA     50

lactoferrin        41 IAEKKADAVTLDGGMVFEACRDPYKLRPVAAEIYGTKESPQTHYYAVAVV     90
                      |:..:|||||||||:|:||...|..|:||.||.:|||::||||||||||
transferrin        51 ISNNEADAVTLDGGLVYEAGLKPNNLKPVVAEFHGTKDNPQTHYYAVAVV    100

lactoferrin        91 KKGSNFQLDQLQGRKSCHTGLGRSAGWIIPMGILRPYLSWTESLEPLQGA    140
                      ||.::|:|::|:|:||||||||||||.|||..|  |....:..|.:|.|
transferrin       101 KKDTDFKLNELRGKKSCHTGLGRSAGWNIPMAKL--YKELPDPQESIQRA    148

lactoferrin       141 VAKFFSASCVPCIDRQAYPNLCQLCKGEGENQCACSSREPYFGYSGAFKC    190
                      .|.|||||||||.|:.::|.|||||.|:|.::||||:.||||||||||
transferrin       149 AANFFSASCVPCADQSSFPKLCQLCAGKGTDKCACSNHEPYFGYSGAFKC    198

lactoferrin       191 LQDGAGDVAFVKETTVFENLPEKADRDQYELLCLNNSRAPVDAFKECHLA    240
                      |.:|||||||||.:|||:|||...||..||||||.:|:|:..||.::||:||
transferrin       199 LMEGAGDVAFVKHSTVFDNLPNPEDRKNYELLCGDNTRKSVDDYQECYLA    248

lactoferrin       241 QVPSHAVVARSVDGKEDLIWKLLSKAQEKFGKNKSRSFQLFGSPPGQRDL    290
                      .|||||||||:|.||||:||:||:.|||.|||:|..:|||||.||.| :||
transferrin       249 MVPSHAVVARTVGGKEDVIWELLNHAQEHFGKDKPDNFQLFQSPHG-KDL    297

lactoferrin       291 LFKDSALGFLRIPSKVDSALYLGSRYLTTLKNLRETAEEVKAR-YTRVVW    339
                      ||||||.|||:||||:|..||||..|:|.|:||||:.....:: ...|.|
transferrin       298 LFKDSADGFLKIPSKMDFELYLGYEYVTALQNLRESKPPDSSKDECMVKW    347

lactoferrin       340 CAVGPEEQKKCQQWSQQSGQNVTCATASTTDDCIVLVLKGEADALNLDGG    389
                      ||:|.:|:.||.:||..||..:.|.||..|::||..::|||||::||||
transferrin       348 CAIGHQERTKCDRWSGFSGGAIECETAENTEECIAKIMKGEADAMSLDGG    397

lactoferrin       390 YIYTAGKCGLVPVLAENRKSSKHSSLDCVLRPTEGYLAVAVVKKANEGLT    439
                      |:|.||||||||||||.|:...|   |...|.:|||||||||.:::..:.
transferrin       398 YLYIAGKCGLVPVLAENYKTEGES---CKNTPEKGYLAVAVVKTSDANIN    444
```

```
lactoferrin      440 WNSLKDKKSCHTAVDRTAGWNIPMGLIVNQTGSCAFDEFFSQSCAPGADP    489
                     ||:||||||||||||||||||||||:.::..:|.||||||..||||:..
transferrin      445 WNNLKDKKSCHTAVDRTAGWNIPMGLLYSKINNCKFDEFFSAGCAPGSPR    494

lactoferrin      490 KSRLCALCAGDDQGLDK-CVPNSKEKYYGYTGAFRCLAEDVGDVAFVKND    538
                     .|.|||||.|.::|..| |||||.|:||||||||||.|. |||||||:.
transferrin      495 NSSLCALCIGSEKGTGKECVPNSNERYYGYTGAFRCLVEK-GDVAFVKDQ    543

lactoferrin      539 TVWENTNGESTADWAKNLNREDFRLLCLDGTRKPVTEAQSCHLAVAPNHA    588
                     ||.:||:|.:...|||||.:|:|.:||.||||||||:|::||||..||||
transferrin      544 TVIQNTDGNNNEAWAKNLKKENFEVLCKDGTRKPVTDAENCHLARGPNHA    593

lactoferrin      589 VVSRSDRAAHVKQVLLHQQALFGKNGKNCPDKFCLFKSETKNLLFNDNTE    638
                     ||||.|:|..|:::|..||..|||:..:|...||||:|.:|:|||.|:|:
transferrin      594 VVSRKDKATCVEKILNKQQDDFGKSVTDCTSNFCLFQSNSKDLLFRDDTK    643

lactoferrin      639 CLAKLGGRPTYEEYLGTEYVTAIANLKKCSTSPLLEACAFLTR    681
                     |||.: .:.||:.|||.:||.|:.||::||||.|||||.|...
transferrin      644 CLASI-AKKTYDSYLGDDYVRAMTNLRQCSTSKLLEACTFHKP    685
```

Appendix 12: List of primer sequences (5'-3') used in this project.

Primers used for sequencing the mammalian expression vector (pDest12) and the scFv used for phage display.

pDest12 forward	GCTGCCTCTGCTGTGGGCTG
pDest12 reverse	GTCACACCACAGAAGTAAGG
scFv forward	GATTACGCCAAGCTTTGGAGC
scFv reverse	CTCTTCTGAGATGAGTTTTTG

Primers used to add an N-terminal TEV protease site to Bn2 ESAG6 and remove affinity tags from Bn2 ESAG7. Primers for multi site-directed mutagenesis to convert N-linked glycosylation sites to aspartic acid are also listed. Fw = forward and Rv = reverse.

