MUTATIONAL ANALYSIS OF A GENE REQUIRED FOR FLAGELLAR MOTILITY IN THE AFRICAN SLEEPING SICKNESS PARASITE

by

Sonia N. Dantas

A Paper Presented to the

Faculty of Mount Holyoke College in

Partial Fulfillment of the Requirements for

the Degree of Bachelor of Arts with

Honor

Department of Biological Sciences,

South Hadley, MA 01075

May, 2008

This paper was prepared

under the direction of

Professor Amy Springer

for eight credits

ACKNOWLEDGMENTS

I would like to thank my advisor, Amy Springer, for all of the invaluable assistance and advice with which she has provided me. She is an exceptional scientist and mentor. It was an honor for me to have been given the opportunity to work with her.

I would like to my committee members, Sheila Browne and Robin Herlands. Sheila's love and dedication to her students is remarkable. I thoroughly enjoyed being her pupil and research assistant and would not be the chemist I am today without her wisdom and guidance. Robin is one of the best professors I have had at Mount Holyoke, and she has truly instilled a love of immunology in me. I am very appreciative of her support and all of her anecdotes that kept me entertained during the long hours I spent working in Carr.

A special thanks to Mary Glackin for all of her assistance in all of the predictions performed in this experiment and to our collaborators at UCLA for making this research possible.

Thank you Angie Diciccio, Amy Dubuc, Lindsay Goodale, Laura Fernandez, Julie Hare, Lindsey Poole, and Jessi Wiwczar for your friendship and support. You have made these past four years truly wonderful. Lastly, I would like to thank my family for all of the sacrifices they have made to help me fulfill my dreams.

TABLE OF CONTENTS

Page No.

LIST OF FIGURES

Page No.

LIST OF TABLES

ABSTRACT

Trypanosoma brucei is a flagellated, unicellular protozoan that causes African Sleeping Sickness, a disease infecting humans and livestock in sub-Saharan Africa. The flagellum plays an integral role in pathogenicity of the organism. It is comprised of elements conserved in many organisms, making *T. brucei* a popular model for understanding flagellar components.

Components of motile flagella (CMF) genes are found in all organisms with motile flagella. The knock down of CMF63 expression demonstrated that loss of the gene has moderate effects on flagellar motility. The gene has an EF-hand region, indicating that this protein may play a role in intracellular signaling.

The purpose of my research is to better understand the function of CMF63 in *T. brucei*. Through homology searches, secondary structure, and helical wheel predictions, it was determined that a predicted α-helix from 393 – 404 amino acids may be involved in interactions with other secondary structures within the CMF63 protein or with other proteins. *T. brucei* was transfected with three different mutant CMF63 genes that were generated to perturb the potential interactions. By viewing the phenotypes of the mutants, it may be possible to ascertain if the α-helix plays an integral role in the function of CMF63 by interacting with another body.

INTRODUCTION

Human African Trypanosomiasis

Endemicity

Human African Trypanosomiasis (HAT), more commonly known as African Sleeping Sickness, is a tropical, vector-borne disease caused by haemoflagellates belonging to the genus *Trypanosoma* (Pépin and Meda, 2001). More specifically, three subspecies (ssp) of *Trypanosoma- T. brucei brucei*, *T. b. rhodesiense*, and *T. b. gambiense*, are responsible for infecting humans and livestock in 36 countries of sub-Saharan Africa (Matthews, 2005; Pépin and Meda, 2001). As of 1998, greater than 60 million people live within the "tsetse fly 'belt,'" which lies between the 15º north and south latitudes, comprising a geographical area the size of the United States, India, and Western Europe combined (Blum and Burri, 2002; Pépin and Meda, 2001). Within this area, 300,000 – 500,000 cases and 70,000 deaths are reported annually (Blum and Burri, 2002; Matthews, 2005). It is believed that these numbers are grossly underestimated, as this region is bereft of proper communications systems and medical care, and there is great political instability in the region (Pépin and Meda, 2001).

Trypanosome life cycle

African trypanosomes have a complex life cycle involving both a tsetse fly (*Glossina* ssp) and a vertebrate vector (Matthews, 2005). As it is an extracellular pathogen, cell motility is essential for its viability throughout their life cycle. While it is in the insect vector, *T. brucei* must make a series of migrations to different compartments within the host, starting the midgut and ending in the salivary gland, to undergo different developmental changes that will enable it to survive in a mammalian host (Matthews, 2005). Once in the mammalian host, it initially replicates in the bloodstream. Then it infects the connective tissues and central nervous system (Matthews, 2005).

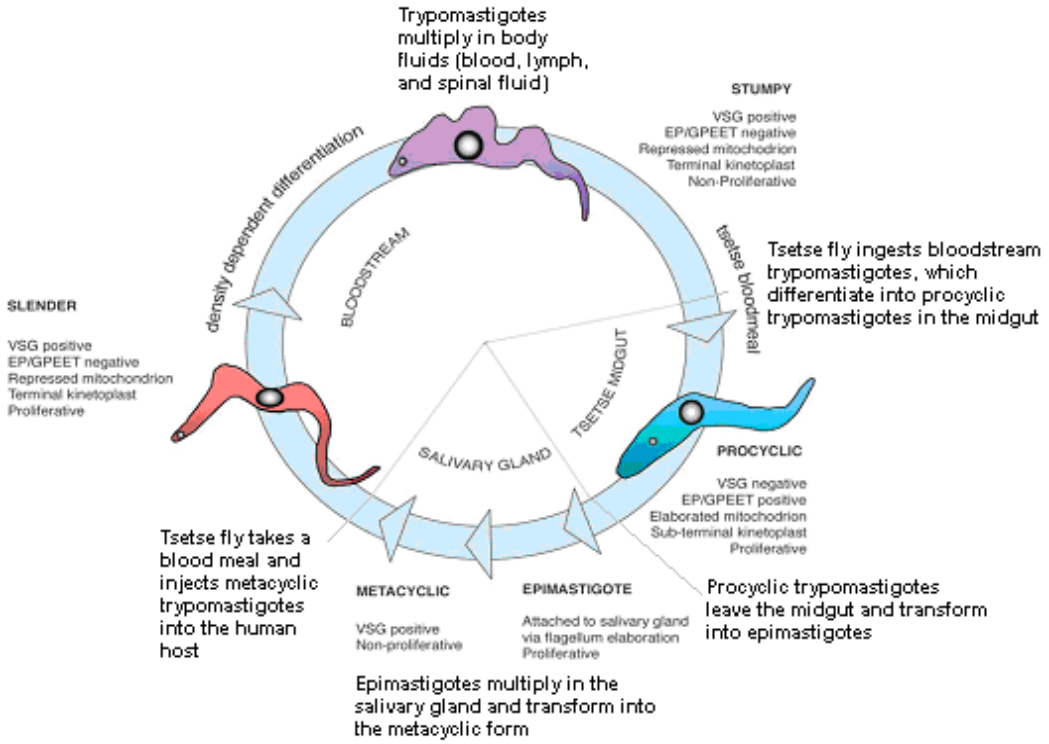


Figure 1. Representation of the stages in the life cycle of *T. brucei*. In response to environmental stimuli generated by the host (insect or mammalian vector), trypanosomes undergo different morphological changes that are better suited for proliferating within the host. Once they have infect the midgut of the tsetse fly, trypanosomes stop dividing and migrate to the salivary gland, where they differentiate into the proliferative epimastigote form and attach to the wall of the salivary gland. Eventually, they stop dividing and morph into the non-proliferative metacyclic form. When tsetse fly takes a blood meal from a mammalian host, the metacyclic trypanosomes differentiate into the slender, bloodstream form and eventually into the stumpy, bloodstream form, which are proliferative and non-proliferative, respectively. Then, these trypanosomes can either infect invade the central nervous system of the host or ingested by another tsetse fly during a blood meal (Adapted from Matthews, 2005).

In order to adapt to the different compartments within the tsetse fly and the bloodstream of the mammalian host, trypanosomes modify their cell morphology, gene expression and signaling pathways (Fenn and Matthews, 2007). Figure 1 outlines the morphological changes that the trypanosomes undergo. Subsequent to taking a blood meal from an infected mammal, the bloodstream trypomastigotes differentiate into procyclic trypomastigotes within the midgut of

the tsetse fly (Vickerman, et al., 1988). Then the parasites migrate to the ectoperiotrophic space, through the proventriculus, and into the foregut. Within the foregut, they differentiate into the postmesocylic epimastigotes, which are elongated and divide asymmetrically. The epimastigotes travel through the proboscis and hypopharynx and invade the salivary gland, the site of the final stage of development within the insect vector (Vickerman, et al., 1988). Small epimastigotes are generated and attach to the gland epithelium using their flagella. Finally, they differentiate into metacyclic trypomastigotes coated with variant-surface glycoproteins (VSG). After this final transformation, the trypomastigotes are equipped to survive in the bloodstream of the mammalian host (Vickerman, et al., 1988).

When the trypanosomes enter the bloodstream, they differentiate into long, slender forms and multiply extracellularly for a period that can range from weeks to months (Vincedeau and Bouteille, 2006). Once the population size increases to a certain size, an external signal, stumpy induction factor, is released to inhibit proliferation and promote the formation of a stumpy cell body (Fenn and Matthews, 2007). Then they penetrate the blood vessel endothelium, disseminate in the connective tissues, and invade the central nervous system (CNS) (Vincedeau and Bouteille, 2006).

After being bitten by a tsetse fly, the human host develops a chancre, or skin lesion, that disappears after two to three weeks (Vincedeau and Bouteille, 2006). It is characterized by local erythema, edema, heat, tenderness, and the

absence of pus. There are two phases of infection within the mammalian host: the haemolymphatic stage and the meningoencephalitic stage (Vincedeau and Bouteille, 2006).

Stage I: Haemolymphatic stage

The haemolymphatic stage begins within a few days after the bite of the tsetse fly (Vincedeau and Bouteille, 2006). The occurrence of sporadic fevers, accompanied by headaches, malaise, fatigue, anorexia, muscle and joint pain, anemia, rashes, and deep hyperesthesia (increased sensitivity to sensory stimuli), results from the invasions of trypanosomes into the blood, arise in successive waves. When the trypanosomes invade the reticuloendothelial system, glands, particularly the lymphatic glands, and the spleen become enlarged. Skin eruptions and severe itching are commonly seen among patients (Vincedeau and Bouteille, 2006).

Hosts develop intermittent fevers, weakness, and weight loss (Kennedy, 2004). Endocrine disease is marked by a constant feeling of coldness, loss or increase of appetite, excessive thirst, impotence, absence of menstruation, and infertility (Blum and Burri, 2002; Vincedeau and Bouteille, 2006).

Stage II: Meningoencephalitic Stage

The meningoencephalitic stage develops over the course of months to years (Vincedeau and Bouteille, 2006). Neurological symptoms appear gradually, and a multiplicity of symptoms, predominately somnolence and insomnia, are seen. The disruptions in the circadian rhythms of the sleep-wake cycle can be

coupled with headaches, hyperesthesia, psychiatric disorders (including confusion, mood swings, agitation, aggression, euphoria, mutism, and indifference), and tremors (Blum and Burri, 2006; Vincedeau and Bouteille, 2006).

Evasion of the Immune System

Variable Surface Glycoproteins (VSGs)

Most flagellated, pathogenic protozoans are able to evade the innate and adaptive immune responses infecting the host as intracellular parasites (Stijlemans et al., 2007). Trypanosomes are strictly extracellular parasites, making them susceptible to attack by the immune system of the host (Stijlemans et al., 2007). These organisms are able to elude the immune system of their human host via antigenic variation, a process involving the sequential expression of VSGs that are antigenically distinct from one another (Matthews, 2005; Stijlemans et al., 2007).

The immune system is capable of detecting the VSGs on the surface of the trypanosome and inciting an immune response against the parasites by targeting these VSGs (Bisser et al., 2006). This prevents the trypanosomes from overpopulating the bloodstream, enabling them to survive in the host for an extended period of time (Bisser et al., 2006). Trypanosomes escape complete decimation by periodically switching the VSG presented on the cell surface, allowing them to avoid immunoglobulin recognition. VSGs are the principal surface antigens of African trypanosomes (Bisser et al., 2006; Vincedeau and Bouteille, 2006); approximately 10^7 VSGs blanket the outer membrane of the

procyclic and epimastigote forms (Donelson, 2002). Due to the presence of this substantial number of VSGs, other invariant antigens and epitopes on the cell surface are hidden from detection for them immune system. This also increases the trypanosomes' chance of survival (Donelson et al., 1998; Bisser et al., 2006).

There are greater than 1000 VSG genes and pseudogenes with approximately 20 telomeric VSG expression sites, but only one gene or psuedogene is expressed at a given time (Taylor and Rudenko, 2006). African trypanosomes are capable of switching VSGs at a rate of 10^{-2} to 10^{-7} switches per doubling time, which ranges between 5-10 hours (Donelson, 2002). During switching, duplicative transpositions or other VSG gene rearrangements can occur, increasing the variability in the structure of the glycoprotein (Donelson et al., 1998).

Cytokines and chemokines

In addition to protecting trypanosomes from the immune system, VSGs can have deleterious effects on the host (Vincedeau and Bouteille, 2006). As seen in Figure 2, VSGs can negatively impact the integrity of the immune system. They can inhibit trypanosome lysis via complement (a cascade of protein interactions occurring extracellularly that ultimately lead to phagocytosis of the pathogen), stimulate the constant production of cytokines IL-1 and TNFα (cytokines that promote an inflammatory response), and cause the generation of autoantibodies using molecular mimicry of host tissues. Some autoantibodies that

have been reported were directed against red blood cells, liver and cardiolipids, and components of CNS myelin (Vincedeau and Bouteille, 2006).

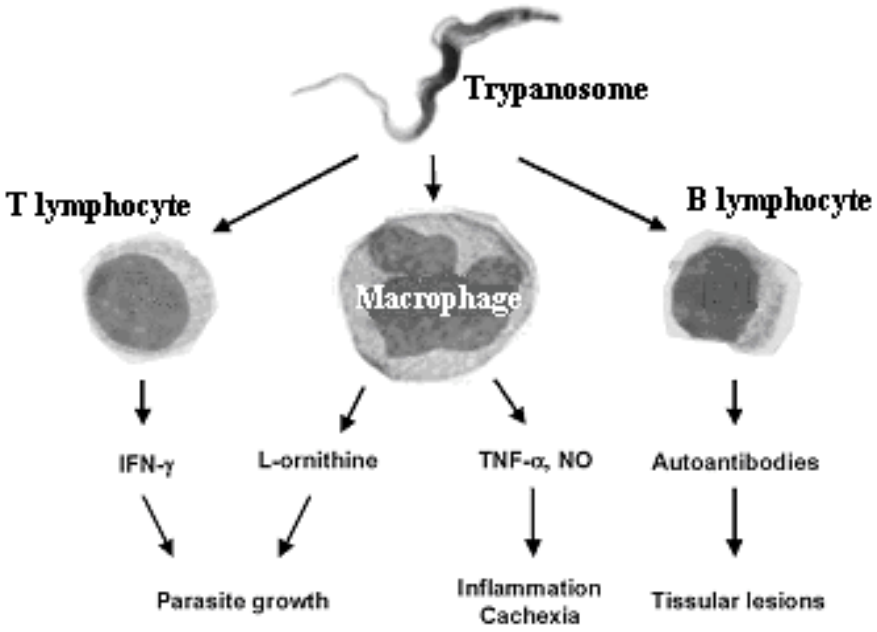


Figure 2. The effects of VSGs on the immune system of a human host. Trypanosomes induce immune cells to secrete various molecules that are harmful to host cells and/or promote the growth of the parasites. Adapted from Vincedeau and Bouteille, 2006.

