CLONING AND FUNCTIONAL ANALYSIS OF TRANS-SPLICING FACTORS IN *Trypanosoma brucei*

By

OWINO V. A. (B.Sc.)

A research thesis submitted to the Graduate School in partial fulfilment for the requirements of the Master of Science Degree in Biochemistry of Egerton University.

EGERTON UNIVERSITY

August 2006

DECLARATION AND RECOMMENDATION

I declare that this thesis is my own original work and has not been presented before for the award of any degree.

STUDENT:
SIGNATURE:............... DATE: 30th August 2006
OWINO V. A.

This thesis has been submitted with our approval as supervisors.

SUPERVISORS:

Dr. DANIEL K. MASIGA.
Biochemistry and Biotechnology Department
Kenyatta University

SIGNATURE:.................... DATE: 30/08/2006

PROF. MOSES K. LIMO.
Department of Biochemistry and Molecular Biology
Egerton University

SIGNATURE:....MKLimo..... DATE:.........1/9/2006

DEDICATION

This work is dedicated to Lord God Almighty and my mother, Anne Awino.

ACKNOWLEDGEMENTS

I give God the Glory for the great opportunities He gave me during this thesis work and realization of my dream.

In a very special way, I would like to thank my supervisors, Dr. Daniel K. Masiga and Prof. Moses K. Limo for their professional advice during this work. Special thanks to Dr. Ellie O. Osir of International Centre Insect Physiology and Ecology (ICIPE) for his material support.

Many thanks also to the Director Research and Partnership, Prof Onesmo K. Ole-MoiYoi for providing me with a Dissertation Research Internship Programme (DRIP) scholarship. Special thanks to Centre Director and staff at Kenya Agricultural Research Institute–Trypanosome Research Centre (KARI-TRC) tissue culture especially Joana Auma, Jane Hanya and Steve Okaye for their help during *in vitro* trypanosome propagation.

Special thanks to Benson Nyambegah for guidance in application of bioinformatics approaches during the study.

I am also grateful to each member of Molecular Biology and Biotechnology Department at ICIPE especially Violet Jepchumba, Paul Mireji, James Mutunga, Marion Warigia, Betty Mbatia, Harrison Kibogo, Ishmael Rabbi, Judith Adhiambo and Daniel Amin for their encouragement and support.

Lastly, my heartfelt gratitude, thanks and appreciation go to my family and Amos Awendo for their patience, moral and material support during the entire period of this work.

ABSTRACT

Trypanosomes are parasitic protozoa that cause trypanosomosis. The disease is a threat to human population and a major impediment to livestock production and economic development in many countries in sub-Sahara Africa, where it is endemic. Presently, it is controlled by reduction of tsetse fly vector population, chemotherapy and chemoprophylaxis. However, none of these approaches is completely effective and thus the need for development of new approaches. The splicing machinery in trypanosomes presents a potential target for anti-trypanosome drug development. The development of new drugs is needed because of increasing incidents of resistance to available drugs. In trypanosomes, the mature messenger RNA is derived from independent pre-mRNA molecules in a process called trans-splicing. This is a variation from the more common cis-splicing which occurs in the mammalian hosts of trypanosomes, where the mature RNA is derived from one pre-mRNA molecule. The process is important in regulation of gene expression in trypanosomes that is predominantly post-transcriptional. In this study, thirteen homologs of *Trypanosoma brucei* genes for trans-splicing and polyadenylation were identified *in silico* using *Trypanosoma cruzi,* yeast and/or human splicing and polyadenylation factors to query GeneDB, the repository of genome data for *Trypanosoma brucei, Trypanosoma cruzi* and *Leishmania major.* Degenerate PCR approach was used to clone the factors, which were subsequently sequenced. The amino acid sequences generated were used to query public protein databases and were also compared to homologous sequences from *T. cruzi, L. major* and *Homo sapiens.* Conserved RNA binding proteins domains and domains of proteins involved in multi-protein complex assemblies were identified. The kinetoplastid sequences were similar to each other, but were individually significantly different from human homologs. Sequence specific gene silencing of three factors (P14, CPSF 30 and U1-70k) using RNA interference (RNAi) technique was lethal suggesting the importance of the three factors in viability. Protein depletion was detected in the silencing of the P14 gene. Significant variations of the kinetoplastid sequences from human and importance of the factors in parasite viability suggest that some components of the trypanosome spliceosome are targets for the design of novel drugs.

TABLE OF CONTENTS

Title

CHAPTER ONE

INTRODUCTION

CHAPTER TWO

LITERATURE REVIEW

CHAPTER THREE

MATERIALS AND METHODS

CHAPTER FOUR

RESULTS AND DSCUSSION

CHAPTER FIVE

Appendix 1

Appendix 2

Appendix 3

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAT:	African Animal Trypanosomosis
BCIP:	5-Bromo-4-chloro-3-indolyl phosphate
BPS:	Branch Point Sequence
CPSF:	Cleavage and Polyadenylation Splicing Factor
CstF:	Cleavage Stimulating Factor
DALYs:	Disability-Adjusted Life Years
DDT:	Dichloro-diphenyl-trichloroethane
DEAE:	Diethylaminoethyl
DFID:	Department For International Development
DFMO:	Difluoromethylornithine
dsRNA:	Double Strand RNA
HAT:	Human African Typanosomosis
IPTG:	Isopropyl-β-D- thiogalactopyranoside
LB:	Luria-Bertani
miRNA:	microRNA
NBT:	Nitro Blue Tetrazolium
ORFs:	Open Reading Frames
PAP:	polyA polymerase
PCR:	Polymerase Chain Reaction
PTGS:	Post-Transcriptional Gene Silencing
Py:	Pyrimidine Tract
RdRP:	RNA-dependent RNA Polymerase
RISC:	RNA-induced Silencing Complex
RNAi:	RNA Interference
RRM:	RNA Recognition Motif
SAP:	Spliceosome Associated Proteins
SF:	Splicing Factor
siRNA:	Small Interfering RNA
SIT:	Sterile Insect Technique
SL RNA:	Spliced Leader RNA
SL RNP:	Spliced Leader Ribonucleoprotein
snRNA:	Small Nuclear RNA

snRNP:	Small Nuclear Ribonucleoprotein
SR:	Serine- Arginine
U2AF:	U2 Auxiliary Factor
WHO:	World Health Organization
X-Gal:	5-bromo-4-chloro-3-indoyl-β-D- galactopyranoside
ZPFM:	Zimmerman Post-Fusion Medium

CHAPTER ONE

INTRODUCTION

1.1 Background information

African trypanosomes, members of genus *Trypanosoma*, cause diseases that affect man and domestic animals. The disease is known as sleeping sickness in human (or Human African Trypanosomosis, HAT) and nagana in cattle. Tsetse-transmitted trypanosomosis affects 36 countries in sub-Saharan Africa where more than 48 million cattle and 60 million people are at risk (WHO, 2002). It has been recognised as a cause of severe morbidity and mortality throughout sub-Saharan Africa and a major constrain to livestock production (Allsopp, 2001). Annual losses in cattle production alone are estimated at US $1.2 billion (FAO, 2002). Despite its impact, trypanosomosis remains among the most neglected tropical diseases (Morel, 2003).