Bn2 ESAG6 N-terminal TEV Fw	GCACGCTAGCGAGAACCTCTATTTCCAAGGGTCTGGCTACGAGAACGAGAGAAACGC
Bn2 ESAG6 C-terminal stop Rv	GCACGCTAGCTTATCAGGCGCCAGCCACTGTAAAAG
Bn2 ESAG7 Fw	GCACGCGGCCGCATACGAGAACGAG
Bn2 ESAG7 C-terminal stop Rv	GCACGCTAGCTTATCAGCTGCCCAGGAACCACAGG
Bn2 N26D Fw	GAACGCCCTGGATGCCACCGCCGCCAACAAAG
Bn2 N110D Fw	GGAACTGGAAGAGATGAGAGACGCCAGCGCCCTGGCC
Bn2 ESAG6 N235D Fw	GCAGCCCCACAAGACACGACCTGACATGGGGCGG
Bn2 ESAG7 N234D Fw	GGACGTGTGCCTGAACCGGGACTTCACCTGGGGAGG
Bn2 ESAG6 N250D Fw	GGCAGCTACCAGGACGGCTCTATGTACGTGGAAGGGGGCG
Bn2 ESAG6 N360D Fw	GGAAGAGACAATCGTGAAGTTCGACTACACCGCCGAGCCCGTG

Primers for site-directed mutagenesis in BES1-TfR. Residues are numbered according to the crystal structure (Bn2-TfR).

BES1 ESAG6 G139R forward	CGTGTTCGCCCAGGCCGCAGGCCGCAGCTCCCAGTTCTGTATCAGC
BES1 ESAG6 G139R reverse	GCTGATACAGAACTGGGAGCTGCGGCCTGCGGCCTGGGCGAACACG
BES1 ESAG7 S246Y forward	GGAGGCGTGATGAACTTCGGCTATTGCGTGGCCGGCAATCTGAAG
BES1 ESAG7 S246Y reverse	CTTCAGATTGCCGGCCACGCAATAGCCGAAGTTCATCACGCCTCC
BES1 ESAG7 I229V forward	GATGGCGGCCTGCTGAAGGACGTCTGCCTGAACTGCAACTTCACC
BES1 ESAG7 I229V reverse	GGTGAAGTTGCAGTTCAGGCAGACGTCCTTCAGCAGGCCGCCATC
BES1 ESAG7 C233R forward	CTGCTGAAGGACGTCTGCCTGAACCGCAACTTCACCTGGGGCGG
BES1 ESAG7 C233R reverse	CCGCCCCAGGTGAAGTTGCGGTTCAGGCAGACGTCCTTCAGCAG

Primers used for BES1-TfR RT-PCR.

BES1 ESAG6 ESAG7 forward	CACTCTTGAAAACAGTATTG
BES1 ESAG6 reverse	TCTCTTTGACAGCTCCAATC
BES1 ESAG7 reverse	TAGYACTTTTGTATTGCCTC

References

Alsford, S., Eckert, S., Baker, N., Glover, L., Sanchez-Flores, A., Leung, K. F., … Horn, D. (2012). High-throughput decoding of antitrypanosomal drug efficacy and resistance. *Nature*, *482*(7384), 232–236. https://doi.org/10.1038/nature10771

Balmer, O., Beadell, J. S., Gibson, W., & Caccone, A. (2011). Phylogeography and taxonomy of Trypanosoma brucei. *PLoS Neglected Tropical Diseases*, *5*(2). https://doi.org/10.1371/journal.pntd.0000961

Bartossek, T., Jones, N. G., Schäfer, C., Cvitković, M., Glogger, M., Mott, H. R., … Engstler, M. (2017). Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. *Nature Microbiology*, *2*(11), 1523–1532. https://doi.org/10.1038/s41564-017-0013-6

Becker, M., Aitcheson, N., Byles, E., Wickstead, B., Louis, E., & Rudenko, G. (2004). Isolation of the repertoire of VSG expression site containing telomeres of Trypanosoma brucei 427 using transformation-associated recombination in yeast. *Genome Research*, *14*(11), 2319–2329. https://doi.org/10.1101/gr.2955304

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., … El-Sayed, N. M. (2005). The genome of the African trypanosome Trypanosoma brucei. *Science*, *309*(5733), 416–422. https://doi.org/10.1126/science.1112642

Berriman, M., Hall, N., Sheader, K., Bringaud, F., Tiwari, B., Isobe, T., … Rudenko, G. (2002). The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Molecular and Biochemical Parasitology*, *122*(2), 131–140. https://doi.org/10.1016/s0166-6851(02)00092-0

Bitter, W., Gerrits, H., Kieft, R., & Borst, P. (1998). The role of transferrin-receptor variation in the host range of Trypanosoma brucei. *Nature*, *391*(6666), 499–502. https://doi.org/10.1038/35166

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M., & Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12 I), 2210–2221. https://doi.org/10.1107/S0907444904016427

Blum, M. L., Down, J. A., Gurnett, A. M., Carrington, M., Turner, M. J., & Wiley, D. C. (1993). A structural motif in the variant surface glycoproteins of Trypanosoma brucei. *Nature*, *362*(6421), 603–609. https://doi.org/10.1038/362603a0

Bonnet, J., Boudot, C., Courtioux, B., & Karl, S. (2015). Overview of the Diagnostic Methods Used in the Field for Human African Trypanosomiasis: What Could Change in the Next Years? *BioMed Research International*. https://doi.org/10.1155/2015/583262

Borst, P. (1991). Transferrin receptor, antigenic variation and the prospect of a trypanosome vaccine. *Trends in Genetics*, *7*(10), 307–309.

Breidbach, T., Scory, S., Krauth-Siegel, R., Steverding, D. (2002). Growth inhibition of

bloodstream forms of Trypanosoma brucei by the iron chelator deferoxamine. *International Journal for Parasitology*, 32(4), 473-479.