Trypanosomes cause the levels of IFN-γ, a cytokine that promotes the inflammatory response, to become elevated in the mammalian host (Donelson et al., 1998). Trypanosomes stimulate $CD8^+$ T cells to produce IFN-γ by secreting a protein called T lymphocyte triggering factor (TLTF). IFN-γ activates the production of MAP kinase in African trypanosomes, a mitogen-activated protein that may promote the proliferation of the bloodstream form of trypanosomes (Donelson et al., 1998).

Elevated levels of TNF-α have been found to cause inflammatory lesions of the CNS. Chemokines recruit macrophages and lymphocytes to the CNS,

which can lead to sleep and endocrine disorders (Vincedeau and Bouteille, 2006). Cytokines promote the production of NO, which can attack both trypanosomes and the host cells. Nitrosylated compounds affect tissue functions, such as neurotransmission, depending upon their location (spleen, liver, CNS, or other parts of the body) (Vincedeau and Bouteille, 2006).

Treatment

Current treatment of HAT is not sufficient, as there are many problems with the medications that have yet to be resolved (Keiser et al., 2001). Due to the presence of VSGs, efforts to produce a vaccine have been unsuccessful, and the only treatment available is the use of chemotherapy (Delespaux and Koning, 2007). The effectiveness of trypanocides is dependent upon whether they can reach the areas of the host in which the trypanosomes reside and on the ability of trypanosomes to uptake the drugs. Most of the drugs that are available cannot pass the blood brain barrier (BBB), making them ineffective in treating late stage Trypanosomiasis, when the trypanosomes invade the central nervous system and reside in the cerebral spinal fluid. If the trypanosomes do not take up the drugs, then they are considered to be drug resistant. (Delespaux and Koning, 2007; Nok, 2003). According to the authors, an additional obstacle is that there has been little interest in developing new compounds because it is not seen as a profitable venture; HAT infects the populations belonging to the poorest nations (Keiser et al., 2001; Barrett, et al., 2007).

Pentamidine and suramin are drugs that primarily remain in the lymphatic system (Nok, 2003). Use of these drugs to treat early stage HAT has been very successful. Melarsoprol, an organo-asenical compound, possesses the ability to cross the BBB and kill both *T. b. rhodesiense* and *T. b. gambiense*, making it a favorable drug for treatment (Nok, 2003 and Fairlamb, 2003). Although these drugs have been effective in treating HAT, they are highly toxic and can have deleterious side effects (Fairlamb, 2003).

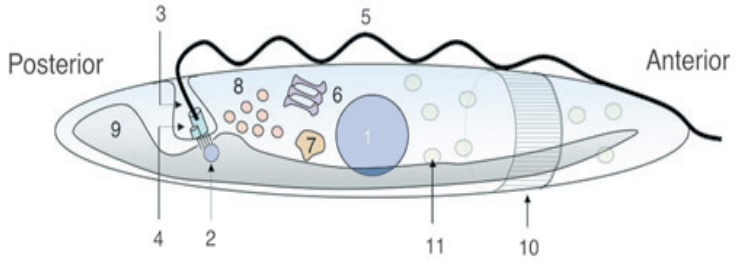
Pentamidine, an aromatic diamine, inhibits the accumulation of lysine and the transport of several amino acids and polyamines (Nok, 2003). After it is taken up by the trypanosome, the drug hinders nucleic acid replication in the parasite. Intracellular concentrations of pentamidine must be greater than 1 mM within the trypanosomes for it to be lethal to the organisms. As only low levels of pentamidine are found in the cerebral spinal fluid (CSF), the concentrations are not high enough to clear the infection from the CNS (Nok, 2003). Consequently this drug is only effective in the first stage of infection. Pentamidine can generate nephrotoxicity and diabetes mellitus and binds strongly to imidazole receptors, leading to hypertension and a few other side effects (Nok, 2003).

Suramin is a derivative of naphthalene that is symmetrical and polysulphonated (Nok, 2003). As it has six negative charges at physiological pH, it cannot diffuse across biological membranes, making it impossible for the drug to cross the BBB. Suramin is commonly used in early stage HAT, as it is particularly effective in combating *T. b. rhodesiense* infections (Nok, 2003). The

compound's charge gives it the ability to bind easily to many serum proteins. It is possible that the drug enters trypanosomes by binding to low-density lipoproteins (LDL), which are taken up by the parasites via endocytosis (Nok, 2003). The drug can generate some life-threatening side effects, such as vomiting, shock, kidney damage, agranulocytosis, hemolytic anemia, jaundice, and diarrhea (Fairlamb, 2003).

Melarsoprol is a trivalent arsenical derivative (Fairlamb, 2003). When it is administered to patients, it is rapidly converted to melarsin oxide, which can rapidly and reversibly bind to serum proteins and acts as an enzyme inhibitor. Once it is taken up by a trypanosome, the drug inhibits motility and lyses the parasite (Delespaux and Koning, 2007; Fairlamb, 2003). The use of melarsoprol generates adverse effects, such as the development of encephalopathy, as well as polyneuropathy, fever, abdominal pain, diarrhea, and death (Nok, 2003; Fairlamb, 2003).

Trypanosome cell structure



1. Nucleus
2. Kinetoplast
3. Flagellar pocket
4. Basal body and probasal body
5. Axoneme and paraflagellar rod
6. Golgi
7. Lysosome
8. Endosomes
9. Mitochondrion
10. Microtubule cytoskeleton
11. Glycosomes

Figure 3. Trypanosome cell structure. An illustration highlighting the major features of a trypanosome cell. The (3) flagellar pocket is the site where the (5) flagellum exits the cell and the only place where endocytosis and exocytosis occur. (10) Shows the microtubule cytoskeleton that lies beneath the cell membrane (Matthews, 2005).

Figure 3 illustrates the cell structure of a trypanosome. The cell body of a trypanosome is somewhat cylindrical in shape and has tapered anterior and posterior ends (Matthews, 2005). It is approximately 10 to 20 μm in length but can be much longer in different stages of development within the tsetse fly (Landfear and Ignatushchenko, 2001; Hill, 2003). The cell shape is defined by its microtubule cytoskeleton, which has a definite polarity, where the minus ends are at the anterior end, while the plus ends are at the posterior end (Matthews, 2005). The organelles, including the flagellar pocket, kinetoplast, mitochondrion, and nucleus are primarily localized between the posterior end and the center of the cell (Matthews, 2005).

Flagellar structure

T. brucei propels itself with a single flagellum, which starts at the basal body near the posterior end of the cell and extends along the length of the cell body (Matthews, 2005; Hill 2003). Beating is generally propagated from base to tip in flagellated organisms (Holwill, 1979). However, flagellar beating is unusual in *T. brucei*, as it has two waveforms, propagating from the tip of the flagellum to its base and propagating in the reverse motion (base to tip). By reversing the direction of flagellar beat, the trypanosome is able to change direction (Branche et al., 2006; Ralston et al., 2006). They change waveforms depending on environmental cues (Branche et al., 2006; Ralston et al., 2006).

The flagellum has a distinct membrane that is adjacent to the plasma membrane (Hill, 2003). At the posterior end of the cell lies the flagellar pocket, a specialized compartment formed by an invagination of the plasma membrane and the location from which the flagellum exits the cell (Matthews, 2005; Hill 2003). It is the only site for endocytosis and exocytosis (Matthews, 2005).

The flagellum has a typical axoneme, which consists of nine outer doublet microtubules that encircle a central pair of microtubules (9 + 2 configuration). Additionally, the flagellum has a paraflagellar rod (PFR) and a flagellum attachment zone (FAZ), which are found only in trypanosomes and a few other protozoa. PFR is a fibrous body composed of filaments that extend along the length of the body (Landfear and Ignatushchenko, 2000). FAZ consists of an electron-dense filament (composition unknown) and four specialized

microtubules that have distinct biochemical characteristics. It links the axoneme and the paraflagellar rod together and links both of these structures to the junction between the flagellum and plasma membrane. A defect in any of these components may affect motility (Rocha et al., 2006; Hutchings et. al., 2002). ATP-dependent dynein motors are used to power flagellar movement. Nexin links connect the nine outer doublet microtubules together, and radial spokes attach the outer doublets to the central pair; these attachments limit sliding, causing the flagellum to bend (Alberts et al., 2001).

The importance of cilia and flagella

Flagella and cilia are highly conserved organelles, which are identical in eukaryotes, that possess a variety of motility and sensory functions (Bisgrove and Yost, 2006). Some of the motility functions of cilia include moving fluids and particles over epithelial surfaces, and propelling sperm and unicellular algae (Bisgrove and Yost, 2006). Motility plays a big role of virulence of many pathogenic organisms, such as bacteria (Ramos et al., 2004). It is involved in the migration of organisms within a host, helping them survive in different environments, host cell attachment, intracellular movement, and colonizing the host (Ramos et al., 2004). Though the notion has not been thoroughly tested in protozoan pathogens, including *Plasmodium* spp., *Toxoplasma gondii*, *Leishmania* spp, and trypanosomes, it is believed that motility is required for their survival (Hill, 2003).

Sensory functions of cilia comprise detecting light, odors, and fluid flow in vertebrates, as well as identifying changes in osmolarity, chemoattractants and repellants, and sound in invertebrates (Bisgrove and Yost, 2006). Additionally, cilia play important roles in cell morphogenesis, cytokinesis, and host-parasite interactions in *T. brucei* (Hill, 2003; Kohl et al., 2003).

Cilia are found in most of the organs in the human body and are involved in tissue function (Ibañez-Tallon et al., 2003; Pan et al., 2005). They are employed in sensory functions (hearing and smelling), motility of sperm, and the transport of fluids in the respiratory and gastrointestinal tracts, as well as in the central nervous system. Moreover, their sensory roles are vital for proper kidney function (Pan et al., 2005). Defects in ciliary function have been attributed to a number of disorders, such as immotile cilia syndrome, Bardet-Biedl syndrome, and Polycystic kidney disease (Bush et al., 2007; Tobin and Beales, 2007; Vora et al., 2008). These disorders can include one or several of the following symptoms: respiratory tract infections, retinopathy, hydrocephalus, and renal failure, indicating that the proper function of cilia play a large role in human health and disease. Therefore, learning more about ciliary function will provide insights into how diseases are generated and can be prevented (Pan et al., 2005; Ibañez-Tallon et al., 2003).

Components of Motile Flagella (CMF)

As having motile flagella is necessary for normal human development and the virulence of many pathogenic organisms, Baron et al. (2007) searched for

homologues that are unique to organisms with motile flagella. They dubbed these homologues components of motile flagella (CMF) genes.

To better understand the mechanism and regulation of flagellar motility, Kent Hill and his colleagues at UCLA used *T. brucei* as a model for comparative genomics and functional analysis to identify motility genes. They first identified genes in *T. brucei* that are conserved among organisms with motile flagella. Genes that are found in organisms with non-motile flagella or lacking flagella were subtracted from their dataset before they subjected the dataset to homology predictions. It was found that the predicted proteins were homologues of flagellar component proteins, indicating that their dataset was enriched with flagellar proteins. The final dataset consists 50 CMFs, 30 of which have never been characterized (Baron et al., 2007).

Baron et al. selected 41 CMFs that they were interested in investigating further and knocked them down using tetracycline-induced RNA interference. The mutants were characterized using a sedimentation assay, which allowed them to measure the degree to which motility of the trypanosomes was impaired using optical density. Most mutants demonstrated various degrees of motility defects, which were divided into four classes. Class 1 mutants were classified as unaffected, as they did not exhibit any phenotypes that were distinguishable from the uninduced trypanosomes. Class 2 mutants were characterized by the presence of small clusters of trypanosomes. Both Class 3 and Class 4 mutants generated

large clumps, but Class 3 mutants had a reduced growth rate, while the Class 4 mutants were ultimately lethal (Baron et al., 2007).

Seventy-six percent of the mutants, comprising Class 2, Class 3, and Class 4 mutants, exhibited a mutant phenotype (Baron et al., 2007). As the severity of the phenotype increased, the severity of the motility defect increased; Class 3 and Class 4 mutants, before they died, had severely impaired wave propagation, resulting from the loss of propulsive movement. Since the knockdown of a large percentage of CMFs generated motility defects, it is important to characterize these genes to determine the role they play in flagellar motility (Baron et al., 2007).

The gene selected for this study, CMF63, exhibits the phenotype of a Class 3 category gene when it is knocked down. In recent genomic and flagellar proteomic analyses, it was predicted by Baron et al. (2006) that the CMF63 protein has an EF-hand, or calcium-binding, region, indicating that this protein may play a role in regulating intracellular calcium concentrations or in signal transduction.

Calcium-binding proteins

Calcium is an intracellular signaling ion that plays an important role in cell division, differentiation and proliferation, transportation of intracellular molecules, apoptosis, fertilization, and cell motility (Gifford et al., 2007; Grabarek et al., 2006). It has also been found that the regulation of calcium plays an important role in the flagellar and ciliary bending patterns, but the mechanism

is unknown (Bannai et al., 2000; Smith et al., 2002). EF-hand calcium-binding proteins (CaBP) are commonly used in calcium intracellular signaling pathways (Grabarek et al., 2006). They respond to intracellular calcium concentrations and are involved in signal transduction (Grabarek et al., 2006). These proteins can be divided into two classes: calcium buffers and calcium sensors (Gifford et al, 2007). The calcium buffers modify a calcium signal by binding to free calcium and transmitting the signal throughout the cell or removing it from the cytoplasm. The calcium sensors act as intermediates in a signal induction pathway that transduce a signal by changing their conformation (Gifford et al, 2007).

The EF-hand domain has a distinct secondary structure, consisting of two helix-loop-helix structures and a linker loop connecting the two structures (Gifford, 2007). Each loop contains twelve amino acid residues, which form a pentagonal bipyramidal shape (Santamaria-Kisiel et al., 2006). Both loops chelate a calcium ion, which is attracted to the oxygen atoms on the loop. Chelation can be assisted by hydrogen bonds and/or salt bridges between the loops (Gifford, 2007; Grabarek, 2006). The linker region is very different between EF-hand proteins, varying in amino acid composition and length. Structural stability of the domain is provided by a backbone of hydrogen bonds that connect the calcium-binding loops to form an antiparallel β-sheet (Garbarek, 2006).

Figure 4 illustrates that when an EF-hand CaBP is bound to calcium, it undergoes a conformational change from a closed domain to an open domain (Grabarek et al., 2006). It was found that during a conformational change, the α-

helices can undergo a rearrangement to generate and expose a hydrophobic surface to bind to target proteins (Santamaria-Kisiel et al., 2006). At this moment, there is insufficient knowledge of EF-hand protein structures (apo and calcium-bound) to know how prevalent this phenomenon is among EF-hand proteins (Grabarek, 2006).

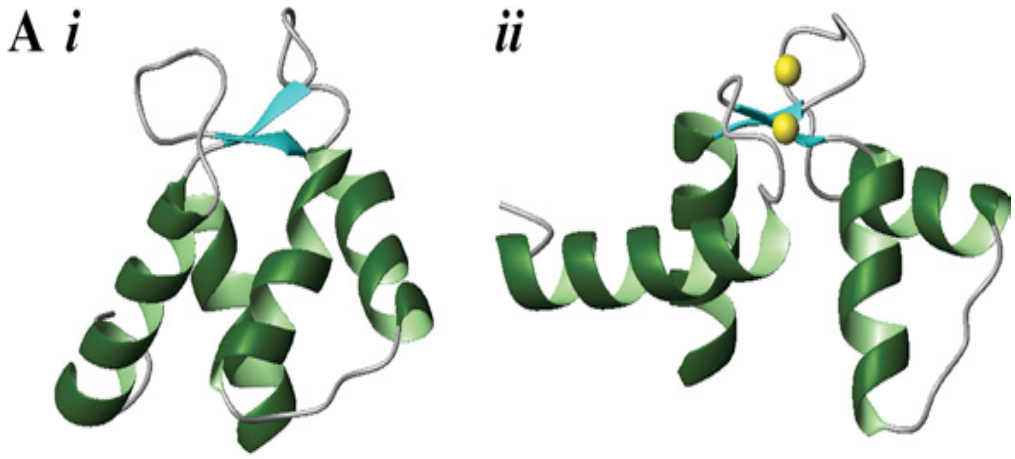


Figure 4. Representation of an EF-hand conformational change. When calcium binds to the domain, it undergoes a conformational change from a i) closed state to an ii) open state to expose a hydrophobic pocket that can interact with a target to transduce a calcium signal (Gifford et al., 2007).