Trypanosoma brucei has three subspecies, two of which cause sleeping sickness. *Trypanosoma brucei gambiense* causes a chronic form of the disease and is found mainly in Western and Central Africa. *Trypanosoma brucei rhodesiense* causes a more acute form of sleeping sickness and occurs in Eastern and Central Africa while *Trypanosoma brucei brucei* does cause the disease in man. One of the control efforts of the disease is the use of drugs. In the recent years increased incidents of resistance to available drugs has limited their value. Tsetse control methods such as the use of trapping and bait technology, insecticide spraying and the Sterile Insect Technique (SIT) have also been applied with varying levels of success. The efforts to develop more efficient and sustainable methods and tools for the control of African trypanosomosis must therefore continue.

Drug development depends on the understanding of parasite biology. Parasites have developed mechanisms that enable them to survive and thrive in their hosts. Rational drug development is based on differences that allow the development of products that are selectively toxic to parasites. Trypanosomes have unique molecular mechanisms that are different from their human and animal hosts, which should be investigated and exploited for effective chemotherapeutic control.

In eukaryotic organisms, genes occur as coding regions (exons) interrupted by non-coding regions (introns) in genomes. They are decoded into precursor messenger ribonucleic acid (pre-mRNA) by the process known as transcription. Synthesis of mRNA occurs in the spliceosome, which is a macromolecular complex consisting of small nuclear RNA (snRNA) and proteins (Jurica and Moore, 2003). The spliceosome contains various macromolecular interactions: RNA-RNA, RNA-protein and protein-protein interactions and the complex facilitates the process of splicing, which entails the removal of introns and joining of exons. The majority of trypanosome transcript lack introns and are transcribed as polycistronic units (Lücke *et al.*, 1997; Denker *et al.*, 2002; Liang *et al.*, 2003).

In trypanosomes, the mature mRNA is derived from independent pre-mRNA molecules in a process known as trans-splicing (Sutton and Boothroyd, 1986; Denker *et al.*, 2002; Garcia-Blanco, 2003). Trans-splicing in trypanosomes appears to be linked to polyadenylation, the addition of a poly-adenosine tail to the 3'-end of pre-mRNA (Clayton, 2002; Jurica and Moore, 2003). In mammals, however, trans-splicing of conventional pre-mRNAs appears to be exceedingly rare due to the presence of trans-acting inhibitors or lack of specific trans-activators (Garcia-Blanco, 2003). Proteins that are essential for trans-splicing but not for cis-splicing have also been recorded to be absent in human, fly and plant genomes (Denker *et al.*, 2002). Cis-splicing involves formation of mature mRNA from a single pre-mRNA. The presence of trans-splicing-specific factors indicates that this process can be explored as a possible target for therapeutic intervention. Thus, in this study, insights to the *Trypanosoma brucei* trans-splicing and polyadenylation factors are sought.

1.2 The statement of the problem

Trypanosomosis is a major constrain to livestock farming in sub-Saharan Africa and limits the full potential of agricultural development in the 36 countries that hold the continent's greatest potential for expanded agricultural production (Swallow, 1999). Despite the annual loss of US $1.2 billion in cattle production alone and an estimated US$ 4.7 billion in agricultural gross domestic production (FAO, 2002), the disease remains among the most neglected in terms of drug development. Most available drugs are highly toxic and increasingly encounter parasite resistance. Since we may soon be confronted with complete absence of effective drugs, there is an urgent need for research geared towards identification of unique biological processes that will facilitate clinical development of promising compounds. Investigation on trans-splicing and polyadenylation factors and their potential functions thus provides a foundation for appropriate target identification in efforts to develop anti-parasitic drugs. Improved chemotherapeutic control strategy will enhance the realization of Africa's optimum agricultural potential that would in turn support her economically disadvantaged inhabitants.

1.3 General objective

To clone and undertake a functional study of factors involved in trans-splicing and polyadenylation in *Trypanosoma brucei.*

1.3.1 Specific objectives

i. To identify *Trypanosoma brucei* homologs of the *Trypanosoma cruzi, Leishmania major, Homo sapiens* and *Saccharomyces cerevisiae* splicing and polyadenylation factors in public database.

ii. To amplify, clone and sequence genes encoding these factors.

iii. To analyse the nucleotide and amino acid sequences of the cloned factors.

iv. To express the open reading frames (ORFs) of selected factors in bacterial expression system.

v. Undertake functional analysis of selected factors using RNA interference technique.

1.4 Justification

Kinetoplastid parasites such as trypanosomes have developed specific variations from common eukaryote mechanisms such as trans-splicing and translation (Denker *et al.,* 2002). The variations are determined at least partially by the protein components. Hence, it is reasonable to propose that trypanosome proteins can harbour some fundamental variations when compared to homologous proteins in the mammalian host. These variations are potential chemotherapeutic targets.

Trans-splicing is an essential step in the expression of all protein-coding genes in trypanosomes (Lücke *et al.,* 1997; Sturm and Campbell, 1999; Palfi *et al.,* 2000). The factors involved in trans-splicing have also been shown to provide crucial communication links between splicing machinery and other processes such as transcription, capping and polyadenylation (Jurica and Moore, 2003). Analysis of factors involved in trypanosome trans-splicing will therefore provide insights into the machinery employed by them to carry out their functions. The investigations of trans-splicing and polyadenylation factors and their potential interactions will provide insights into the linkage between trans-splicing and polyadenylation, which appear to be closely coordinated. It also lays a foundation the discovery of potential drug target.

CHAPTER TWO

LITERATURE REVIEW

2.1 Impact of trypanosomosis in Africa

African trypanosomosis affects man and his livestock. The disease limits Africas` agricultural potential since it adversely restricts livestock keeping and affects, directly or indirectly, other mans` economic activities. The economic deprivation is exacerbated by losses in milk production, tractive power, waste products that provide natural fuel and fertilizer and secondary products such as clothing and hides. This denies Africa a great economic benefit since affected regions are the continent's greatest potential for expanded agricultural production (Swallow, 1999).

2.1.1 Human African trypanosomosis (HAT)

Human African trypanosomosis (HAT) also known as sleeping sickness occurs only in Africa. This devastating disease kills all those infected unless treated (Kioyi and Mattock, 2005). Although accurate statistics for HAT are not available, it is estimated that between 300,000 and 500,000 new infections occur per year, and more than 60 million people are at risk of contacting the disease (WHO, 2000; Fairlamb, 2003; Kubata *et al.*, 2005). An estimated 50,000 people die every year with 100% fatality rate in untreated patients (Table 1). However, these numbers are considered underestimates due to difficulties in diagnosis and remoteness in endemic regions (WHO, 2000).

HAT has been a major cause of depopulation of large tracts of Africa and mainly strikes the active adult population (WHO, 2002). This has led to abandonment of fertile lands in tsetse infested area by farmers due to fear of contracting the disease. It is estimated that the average rural inhabitant would lose 10 years of income, or $615 (1986 U.S. dollars), because of premature death caused by the disease (WHO, 2002). With the scourge making a comeback, greater risks in human health are unavoidable particularly in regions experiencing large-scale migrations and inadequate or non-existent health systems, such as in southern Sudan, Democratic Republic of Congo and Angola. These regions also have suffered political instability, displacement of populations and war leading to extreme poverty in a large proportion of the population. With these limitations and poor monitoring structures, the exact impact of the disease may be more severe than is estimated.

Table 1. Fact file: African trypanosomosis statistics

Statistic/Information	**Statistical information in sub-Saharan Africa**
Geographical location	36 countries
Population at risk	60 million
Number of deaths at 2001	50, 000
Global case fatality rate in 2002	N/A
Number of people infected in 2003	300,000-500,000
Global disease burden in 2001 (DALYs)	1.6 million

Abbreviations: DALYs, Disability-Adjusted Life Years; N/A, Not Applicable.