Brun, R., Blum, J., Chappuis, F., & Burri, C. (2010). Human African trypanosomiasis. *The Lancet*, *375*(9709), 148–159. https://doi.org/10.1016/S0140-6736(09)60829-1

Budzak, J., Kerry, L. E., Aristodemou, A., Hall, B. S., Witmer, K., Kushwaha, M., … Rudenko, G. (2019). Dynamic colocalization of 2 simultaneously active VSG expression sites within a single expression-site body in Trypanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(33), 16561–16570. https://doi.org/10.1073/pnas.1905552116

Bullard, W., Kieft, R., Capewell, P., Veitch, N. J., Macleod, A., & Hajduk, S. L. (2012). Haptoglobin-hemoglobin receptor independent killing of African trypanosomes by human serum and trypanosome lytic factors. *Virulence*, *3*(1), 72–76. https://doi.org/10.4161/viru.3.1.18295

Capewell, P., Clucas, C., DeJesus, E., Kieft, R., Hajduk, S., Veitch, N., … MacLeod, A. (2013). The TgsGP Gene Is Essential for Resistance to Human Serum in Trypanosoma brucei gambiense. *PLoS Pathogens*, *9*(10), e1003686. https://doi.org/10.1371/journal.ppat.1003686

Capewell, P., Veitch, N. J., Turner, C. M. R., Raper, J., Berriman, M., Hajduk, S. L., & MacLeod, A. (2011). Differences between Trypanosoma brucei gambiense groups 1 and 2 in their resistance to killing by trypanolytic factor 1. *PLoS Neglected Tropical Diseases*, *5*(9). https://doi.org/10.1371/journal.pntd.0001287

Carrington, M., & Boothroyd, J. (1996). Implications of conserved structural motifs trypanosome surface proteins. *Molecular and Biochemical Parasitology*, *81*, 119–126.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., & Turner, M. (1991). Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. *Journal of Molecular Biology*, *221*(3), 823–835. https://doi.org/10.1016/0022-2836(91)80178-W

Chaudhri, M., Steverding, D., Kittelberger, D., Tjia, S., & Overath, P. (1994). Expression of a glycosylphosphatidylinositol-anchored Trypanosoma brucei transferrin-binding protein complex in insect cells. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(14), 6443–6447. https://doi.org/10.1073/pnas.91.14.6443

Checchi, F., Filipe, J. A. N., Haydon, D. T., Chandramohan, D., & Chappuis, F. (2008). Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC Infectious Diseases*, *8*(1), 16. https://doi.org/10.1186/1471-2334-8-16

Coppens, I., Opperdoes, F. R., Courtoy, P. J., & Baudhuin, P. (1987). Receptor-Mediated Endocytosis in the Bloodstream Form of Trypanosoma brucei. *The Journal of Protozoology*, *34*(4), 465–473. https://doi.org/10.1111/j.1550-7408.1987.tb03216.x

Cordon-Obras, C., Cano, J., González-Pacanowska, D., Benito, A., Navarro, M., & Bart, J. M. (2013). Trypanosoma brucei gambiense adaptation to different Mammalian Sera is associated with VSG expression site plasticity. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0085072

Cross, G. A. M., Kim, H. S., & Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Molecular and Biochemical Parasitology*. https://doi.org/10.1016/j.molbiopara.2014.06.004

Daramola, O., Stevenson, J., Dean, G., Hatton, D., Pettman, G., Holmes, W., & Field, R. (2014). A high-yielding CHO transient system: Coexpression of genes encoding EBNA-1 and GS enhances transient protein expression. *Biotechnology Progress*, *30*(1), 132–141. https://doi.org/10.1002/btpr.1809

DeJesus, E., Kieft, R., Albright, B., Stephens, N. A., & Hajduk, S. L. (2013). A Single Amino Acid Substitution in the Group 1 Trypanosoma brucei gambiense Haptoglobin-Hemoglobin Receptor Abolishes TLF-1 Binding. *PLoS Pathogens*, *9*(4), e1003317. https://doi.org/10.1371/journal.ppat.1003317

Eckenroth, B. E., Steere, A. N., Chasteen, N. D., Everse, S. J., & Mason, A. B. (2011). How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomal pH. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13089–13094. https://doi.org/10.1073/pnas.1105786108

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

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*(3), 505–515. https://doi.org/10.1016/j.cell.2007.08.046

Fairhead, M., & Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. *Methods in Molecular Biology*, *1266*, 171–184. https://doi.org/10.1007/978-1-4939-2272-7_12

Ferguson, M. A. J. (1997). The Surface Glycoconjugates of Trypanosomatid Parasites. In *Source: Philosophical Transactions: Biological Sciences* (Vol. 352).

Fevre, E. M., Wissmann, B. V., Welburn, S. C., & Lutumba, P. (2008). The burden of human African Trypanosomiasis. *PLoS Neglected Tropical Diseases*, *2*(12). https://doi.org/10.1371/journal.pntd.0000333

Field, M. C., & Carrington, M. (2009, October 6). The trypanosome flagellar pocket. *Nature Reviews Microbiology*, 7, 775–786. https://doi.org/10.1038/nrmicro2221

Franke, E. (1905). Ueber trypanosomentherapie. *Muenchener Med. Wochenschr*, *42*, 2059–2060.

Freymann, D., Down, J., Carrington, M., Roditi, I., Turner, M., & Wiley, D. (1990). 2.9 A Resolution Structure of the N-terminal Domain of a Variant Surface Glycoprotein from Trypanosoma brucei. In *Journal of Molecular Biol*ogy. 216(1) 141-160.