It has been hypothesized that the structure connecting the Ca^{2+}-binding loops in the EF-hand domain may be responsible for giving the EF-hand proteins a large variety of conformations (Grabarek, 2006; Gifford et al., 2007). Additionally, the interhelical contacts, length, and the structure and flexibility of the linker region may also play a role in the conformational change (Grabarek, 2006).

A representation of the predicted calcium-binding region of CMF63 can be seen in Figure 5 (Hare, 2007).

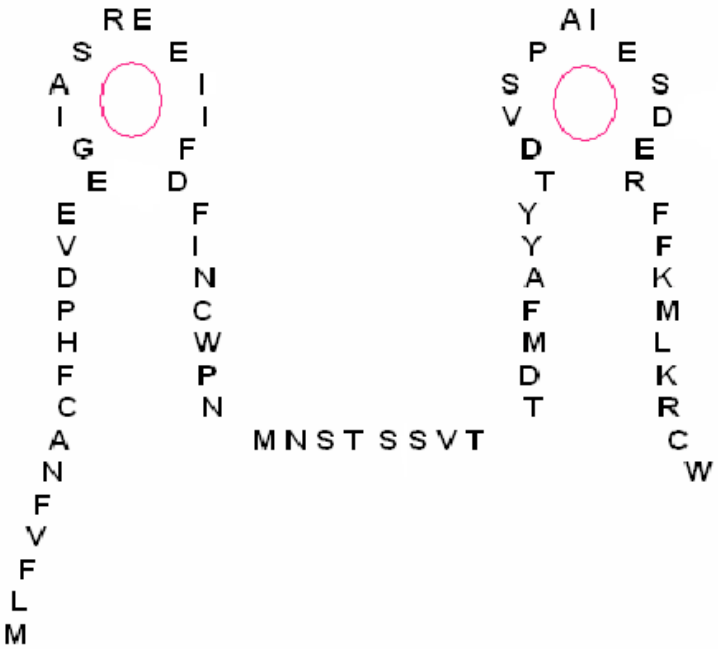


Figure 5. Representation of the structure of the calcium-binding region of CMF63. The vertical strands represent α-helices and the horizontal strand is a linker loop adjoining the two calcium-binding loops. The red circles indicate where the calcium cations are chelated to the binding loops (Hare, 2007).

In a previous study, Hare predicted the secondary structure of the EF-hand domain of CMF63. Using BLAST searches, she mapped out which residues in the CMF63 protein sequence corresponded to the EF-hand domain (from 340 – 489) and the calcium-binding region (from 419 – 489 amino acids) within it. The characterization of the calcium-binding region is a work in progress.

RNAi

RNA interference (RNAi) is a process used by many organisms, including fungi, plants and animals, to silence gene expression (Rana, 2007). RNA viruses, transposons, exogenously introduced dsRNAs (small interfering RNAs, or siRNAs), and endogenous small non-coding micro RNAs (miRNAs) are capable of initiating RNAi (Rana, 2007). In the mammalian model, illustrated in Figure 5,

RNAi is initiated when Dicer, an RNase III-like enzyme, cuts long dsRNA molecules into siRNAs (siRNAs) that are 21 – 23 nucleotides (nt) in length. The sequences of the sense and antisense strands of the siRNAs are complementary to the target mRNA. They have 2-nt overhangs at their 3' ends and phosphate groups at their 5' ends. The antisense strand, also known as the guide strand, of the siRNA is a template for sequence-specific gene silencing (Rana, 2007). It becomes integrated into the RNA-induced silencing complex (RISC) complex, forming an RNA-protein complex called siRISC. This complex binds to the target mRNA and cleaves it 10 – 11 nt from the 5' end of the guide strand, silencing gene expression. When there is perfect complementarity between the target and guide strands, the target strand is cleaved and degraded (Rana, 2007).

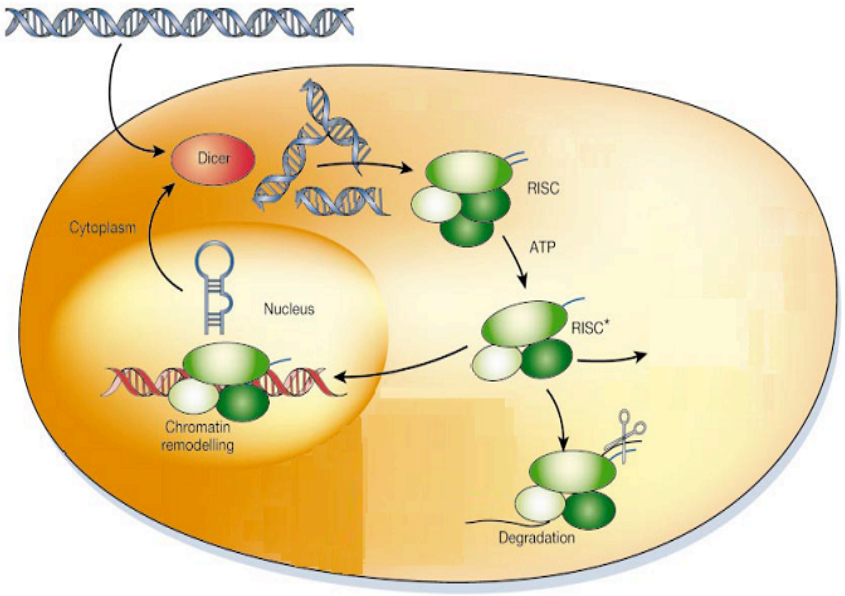


Figure 6. Mechanism of RNA interference. DsRNA is recognized and cleaved by Dicer into fragments between 21-23 nucleotides in length. The fragments form a complex with RISC, and the fragments are unwound and cleaved, so that only the antisense strand of the fragment is bound to RISC. RISC searches for mRNA that is complementary to the RNA fragment. When the target is found, it is cleaved, silencing the expression of the gene (Hannon, 2002).

Much is still unknown about the RNAi pathway in *T. brucei* (Ullu et al., 2004). The identity of the Dicer homologue still needs to be established. AGO1, a component of RISC, is the only component of RNAi machinery that has been characterized in *T. brucei*. It is believed that AGO1, associates with additional proteins, but no proteins have been found (Ullu et al., 2004). It has been proposed that AGO1 and siRNA form a complex, and polyribosomes may play a role in the recognition and degradation of the target strand in the cytoplasm. It has also been proposed that the AGO-siRNA complex cleaves mRNA without any interaction between the machineries used for RNAi and translation (Ullu et al., 2004).

RNAi induction in T. brucei

In the first RNAi experiments performed with *T. brucei*, it was found that RNAi could be induced in the trypanosomes by electroporating dsRNA into the cells, but the phenotypes that were generated were transitory (Motyka and Englund, 2004). Vectors that could be stably integrated into the genome of *T. brucei* were subsequently produced, and the efficiency of RNAi was greatly improved. One class of vectors contained a pair of opposing tetracycline-inducible T7 promoters (pT7) that flanked a fragment of a target gene (Motyka, S. A. and Englund, P. T., 2004). pT7 is a particular type of promoter to which only a T7 RNA polymerase (T7RNAP) can bind. When the vectors were expressed in a strain of *T. brucei* expressing T7 RNA polymerase and a tetracycline repressor, RNA molecules that were complementary to the target RNA were constitutively synthesized. These opposing promoters are commonly used, as RNAi can be

induced in both the bloodstream and procyclic form of *T. brucei* with these primers. Additionally, they can simultaneously target multiple genes with a one construct containing a chimeric insert (Motyka, S. A. and Englund, P. T., 2004).

Statement of Purpose

The purpose of this study is to better understand the function of CMF63 in a non-pathogenic strain of *T. brucei*. As the calcium-binding region is currently being probed, homology searches, secondary structure predictions and helical wheel prediction programs were used to select a different region in the protein sequence for further investigation. From these predictions, a region in the CMF63 protein sequence, α-helix from 393 – 404 amino acids, that could be involved in potential interactions was identified. There are three interactions in which the helix could potentially be involved. Its hydrophobic face could be interacting with the hydrophobic face of the α-helix from 378 – 447 amino acids, or its hydrophilic face could be interacting with an α-helix in the calcium-binding region or with a different protein.

Three different mutant CMF63 genes were generated using site-specific mutagenesis: a double mutant in which two bulky amino acids were changed to a small amino acid, a mutant in which a small amino acid was changed to a bulky amino acid, and a mutant in which a bulky amino acid was changed to a small amino acid. *T. brucei* was transfected with the mutant CMF63 genes and will be expressed in the organisms. After the expression of the genes is induced, the phenotypes of the mutants can be characterized. By viewing the phenotypes of the

mutants, it may be possible to establish if the target helix plays an important role in the proper function of the CMF63 protein. If a motility defect is seen, then the target helix is needed for proper protein function. If no defects are seen, then endogenous CMF63 may be compensating for any motility defects generated by mutant CMF63, or the helix may not be needed for proper protein function.

Overview of mutant CMF63 expression

In this experiment, 29-13, a recombinant strain of *T. brucei*, was transfected with CMF63 mutants. To generate this strain, two plasmids were cloned into the α-β-TUBULIN locus of the genomic DNA in the wild-type T. brucei 427 strain (Wirtz et al., 1999). These plasmids contain T7 RNA polymerase (T7RNAP) and a tetracycline repressor (Tet^R), as well as G418 and hygromycin resistance genes that serve as reporters (Wirtz et al., 1999).

The 29-13 strain was transfected with an expression vector: a pKRO8 vector containing the mutant CMF63 gene with a 3' UTR different from the endogenous 3' UTR of *T. brucei*, which will prevent the mutant gene from being silenced via RNAi, and a puromycin resistance gene. In the absence of tetracycline, Tet^R binds to a tetracycline operator in encoded in the pKRO8 vector. This prevents T7RNAP from binding to the T7 promoter region and transcribing the mutant CMF63. When tetracycline is introduced, it binds to Tet^R, allowing T7RNAP to bind to the promoter region and initiate transcription. The process is outlined in Figure 7.

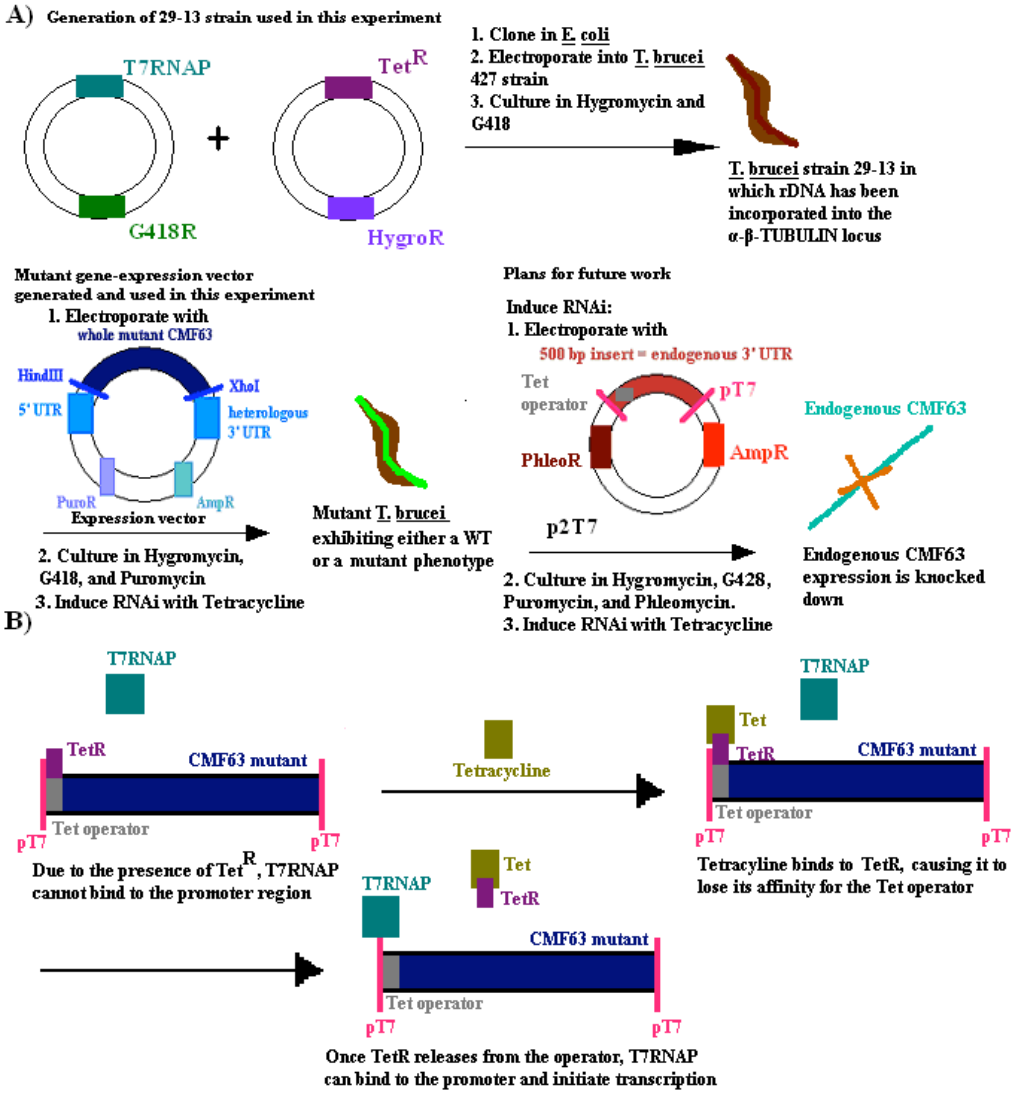


Figure 7. Overview of mutant CMF63 expression. A) The 29-13 strain of *T. brucei* used in this experiment was generated by transfecting the *T. brucei* 427 strain with two vectors: a vector with a T7 RNA polymerase gene and G418 reporter and a second vector with a tetracycline repressor and hygromycin reporter (Wirtz et al., 1999). In this experiment, the 29-13 strain was transfected with a vector containing a mutant CMF63 gene, heterologous 3' UTR, and puromycin reporter, and gene expression of the CMF63 mutant is induced using tetracycline. Once the gene is expressed, it could generate a dominant negative phenotype or no change in phenotype may be seen. For future work, the ultimate goal of the experiment will be to knock down the expression of the endogenous CMF63 by targeting the endogenous 3'UTR using tetracycline induced RNAi so that only the mutant is expressed. Since the 3'UTR in the expression vector is not found in *T. brucei*, the expression of the mutant gene cannot be knocked down. B) CMF63 mutant expression is regulated using tetracycline. The pKRO8 vector has a Tet operator to which Tet^R binds and prevents T7RNAP from binding to the promoter region (pT7). Once TetR is released from the operator, T7RNAP can bind to the promoter and initiate transcription.

In the future, the 29-13 strain will also be transfected with another vector, which contains a 500 bp segment of the endogenous (the cell's copy) 3' UTR flanking CMF63, T7 promoters, and a phleomycin resistance gene. Tetracycline induced RNAi will be used to knock down the expression of the endogenous CMF63 by targeting the 3' UTR flanking the gene. When the production of the 500 bp 3'UTR segment is initiated, only the endogenous CMF63 will be targeted for RNAi. Since the mutant CMF63 has a different 3' UTR, it cannot be targeted for RNAi, allowing the solely mutant gene to be expressed.