Only 3–4 million people are under active surveillance.

Source: Fairlamb, 2003.

2.1.2 African animal trypanosomosis (AAT)

African trypanosomosis remains a major constraint to livestock productivity and continues to impede intensification of crop-livestock systems across vast areas of the humid and sub-humid zones of Africa that hold the greatest potential for increased agricultural production. The main direct economic impact of trypanosomosis is on cattle through reduction of birth rates and increment of abortion and mortality rates. Annual losses here amount to an estimated 3 million deaths, mainly of young stock, with up to 25% mortality in pre-weaning calves (FAO, 2002). Mortality losses are compounded by lower reproduction, milk yields and weight gain. With 94% of the total African cattle population distributed at the fringes of the continental tsetse belt, it is estimated that direct losses amount to US$ 1.2 billion yearly (FAO, 2002). The distorted cattle distribution also affects crop production. In the absence of cattle, there is no draught power for ploughing, less manure to use as fertilizer, less feeding of animals with crop residues and by-products (DFID, 2001; Kabayo, 2002). The productivity of the land therefore remains sub-optimal. Rearing of susceptible animals is uneconomical in areas with high tsetse challenge, but minimal mixed farming is possible in areas with low tsetse challenge. Other valuable livestock such as camels suffer from the disease. However, little information on the economic impact of the disease in small ruminants is documented.

2.2 Control of trypanosomosis

For more than one century, controlling trypanosomosis using a variety of approaches has been a major focus for research and development. These activities include strategies to control the tsetse fly vector and the parasite in livestock and man. These strategies have however recorded varied success.

2.2.1 Vector control

Tsetse control has long been an important option for reducing the incidence of trypanosomosis. The goal is to significantly reduce the vector densities in areas populated by susceptible hosts. Initial attempts to control tsetse involved widespread bush clearing to destroy the flys` breeding habitats. This method is no longer employed since it poses a serious threat to biodiversity. Other methods used to control the vector population include spraying with insecticides, SIT, targets, traps and bait techniques (Aksoy *et al.*, 2001).

The discovery of dichloro-diphenyl-trichloroethane (DDT) and other persistent insecticides in 1945 paved way for ground spraying as the main line of defence against tsetse fly throughout Africa (Dransfield *et al.*, 1991; Allsopp, 2001). Other chemical compounds that have been used include chlorobenzene derivatives, halobenzene insecticides and synthetic pyrethrins. Aerial applications of insecticides using fixed wing aircraft were used with considerable success in savannah regions and from helicopters in forested areas (Aksoy *et al.*, 2001). Although quite successful, this method is environmentally damaging and has resulted in the spread of insecticide resistance in many vectors. The problem of reinvasion of previously cleared areas has also been exhibited (Schofield and Maudlin, 2001). Due to this shortfall in addition to high costs, there is minimal reliance to this approach but the dependence still exists.

Bait technology became widely used in the 20th century and is an attractive tactic for reducing tsetse fly population (Allsopp, 2001). It involves attracting tsetse with visual or olfactory baits to allow capture and killing. Improvement of this technique is the use of live baits (i.e. cattle) to which insecticide has been applied by dipping or spraying or with pour-on formulations (Allsopp, 2001). It has become increasingly popular since it is suited to use by individual farmers or communities. Recent availability of synthetic pyrethroid insecticides, which are highly effective against insect pests but have a low mammalian toxicity, has boosted its use.

Additional environmentally acceptable methods for eradicating tsetse consist of an integrated campaign using insecticide-treated screens or traps. They have been made simpler into cloth screens impregnated with insecticides, called targets. The targets and traps developed have been shown to reduce fly population to low levels but only in some areas and species of tsetse (Allsopp, 2001). Moreover, to be efficacious, this control method demands regular target and trap maintenance and the active participation of the livestock keeping communities where traps are deployed (Allsopp, 2001).

Sterile insect technique (SIT) involves sustained systematic release of sterile male insects to fertilise with the wild females among the wild population, which are then unable to produce viable progeny. SIT does not require insecticides and is environmentally benign. It has been used successfully in Burkina Faso, Tanzania, Nigeria and most recently, in Zanzibar (Allsop and Phillemon-Motsu, 2000). However, the problem of reinvasion has been reported in the three countries except Zanzibar (WHO, 2004). While the low reproductive rate of tsetse makes this a highly desirable approach, it has been criticized due to the relatively large upfront costs that would be associated with its implementation (Aksoy, 2003). Unfortunately, the overall control of the vector is made more complex since tsetse flies have adapted to a wide range of habitats, from central African humid rain forests to the vast, semi-arid, open savannahs of Eastern Africa (Allsopp, 2001).

2.2.2 Chemotherapy

Chemotherapy is one of the methods heavily relied on in trypanosomosis control. This method has the advantage that it can be used under any production system, in any ecological zone and farmers can individually apply trypanocides to their cattle. Various compounds have been developed for human and animal treatment. However, set against the spectacular advances in other areas of chemotherapy over the past, trypanosomosis in particular has fared extremely badly. This is because of low profit margins expected by pharmaceutical companies since people most at risk are among the poorest in the world. The emergence of chemoresistance and toxicity highlights the limits of chemotherapy.

2.2.2.1 Chemotherapy in human

Sleeping sickness evolves through clinically distinct stages namely early/stage 1 and late/stage 2. The early stage is a haemolymphatic stage in which parasites develop in the blood, lymph and peripheral organs. The parasites then spread to the central nervous system (late/encephalitic stage) where they cause serious neurological disorders. The type of treatment depends on the phase of the disease (Table 2). Pentamidine isethionate (Lomidine^{TM}) is used only for the early phase of the disease since it does not cross the blood-brain barrier. It is active against *T. b. gambiense*, but is not used against *T. b. rhodesiense*, since primary resistance to it has been found in some areas. Other limitations include inactivity when dosed orally, slow elimination and toxicity. It binds negatively charged cellular components, such as phospholipids and nucleic acids and disrupts the structure of kinetoplast DNA (kDNA). Suramin (Germanin^{TM}) is used for treating the early stages against *T. b. rhodesiense.* The mode of action is still not well understood but it is suggested to inhibit various glycolytic enzymes (Fairlamb and Bowman, 1980; Fairlamb, 2003). All of these drugs are expensive and have serious side effects.

The regimen for treatment of advanced stage (late phase/stage 2) of HAT where the central nervous system is affected involves the use of melarsoprol (Mel B, Arsobal^{TM}). It inhibits trypanothione reductase, an enzyme essential for maintaining the correct thiol-redox balance for biochemical processes. Its use is limited since it causes encephalopathy in 5-10% of cases, half of which are fatal (Fairlamb, 2003). Other side effects include vomiting, abdominal colic, peripheral neuropathy, arthralgia and thrombophlebitis. A new drug eflornithine (Difluoromethylornithine/DFMO, Ornidyl^{TM}) has been developed that is effective in the treatment of both the early and the late stage of West African sleeping sickness. However, it is inactive against East African sleeping sickness due to innate tolerance of *T. b. rhodesiense* to the drug (Matovu *et al.*, 2001). It is currently mainly used as back-up drug for melarsoprol refractory *T. b. gambiense* cases, but recently also as first line treatment (EANETT, 2003). It is an ornithine decarboxylase inhibitor, which involves a 14-day treatment regime at the cost of

US$ 250 per patient for the drug alone. Like melarsoprol, eflornithine is expensive and requires weeks of hospitalisation. At present, only two drugs (DB 289, diamidine derivative and Megazol, nitro-imidazole) are under development (Legros *et al.*, 2002).