Furger, A., Schürch, N., Kurath, U., & Roditi, I. (1997). Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of Trypanosoma brucei by modulating RNA stability and translation. *Molecular and Cellular Biology*, *17*(8), 4372–

4380. https://doi.org/10.1128/mcb.17.8.4372

Garman, E. F., & Grime, G. W. (2005, October 1). Elemental analysis of proteins by microPIXE. *Progress in Biophysics and Molecular Biology*, 89, 173–205. https://doi.org/10.1016/j.pbiomolbio.2004.09.005

Gavel, Y., & Heijne, G. Von. (1990). Sequence differences between glycosylated and non-glycosylated asn-x-thr/ser acceptor sites: Implications for protein engineering. *Protein Engineering, Design and Selection*, *3*(5), 433–442. https://doi.org/10.1093/protein/3.5.433

Gerrits, H., Mussmann, R., Bitter, W., Kieft, R., & Borst, P. (2002). The physiological significance of transferrin receptor variations in Trypanosoma brucei. *Molecular and Biochemical Parasitology*, *119*(2), 237–247. https://doi.org/10.1016/s0166-6851(01)00417-0

Giannetti, A. M., Halbrooks, P. J., Mason, A. B., Vogt, T. M., Enns, C. A., & Björkman, P. J. (2005). The molecular mechanism for receptor-stimulated iron release from the plasma iron transport protein transferrin. *Structure*, *13*(11), 1613–1623. https://doi.org/10.1016/j.str.2005.07.016

Giannetti, A. M., Snow, P. M., Zak, O., & Bjö Rkman, P. J. (2003). Mechanism for Multiple Ligand Recognition by the Human Transferrin Receptor. *PLoS Biology*, *1*(3), 341–350. https://doi.org/10.1371/journal.pbio.0000051

Graphodatsky, A. S., Trifonov, V. A., & Stanyon, R. (2011). The genome diversity and karyotype evolution of mammals. *Molecular Cytogenetics*, *4*(1), 22. https://doi.org/10.1186/1755-8166-4-22

Grünfelder, C. G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.-D., Boshart, M., & Overath, P. (2002). Accumulation of a GPI-Anchored Protein at the Cell Surface Requires Sorting at Multiple Intracellular Levels. *Traffic*, *3*, 547–559.

Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., … Lee, M. G. S. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryotic Cell*, *2*(3), 542–551. https://doi.org/10.1128/EC.2.3.542-551.2003

Hannaert, V., Bringaud, F., Opperdoes, F. R., & Michels, P. A. M. (2003, October 28). Evolution of energy metabolism and its compartmentation in Kinetoplastida. *Kinetoplastid Biology and Disease*, 2, 1–30. https://doi.org/10.1186/1475-9292-2-11

Haqqani, A. S., Thom, G., Burrell, M., Delaney, C. E., Brunette, E., Baumann, E., … Stanimirovic, D. B. (2018). Intracellular sorting and transcytosis of the rat transferrin receptor antibody OX26 across the blood-brain barrier *in vitro* is dependent on its binding affinity. *Journal of Neurochemistry*, *146*(6), 735–752. https://doi.org/10.1111/jnc.14482

Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., … Berriman, M. (2008). Telomeric Expression Sites Are Highly Conserved in Trypanosoma brucei. *PLoS ONE*, *3*(10), e3527. https://doi.org/10.1371/journal.pone.0003527

Higgins, M. K., Tkachenko, O., Brown, A., Reed, J., Raper, J., & Carrington, M. (2013). Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(5), 1905–1910. https://doi.org/10.1073/pnas.1214943110

Higgins, M. K., Tkachenko, O., Brown, A., Reed, J., Raper, J., & Carrington, M. (2013b). Structure of the trypanosome haptoglobin–hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(5), 1905–1910. https://doi.org/10.1073/pnas.1214943110

Hirumi, H., & Hirumi, K. (1989). Continuous Cultivation of Trypanosoma brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers. *The Journal of Parasitology*, *75*(6), 985. https://doi.org/10.2307/3282883

Isobe, T., Holmes, E. C., & Rudenko, G. (2003). The transferrin receptor genes of Trypanosoma equiperdum are less diverse in their transferrin binding site than those of the broad-host range Trypanosoma brucei. *Journal of Molecular Evolution*, *56*(4), 377–386. https://doi.org/10.1007/s00239-002-2408-z

Jeacock, L., Faria, J., & Horn, D. (2018). Codon usage bias controls mRNA and protein abundance in trypanosomatids. *ELife*, *7*. https://doi.org/10.7554/eLife.32496

Kabiri, M., & Steverding, D. (2000). Studies on the recycling of the transferrin receptor in Trypanosoma brucei using an inducible gene expression system. *European Journal of Biochemistry*, *267*(11), 3309–3314. https://doi.org/10.1046/j.1432-1327.2000.01361.x

Kabiri, M., & Steverding, D. (2001). Identification of a developmentally regulated iron superoxide dismutase of Trypanosoma brucei. *Biochemical Journal*, *360*(1), 173–177. https://doi.org/10.1042/bj3600173

Kennedy, P. G. E. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology*, *12*(2), 186–194. https://doi.org/10.1016/S1474-4422(12)70296-X

Kieft, R., Capewell, P., Turner, C. M. R., Veitch, N. J., MacLeod, A., & Hajduk, S. (2010). Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16137–16141. https://doi.org/10.1073/pnas.1007074107

Kooter, J. M., & Borst, P. (1984). Alpha-Amanitin-insensitive transcrption of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Research*, *12*(24), 9457–9472.