EXPERIMENTAL

Identifying a region in CMF63 to target

The endogenous CMF63 protein sequence was obtained from *Trypanosoma brucei* GeneDB (Hertz-Fowler et al., 2004). The GeneDB number of CMF63 is Tb11.01.1210. A GenomeNet prediction program was used to search for BLOCKS in the GeneDB CMF63 protein sequence and the motif library documented in the PROSITE Database (Kyoto University Bioinformatics Center, http://motif.genome.jp). BLOCKS are highly conserved amino acid regions of related proteins. The regions of the CMF63 protein where the most BLOCKS were seen were identified. A Jpred secondary structure prediction program, http://www.compbio.dundee.ac.uk/~www-jpred/ (Barton and Cuff, 2000), was used to predict the locations of α-helices and β-sheets within CMF63. The prediction was compared with the results of the Psipred secondary structure prediction program, http://bioinf.cs.ucl.ac.uk/psipred/psiform.html (McGuffin et al., 2000), (results not shown) to determine the accuracy and consistency of the results.

Selecting a target for mutagenesis

The secondary structure predictions suggest that the α-helix from 393 – 404 amino acids has the potential to interact with the calcium-binding region, a neighboring α-helix consisting of amino acids from 378 – 387, with another

protein, or with the aqueous environment of the cell. To determine if the helix may be interacting with any of these regions, the glutamines at positions 3, 10, and 11, as well as the glycine at position 4 (Figure 14) were targeted for mutagenesis to perturb any potential interactions. A double mutant, in which the two glutamines at positions 3 and 10 were changed to alanines, was generated and named Q395Q403A. Additionally, the glycine at position 4 was changed to a methionine, and the glutamine at position 11 was changed to an alanine, forming G396M and Q403A.

Tables 1 – 4 provide descriptions of the vectors and competent cells used for cloning and mutagenesis and the clones generated.

Table 1. Plasmids used for mutagenesis and cloning

Vector	Description
pET160 containing CMF63	Cloning vector, made by Invitrogen (www.invitrogen.com), in which the CMF63 gene is found between 237 bp and 1897 bp. The vector contains an ampicillin resistance gene to serve as a reporter. There are two EcoRV restriction sites.
PKRO8	Expression vector based on pHD1342 (Alibu et al., 2004) and modified by Katy Ralston, UCLA. The vector contains a PARP promoter (Trypanosome promoter), T7 promoter (*E. coli* bacteriophage promoter), and two T7 promoters to regulate exogenous gene expression. It has a 3' untranslated region (UTR) that is not found in *T. brucei*, as well as ampicillin and puromycin resistance genes that behave as reporters.

Table 2. Competent cells that were transformed following mutagenesis and used for cloning

	Description
XL1- Blue	Strain of *E. coli* that is transformed with the site-specific mutagenesis products. Once transformed, the cells repair nicks in the DNA that arise during mutagenesis.
DH5α	Strain of *E. coli* used for cloning.

Table 3. Mutants generated in this study

	Description
Q395Q402A	The 395th and 402nd amino acids in the CMF63 protein sequence are glutamines, which were each changed to an alanine.
G396M	The 396th amino acid in the CMF63 protein sequence is a glycine, which was changed to a methionine.
Q403A	A glutamate is the 403rd amino acid in the CMF 63 protein sequence, and it was changed to an alanine.

Table 4. Constructs transfected into *T. brucei*

	Description
Wild-type CMF63-pKRO8	Endogenous CMF63 (Springer, personal communication).
Q395Q402A-pKRO8	Double mutant cloned via PCR with XhoI and HindIII primers and ligated to pKRO8 for expression in *T. brucei*.
G396M- pKRO8	Mutant cloned via PCR with XhoI and HindIII primers and ligated to pKRO8 for expression in *T. brucei*.
Q403A- pKRO8	Mutant cloned via PCR with XhoI and HindIII primers and ligated to pKRO8 for expression in *T. brucei*.

Site-Specific Mutagenesis

In order to make the mutant CMF63, a QuikChange Site Directed Mutagenesis Kit was utilized (Stratagene, www.stratagene.com) to generate mutants with a double-stranded template. During this process, a construct, containing a gene with a target site, is denatured. Oligonucleotide primers that have the desired mutations anneal to the construct, and DNA polymerase extends, incorporating the mutagenic primers. This generates nicked circular strands. The methylated, nonmutated parental DNA template is digested with Dpn I before the mutant constructs are transformed into XL1-Blue supercompetent cells, which repairs the nicks in the DNA.

In this experiment, the dsDNA being amplified was the ORF of CMF63 (1659 bp) from purified *T. brucei brucei* strain 29-13 cloned into pET160/GW/D-

TOPO (5839 bp) (Invitrogen, www.invitrogen.com) (Figure 8). PfuTurbo DNA polymerase (Stratagene, www.stratagene.com) and a Bio-Rad Mycycler (Bio-Rad Laboratories, www.biorad.com) were used to replicate both plasmid strands.

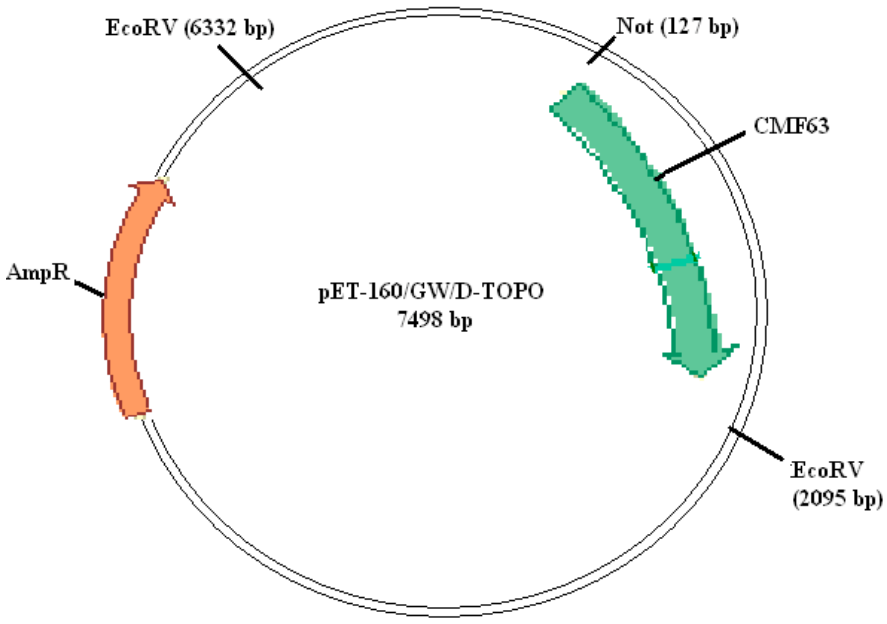


Figure 8. pET160-CMF63 construct. CMF63 was inserted into a pET-160/GW/D-TOPO cloning vector between 237 bp and 1897 bp. CMF63 mutants were generated using this vector. Contains an ampicillin resistance gene to serve as a reporter. There are two EcoRV restriction sites, at 2095 bp and 6332 bp).

To switch the amino acids, the codons that code for the three targeted glutamines (Gln) were changed to a codons of alanine (Ala), and the codon that codes for glycine (Gly) was changed to the codon of methionine (Met). These new codons were incorporated into oligonucleotide primers, which were complementary to the opposite strands of the construct (Figure 9). The primers were designed based on the CMF63 gene and protein sequences posted by GeneDB (http://www.genedb.org/genedb/tryp/index.jsp) and were manufactured by IDT (Integrated DNA Technologies, www.idtdna.com),

a. Gln-Gln double-mutant (Q395Q402A)

F: CAGAGCCGGTTGGCAGGTGTTATATCGGCCTTCCGCCAGCTTGTTATT
R: AATAACAAGCTGGCGGAAGGCCGATATAACACCTGCCAACCGGCTCTG

b. Gly → Met (G396M)

F: AGCCGGTTGCAAATGGTTATATCGGCC
R: GGCCGATATAACCATTTGCAACCGGCT

c. Gln → Ala (Q403A)

F: ACGGCCTTCCAGGCGCTTGTTATTGAG
R: CTCAATAACAAGCGCCTGGAAGGCCGT

Figure 9. Primers designed for mutagenesis. The codons that were changed are highlighted in blue. A) Primer designed for a Gln double-mutant in which Gln 395 and Gln 402 were changed to Ala. B) Primer designed to change Gly 396 to Met. C) Primer designed to change Gln 403 to Ala.

For each mutant, 5 µl of 10x buffer, 125 ng of forward primer, 125 ng of reverse primer, 50 ng of template DNA, 1 µl of dNTP mix, 1 µl of PfuTurbo DNA polymerase (2.5 U/µl), and 43 µl of dH_2O were added together, making a 25 µl reaction mixture. A control reaction was set up. The reaction mixture included 5 µl of 10x buffer, 2 µl of pWhitescript 4.5 kb control plasmid (5 ng/µl), 1.25 µl (125 ng) of each control primer, 38.5 µl of dH_2O, and 1 µl of PfuTurbo DNA polymerase (2.5 U/µl). The conditions under which the temperature cycling was performed in are presented in Table 5.

Table 5. Reaction conditions for mutagenesis

	Temperature (°C)	Duration	Number of cycles
Denaturation (initial)	95	30 sec	1
Denaturation	95	30 sec	16x for control, G396M and Q403A & 18x for Q395Q402A
Annealing	55	60 sec	
Extension	68	7.5 min	
	4	hold	

10 µl of each product was run on a 0.8% agarose gel, and the remainder of the products were digested with 1 µl of Dpn I to eliminate the template DNA. Then, they were transformed into XL1-Blue supercompetent cells and plated on separate plates made of Luria broth and ampicillin.

Gel imaging

All of the agarose gels used in this experiment were stained with ethidium bromide to visualize the DNA. The molecular weight marker used was a 1kb Plus DNA ladder (100bp – 12kb) (Invitrogen, www.invitrogen.com) Gel imaging was completed using a Fujifilm LAS-3000 Intelligent Dark Box with the accompanying Image Reader Fujifilm LAS-3000 version 2.0 software (Fuji Photo Film Co. LTD, http://www.fujifilm.com/). The resolution was set to standard and the exposure type and exposure time used were precision and 1/8 second, respectively. The negative image was used to visualize the bands with more ease.

Preparation of mutant DNA for insertion into the expression vector (pKRO8)

Cloning CMF63 mutants

The ampicillin resistant XL1-Blue colonies were screened for transformants. The mutant DNA was extracted from overnight cultures (3 mL of broth with 3 µl of 100 µg/mL ampicillin, and a selected transformant colony) of tranformants for each mutant using a QIAquick miniprep kit according to manufacturer's instructions (Qiagen Inc, www.Qiagen.com). The identity of the purified DNA was confirmed by digesting it with EcoRV and visualizing the products on a 0.8% agarose gel.

The mutant DNA was PCR amplified with XhoI and HindIII primers to prepare the DNA for ligation into a pKRO8 expression vector (6039 bp). The primers, seen in Figure 10, were prepared by IDT (Integrated DNA Technologies, www.idtdna.com).

CMF63 forward primer (HindIII)
5' CACTCTAGATTACATATTAATTCTAATGT 3'

CMF63 reverse primer (XhoI)
5' CACAAGCTTATGAACGAATTAAAGAACCC 5'

Figure 10. HindIII and XhoI primers used to clone CMF63 mutants. The primers were made by IDT (www.idtdna.com).

2.5 µl of RT-PCR buffer 5 mM Mg^{+}, 1 µl of 50 mM MgCl2, 0.5 µl of 10mM dNTP mix, 0.4 µl of Hifi Taq polymerase (Eppendorf, http://www.eppendorf.com), 1 µl of HindIII primer, 1 µl of XhoI primer, 3 µl of pET160-Q403A or pET160-G396M construct or 4 µl of pET160-Q395Q402A construct, and dH_2O (to bring volume to 25 µl) were included in the reaction mixture. Reaction conditions for the amplification are presented in Table 6.

Table 6. Reaction conditions for PCR amplification of mutant DNA with XhoI and HindIII primers.

	Temperature (°C)	Duration	Number of cycles
Denaturation (initial)	95	5 min	1
Denaturation	95	30 sec	30
Annealing	50	30 sec	
Extension	72	1 min	
Extension (final)	72	5 min	1
	4	hold	

All of the extraction mixture was run on a 0.8% agarose gel with a 12 kb ladder, and the DNA was extracted using a QIAquick gel extraction kit according

to manufacturer's instructions (www.qiagen.com). The concentration of the purified DNA was measured with a Nanodrop nd-1000 spectrophotometer with the accompanying Nanodrop nd-1000 version 3.1.2 software (Nanodrop Technologies, http://www.nanodrop.com); Coleman Technologies Inc for Nanodrop Technologies, http://www.ctiusa.com).

Preparation of the pKRO8 vector for cloning

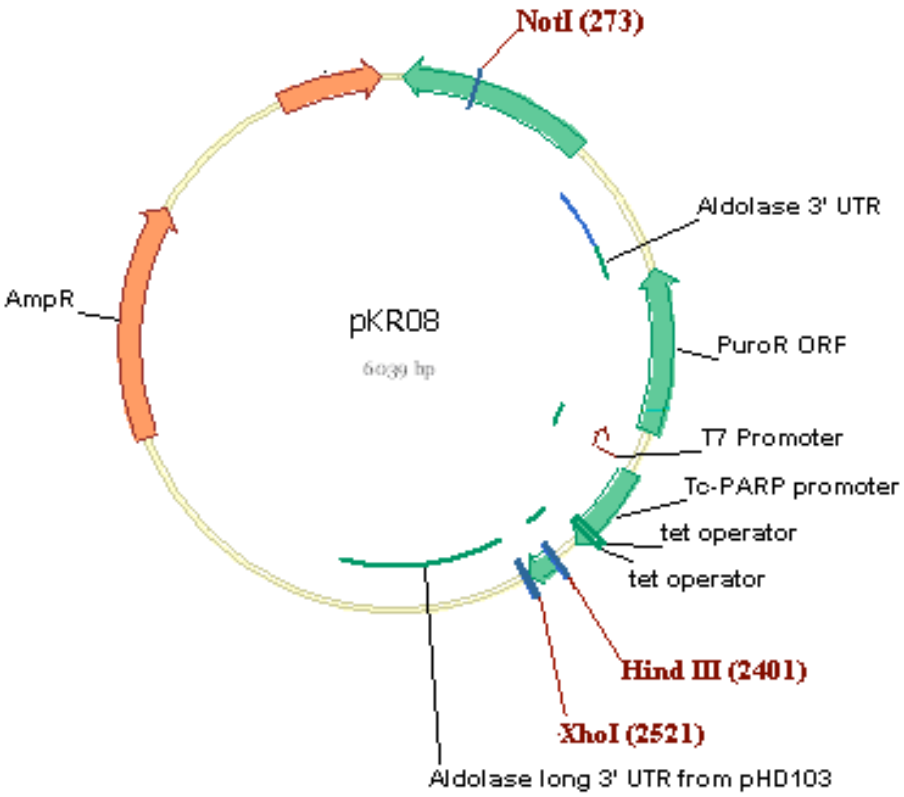


Figure 11. pKRO8 expression vector. The mutants (Q395Q402A, G396M, and Q403A) that were generated via site-specific mutagenesis were cloned with HindIII (5') and XhoI (3') primers using PCR. Ligation of the mutants to the HindIII and XhoI restriction sites of the pKRO8 vector was performed using DNA ligase.