Table 2. Overview of human trypananocidal drugs

Drug	**Species**	**Indication** **½ life**	**Year of first use**	**Comments.**
Pentamidine isethionate	T.b.g	Stage 1 [9.4 hrs - IM] [6.4 hrs-IV]	1940	2nd line drug when suramin therapy is contraindicated
Suramin Sodium	T.b.g T.b.r	Stage 1 [50 days]	Early 1920s	No significant clinical resistance has emerged
Melasoprol (Mel B)	T.b.g T.b.r	Stage 2 [35 hrs]	1949	Increased treatment failure
Eflornithine	T.b.g	Stage 2 [3 hrs]	1981	Difficult use
Nifurtimox¥	T.b.g T.b.r?	Stage 2 [3.5 hrs]	1977	Not registered for HAT Case series only Toxic and action on T.b.r. poorly documented

T.b.g =*Trypanosoma brucei gambiense*, T.b.r =*Trypanosoma brucei rhodesiense*

IM = intramuscular, IV = intravenous

¥ = registered for treatment of *T. cruzi* infections and is used in combination with melarsoprol (EANETT, 2003). The table is modified from Legros *et al.* (2002).

2.2.2.2 Chemotherapy in animals

Diminazene aceturate (Berenil™), isometamidium chloride (Samorin™) and homidium bromide (Novidium™) are the drugs widely used for the treatment of nagana or cattle trypanosomosis. These drugs are effective both for treatment and for prophylaxis. Diminazene aceturate binds to AT-rich regions of DNA and RNA duplexes exhibiting properties characteristic of intercalation and minor groove binding (Portugal, 1994). This unwinding inhibits type II topoisomerase. Isometamidium chloride acts by cleaving kDNA-topoisomerase complexes (Wells *et al.*, 1995) while homidium bromide intercalates into the DNA thus inhibit DNA-primed polymerase. Due to their mode of action, they are mutagenic and tumorigenic and also cause necrosis and sloughing of extensive areas of unpigmented skin. Suramin, used in treatment of sleeping sickness, is also used against *Trypanosoma evansi* in camels. Cattle treated with these drugs may only be slaughtered and used for consumption, several months after treatment, when there is no residual drug left in the animal.

2.2.3 Prophylactic strategies

Prophylaxis concerns all measures that can be undertaken to prevent the onset or transmission of a disease. In trypanosomosis, the prophylactic strategies involve actions taken against vector population and health care for both human and animal populations. Two management approaches are applied in human populations, these are, routine screening for the disease and treatment of the infected persons. In cattle, isometamidium chloride (Samorin^{TM}) is the most commonly used prophylactic drug. Although extensively used in trypanosomosis control, chemoprophylaxis is expensive, and thus unsatisfactory long-term solution to the problem of African trypanosomosis.

2.2.4 Vaccination

The possibility of developing a vaccine against trypanosomosis has been frustrated by the ability of the parasite to vary their surface antigens and thus evade the host immune response (Donelson, 2002). A conventional vaccine that primes animals` immune system against only one or a few antigens will therefore not be broadly effective against trypanosomosis. Lubega *et al.*, (2002) showed that tubulin rich preparation from *T. brucei* confers broad protection against African trypanosomes. This can be through internalisation of the antibodies by the parasite by a mechanism not yet established (Balaban *et al.*, 1995) though flagellar pocket is implicated (Gull, 2002). However, the compound may not be immunogenic enough since it is not exposed on the surface of trypanosomes.

2.2.5 Trypanotolerant animals

The use of animals that are inherently tolerant to the effect of the disease and are able to remain relatively productive even when infected has been used as an approach to farming within the trypanosomosis belt. It is used as a strategy for productive farming since it allows keeping of animals in the infested areas. In West Africa, these tolerant breeds include N`Dama (*Bos taurus*), Baoule, Laguna, Samba, Dahomey and Muturu cattle that have been bred in areas endemic to trypanosomosis and showed considerable resistance (Weits, 1970). In East Africa, the Orma Boran and Maasai zebu have been shown to possess a degree of natural resistance to trypanosomosis (Mwangi *et al.*, 1998). The main drawbacks of this method are that many farmers believe these breeds are less productive than others and few of such animals are available. They also constitute a reservoir for infection to other susceptible livestock.

2.3 RNA splicing

2.3.1 Spliceosome and splice sites

Most mRNAs encoded by nuclear genes are synthesized as precursor RNA molecules, in which the coding sequences (exons) are interrupted by intervening sequences (introns). The mRNA becomes functional when introns are excised and exons precisely joined together by a process called splicing. This splicing reaction is achieved by a dynamic RNA-protein complex called spliceosome (Jurica and Moore, 2003). It consists of five small nuclear RNAs (snRNAs) and more than fifty attached proteins (Ito *et al.*, 1999; Nilsen, 2000). The snRNAs are U1, U2, U4, U5 and U6. They have proteins attached thus exist as small nuclear ribonucleoproteins (snRNPs).

Precise removal of introns (without disturbing the reading frame) requires specific sequences that must be involved in the process. Splicing signals have been identified by examination of the boundaries of introns and exons from a large number of genes in a variety of eukaryotic cells (Padgett *et al.*, 1986; Nilsen, 2000). The signal identified for higher eukaryotes is:

5' **AG|GUAAGU......... YNCURAC.........** $\mathbf{Y_n}$**NAG|G** 3'

where the vertical line (|) denotes the boundary between the exon and the intron, Y indicates a pyrimidine, R a purine and N means any base. Y_n means region of about nine pyrimidines (Py). The red 'A' is the branch point adenosine that is involved in the splicing reaction. These include the three conserved sequence elements identified as the 5' and 3' splice sites and a site within the intron near the 3' splice site known as the branch point sequence (BPS). Consensus sequences from plant introns show strong similarities to the animal consensus elements (Brown *et al.*, 2002). This identifies the sequence of a generic intron as:

Exon…**GU**……….Intron……….**AG**…Exon

2.3.2 Splicing mechanism

2.3.2.1 Cis-splicing

The spliceosome is a multienzyme complex that consists of five snRNPs (U1, U2, U4, U5 and U6), non-snRNPs (helicases, DExD/H box proteins, serine-arginine, SR proteins) and conserved pre-mRNA sequence elements (5' splice site, 3' splice site, Py and BPS) (Jurica and Moore, 2003). The spliceosome assembly proceeds through a series of short-lived intermediate stages (Figure 1) dubbed E/CC, A, B and C (Jurica and Moore, 2003). In summary, the process is initiated upon binding of U1 snRNP to the 5' splice site, the mammalian branch point binding protein or splicing factor 1(mBBP/SF1) to the BPS and U2 auxiliary factor (U2AF) to the Py and 3' AG. This forms the early or commitment complex (E/CC). Associated protein factors such as U2AF are needed to promote the complex formation when the sequence at the BPS varies from the consensus. Subsequently, U2 snRNP binds to the BP to form A complex (a.k.a pre-spliceosome complex) in which ATP is consumed. The addition of U4/U6•U5 tri-snRNPs to pre-spliceosome results into complex B. The C complex is formed from massive rearrangement in which U6 replaces U1 at the 5' splice site, U6 and U2 interact, U5 bridges the splice sites and U1 and U4 becomes destabilized. This rearranged spliceosome is catalytically active and carriers out two trans-esterification reactions.