Kořený, L., Lukeš, J., & Oborník, M. (2010). Evolution of the haem synthetic pathway in kinetoplastid flagellates: An essential pathway that is not essential after all? *International Journal for Parasitology*, *40*(2), 149–156. https://doi.org/10.1016/j.ijpara.2009.11.007

Lake, J. A., De, V. F., Cruz, L. A., Ferreiraf, P. C. G., Morels, C., & Simpson, L. (1988).

Evolution of parasitism: Kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences (Trypanosoma/Leishmania/Leptomonas/Crithidia/kinetoplast DNA). In *Proc. Natl. Acad. Sci. USA*, 85.

Lambert, L. A., Perri, H., Halbrooks, P. J., & Mason, A. B. (2005, October 1). Evolution of the transferrin family: Conservation of residues associated with iron and anion binding. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, Vol. 142, pp. 129–141. https://doi.org/10.1016/j.cbpb.2005.07.007

Lane-Serff, H., MacGregor, P., Lowe, E. D., Carrington, M., & Higgins, M. K. (2014). Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor. *ELife*, *3*, e05553. https://doi.org/10.7554/eLife.05553

Liu, D., Albergante, L., Newman, T. J., & Horn, D. (2018). Faster growth with shorter antigens can explain a VSG hierarchy during African trypanosome infections: a feint attack by parasites. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-29296-8

Lloyd, C., Lowe, D., Edwards, B., Welsh, F., Dilks, T., Hardman, C., & Vaughan, T. (2009). Modelling the human immune response: Performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. *Protein Engineering, Design and Selection*. https://doi.org/10.1093/protein/gzn058

Lukeš, J., Skalický, T., Týč, J., Votýpka, J., & Yurchenko, V. (2014). Evolution of parasitism in kinetoplastid flagellates. *Molecular and Biochemical Parasitology*. https://doi.org/10.1016/j.molbiopara.2014.05.007

MacGregor, P., Gonzalez-Munoz, A. L., Jobe, F., Taylor, M. C., Rust, S., Sandercock, A. M., … Carrington, M. (2019). A single dose of antibody-drug conjugate cures a stage 1 model of African trypanosomiasis. *PLOS Neglected Tropical Diseases*, *13*(5), e0007373. https://doi.org/10.1371/journal.pntd.0007373

Mach, J., Tachezy, J., & Sutak, R. (2013). Efficient Iron Uptake via a Reductive Mechanism in Procyclic Trypanosoma brucei . *Journal of Parasitology*, *99*(2), 363–364. https://doi.org/10.1645/ge-3237.1

Macleod, O. J. S., Bart, J.-M., MacGregor, P., Peacock, L., Savill, N. J., Hester, S., … Carrington, M. (2020). A receptor for the complement regulator factor H increases transmission of trypanosomes to tsetse flies. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-15125-y

Maeda, N. (1985). Nucleotide Sequence of the Haptoglobin and Haptoglobin-related Gene Pair. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, *260*(11), 6698–6709.

Magnus, E., Vervoort, T., & Van Meirvenne, N. (1978). A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of T. b. gambiense trypanosomiasis. *Annales de La Societe Belge de Medecine Tropicale*, *58*(3), 169–176.

Maier, A., & Steverding, D. (1996). Low affinity of Trypanosoma brucei transferrin receptor to apotransferrin at pH 5 explains the fate of the ligand during endocytosis. *FEBS Letters*, *396*(1), 87–89. https://doi.org/10.1016/0014-5793(96)01073-3

Maier, A., & Steverding, D. (2008). Expression and purification of non-glycosylated Trypanosoma brucei transferrin receptor in insect cells. *Experimental Parasitology*, *120*(2), 205–207. https://doi.org/10.1016/j.exppara.2008.07.004

Marvin, D. A., & Hohn1, B. (1969). Filamentous Bacterial Viruses. *Bacteriological Reviews,* 33.

McCafferty, J., Griffiths, A. D., Winter, G., & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*. https://doi.org/10.1038/348552a0

McCann, A. K., Schwartz, K. J., & Bangs, J. D. (2008). A determination of the steady state lysosomal pH of bloodstream stage African trypanosomes. *Molecular and Biochemical Parasitology*, *159*(2), 146–149. https://doi.org/10.1016/j.molbiopara.2008.02.003

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/S0021889807021206

McLintock, L. M. L., Turner, C. M. R., & Vickerman, K. (1993). Comparison of the effects of immune killing mechanisms on Trypanosoma brucei parasites of slender and stumpy morphology. *Parasite Immunology*, *15*(8), 475–480. https://doi.org/10.1111/j.1365-3024.1993.tb00633.x

Mehlert, A., Bond, C. S., & Ferguson, M. a J. (2002). The glycoforms of a Trypanosoma brucei variant surface glycoprotein and molecular modeling of a glycosylated surface coat. *Glycobiology*, *12*(10), 607–612. https://doi.org/10.1093/glycob/cwf079

Mehlert, A., Wormald, M. R., & Ferguson, M. A. J. (2012). Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of Trypanosoma brucei. *PLoS Pathogens*, *8*(4), e1002618. https://doi.org/10.1371/journal.ppat.1002618

Mesu, V. K. B. K., Kalonji, W. M., Bardonneau, C., Mordt, O. V., Blesson, S., Simon, F., … Tarral, A. (2018). Oral fexinidazole for late-stage African Trypanosoma brucei gambiense trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial. *The Lancet*, *391*(10116), 144–154. https://doi.org/10.1016/S0140-6736(17)32758-7