The pKRO8 expression vector (Figure 11) was digested with XhoI and HindIII (Promega, www.promega.com) for 3 hours at 37°C. The reaction mixture

consisted of 1 µl of NEBuffer2 (New England BioLabs, http://www.neb.com/nebecomm/default.asp), 4 µl of pKRO8, 0.5 µl of XhoI, 0.5 µl of HindIII, and 4 µl of dH_2O. The reaction mixture was heat inactivated at 65°C, run on an 0.8% gel with a 12 kb ladder, and extracted and purified with a QIAquick gel extraction kit. A Nanodrop nd-spectrophotometer was used to measure the concentration of the vector.

Ligation of mutant DNA to pKRO8

20 µl reaction mixtures were made with pKRO8, XhoI/HindIII digested mutant DNA, 2 µl of ligase buffer, 1 µl of DNA ligase, and dH_2O. The molar ratio of insert to vector used was 3:1, and less than 20 ng of DNA was used in the reaction mixture. The reaction mixtures were incubated at room temperature for 15 minutes. 10 µl of the reaction mixtures for each mutant was transformed into competent DH5α *E. coli* cells and plated on Luria broth and ampicillin.

Screening for mutant DNA-pKRO8 transformants

Many ampicillin resistant colonies were picked from the plated transformed cells to determine which cells had been transformed with the mutant DNA-pKRO8 constructs. Overnights were prepared for the constructs in LB with ampicillin. The DNA was extracted and double-digested with XhoI and HindIII at 37°C for two hours. The reaction mixtures were run on a 0.8% agarose gel with a 12 kb ladder. It was anticipated that two bands representing 1.6 kb and 6.0 kb would be generated, as they are the sizes of the CMF63 gene and the pKRO8 vector, respectively. Therefore, samples that generated bands representing DNA

1.6 kb and 6.0 kb in size were sent to Elim Biopharmaceuticals (Elim Biopharmaceuticals, Inc. www.elimbio.com) for sequencing. The sequencing results were screened for the designed mutations (Q395Q402A, G396M, and 403A) using Vector NTI software (Invitrogen, www.invitrogen.com). One of each sample with the designed mutations were selected for transfection. The following primers were used: CMF63 for (HindIII) & CMF63 rev (XhoI) and CMF63 seq & CMF63 seq2. The primer sequences are presented in Table 7.

Table 7. Sequences of the sequencing primers.

Primer	Sequence
CMF63 for (HindIII)	5' CACAAGCTTATGAACGAATTAAAGAACCC 3'
CMF63 rev (XhoI)	5' CACCTCGAGTTACATATTTAATCTAATGTCCTTG 3'
CMF63 seq	5' GTCAGAGAGCTACGAAGTGG 3'
CMF63 seq2	5' CGAGCTGAGCGCCTGACTGGC3'

A midiprep was performed for each sample of DNA that was selected for transfection with Promega Wizard Plus Midipreps DNA Purification System (Promega Co., www.promega.com). The concentration of the DNA was measured with a Nanodrop nd-1000 spectrophotometer.

Preparation of Trypanosomes

A 1 mL sample containing 5.7 X 10^6 the non-pathogenic *T. brucei brucei* recombinant 29-13 cells (Wirtz et al., 1999) that were frozen in 5% glycerol was thawed. The cells were cultured at 28°C in SDM-79 medium (SAFC Biosciences, www.sigmaaldrich.com/SAFC/Biosciences.html), pH 7.4 and buffered with 0.67 mg/mL of sodium bicarbonate. The medium was supplemented with 15% heat inactivated Fetal Bovine Serum (FBS) (Hyclone, www.hyclone.com), 6 μg/ml

hemin (Sigma-Aldrich, www.sigmaaldrich.com), 15 μg/ml geneticin (G418) (MP Biomedicals, www.mpbio.com), and 50 μg/ml hygromycin (MP Biomedicals, www.mpbio.com).

Transfection of Trypanosomes

Five transfections were performed. These included a no DNA control, wild-type CMF63-pKRO8, as well as the Q395Q402A-pKRO8, G396M-pKRO8, and Q403A-pKRO8 clones that were verified by sequencing. The night before the transfections were performed. 15 μg of DNA (wild-type CMF63, Q395Q402A, G396M, and Q403A) was linearized in 10x buffer and 20 – 30 units of NotI (Promega, www.promega.com), and dH_20 was used to bring the final volume to 30 – 50 μl. The reaction mixtures were incubated overnight at 37 °C. On the day of the transfection, 2.5 volumes of 100% molecular biology grade ethanol (Pharmco, www.pharmco-prod.com) and 0.1 volume sodium acetate were added to the linearized DNA, which were then incubated at -20°C for 30 minutes to precipitate. The DNA was centrifuged at 14,000 rpm for 30 minutes at room temperature, washed with 70% ethanol, resuspended in 500 μl cytomix, and heat solublized at 37 °C for a few minutes. The cytomix was made with 25mM HEPES, 120mM KCl, 0.15mM CaCl2, 10mM K2HPO4/KH2PO4, pH 7.6, 2mM EDTA, 5mM MgCl2, 1% Glucose, and 0.1% BSA, pH to 7.6. The solution was filtered to sterilize and stored at 4°C. To confirm linearization, one of the products was visualized on a 0.8% agarose gel.

To promote the growth of the transfectants, conditioned medium was made. Spent medium (SDM-79 medium with 15% FBS, hygromycin, and G418 in which the trypanosomes were cultured after being thawed) was collected and filtered over time. To make the conditioned medium, 1 volume of spent medium, 1 volume of fresh medium (SDM-79, pH 7.4 supplemented with 20% FBS), and 0.4 volumes of FBS were mixed together.

10 mL aliquots containing 10^7 cells/mL were used for each transfection. The cells in each aliquot were pelleted in separate tubes by centrifuging them at 2400x g at room temperature for 12 minutes. The supernatant was removed quickly. The cells were washed by resuspending them in 5 mL of cytomix, centrifuging them for an additional 12 minutes, quickly removing the supernatant, and resuspending them with the 500 µl of DNA/cytomix mixtures to separate tubes. The no DNA control was resuspended in 500 µl of cytomix. Each sample was transferred to a separate 4 mm gap cuvette and electroporated at 1.6 kV, 25 Ω with a time constant of 0.8. Each sample was added to separate flasks containing 9.5 mL of conditioned medium, 15 µg/ml G418, and 50µg/ml hygromycin.

24 hours post transfection, 2 µg /mL puromycin (Acros Organics, www.acros.be) was added to each flask. Three days post transfection, the flasks were split 1:2. Over the course of two weeks, the flasks were checked periodically to monitor cell growth and death. Conditioned medium was added to the flasks to renew the supply of food and growth factors.

Monitoring the growth of the transfectants

The growth of the trypanosomes was monitored using a Nikon Eclipse TS100 inverted microscope (Nikon Instruments Inc, www.nikoninstruments.com) in phase contrast at 100X and 200X magnification. Conditioned medium and FBS were occasionally added to promote the growth of the transfectants. With the exception of the flask containing the pKRO8-G396M transfectants, the flasks exhibited no signs of life ten days following the transfection. Six weeks after the transfection, the pKRO8-G396M transfectants were split 1:10 with complete medium in a new flask. 2 mL of FBS was added to the flasks three days after they were split to encourage growth. Cells were maintained at titers between 2×10^5 – 1×10^7 cells/mL.

RESULTS

Identifying conserved amino acid regions

In order to identify a target for mutagenesis in the CMF63 protein, a BLOCKS search was performed to find amino acid regions that are highly conserved in other proteins. 85 BLOCKS were found within the amino acid sequence of the CMF63 protein sequence and the documented in the PROSITE Database. As seen in Figure 12, the majority of the BLOCKS found in within the amino acid sequence of CMF63 lie between 121 – 480 amino acids, suggesting that the sequence is more conserved within this range. Within this range, 55% of the BLOCKS were found in the EF-hand region (340 – 489 amino acids).

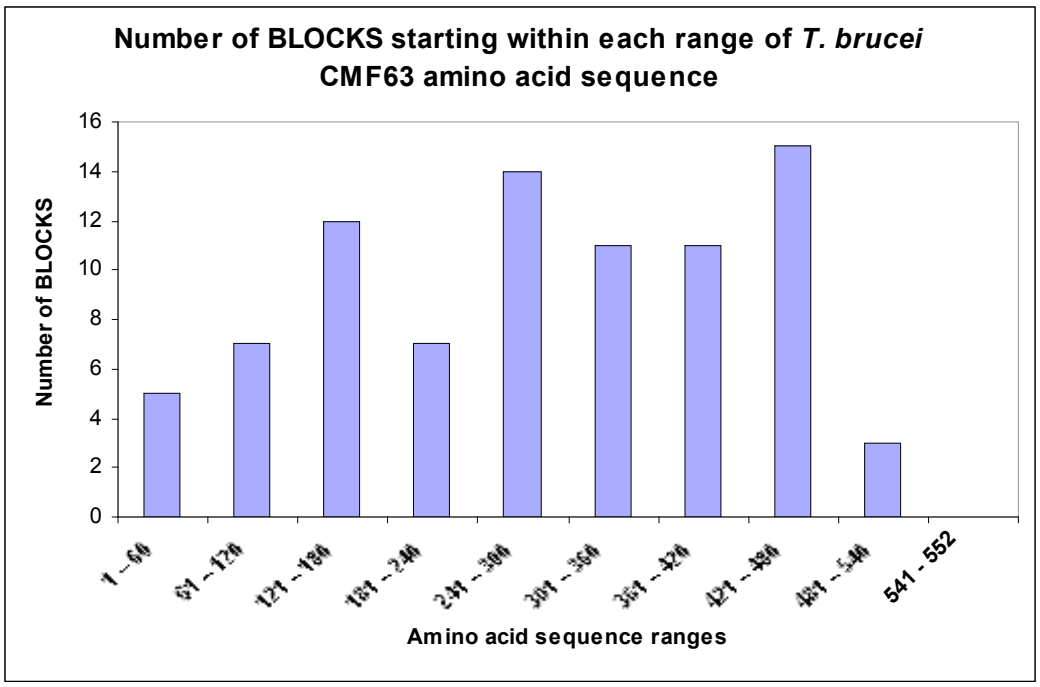


Figure 12. Histogram illustrating where most of the BLOCKS were found within the amino acid sequence of the CMF63 protein. The data were obtained by performing a BLOCKS search with the amino acid sequence of the CMF63 protein, dividing the sequence into segments that were 60 amino acids in length, and organizing the BLOCKS based upon in which segment they were found.

The BLOCKS that lie within 121 – 480 amino acid range are presented in Table 8. The BLOCKS that were identified, such as the VDJ recombination activating protein 2 and tryptophan synthase, β chain, are proteins that form complexes with other proteins, suggesting that CMF63 may also interact with other proteins.

The length of the BLOCKS found ranges from 15 to 55 amino acids in length with a percent identity from 10 – 60%. Most BLOCKS had a percent identity below 25%, indicating that very few amino acids are conserved among different proteins.

Table 8. BLOCKS sequences starting within 121 – 480 amino acids

aa Range	Protein	aa Range	Protein
130 – 162	Influenza virus matrix protein MZ	309 – 349	Corona nucleocapsid I
124 – 155	Hox 9 activation region	359 – 377	Flagellar Ca^{2+}-binding protein (calflagin) sign
166 –199	Photosystem II reaction center protein pbsL	305 – 357	Phage NinH
155 – 190	FXYD protein	315 – 351	Photosystem II reaction center protein PsbJ
142 – 182	Nitric oxide synthase, NOS	352 – 391	Malate dehydrogenase
168 – 205	Flagellar basal body FlaE	346 – 397	NADH dehydrogenase (ubiquinone)
165 – 189	Fusion glycoproteins FO	308 – 355	B tubulin
159 – 177	TNF-α (calchectin) signature	348 – 384	CemA family
163 – 186	Influenza matrix protein (MI)	365 – 381	Flagellar Ca^{2+}-binding protein (calflagin)
137 – 181	VDJ recombination activating protein 2 (RAG2)	364 – 385	Recoverin family signature
143 – 192	PSI psaA and psaB	367 – 417	Protein of unknown function UPF0154
223 – 277	Paramyxovirus non-structural protein C	361 - 393	Ca^{2+}-binding protein, S-100/IcaBP type
234 – 271	Pro-opiomelanocortin N-terminal	405 – 505	Ca^{2+}/calmodulin dependent Protein Kinase II
234 – 273	Cleaved adhesion	368 – 418	Arterivirus nucleocapsid protein
206 – 224	Plexin cytoplasmic region	361 – 406	Polyhedron
205 – 252	Shiga-like toxin, β subunit	365 – 378	Diaminopimelate epimerase
236 – 265	CLN3 Batten's disease protein (battenin), signature	384 – 438	Daxx protein
210 – 231	Adhesion molecule CD36 family signature	388 – 427	Poxvirus A28 protein
273 – 318	Parvovirus coat protein VPI, N-terminal	377 – 387	Lyphotactin precursor signature
243 – 296	ROK, N-terminal	442 – 463	Cytolethal distending toxin B signature
264 – 317	Photosystem II phosphoprotein PsbH	443 – 467	Apical membrane antigen 1 signature
295 - 328	Vibrio thermostable direct hemolysin	479 – 533	Mitochondrial protein of unknown function DUF1082
274 – 316	Pro-opiomelanocartin, N-terminal	447 – 488	Reovirus RNA-dependent RNA polymerase λ_3
269 – 316	Pro-opiomelanocortin, N-terminal	421 – 454	Coronavirus nonstructural protein NS2
250 – 284	Coronavirus protein 7	478 – 494	Bontoxilysin
291 – 322	Photosystem I reaction center subunit VI	463 – 493	Protein of unknown function DUF1706
270 – 304	Progressive ankylosis	463 – 490	Core – binding factor, β subunit
265 – 306	Polyhedrin	453 – 490	Porin, opacity type
280 – 292	Tryptophan synthase, β chain	454 – 504	CTF/NF-1 family
264 – 274	Aspartate kinase	423 – 474	FEZ-like
248 - 259	Na^+/H^+ exchanger isoform I (NHEI) signature	455 – 501	SBP plant protein
276 – 328	TraM	422 – 447	Glycoside hydrolase, family 34
321 – 344	Erythropoietin signature	479 – 493	Clostridium neurotoxin, translocation
322 – 372	Ephrin receptor, ligand binding	476 – 501	Pro-opiomelanocortin, N-terminal
304 – 338	Plasmodium histidine-rich		

The results were obtained on October 12, 2007 using the motif search program generated by the Kyoto University Bioinformatics Center.