During the ATP dependent transition of complex E to pre-splicing complex A, a short helix results between U2 snRNP and the branch site, defining the adenosine that functions as the nucleophile in the first catalytic step. The U2 snRNP and U6 snRNP in the C complex leads to bulging of the adenosine exposing its nucleophilic 2' hydroxyl (OH) group that attacks the 5' splice site guanosine bound to U5 snRNP. The reaction results into cleavage of pre-mRNA at the 5' splice site, generating an exon RNA fragment that contains a 3' hydroxyl group and an intron-exon fragment with the intron in the form of a lariat due to a 5'-2' bond (Padgett *et al.*, 1986). The two RNAs are held together non-covalently. A further rearrangement in complex C brings the 3'OH of exon fragment into close proximity to 5' phosphate of intron-exon fragment enabling the second trans-esterification reaction. This results into 3' splice site excision, ligation of the exons and release of the free intron. This chemical process is cis-splicing (Figure 2).

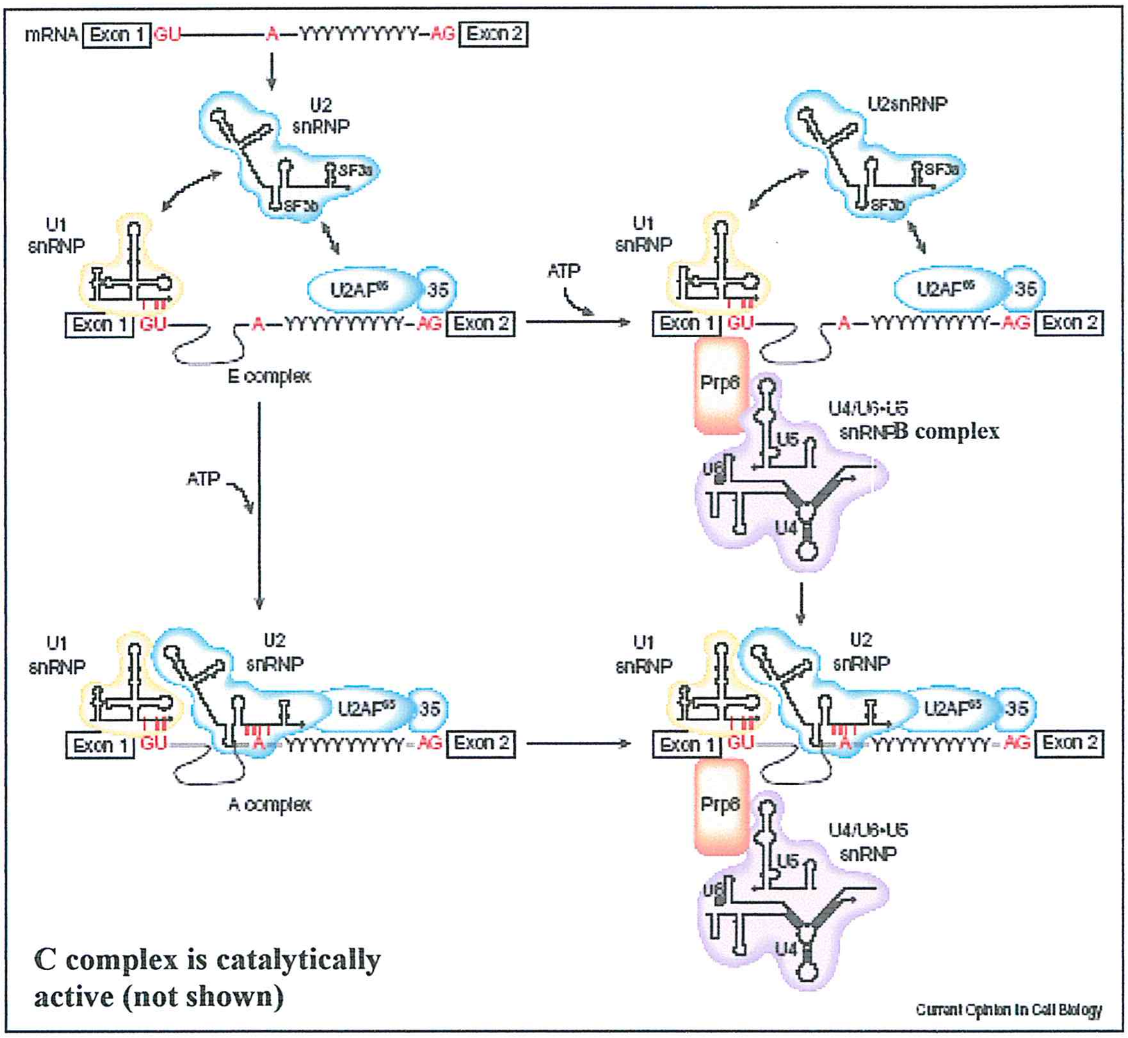


Figure 1. Schematic representation of intermediates of spliceosome assembly. In the E complex, U1 snRNP binds 5'ss (red GU), and U2 snRNP loosely binds to the pre-mRNA near the 3'ss (red AG) in an ATP dependent process (A complex). U4/U6•U5 tri-snRNP binds to the 5'ss, in part through interaction between Prp8 and pre-mRNA forming B complex. Red vertical lines show RNA base-pairing interactions, and filled circles and squares depict snRNA 5' cap structures. Modified from Hastings and Krainer (2001).

Self-splicing introns have shown that RNA can catalyse transesterification reaction by a mechanism in which metal ions that interact with specific atoms in the RNA activate the attacking nucleophile and stabilize the leaving group (Sontheimer *et al.*, 1997; Nilsen, 2000; Gordon *et al.*, 2000). This, a form of autocatalysis is known as self-splicing.

Alternative mRNA splicing is the term used to describe the regulated process of differential inclusion or exclusion of regions of the pre-mRNA. It generates multiple mRNA isoforms (Breitbart *et al.*, 1987; Gascard *et al.*, 1998) hence proteins with different functions from a single gene. Multiple transcripts from a single gene can result from exon skipping, mutual exclusion of exons and retention of introns and/or selection of an alternative 5' or 3' site (Smith and Valcárcel, 2000; Maniatis and Tasic, 2002; Roberts and Smith, 2002; Lee and Irizarry, 2003; Reddy, 2004). It has been documented in plants (Reddy, 2004), Drosophila, human and most vertebrates (Smith and Valcárcel, 2000).