Morgan, E. H. (1979). Studies on the mechanism of iron release from transferrin. *BBA - Protein Structure*, *580*(2), 312–326. https://doi.org/10.1016/0005-2795(79)90144-2

Morgan, E. H. (1983). Effect of pH and iron content of transferrin on its binding to reticulocyte receptors. *Biochimica et Biophysica Acta*, *762*(4), 498–502. https://doi.org/10.1016/0167-4889(83)90052-6

Mugnier, M. R., Cross, G. A. M., & Papavasiliou, F. N. (2015). The in vivo dynamics of antigenic variation in Trypanosoma brucei. *Science*, *347*(6229), 1470–1473. https://doi.org/10.1126/science.aaa4502

Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., … Siegel, T. N. (2018, November 1). Genome organization and DNA accessibility control

antigenic variation in trypanosomes. *Nature*, Vol. 563, pp. 121–125. https://doi.org/10.1038/s41586-018-0619-8

Nascimento, J. de F., Kelly, S., Sunter, J., & Carrington, M. (2018). Codon choice directs constitutive mRNA levels in trypanosomes. *ELife*, *7*. https://doi.org/10.7554/eLife.32467

Navarro, M., & Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature*, *414*(6865), 759–763. https://doi.org/10.1038/414759a

Newman, J., Swinney, H. L., & Day, L. A. (1977). Hydrodynamic properties and structure of fd virus. *Journal of Molecular Biology*, *116*(3), 593–603. https://doi.org/10.1016/0022-2836(77)90086-9

Nielsen, M. J., Petersen, S. V., Jacobsen, C., Oxvig, C., Rees, D., Møller, H. J., & Moestrup, S. K. (2006). Haptoglobin-related protein is a high-affinity hemoglobin-binding plasma protein. *Blood*, *108*(8), 2846–2849. https://doi.org/10.1182/blood-2006-05-022327

Noinaj, N., Easley, N. C., Oke, M., Mizuno, N., Gumbart, J., Boura, E., … Buchanan, S. K. (2012). Structural basis for iron piracy by pathogenic Neisseria. *Nature*, *483*(7387), 53–61. https://doi.org/10.1038/nature10823

Odiit, M., Kansiime, F., & Enyaru, J. C. K. (1997). Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda. *East African Medical Journal*, *74*(12), 792–795.

Pal, A., Hall, B. S., Jeffries, T. R., & Field, M. C. (2003). Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in Trypanosoma brucei. *Biochemical Journal*, *374*(2), 443–451. https://doi.org/10.1042/BJ20030469

Pearse, B. M. F., & Robinson, M. S. (1990). Clathrin, Adaptors, and Sorting. *Annual Review of Cell Biology*, *6*(1), 151–171. https://doi.org/10.1146/annurev.cb.06.110190.001055

Pérez-Morga, D., Vanhollebeke, B., Paturiaux-Hanocq, F., Nolan, D. P., Lins, L., Homblé, F., … Pays, E. (2005). Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science*, *309*(5733), 469–472. https://doi.org/10.1126/science.1114566

Pinger, J., Nešić, D., Ali, L., Aresta-Branco, F., Lilic, M., Chowdhury, S., … Stebbins, C. E. (2018). African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. *Nature Microbiology*, *3*(8), 932–938. https://doi.org/10.1038/s41564-018-0187-6

Povelones, M. L., Gluenz, E., Dembek, M., Gull, K., & Rudenko, G. (2012). Histone H1 Plays a Role in Heterochromatin Formation and VSG Expression Site Silencing in Trypanosoma brucei. *PLoS Pathogens*, *8*(11). https://doi.org/10.1371/journal.ppat.1003010

Raper, J., Fung, R., Ghiso, J., Nussenzweig, V., & Tomlinson, S. (1999). Characterization of a novel trypanosome lytic factor from human serum. *Infection and Immunity*, *67*(4), 1910–1916.

Ross, R., & Thomson, D. (1910). A Case of Sleeping Sickness Studied by Precise Enumerative Methods: Further Observations. *Proceedings of the Royal Society B: Biological Sciences*, Vol. 83, pp. 187–205. https://doi.org/10.1098/rspb.1911.0003

Salmon, D., Geuskens, M., Hanocq, F., Hanocq-Quertier, J., Nolan, D., Ruben, L., & Pays, E. (1994). A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in T. brucei. *Cell*, *78*(1), 75–86. https://doi.org/10.1016/0092-8674(94)90574-6

Salmon, D., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Pays, A., Tebabi, P., Nolan, D. P., … Pays, E. (1997). Characterization of the ligand-binding site of the transferrin receptor in Trypanosoma brucei demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. *EMBO Journal*, *16*(24), 7272–7278. https://doi.org/10.1093/emboj/16.24.7272

Salmon, D., Paturiaux-Hanocq, F., Poelvoorde, P., Vanhamme, L., & Pays, E. (2005). Trypanosoma brucei: Growth differences in different mammalian sera are not due to the species-specificity of transferrin. *Experimental Parasitology*, *109*(3), 188–194. https://doi.org/10.1016/j.exppara.2004.11.010

Schell, D., Borowy, N. K., & Overath, P. (1991). Transferrin is a growth factor for the bloodstream form of Trypanosoma brucei. *Parasitology Research*, *77*(7), 558–560. https://doi.org/10.1007/BF00931012

Schell, D., Evers, R., Preis, D., Ziegelbauer, K., Kiefer, H., Lottspeich, F., … Overath, P. (1991). A transferrin-binding protein of *Trypanosoma brucei* is encoded by one of the genes in the variant surface glycoprotein gene expression site. *The EMBO Journal*, *10*(5), 1061–1066.