Predicting the secondary structure of CMF63

Two secondary structure prediction programs, Jpred and Psipred, were used to identify where α-helices and β-sheets may lie within the protein sequence of CMF63. These prediction programs used pattern recognition techniques and analyze structural features and residue compositions to locate possible secondary structures in a protein (Wilson et al., 2002). The information is derived from known protein structures, so the accuracy of the predictions is dependent upon the information posted in current structural databases (Wilson et al., 2002). The predictions made by the two programs (Psipred results not shown) were comparable, so the Jpred prediction was used to select a target for mutagenesis.

```
1   MNELKNPMRY SPGVLVGNWY EDMRVAEDKI KSYRSKMGHD ADILGSQRLE ESCVDADTHS 60
    ---------- ---EE----- HHHHHHHHHH HHHHHH---- -HHH--HHHH ----------

61  MTMGDVIMLG KPLRLLNVAT EAVLAVDTAW THPQRLPHQF LLTATGNTAP RQRVEWVLMR 120
    -----EEE-- --EEEE---- ---EEE---- ---------E EEE------- ---EEEEEEE

121 AEDENNVGYT KQLKEENVLH YGQHIRIANE AAHSEGFLYL HSSIRDVGQS GAQLAVASLG 180
    E--------- -------EEE E---HHHH-- -------EEE ---------- ---HHHHH--

181 TSKDNIFVVA KPGEKRDDIR YGAPVRVGDR FVLYHAATNQ PLRCIKKLQR TSFGFEYGMD 240
    -----EEEEE ---------- ---------E EEEEE----- ---------E E---------

241 CSFAGDNHSR SVAAVTTEPT NLFVVVAANY GVTNTMSVSS LRNLTGRGNV GVLSESYEVD 300
    ---------E ---------- -EEEEEE--- ---------- ---------- ----------

301 LSAIISLIRE GVLYFGGRLG FRLLSKVLGV ACNEQCVTPV RRQDIFHGIS LMGVTIHPGE 360
    HHHHHHHHHH HHH----HHH HHHHHHHHHH HH-------H HHHHHHHHHH H------HHH

361 LDVIFKKLDR VGNGFVVAQE FLRELRCELP QSRLQGVISA FQQLVIEGGG SVDYKDMLNL 420
    HHHHHHHH-- -------HHH HHHHHHH--- --HHHHHHHH HHHH------ EEHHHHHHHH

421 FVFNACFHPD VEEGIASREE IIFDFINCWP NMNSTSSVTT DMFVAYYTDV SPAIESDERF 480
    HHHH------ -------HHH HHHHHHHHH- ---------H HHHHHHHHH- ------HHHH

481 FKMLKRCWKI PETDAYKSMK PCRSVTVFRS DNTSSIIYLP DSSVLNIKDL SSVRRFLTQC 540
    HHHHHHHHH- ---------- ---EEEEEEE ---EEEEE-- ---------H HHHHHHHHH-

541 GVKDIKDIRL NM                                                    552
    ------EEEE --
```

Figure 13. Secondary structure prediction for the amino acid sequence of CMF63. The data were obtained using the Jpred secondary structure prediction program. H denotes α-helices, E denotes β-sheets, and – denotes coiled coils. The calcium-binding region is highlighted in blue. The bold amino acids indicate the helix that is being targeted for mutagenesis. The region highlighted in green indicates a helix with which the target helix could be interacting.

Based on the secondary structure prediction, it is possible to predict which secondary structures could potentially interact with each other. Hydrophobicity of the amino acids within the secondary structures can give an indication of how the protein may fold or interact with other bodies, such as proteins or the cell membrane, to protect its hydrophobic regions from the aqueous environment of the cell. As seen in Figure 13, there are a large number of α-helices from 300 – 540 amino acids. As there is a preponderance of α-helices in the second half of the protein sequence, it is probable that some of the helices may interact with each other.

Helical wheel analyses were performed to predict the arrangement of the amino acids around the α-helices from 301 – 489 amino acids in the CMF63 protein (data not shown for the α-helices between 301 – 368 amino acids). The four helices lying between 419 – 489 amino acids are components of the calcium-binding region. Based on Figure 13, the helix from 393 – 404 amino acids (the target helix) was selected for further analysis, as it had the potential to interact with both the calcium-binding region and a helix adjacent to it. The amino acids VIEGGGSV could serve as a linker region between the target helix and the first helix of the calcium-binding region, from 413 – 424 amino acids, allowing the target helix to interact with this helix. Additionally, ELPQS could serve as a linker region between the target helix and the adjacent helix (378 –387 amino acids), highlighted in green in Figure 13.

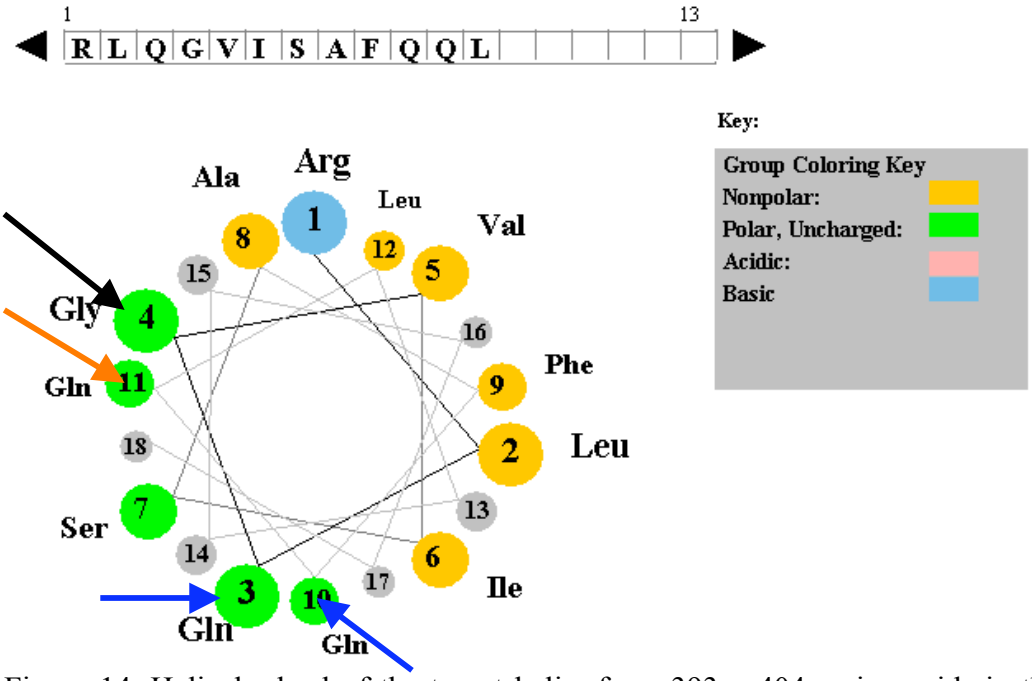


Figure 14. Helical wheel of the target helix, from 393 – 404 amino acids in the CMF63 amino acid sequence. A helical wheel applet (O'Neil and Gisham) was used to make the prediction. A clearly defined hydrophobic face (orange) and hydrophilic face (green) are visible. The helix serves as a good candidate for interacting with other bodies, particularly those with hydrophobic surfaces to protect its hydrophobic face. Since there is more uncertainty in what other bodies that hydrophilic face could be interacting, three different mutants were generated by targeting specific amino acids for mutagenesis. The arrows indicate the sites being targeted for mutagenesis. The blue arrows indicate the Gln being targeted in the Gln double mutant (Q395Q402A). The orange arrow indicates the Gln being changed to Ala (Q403A). The black arrow indicates the Gly that will be changed to Met (G396M).

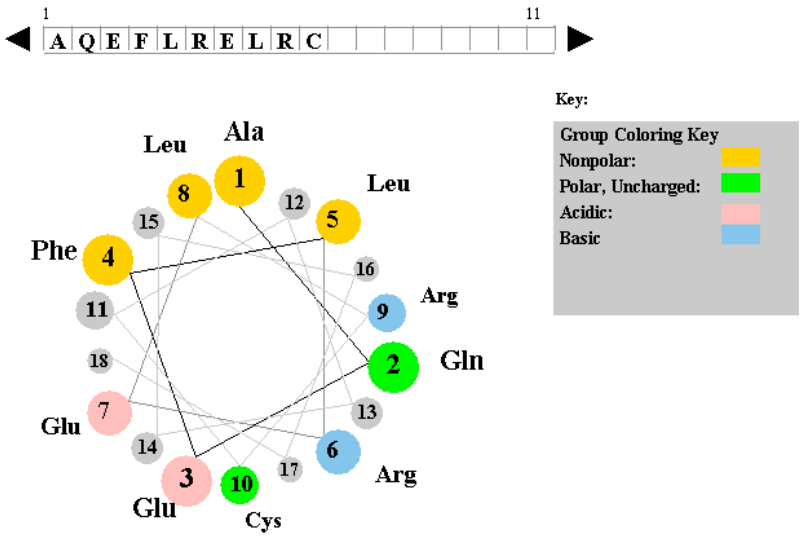


Figure 15 Helical wheel (from 378 – 387 amino acids) with which the target wheel could be interacting in the CMF63 amino acid sequence. The prediction was made using a helical wheel applet (O'Neil and Gisham). A clearly defined hydrophobic face (orange) and hydrophilic face (green, blue, and pink) are visible. Both faces have the potential for interacting with other bodies, particularly the hydrophobic face, which needs to be protected from the aqueous environment of the cell.

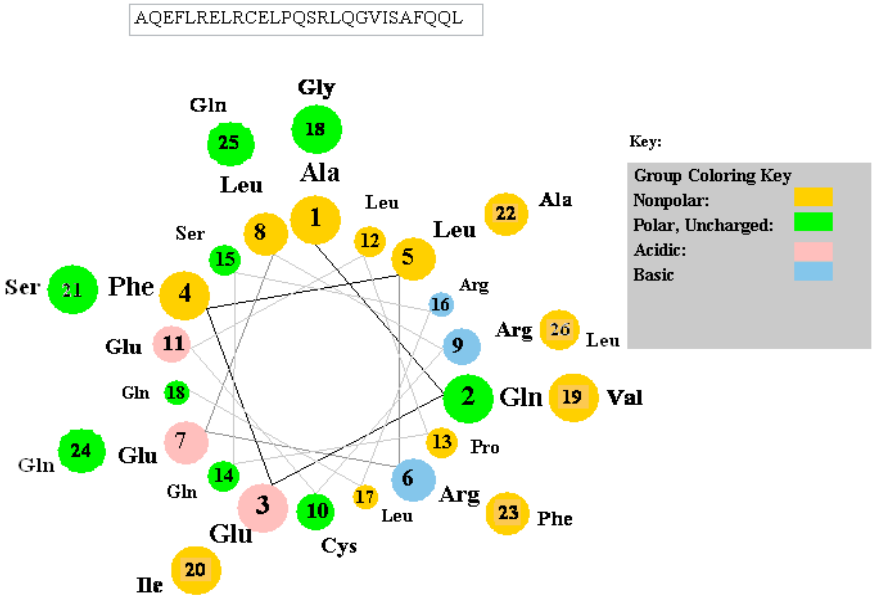


Figure 16. Sequence from 378 – 404 amino acids cannot comprise a single helix. The helical wheel prediction was made using a helical wheel applet (O'Neil and Gisham). There are not any defined hydrophobic and hydrophilic faces; hydrophobic amino acids (orange) overlap hydrophilic (green, pink, and blue) and vice versa. This suggests that it is more probable that two separate helices are formed by the amino acids within this region.

Distinct hydrophobic (orange) and hydrophilic (green, pink, and blue) faces can be seen on the helices in Figure 14 and Figure 15. Since they are adjacent to one another in the sequence, it is possible that the hydrophobic faces could interact with one another to protect these regions from the aqueous environment of the cell. The defined hydrophilic faces have the potential to interact with other proteins or other regions within CMF63. Figure 16 was generated to give a better indication of whether the helices in Figure 14 and Figure 15 are more likely to be interacting with each other or if the amino acids can be arranged to form a larger helix. In Figure 16, a hydrophobic face is formed on the same side as a hydrophilic face and vice versa, affirming that two separate helices are formed. This also provides further support that hydrophobic faces of the helices could be interacting and the amino acids ELPQS (seen in Figure 13) form a linker region between the two helices.

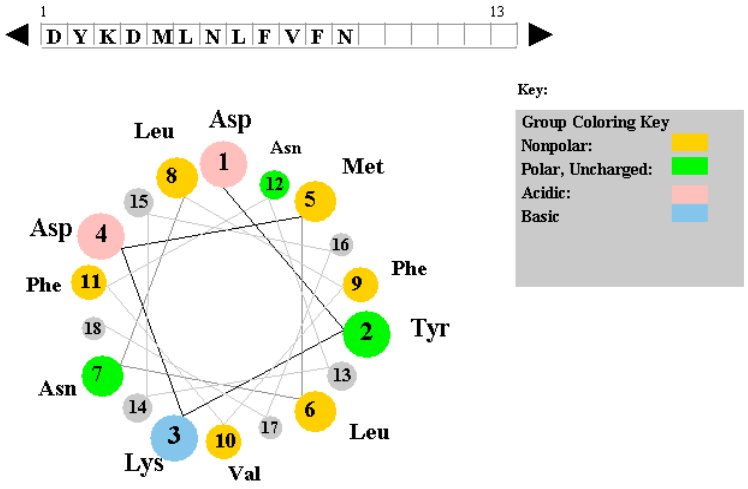


Figure 17. Helical wheel prediction for the first α-helix of the calcium-binding region. The helix lies from 413 – 424.amino acids. The prediction was made using a helical wheel applet (O'Neil and Grisham). No clearly defined hydrophobic and hydrophilic faces are present, so it is not possible to identify which regions could potentially be interacting with the target helix.

Figure 17 shows that the α-helix of the calcium-binding region, which lies from 413 – 424 amino acids, does not appear to have any defined hydrophobic and hydrophilic faces. Based on the secondary structure and helical wheel predictions performed in this study, the absence of defined hydrophobic and hydrophilic faces makes it more difficult to predict where the hydrophilic face of the target helix could potentially interact, if at all.

Site-specific mutagenesis

After generating the three mutants (Q395Q402A, G396M, and Q403A) using site-specific mutagenesis, 10 µl of each product was run on a 0.8% agarose gel to check amplification.

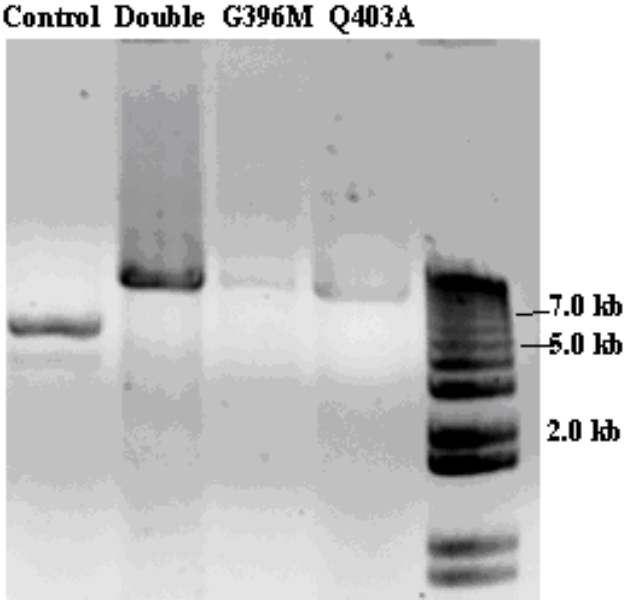


Figure 18. Site-specific mutagenesis products. A 4.5 kb control provided by the kit was used during the amplification of the three mutants Q395Q402A (Double), G396M, and Q403A with their respective primers and the CMF63-pET160 construct. The products were run on a 0.8% agarose gel stained with ethidium bromide. A 1kb Plus DNA ladder (100bp – 12kb) was used. A negative image of the gel was captured with a Fuji Film LAS-3000 Intelligent Dark Box with accompanying image reader software, version 2.0. The control band represents approximately 4.5 kb and those of the mutants represent about 7.5 kb.

As seen in Figure 18, the ladder suggests that 4.5 kb control is the correct size. The actual size of the pET160-CMF63 mutant construct was 7498 bp. Based on the migration of the bands, it appears as though the products were the correct size.

Verification of the identity of the cloned pET160-CMF63 mutant constructs

To verify the identity of the amplified pET160-CMF63 mutant constructs, the constructs were digested with EcoRV.