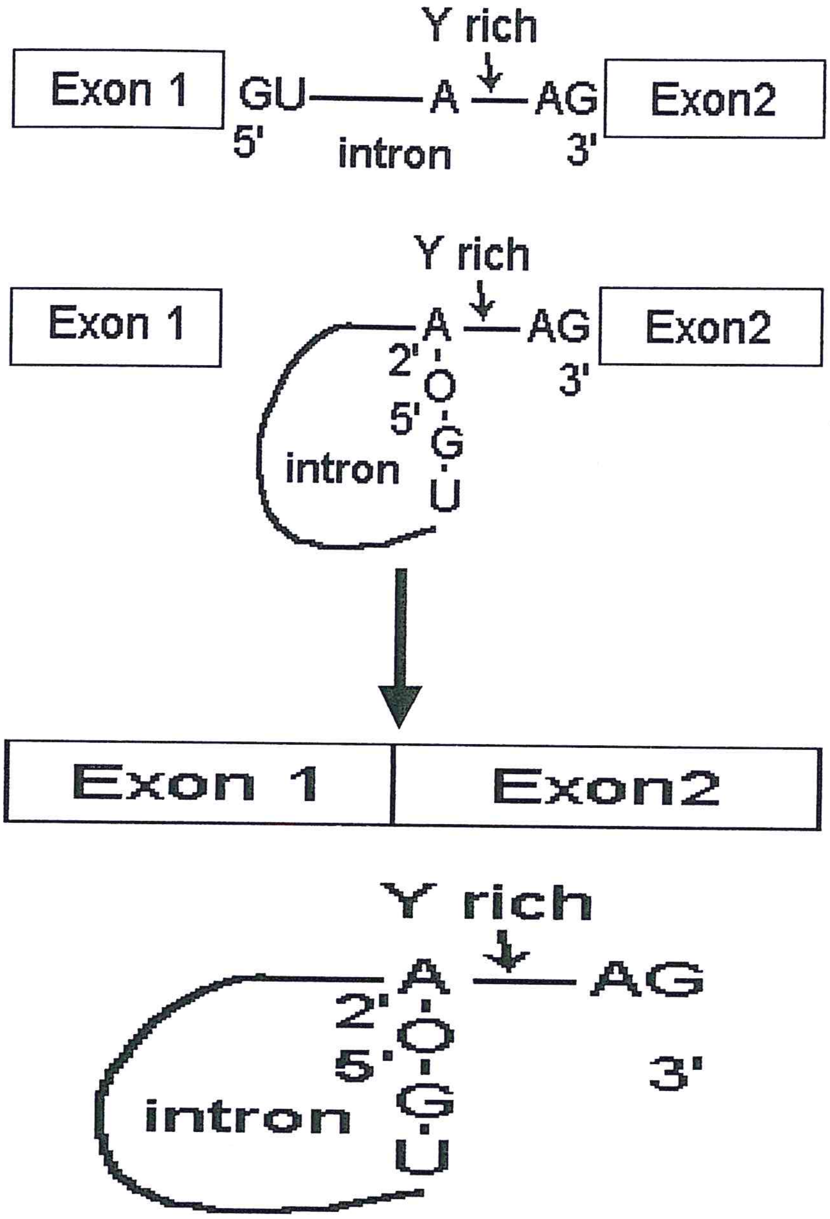


Figure 2. Two steps splicing (cis-splicing). Step one is the first trans-esterification reaction involving the 2' OH of the adenosine residue at the BPS and phosphate of the 5' splice site leading to lariat formation. Step two represents nucleophilic attack of 3'OH of the 5' splice site to the 5' phosphate of the second exon leading to exon ligation and intron excision.

2.3.2.2 Trans-splicing in trypanosomes

Trypanosomes belong to a group of eukaryotic organisms in which the splicing mechanism involves independent pre-mRNAs, a process known as trans-splicing or intermolecular exon-ligation (Figure 3). Surprisingly, cis-splicing has been reported in poly (A) polymerase genes (PAP) genes in *T. brucei* and *T. cruzi* (Mair *et al.*, 2000) against the almost two-decade-old tenets that trypanosomes exhibit only trans-splicing. Nonetheless, trans-splicing process is an essential step in the expression of all protein coding genes in trypanosomes that form polycistronic transcripts (Mandelboim *et al.*, 2002). It involves interaction between 5' and 3' splice sites on separate transcripts and occurs in a variety of eukaryotic organisms including trypanosomes (Sutton and Boothroyd, 1986), euglena, nematodes, trematodes and chordates (Lücke *et al.*, 1997; Mandelboim *et al.*, 2002). In this process, a Y branched intermediate is formed as opposed to a lariat in cis-splicing. This occurs by addition of a short non-coding miniexon sequence derived from the splice leader (SL) RNA onto each protein-coding exon sequence present within polycistronic precursor transcripts (Sutton and Boothroyd, 1986; Lücke *et al.*, 1997; Li *et al.*, 2000). The splicing complex that carries out this process is known as trans-spliceosome. The SL sequence is derived from a large transcript called the SL RNA (Li *et al.*, 2000; Landfear, 2003). The SL RNA is transcribed from arrays of tandemly repeated genes of 10-11 copies per haploid genome (Roberts *et al.*, 1996) and is present in the cell in the form of a SL ribonucleoprotein, the SL RNP (Goncharov *et al.*, 1999; Evans *et al.*, 2001).

The SL RNA of all organisms has two domains (Mandelboim *et al.*, 2002). The SL sequence or miniexon, which in trypanosomes is thirty-five nucleotides long (Sutton and Boothroyd, 1986) followed by the intron of variable length and can be folded into a conserved secondary structure with three stem-loops (Mandelboin *et al.*, 2002). The primary sequence of SL RNA is not conserved among the trypanomastids, but the secondary structure is consistent (Sturm and Campbell, 1999) and fulfils a dual function of recruiting splicing co-factors and serving as a substrate for splicing.

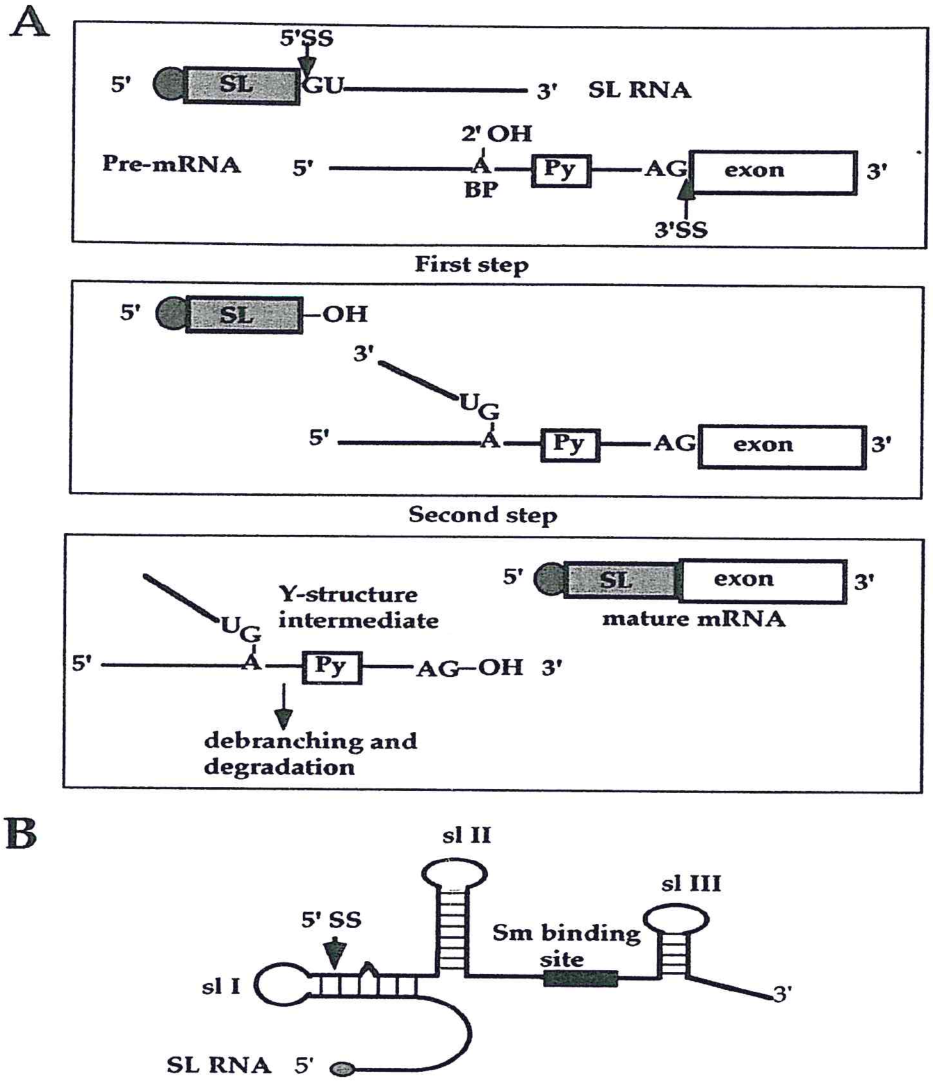


Figure 3. Mechanism of trans-splicing. (A) Schematic representation of trans-splicing. The 5' splice site GU on the SL RNA and the 3' splice site AG on the pre-mRNA are indicated. Y branched intermediate is shown. BP, branch point; Py, polypyrimidine tract. (B) Secondary structure of SL RNA. The three stem-loop structures (sl I, II, and III), the 5' splice site, and the Sm-binding site are indicated. The shadowed dot at the 5' end of SL RNA indicates the cap 4 structures. (Adapted from Liang *et al.*, 2003)