Schwartz, K. J., Peck, R. F., Tazeh, N. N., & Bangs, J. D. (2005). GPI valence and the fate of secretory membrane proteins in African trypanosomes. *Journal of Cell Science*, *118*(23), 5499–5511. https://doi.org/10.1242/jcs.02667

Schwede, A., Jones, N., Engstler, M., & Carrington, M. (2011). The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. *Molecular and Biochemical Parasitology*, *175*(2), 201–204. https://doi.org/10.1016/j.molbiopara.2010.11.004

Scott, J. K., & Smith, G. P. (1990). Searching for peptide ligands with an epitope library. *Science*, *249*(4967), 386–390. https://doi.org/10.1126/science.1696028

Shih, Y. J., Baynes, R. D., Hudson, B. G., Flowers, C. H., Skikne, B. S., & Cook, J. D. (1990). Serum Transferrin Receptor Is a Truncated Form of Tissue Receptor. *The Journal of Biological Chemistry,* 265.

Simarro, P. P., Jannin, J., & Cattand, P. (2008, February). Eliminating human African trypanosomiasis: Where do we stand and what comes next? *PLoS Medicine*, 5, 174-180. https://doi.org/10.1371/journal.pmed.0050055

Smith, A. B., Esko, J. D., & Hajduk, S. L. (1995). Killing of trypanosomes by the human haptoglobin-related protein. *Science*, *268*(5208), 284–286. https://doi.org/10.1126/science.7716520

Steverding, D. (1998). Bloodstream forms of Trypanosoma brucei require only small amounts of iron for growth. *Parasitology Research*, *84*(1), 59–62.

Steverding, D. (2003, March 1). The significance of transferrin receptor variation in Trypanosoma brucei. *Trends in Parasitology*, Vol. 19, pp. 125–127. https://doi.org/10.1016/S1471-4922(03)00006-0

Steverding, D., Sexton, D. W., Chrysochoidi, N., & Cao, F. (2012). Trypanosoma brucei transferrin receptor can bind C-lobe and N-lobe fragments of transferrin. *Molecular and Biochemical Parasitology*, *185*(2), 99–105. https://doi.org/10.1016/j.molbiopara.2012.06.007

Steverding, D., Stierhof, Y.-D., Fuchs, H., Tauber, R., & Overath, P. (1995). Transferrin-binding Protein Complex Is the Receptor for Transferrin Uptake in Trypanosoma brucei. *The Journal of Cell Biology*, *131*(5), 1173–1182.

Steverding, D., Stierhof, Y. D., Chaudhri, M., Ligtenberg, M., Schell, D., Beck-Sickinger, A. G., & Overath, P. (1994). ESAG 6 and 7 products of Trypanosoma brucei form a transferrin binding protein complex. *European Journal of Cell Biology*, *64*(1), 78–87.

Stewart, E. J., Åslund, F., & Beckwith, J. (1998). Disulfide bond formation in the Escherichia coli cytoplasm: An in vivo role reversal for the thioredoxins. *EMBO Journal*, *17*(19), 5543–5550. https://doi.org/10.1093/emboj/17.19.5543

Stijlemans, B., Baetselier, P. De, Caljon, G., Van Den Abbeele, J., Van Ginderachter, J. A., & Magez, S. (2017, June 30). Nanobodies As tools to Understand, diagnose, and treat African trypanosomiasis. *Frontiers in Immunology*, Vol. 8. https://doi.org/10.3389/fimmu.2017.00724

Stijlemans, B., Beschin, A., Magez, S., Van Ginderachter, J. A., & De Baetselier, P. (2015). *Iron Homeostasis and Trypanosoma brucei Associated Immunopathogenicity Development: A Battle/Quest for Iron*. https://doi.org/10.1155/2015/819389

Sutherland, C. S., Yukich, J., Goeree, R., & Tediosi, F. (2015). A Literature Review of Economic Evaluations for a Neglected Tropical Disease: Human African Trypanosomiasis (Sleeping Sickness). *PLoS Neglected Tropical Diseases*, *9*(2), 1–22. https://doi.org/10.1371/journal.pntd.0003397

Symula, R. E., Beadell, J. S., Sistrom, M., Agbebakun, K., Balmer, O., Gibson, W., … Caccone, A. (2012). Trypanosoma brucei gambiense Group 1 Is Distinguished by a Unique Amino Acid Substitution in the HpHb Receptor Implicated in Human Serum Resistance. *PLoS Neglected Tropical Diseases*, *6*(7), e1728. https://doi.org/10.1371/journal.pntd.0001728

Taylor, M. C., & Kelly, J. M. (2010). Iron metabolism in trypanosomatids, and its crucial role in infection. *Parasitology*, *137*(6), 899–917. https://doi.org/10.1017/S0031182009991880

Taylor, M. C., Mclatchie, A. P., & Kelly, J. M. (2013). Evidence that transport of iron from the lysosome to the cytosol in African trypanosomes is mediated by a mucolipin orthologue. *Molecular Microbiology*, *89*(3), 420–432. https://doi.org/10.1111/mmi.12285

Thomson, R., & Finkelstein, A. (2015). Human trypanolytic factor APOL1 forms pH-gated cation-selective channels in planar lipid bilayers: Relevance to trypanosome lysis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(9), 2894–2899. https://doi.org/10.1073/pnas.1421953112

Tiengwe, C., Bush, P. J., & Bangs, J. D. (2017). Controlling transferrin receptor trafficking with GPI-valence in bloodstream stage African trypanosomes. *PLOS Pathogens*, *13*(5), e1006366. https://doi.org/10.1371/journal.ppat.1006366

Tiengwe, C., Muratore, K. A., & Bangs, J. D. (2016a). Surface proteins, ERAD and antigenic variation in Trypanosoma brucei. *Cellular Microbiology*, *18*(11), 1673–1688. https://doi.org/10.1111/cmi.12605

Tóthová, C., Nagy, O., Nagyová, V., & Kováč, G. (2016). The concentrations of selected blood serum proteins in calves during the first three months of life. *Acta Veterinaria Brno*, *85*, 33–40. https://doi.org/10.2754/avb201685010033

Trevor, C. E., Gonzalez-Munoz, A. L., Macleod, O. J. S., Woodcock, P. G., Rust, S., Vaughan, T. J., … Higgins, M. K. (2019). Structure of the trypanosome transferrin receptor reveals mechanisms of ligand recognition and immune evasion. *Nature Microbiology*, Vol. 4, pp. 2074–2081. https://doi.org/10.1038/s41564-019-0589-0

Tsuji, S., Kato, H., Matsuoka, Y., & Fukushima, T. (1984). Molecular weight heterogeneity of bovine serum transferrin. *Biochemical Genetics*, *22*(11–12), 1145–1159. https://doi.org/10.1007/bf00499638

Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homblé, F., … Pays, E. (2013). Mechanism of Trypanosoma brucei gambiense resistance to human serum. *Nature*, *501*(7467), 430–434. https://doi.org/10.1038/nature12516

Van Luenen, H. G. A. M., Kieft, R., Mußmann, R., Engstler, M., Ter Riet, B., & Borst, P. (2005). Trypanosomes change their transferrin receptor expression to allow effective uptake of host transferrin. *Molecular Microbiology*, *58*(1), 151–165. https://doi.org/10.1111/j.1365-2958.2005.04831.x

Van Xong, H., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van Den Abbeele, J., Pays, A., … Pays, E. (1998). A VSG expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell*, *95*(6), 839–846. https://doi.org/10.1016/S0092-8674(00)81706-7

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D. P., Lins, L., Abbeele, J. Van Den, … Pays, E. (2003). Apolipoprotein L-1 is the trypanosome lytic factor of human serum. *Nature*, *422*, 83–87. https://doi.org/10.1038/nature01457.1.