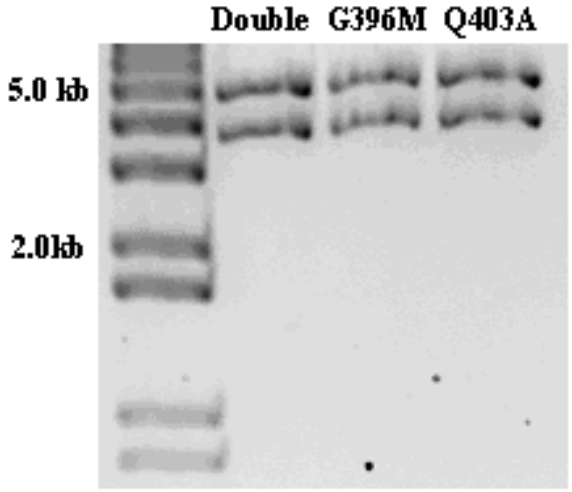


Figure 19. Results from the EcoRV digestion of the mutagenesis products. Site-specific mutagenesis was used to produce pET160-Q395Q402A (double), pET160-G396M, and pET160-Q403A from pET160–endogenous CMF63 and the oligonucleotide primers presented in Figure 8. The products were digested with EcoRV and run on a 0.8% agarose gel, stained with ethidium bromide, with a 1kb Plus DNA ladder (100bp – 12kb) to confirm their identity. A negative image of the gel was captured with a Fuji Film LAS-3000 Intelligent Dark Box with accompanying image reader software, version 2.0.

There are two EcoRV cut sites in the pET160-endogenous CMF63 construct. When digested with the enzyme, the construct should generate two fragments 4237 bp and 3261 bp in size. As expected, two bands were generated on the gel (Figure 19) for each mutant, and they indicate that the fragments were

approximately 3 kb and 4 kb in size, suggesting that the right constructs had been amplified. Based on the results, it was decided that CMF63 mutants should be cloned and ligated to the pKRO8 vector before sending them for sequencing to determine if the desired mutations had been generated.

Preparation of mutants and pKRO8 for ligation

Since the EcoRV digestion confirmed that the correct constructs had been amplified, the mutant genes (Q395Q402A, G396M, and Q403A) were cloned with the XhoI and HindIII primers using PCR, and the products were run on a 0.8% agarose gel to ascertain if the mutant genes had been successfully amplified.

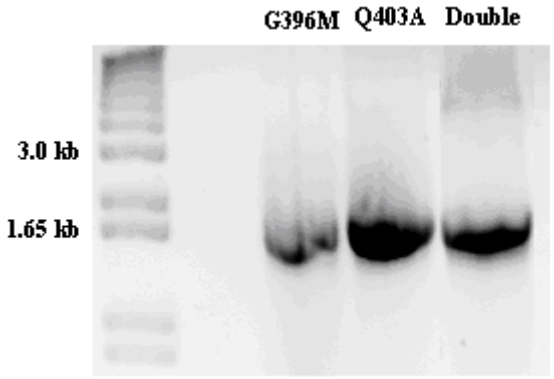


Figure 20. Cloned CMF63 mutants. G396M, Q403A, and Q395Q402A (Double) were PCR amplified with XhoI and HindIII primers. The products were run on a .8% agarose gel, stained with ethidium bromide, with a 1kb Plus DNA ladder (100bp – 12kb). A negative image of the gel was captured with a Fuji Film LAS-3000 Intelligent Dark Box with anaccompanying image reader software, version 2.0. The products were approximately 1.6 kb.

As seen in Figure 20, after being amplified with XhoI and HindIII primers, the Q395Q402A, G396M, and Q403A PCR products are approximately 1.6 kb, which was the anticipated size of the products, as it is about the size of the endogenous CMF63 gene, 1659 bp. Since the products were the correct size, it

suggested that the mutants had been successfully cloned and could be ligated to the expression vector (pKRO8).

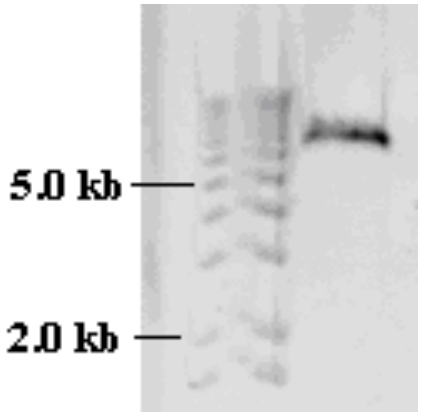


Figure 21. pKRO8 double-digested with XhoI and HindIII. The pKRO8 vector was digested with XhoI and HindIII for 3 hours at 37°C, and the products were run on a 0.8% agarose gel stained with ethidium bromide. A 1kb Plus DNA ladder (100bp – 12kb) was used. A negative image of the gel was captured with a Fuji Film LAS-3000 Intelligent Dark Box with anaccompanying image reader software, version 2.0. The vector was successfully linearized.

Before the cloned mutants could be ligated to the pKRO8 expression vector, the vector had to be cut with HindIII and XhoI, linearizing it and making the correct ends available for ligation. The distance that the band migrates gives an indication of whether a vector is linear or supercoiled. Supercoiled vectors tend to migrate faster than linearized vectors, as the rate of their migration is limited by the molecular weight of the DNA. The band in Figure 21 represents about 6.5 kb, indicating that the pKRO8 vector was successfully linearized by XhoI and HindIII and ready to be ligated to the CMF63 mutants.

Screening for pKRO8-CMF63 mutant transformants

Once the mutant genes had been ligated to the pKRO8 vector, the constructs were digested with HindIII and XhoI restriction enzymes to confirm if the ligation had been successful. Since the mutant CMF63 is approximately 1.6 kb and the vector is about 6.0 kb, it was anticipated that two bands representing these sizes would be generated if the ligation was successful. Figure 22 shows an example of a gel image that represents a successful ligation.

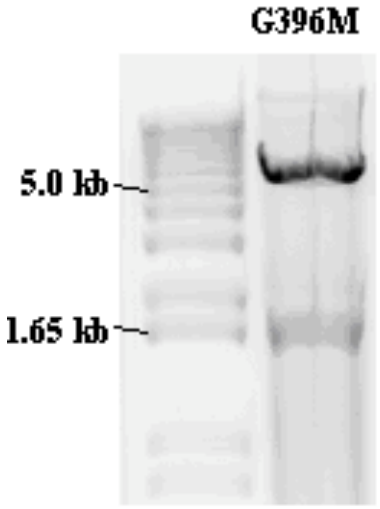


Figure 22. Sample gel image of the results from screening for pKRO8-CMF63 mutant constructs. DH5α transformants were screened for pKRO8-Q395Q402A, pKRO8-Q403A, and pKRO8-G396M by double-digesting them with XhoI and HindIII at 37°C for two hours. A1kb Plus DNA ladder (100bp – 12kb) was used. A negative image of the gel was captured with a Fuji Film LAS-3000 Intelligent Dark Box with anaccompanying image reader software, version 2.0. The bands represent DNA 1.6 kb and 6 kb in size.

Six of the digested samples (data not shown) solely contained the vector; only one band, representing 6 kb, was generated, suggesting that the ends of the linearized vector had been spliced back together.

There were six successful ligations (data not shown); two bands were seen, representing DNA fragments 1.6 kb and 6 kb in size. One of each of the pKRO8-Q395Q402A, pKRO8-Q403A, and pKRO8-G396M samples that were XhoI/HindIII double digested and generated the bands representing 1.6 kb and 6.0 kb were sent for sequencing to ascertain if they possessed the desired mutations.

Sequencing

After confirming that the ligation of the mutant genes to the pKRO8 vector was successful, one of each sample (pKRO8-Q395Q402A, pKRO8-Q403A, and pKRO8-G396M) was sent for sequencing to determine if the engineered amino acid mutations had been incorporated into the mutant genes.

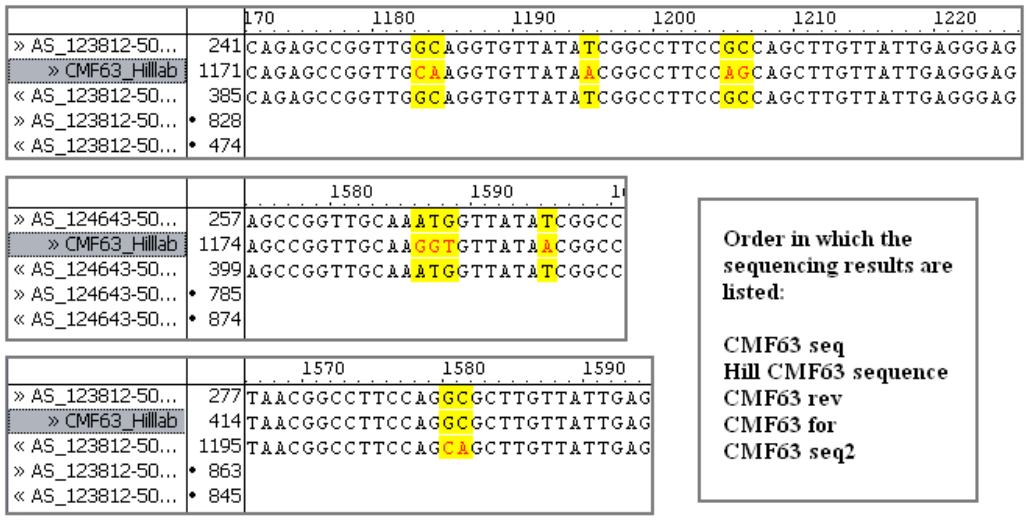


Figure 23. Sequencing results for a) pKRO8-Q395Q402A, b) pKRO8-G396M, and c) pKRO8-Q403A. Four primers- CMF63 seq, CMF63 rev, CMF63 for, and CMF63 seq2- were used for sequencing. The Hill CMF63 sequence for the 29-13 *T. brucei* strain was provided by collaborators at UCLA, who have more accurately sequenced the CMF63 gene. A discrepancy between the Hill sequence and the GeneDB sequence is located at 1196 bp, marked by the TAT highlighted in yellow. All of the designed mutants, highlighted by the remaining yellow boxes, are present.

The sequencing results, presented in Figure 23, were compared to the endogenous CMF63 sequence of the 29-13 strain of *T. brucei*, which was provided by collaborators at UCLA (Hill CMF63 sequence), who have more accurately sequenced the CMF63 gene. The results demonstrate that the mutations designed in the oligonucleotide primers (Figure 8) were present in the samples sent for sequencing. There was a discrepancy between the Hill sequence and the GeneDB sequence at 1196 bp that was not accounted for in the (a) pKRO8-Q395Q492A and (b) pKRO8- G396M; an adenine is present at 1196 bp, instead of a thymine, which was in the GeneDB sequence. When the Hill sequence was inserted into Artemis (Rutherford et al., 2000), a DNA sequence viewer, it was found that the same amino acid was generated, despite not changing thymine to adenine, indicating that a silent mutation was created. The change from thymine to adenine was designed in the sequence of the (c) pKRO8-Q396M mutant, so it matches the Hill sequence. Since no other mutations had been generated in the samples that were sequenced, they were prepared for transfection.

Monitoring the growth of the transfectants

Once the success of the site-specific mutagenesis had been confirmed through sequencing, the three constructs and two controls (pKRO8-Q395Q402A, pKRO8-Q403A, and pKRO8-G396M and the pKRO8-wild-type CMF63 and no DNA controls) were transfected into the non-pathogenic 29-13 recombinant strain of *T. brucei brucei*. The growth of the trypanosomes was carefully monitored after the transfection was performed. The transfectants looked healthy a week

after the transfection was completed. Most of the cells had died ten days post transfection, which was anticipated; the trypanosomes were being cultured with three reporter drugs (hygromycin, G428, and puromycin), and if they had not successfully taken up a construct, which contains a puromycin resistance gene, then they would die without the protection conferred by the gene. Twelve days post-transfection, only one trypanosome was seen in the pKRO8-G396M flask, and no sign of life was seen in any of the other flasks. Since one trypanosome survived the transfection, it was able to multiply via binary fission. Six weeks after the transfection, the cell titer was high enough to split the pKRO8-G396M transfectants. A few days after they were split, the titer did not increase, so some FBS was added to the flask to encourage their proliferation. As a result, the doubling time of the trypanosomes increased, but it was solely 36 hours. However, the normal doubling time of these trypanosomes before they were transfected was approximately 15 – 16 hours. Therefore, the 36-hour doubling time was not fast enough for characterization using a growth curve or a sedimentation assay, so the G396M gene expression could not be tetracycline-induced and the mutants could not be characterized.

DISCUSSION

The goal of my study was to learn more about how CMF63 functions. I performed a BLOCKS search and secondary structure and helical wheel predictions to identify a target that may play a role in proper protein function. I used site-specific mutagenesis to generate amino acid changes in the target that would perturb any potential interactions in which the target is involved. These include interactions with the calcium-binding region of the protein or with another protein. I generated three mutant CMF63 genes and ligated them into separate pKRO8 expression vectors. A recombinant strain of *T. brucei*, 29-13, was transfected with the constructs.

Selecting a target for mutagenesis

Initially, I performed a BLOCKS search to find regions of conserved amino acid sequences in the CMF63 protein. BLOCKS searches for motifs, which are protein sequence patterns, found in different protein sequences (Bork and Koonin, 1996). Motifs are typically characteristic of protein families, domains, or highly conserved residues that may play an important role in protein function; therefore these motifs can be used to predict protein function. A protein family can be assembled as the number of related sequences increases (Bork and Koonin, 1996).

During the BLOCKS search, the CMF63 sequence was compared to the sequences listed in the PROSITE motif library. 81 BLOCKS were found in the entire sequence, 71 of which were found within 121 – 480 amino acids (Table 8). 55% of the BLOCKS within the 121 – 480 amino acids range are found within the calcium-binding region. This suggests that the EF-hand domain is highly conserved among different proteins. The BLOCKS identified were found in proteins that form complexes with other proteins, suggesting that the CMF63 protein may also interact with other proteins.

I used Jpred and Psipred to predict the secondary structure of the CMF63 protein. They analyze the structural features and residue compositions to identify potential secondary structures in a protein (Wilson et al., 2002). Both the BLOCKS and the secondary structure predictions gave me an indication of what region of the protein I should focus on to find a target. Since the majority of the BLOCKS were found within 121 – 480 amino acids, this region is more conserved among proteins and has a greater potential to play an important role in proper protein function. As seen in Figure 13, Jpred predicted that there are predominately α-helices from 301-552 amino acids. I thought it was interesting that there was a preponderance of helices in this region of the protein, so I decided to predict the arrangement of the residues around the helices to try to ascertain if any of the helices could potentially interact with each other.

Around the periphery of the helix, each amino acid in the sequence is spaced 100º apart, comprising 3.6 residues in each turn. Hydrogen bonds connect

the residues; the bonds are formed perpendicular to the helix. It was previously determined that the α-helices from 418 – 489 amino acids comprise the calcium-binding region, so I performed helical wheel analyses on the remaining helices. By generating helical wheels, it is possible to visualize potential side-chain interactions and general characteristics of the helix, such as the hydrophobicity.

From the helical wheel analyses, it was predicted that the α-helices from 393 – 404 amino acids (Figure 14) and from 378 – 387 amino acids (Figure 15) have distinct hydrophobic and hydrophilic faces. These helices are separated by five residues, ELPQS, which could serve as a linker region between them. Since there were only five amino acids separating the α-helices, I wanted to ascertain if there was a possibility that the residues of the two α-helices may form a large α-helix. To do so, I performed an additional helical wheel analysis, in which I inserted the sequence from 378 – 404 amino acids into the applet. The program yielded the helical wheel in Figure 16. This helix does not appear to have a distinct hydrophobic and hydrophilic face; the program predicted that some of the hydrophobic residues would lie on the same face as some hydrophilic residues and vice versa. Therefore, the helical wheel analyses are consistent with the secondary structure predictions. Due to the proximity of the two α-helices, it is highly probable that the hydrophobic faces interact with each other to protect them from the aqueous environment of the cell.

There is a possibility that the hydrophilic face of the target could interact with the first helix in the calcium-binding region, other proteins in the cytosol or

solely with cytosol. Figure 13 shows that the target α-helix from 393 – 404 amino acids also neighbors the first helix of the calcium-binding region (413 – 424 amino acids). The α-helices are separated by eight residues, VIEGGGSV, which could also serve as a linker. A helical wheel (Figure 17) was generated to ascertain with which region of the helix the target could interact. The applet indicated that there is not a clearly defined hydrophobic or hydrophilic region. Consequently, based on this prediction, it was not possible to predict where the interaction could occur.