2.4 Linkage between trans-splicing and polyadenylation

Factors involved in capping, splicing and polyadenylation interact with the carboxy-terminal domain of RNA polymerase II at an early stage of mRNA production in cis-splicing. These processes are therefore coupled to transcription (Jurica and Moore, 2003; Liang *et al.*, 2003). This linkage involves specific sequence, such as the signal for polyadenylation (AAUAAA) in higher eukaryotes. Such sequences are absent in trypanomastid mRNAs (Liang *et al.*, 2003). In trypanosomes, evidence suggests that maturation of polycistronic pre-mRNAs involves a temporal and mechanistic relationship between trans-splicing and 3'-end formation (Ito *et al.*, 1999; Shepard *et al.*, 2002; Hendriks *et al.*, 2003; Liang *et al.*, 2003). It has been shown that the pyrimidine rich sequences immediately upstream of the α-tubulin trans-splice site are a major determinant for β-tubulin mRNA 3'-end formation in trypanosomes (Matthews *et al.*, 1994). The Py thus plays a role in the coupling of these processes since its maturation leads to aberrant poly (A) site choice (Hug *et al.*, 1994). This model is supported by the finding that the location of poly (A) site moves in concert with the 3' splice site AG in *Leishmania* (LeBowitz *et al.*, 1993), a close relative of trypanosome. In *Leishmania,* the factor(s) responsible for cleaving the pre mRNA at the poly (A) site bind only a certain short distance away from the 3' splice site. The factors required for 3'-end cleavage and polyadenylation therefore associate with the pre-mRNA after the machinery of the spliceosome has marked the 3' splice site region. The factors shared between 3'-end formation and trans-splicing couple the two processes. U1 snRNP is involved in coupling cis-splicing to polyadenylation. Similarly, an SL RNP specific protein may coordinate coupling of polyadenylation to trans-splicing in trypanosomes (Liang *et al.*, 2003).

2.5 Splicing, cleavage and polyadenylation factors

The spliceosome assembly proceeds through a series of short-lived intermediate stages dubbed E/CC, A, B and C (Jurica and Moore, 2003). The E complex consists of U1 snRNP base paired to the 5' splice site, U2 auxiliary factor heterodimer (U2AF) which binds to Py and 3' splice site, and SF1/mBBP. Associated protein factors such as U2AF are needed to promote the complex formation when the sequence at the BPS deviates from the consensus. The U2AF heterodimer consists of 65-Kda (U2AF^{65}) and 35-Kda (U2AF^{35}) subunits (Ito *et al.*, 1999; Shepard *et al.*, 2002) and is essential for viability in eukaryotes (Vàzquez *et al.*, 2003). U2AF^{65}

contains an N-terminal arginine-serine-rich (RS) domain and three RNA recognition motifs (RRM) in mammals (Shepard *et al.*, 2002). It interacts directly with the Py and is involved in stabilizing the interaction of U2 snRNP with the BP (Gozani *et al.*, 1998). This activity requires the RS domain, which is thought to assist in the U2 snRNP-pre-mRNA duplex and the third RRM that interacts with SF3b155, a component of U2 snRNP (Shepard *et al.*, 2002). SF1 and U2AF^{65} interact with each other facilitating cooperative recognition of the BP and Py tract that are adjacent to each other (Selenko *et al.*, 2003) (Figure 4). SF1 contains at least five distinct structural domains in which the amino-terminal region contacts the U2AF^{65} (Guth and Valcárcel, 2000).

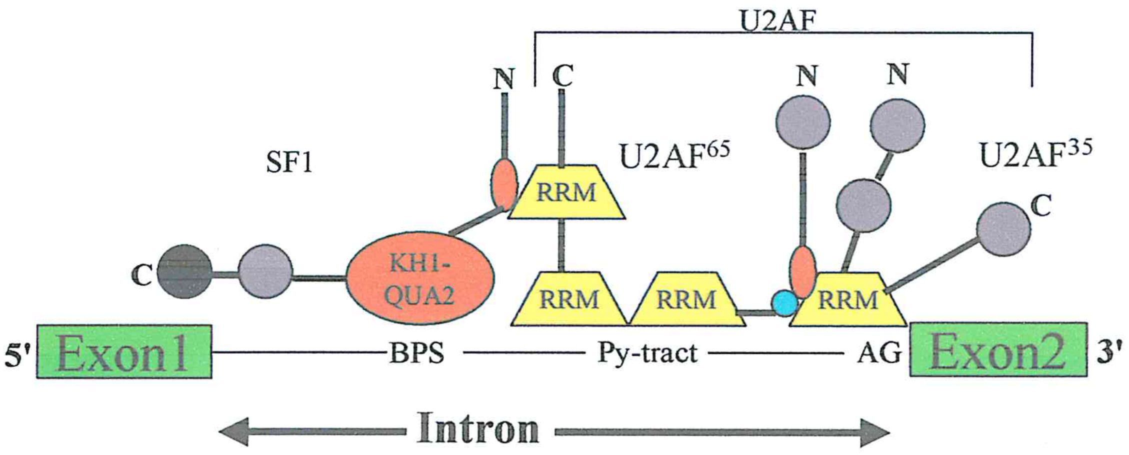


Figure 4. Recognition of the 3' splice site. SF1 and the U2AF heterodimer recognize the branchpoint sequence (BPS), polypyrimidine tract (Py), and 3' splice site (AG) by forming a network of protein-RNA and protein-protein interactions mediated by conserved KH and RRM domain (Selenko *et al.*, 2003).

The A complex also known as the pre-spliceosome is the second step in spliceosome assembly. It is formed when U2AF recruits U2 snRNP to the BP, which displaces SF1 in a process in which ATP is consumed. Two multisubunit splicing factors, SF3a and SF3b, are components of U2 snRNP and are required in the binding to the BP. SF3a consists of three subunits (SF3a 60, SF3a 66 and SF3a 90) and SF3b consists of eight subunits (P14, P14b, SF3b 10, SF3b 49, SF3b 125, SF3b 130, SF3b 145, and SF3b 155). They anchor U2 snRNP tightly to the pre-mRNA.

Addition of U2, U5 and U6 snRNPs to complex A as a preassembled tri-snRNP results in complex B. Subsequent formation of competent C complex involves the recruitment of additional protein factors along with significant structural rearrangement that destabilize the association of U1 and U4 snRNPs with the catalytic core (Jurica and Moore, 2003).