Vanhamme, L., & Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiological Reviews*, 59, pp. 223–240. https://doi.org/10.1128/mmbr.59.2.223-240.1995

Vanhollebeke, B., De Muylder, G., Nielsen, M. J., Pays, A., Tebabi, P., Dieu, M., … Pays, E. (2008). A haptoglobin-hemoglobin receptor conveys innate immunity to Trypanosoma brucei in humans. *Science*, *320*(5876), 677–681. https://doi.org/10.1126/science.1156296

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., … Johnson, K. S. (1996). Human Antibodies With Sub-Nanomolar Affinities Isolated From A Large Non-Immunized Phage Display Library. *Nature Biotechnology*. https://doi.org/10.1038/nbt0396-309

Vickerman, K. (1969). On The Surface Coat and Flagellar Adhesion in Trypanosomes. *Journal of Cell Science*, *5*(1), 163–193.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin*, *41*(2), 105–114. https://doi.org/10.1093/oxfordjournals.bmb.a072036

Vidarsson, G., Dekkers, G., & Rispens, T. (2014). IgG subclasses and allotypes: From structure to effector functions. *Frontiers in Immunology*, *5*(OCT). https://doi.org/10.3389/fimmu.2014.00520

Wally, J., Halbrooks, P. J., Vonrhein, C., Rould, M. A., Everse, S. J., Mason, A. B., & Buchanan, S. K. (2006). The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. *Journal of Biological Chemistry*, *281*(34), 24934–24944. https://doi.org/10.1074/jbc.M604592200

Wang, M., Lai, T. P., Wang, L., Zhang, H., Yang, N., Sadler, P. J., & Sun, H. (2015). “Anion clamp” allows flexible protein to impose coordination geometry on metal ions. *Chemical Communications (Cambridge, England)*, *51*(37), 7867–7870. https://doi.org/10.1039/c4cc09642h

Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T., & Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. *Nature*. https://doi.org/10.1038/341544a0

Webb, H., Burns, R., Kimblin, N., Ellis, L., & Carrington, M. (2005). A novel strategy to identify the location of necessary and sufficient cis-acting regulatory mRNA elements in trypanosomes. *RNA*, *11*(7), 1108–1116. https://doi.org/10.1261/rna.2510505

Williams, J., & Moreton, K. (1980). The distribution of iron between the metal-binding sites of transferrin in human serum. *Biochemical Journal*, *185*(2), 483–488. https://doi.org/10.1042/bj1850483

Yang, N., Zhang, H., Wang, M., Hao, Q., & Sun, H. (2012). Iron and bismuth bound human serum transferrin reveals a partially-opened conformation in the N-lobe. *Scientific Reports*, *2*. https://doi.org/10.1038/srep00999

Young, R., Taylor, J. E., Kurioka, A., Becker, M., Louis, E. J., & Rudenko, G. (2008). Isolation and analysis of the genetic diversity of repertoires of VSG expression site containing telomeres from Trypanosoma brucei gambiense, T. b. brucei and T. equiperdum. *BMC Genomics*, *9*, 385. https://doi.org/10.1186/1471-2164-9-385

Yu, C., Crispin, M., Sonnen, A. F. P., Harvey, D. J., Chang, V. T., Evans, E. J., … Davis, S. J. (2011). Use of the α-mannosidase I inhibitor kifunensine allows the crystallization of apo CTLA-4 homodimer produced in long-term cultures of Chinese hamster ovary cells. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *67*(7), 785–789. https://doi.org/10.1107/S1744309111017672

Ziegelbauer, K., & Overath, P. (1992). Identification of invariant surface glycoproteins in the bloodstream stage of Trypanosoma brucei. *Journal of Biological Chemistry*, *267*(15), 10791–10796.

Zoll, S., Lane-Serff, H., Mehmood, S., Schneider, J., Robinson, C. V., Carrington, M., & Higgins, M. K. (2018). The structure of serum resistance-associated protein and its implications for human African trypanosomiasis. *Nature Microbiology*, *3*(3), 295–301. https://doi.org/10.1038/s41564-017-0085-3