X-ray crystallography and NMR spectroscopy have revealed that there is great variability in the organization and conformations of the EF-hand motifs (Nelson and Chazin, 1998). As there are numerous potential orientations that the EF-hand domain could potentially have, it is difficult to predict how the two helices would orient themselves with respect to each other and how they could potentially interact. As there is greater uncertainty in interactions that the hydrophilic face of the target helix could be having, I decided to target residues on this face for site-specific mutagenesis. In doing so, I hoped to gain more insights into whether the hydrophilic face is interacting with another body and if the target plays a role in the proper function of the CMF63 protein.

Site-specific mutagenesis

In order to study protein structure and function, amino acid substitutions can be performed in a protein (Dougherty, 2000). The amino acid changes can

give rise to phenotypes that are dramatically different or exhibit little or no change from the natural phenotype (Dougherty, 2000).

Site-specific mutagenesis can give insights into the roles that particular residues play in protein function. In this experiment, site-specific mutagenesis was used to probe potential interactions in which the hydrophilic face of the target helix may be involved. Bulky amino acids were changed to small amino acids and vice versa to try to perturb any potential interactions. As seen in Figure 14, three mutants were generated. The first mutant (Q395Q403A) was a double mutant in which the 395th and 402nd residues, both of which are glutamines (bulky amino acids), were each substituted with an alanine, a small amino acid. The second mutant (G396M) was generated by changing the 396th residue, a small glycine, to a bulky methionine. The third mutant (Q403A) was made by changing the 403rd residue, a glutamine, to an alanine. By generating these mutations, it is possible that the proper function of the protein may be affected, which could give rise to motility defects.

At this moment, it is unknown what motility defects may be seen. In a study performed by Ralston et al. (2006), it was determined that knocking down CMF63 would promote the generation of large clusters and reduce the growth rate of *T. brucei*. As I am not completely silencing gene expression, it is probable that such severe defects will not be seen.

Characterization of the mutants

In order to characterize the mutants, a growth curve and sedimentation assay would be performed. Once the expression of the mutant genes is induced with tetracycline, both of these assays would give an indication if the mutant genes caused any motility defects. As CMF63 has been shown to play a role in motility, a perturbation in a potential interaction in which the hydrophilic face of the target helix may be involved (with its own calcium-binding region or with another protein) may lead to a motility defect. This may, in turn, affect the growth rate of the transfectants. Generating a growth curve would indicate if any of the mutations had affected the ability of the trypanosomes to multiply by showing any shifts in the growth rate from the uninduced transfectants. A sedimentation assay is used to measure motility defects. The procedure for performing the assay is described by Bastin et al. (1999). By growing the induced and uninduced transfectants in cuvettes, the optical density of the transfectants can be compared over time. If the optical density of the transfectants of the induced transfectants decreases, then it can be concluded that the mutations generated abnormalities in motility.

Since no growth has been seen in the flasks with the pKRO8-Q395Q402A, pKRO8-Q403A, and pKRO8-wild-type CMF63 transfectants, the transfections do not appear to have been successful. Consequently, it was not possible to characterize the pKRO8-Q395Q402A and pKRO8-Q403A mutants. After splitting the pKRO8-G396M transfectants six weeks post transfection, their

doubling time was 36 hours, and the normal doubling time of the trypanosomes before they were transfected was approximately 15 – 16 hours. Measurements for a growth curve are taken in 24-hour intervals. Since the transfectants are growing too slowly to perform the assay, it was not possible to induce the transfectants to express the mutant gene with the G396M mutation.

Once the mutants are expressed, motility defects may or may not be seen. Possible motility defects include the formation of clusters and/or a reduction in growth rate. If motility defects are generated, it is an indication that the target plays a role in proper CMF63 protein function. If no motility defects are generated, it is possible that the target may not be needed for the proper function of the protein, or the endogenous CMF63, which would be concomitantly expressed with the mutant, may be able to compensate for any motility defects generated by the mutant; therefore, the mutant phenotype may be masked.

The next steps in this experiment are to perform the transfections with the pKRO8-Q395Q402A, pKRO8-Q403A, and pKRO8-wild-type CMF63 constructs and to characterize all of the mutants by performing a sedimentation assay and generating a growth curve. The ultimate goal of this experiment is to knock down the endogenous CMF63 gene using tetracycline induced RNA interference, allowing solely the mutant CMF63 genes to be expressed.

Conclusion

Using site-specific mutagenesis, I was able to successfully generate mutants that can be used to probe potential interactions in which a predicted target

helix (from 393 – 404 amino acids) is involved. Due to unsuccessful transfections and the slow growth rate of the pKRO8-G396M transfectants, it was not possible to induce the expression of the mutant genes and characterize the mutants. After inducing the expression of the mutant genes and ultimately knocking down the expression of endogenous CMF63, it will be possible to ascertain if the target helix plays a role in the proper function of the protein and the hydrophilic face is interacting with another body.

As CMF63 has been identified as an EF-hand calcium-binding protein, it may play an integral role in intracellular signaling pathways. A study performed by Baron et al. (2007) demonstrated that knocking down the gene generates moderate motility defects; therefore, investigating the function of this gene may lead to further understanding of flagellar motility in *T. brucei*. Such studies may have implications in generation of better treatments for Trypanosomiasis and cilia-related diseases.

LITERATURE CITED

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2001). Molecular Biology of the Cell. 4th ed. NY: Garland Science: 966-967.

Alibu, P., Storm, L., Haile, S., Clayton, C., Horn, D. (2004). A doubly inducible system for RNA interference and rapid RNAi plasmid construction in Trypanosoma brucei. *Mol Biochem Parasitol* 139: 75-82.

Bannai, H., Yoshimura, M., Takahashi, K., and Shingyoji, C. (2000). Calcium regulation of microtubule sliding in reactivated sea urchin sperm flagella. *Journal of Cell Science* 113: 831-839.

Baron, D. M., Ralston, K. S., Kabututu, Z. P., & Hill, K. L. (2007). Functional genomics in *Trypanosoma brucei* identifies evolutionarily conserved components of motile flagella. *Journal of Cell Science* 120: 478-491.

Barrett, M. P., Boykin, D. W., Brun, R., and Tidwell, R. R. (2007). Human African Trypanosomiasis: pharmacological re-engagement with a neglected disease. British Journal of Pharmacology 152:1-17.

Barton, G. (2007). Jpred: TNG A method for protein secondary structure prediction. <http://www.compbio.dundee.ac.uk/~www-jpred/>. (Accessed October 12, 2007).

Bastin, P., Pullen, T. J., Sherwin, T. and Gull, K. (1999). Protein transport and flagellum assembly dynamics revealed by analysis of the paralysed trypanosome mutant *snl-1*. *J. Cell Sci.* 112:3769-3777.

Branche, C., Kohl, L., Toutirals, G., Bulsson, J., Cosson, J., and Bastin, P. (2006). Conserved and specific functions of axoneme components in trypanosome motility. *Journal of Cell Science* 119: 3443-3455.

Bisgrove, B. W. and Yost, H. J. (2006). The roles of cilia in developmental disorders and disease. *Development* 133:4131-4143

Blum, J. and Burri, C. (2002). Treatment of late stage sleeping sickness caused by T.b. gambiense: a new approach to the use of an old drug. *Swiss Med Wkly* 135: 51-56.

Bork, P. and Koonin, E. V. (1996). Protein sequence motifs. *Current Opinion in Structural Biology* 6:366-376.

Bush, A., Chodari, R., Collins, N., Copeland, R., Hall, P., Harcourt, J., Hariri, M., Hogg, C., Lucas, J., Mitchison, H. M., O'Callaghan, C., and Phillips, G. (2007). Primary ciliary dyskinesia: current state of the art. *Archives of Disease in Childhood* 92:1136-1140.

Cuff, J. A. and Barton, G. J. (2000) Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction, *PROTEINS: Structure, Function and Genetics* 40:502-511.

Delespaux, V. and de Koning, H. P. (2007). Drugs and drug resistance in African trypanosomiasis. *Drug Resist. Update* 10:30-50.

Donelson, J. E. (2003). Antigenic variation and the African trypanosome genome. *Acta Tropica* 85:391-404.

Donelson, J. E., Hill, K. L., and El0Sayed, M. A. (1998). Multiple mechanisms of immune evasion by African trypanosomes. *Molecular and Biochemical Parasitology*, 91:51-56.

Dougherty, D. A. (2000). Unnatural amino acids as probes of protein structure and function. *Current Opinion in Chemical Biology* 4:645-652.

Fairlamb, A. H. (2003). Chemotherapy of human African Trypanosomiasis: current and future prospects. *TRENDS in Parasitology* 19:488-494.

Fenn, K. and Matthews, K. R., (2007). The cell biology of Trypanosoma brucei differentiation. *Current Opinion in Microbiology* 10:539-546

Gifford, J. L., Walsh, M. P., and Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. *Biochem. J* 405:199-221.

Grabarek, Z. (2006). Structural basis for diversity of the EF-hand calcium-binding proteins. *Journal of Molecular Biology* 359:509-525.

Hannon, G. J. (2002). RNA interference. Nature 418:244-251.

Hare, J. D. (2007). Mutational analysis of *T. Brucei* components of motile flagella (TbCMF) genes in the African trypanosome. Honors thesis, Mount Holyoke College.

Hertz-Fowler, C., Peacock, C. S., Wood, V., Aslett, M., Kerhornou, A., Mooney, P., Tivey, A., Berriman, M., Hall, N., Rutherford, K. et al. (2004). GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res*. 32:339-343.

Hill, K.L. (2003). Biology and mechanism of trypanosome cell motility. *Eukaryote Cell* 2:200-208.

Holwill, M. E., (1974). Some physical aspects of the motility of ciliated and flagellated microorganisms. *Sci. Prog.* 61:63-80.

Hutchings, N. R., Donelson, J. E., Hill, K. L. (2002). Trypanin is a cytoskeletal linker protein and is required for cell motility in African trypanosomes. *Journal of Cell Biology* 156:867-877.

Ibañez-Tallon, I., Heintz, N., and Omran, H. (2003). To beat or not to beat: roles of cilia in development and disease. *Human Molecular Genetics* 12: R27-R35.

Jones, D. T. McGuffin, L. J., and Bryson, K. (2000). The PSIPRED protein structure prediction server. <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>. (Accessed October 12, 2007).

Kohl, L., Robinson, D., and Bastin, P. (2003). Novel roles for the flagellum in cell morphogenesis and cytokinesis of trypanosomes. *The EMBO Journal* 22:5336-5346.

Keiser, J., Stitch, A., and Burri, C. (2001). New drugs for the treatment of human African Trypanosomiasis: research and development. *TRENDS in Parasitology* 17:42-49

Kennedy, P. G. E. (2004). Human African Trypanosomiasis of the CAN: current issues and challenges. *J. Clin. Invest* 113:496-504.

Kyoto University Bioinformatics Center. MOTIF: Searching Protein and Nucleic Acid Sequence Motifs. <http://motif.genome.jp>. Accessed 2007, Oct 12

Lanfear, S. M. and Ignatushchenko, M. (2001). The flagellum and flagellar pocket of trypanosomatids. *Molecular and Biochemical Parasitology* 115:1-17.

Matthews, K. R. (2005). The developmental cell biology of *Trypanosoma brucei*. *J. Cell Sci* 118:283-290.

McGuffin LJ, Bryson K, Jones DT. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16:404-405.

Motyka, S. A. and Englund, P. T. (2004). RNA interference for analysis of gene function in trypanosomatids. *Current Opinion in Microbiology*, 7:382-368.

Nelson, M. R. and Chazin, W. J. (1998). Structures of EF-hand Ca(2+)-binding proteins: Diversity in the organization, packing and response to Ca^{2+} binding. *Biometals: an international journal on the role of metal ions in biology, biochemistry, and medicine* 11:297-318.

Nok, A. J. (2003). Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human African trypanosomiasis. *Parasitol Res* 90:71-79.

O'Neil, E. K. and Gisham, C. M. *Helical Wheel Applet*. University of Virginia. <http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html>. (Accessed November 8, 2007)

Pan, J., Wang, Q., and Snell W. J. (2005). Cilium-generated signaling and cilia-related disorders. *Laboratory Investigation* 85:452-463.

Pépin, J. and Meda, H. A. (2001). The epidemiology and control of human African trypanosomiasis. *Advances in parasitology*, 49, 71-132.

Ralston, K. S., Lerner, A. G., Dierner, D. R., and Hill, K. L. (2006). Flagellar motility contributes to cytokinesis in *Trypanosoma brucei* and is modulated by an evolutionarily conserved dynein regulatory system. *Eukaryotic Cell* 5:696-711.

Ramos, H. C., Rumbo, M., and Sirard, J. (2004). Bacterial flagellins: mediators of the pathogenicity and host immune responses in mucosa. *TRENDS in Microbiology* 12:509-517.

Rana, T. M. (2007). Illuminating the silence: understanding the structure and function of small RNAs. Nature reviews. *Molecular cell biology* 8:23-36.

Rocha, G. M., Brandão, B. A., Mortara, R. A.m Attias, M., de Souza, W., and Carvalho, T. M. U. (2006). The flagellar attachment zone of *Trypanosoma cruzi* epimastigote forms. *Journal of Structural Biology*, 154, 89-99.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16:994-995.

Sanger Institute. Gene DB *T. brucei* BLAST. <http://www.genedb.org/genedb/tryp/blast.jsp>. (Accessed September 26, 2007)

Santamaria-Kisiel, L., A. C. Rintala-Dempsey, and G. S. Shaw. 2006. Calcium-dependent and -independent interactions of the S100 protein family. *The Biochemical Journal* 396:201-214.

Smith, E. F. (2002). Regulation of flagellar dynein by calcium and a role for an axonemal calmodulin and calmodulin-dependent kinase. Molecular Biology of the Cell 13:3303-3313.

Stijlemans, B., Guilliams, M., Raes, G., Beschin, A., Magez, S., and De Baetselier, P. (2007). African trypanosomosis: From immune escape and immunopathology to immune intervention. *Veterinary Parasitology* 148:3-13.

Taylor, J. E., and G. Rudenko. (2006). Switching trypanosome coats: what's in the wardrobe? *Trends in genetics*: TIG 22:614-620.

Tobin, J. I. and Beales, P. I. (2007). Bardet-Biedl sydrome: beyond the cilium. Pediatr Nephrol 22:926-936.

Ullu, E., Tschudi, C., and Chakraborty, T. (2004). RNA interference in protozoan parasites. *Cellular Microbiology* 6:509-519.

Vincendeau, P., and B. Bouteille. 2006. Immunology and immunopathology of African trypanosomiasis. *Anais da Academia Brasileira de Ciencias* 78:645-665.

Vickerman, K., Tetley, L., Hendry, K. A. K., and Tunrer, C. M. R. (1988). Biology of African trypanosomes in the tsetse fly. *Biology of the Cell* 64:109-119.

Vora, N., Rerrone, R., and Bianchi, D. W. (2008). Reproductive issues for adults with Autosomal Dominant Polycystic Kidney Disease. *American Journal of Kidney Diseases* 51:307-318.

Wilson, C. L., Hubbard, S. J., and Dolg, A. J. (2002) A critical assessment of the secondary structure prediction of α-helices and their termini in proteins. *Protein Engineering* 15:545-554.

Wirtz, E., Leal, S., Ochatt, C., Cross, G. A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol Biochem Parasitol* 99:89-101.

Wu, Y., Deford, J., Benjamin, R., Gou-Shu, M., and Ruben, L. (1994). The gene family of EF-hand calcium-binding proteins from the flagellum of *Trypanosoma brucei*. Biochem. J 304:833-841.