Vàzquez *et al.*, (2003) have characterized the *T. cruzi* U2AF^{35}. It is 240 residues long, with four RRMs, one central and three at the C-terminal, and a C-terminal SR segment. In eukaryotes, U2AF^{35} and U2AF^{65} form an intimate heterodimeric complex in which the RRM of U2AF^{35} and a central polyproline segment of U2AF^{65} interact via reciprocal "tongue in groove" tryptophan residues (Vàzquez *et al.*, 2003). Guth and Valcárcel (2000) documented various factors that interact with U2AF^{65} thus suggesting the importance of U2AF in viability in eukaryotes. It is also suspected to interact with poly (A) polymerase (PAP) thus linking trans-splicing to polyadenylation. Poly (A) polymerase (PAP) is the enzyme that adds the poly adenosine (A) tail after the cleavage reaction. Palfi *et al.*, (2000) characterized seven proteins associated with snRNP core complex of *T. brucei.* The seven proteins were found to bind to conserved sequences in U1, U2, U4 and U5 snRNP resembling the Sm sequence in cis-spliceosomal snRNPs.

Members of the serine-arginine (SR) proteins have multiple functions in pre-mRNA splicing reaction. These include the removal of constitutively spliced introns and regulating alternative splicing, which they do both *in vivo* and *in vitro* (Graveley, 2000; Furuyama and Bruzik, 2002). They bridge between other splicing components where interactions are too weak for effective binding of the basal factors U1 and U2 as well as recruitment of other splicing factors. All SR proteins have a modular organization and contain an N-terminal RNA-binding domain that interacts with the pre-mRNA and a C-terminal RS domain that functions as a protein interaction domain (Philipps *et al.*, 2003). Ismaili *et al.*, (2000) characterized *T. brucei* SR domain-containing protein (TSR1IP) bearing homology to cis-spliceosomal U1-70 kDa proteins. They showed that it interacts with the 5' splice region of the SL RNA. However, its identity as a 70-kDa homologue is currently puzzling, since an U1 specific protein whose homology to the U1-70 kDa domain was recently identified (Palfi *et al.*, 2002).

Cleavage and polyadenylation specificity factor (CPSF) is required for both cleavage and polyadenylation reactions. It contains subunits of 160, 100, 70 and 30 kDa referred to as CPSF 160, CPSF 100, CPSF 70 and CPSF 30 respectively. CPSF 30 contains five CCHC zinc knuckles (Hendriks *et al.*, 2003). Both types of motifs have been implicated in binding nucleic acids. Its role is probably to cooperate with CPSF 160 in recognition of pre-RNA substrates and through interaction with poly (A) binding protein II (PAB II) to stabilize the polyadenylation complex (Zhao *et al.*, 1999). Hendriks *et al.* (2003) showed that CPSF 30 depletion by RNA interference is lethal and leads to disruption of RNA processing in *T. brucei*. The 70 kDa and 100 kDa subunits are closely related, with 23% identity and 49% similarity and their function is unknown (Zhao *et al.*, 1999).

Cleavage stimulating factor (CstF) binds downstream of the cleavage site at a GU-rich sequence. It associates with CPSF and is necessary for cleavage but not poly (A) addition. It contains 77, 64 and 50 kDa subunits with the CstF 77 being the central subunit bridging the other two and they are arranged in a linear fashion (Zhao *et al.*, 1999). CstF 64 contains a classical RNA-binding domain (RBD) close to its amino terminus (Evans *et al.*, 2001) while the 50 kDa subunit has a protein interaction domain and binds to RNA Pol II CTD.

Cleavage factors I and II (CF I, CF II) are responsible for actual cleavage reaction. Three polypeptides of 25, 59 and 68 kDa and a possible fourth one of 72 kDa copurify with CF I activity in mammals (Zhao *et al.*, 1999). It increases the CPSF-RNA complex stability, suggesting that this factor interacts with CPSF and contributes to the overall stability of 3'-end-processing complex. CF I contains RS domains similar to splicing factors. CF II has not been purified to homogeneity and its function not known (Zhao *et al.*, 1999).

RNA polymerase II, through its conserved carboxyl-terminal domain (CTD) of its largest subunit, is involved in the cleavage reaction through interaction with CPSF and CstF. Its N-terminal contains a catalytic domain with homology to nucleotidyltransferases (Zhao *et al.*, 1999). CPSF and PAP suffice for poly (A) addition to a pre-cleaved RNA substrate. However, rapid elongation and control of poly (A) tail length requires an additional factor PAB II (Zhao *et al.*, 1999). The poly (A) binding protein in *T. brucei* has been identified and purified by Pitula *et al.* (1998).

More has been done on the understanding of SL RNA but little on trans-splicing and polyadenylation factors in trypanosomes. Few splicing factors have been characterized across the kinetoplastids. This has led to little information regarding the dynamic nature of the spliceosomal complex. Since the information available is highly disjointed, the exact process of spliceosome assembly is not available. Characterization and functional analysis of the factors involved is thus important. This will provide information on important factors and interactions that are essential for RNA maturation, hence identifying potential drug targets.

2.6 RNA interference (RNAi)

RNA interference (RNAi) is a phenomenon leading to post-transcriptional gene silencing (PTGS) after endogenous production or artificial introduction in a cell double-stranded RNAs (dsRNAs) that are complimentary to known mRNAs, specifically destroying that particular mRNA, thereby diminishing or abolishing gene expression. RNAi has been linked to many previously described silencing phenomena such as PTGS in plants and quelling fungi (Dykxhoorn *et al.*, 2003; Pauls and Esté 2004). It is an evolutionarily conserved mechanism of gene silencing that is thought to inhibit the replication and expression of selfish DNA elements and viruses (Caplen *et al.*, 2002; Chi *et al.*, 2003) hence maintains genome integrity. Whereas the transcription of the gene is normal in this process, the translation of the protein is prevented by selective degradation of its encoded mRNA. Although its protective role would be very important in plants and *C. elegans* (Milhavet *et al.*, 2003) which lack an antibody-based immune system analogous to that found in animals, its role in mammalian cells remains unclear. It thus must have evolved as a cellular defence mechanism against foreign DNA and RNA regulating the expression of endogenous genes (Milhavet *et al.*, 2003). This phenomenon has been used to study gene functions *in vitro* by injection of dsRNA that led to an efficient sequence-specific gene silencing in *C. elegans* (Fire *et al.*, 1998).

2.6.1 RNAi mechanism

Double-stranded RNA-mediated interference (RNAi) is a PTGS process in which double-stranded RNA (dsRNA) triggers degradation of homologous mRNA in the cytoplasm of a cell. Studies over the last few years have demonstrated that RNAi is mediated by the generation of 21- to 23-nucleotide dsRNA molecules, termed small interfering RNA (siRNA) (Dillin, 2003; Dykxhoorn *et al.*, 2003). They have a characteristic structure, with 5'-phosphate/3'-hydroxyl ends and a 2-base 3' overhang on each strand of the duplex. Upon entry into the cell, dsRNA is cleaved into siRNAs by the action of RNase-III-type enzyme Dicer (Jones *et al.*, 2004; Pauls and Esté, 2004). These siRNAs are incorporated into a protein-RNA complex, the RNA-induced silencing complex (RISC) (Dykxhoorn *et al.*, 2003). There is a strict requirement for the 3' siRNA to be 5' phosphorylated to enter into RISC, and siRNA that lacks a 5' phosphate are rapidly phosphorylated by an endogenous kinase (Dykxhoorn *et al.*, 2003). The duplex siRNA is unwound, leaving the antisense strand to guide RISC to the homologous target mRNA for endonucleolytic cleavage (Dykxhoorn *et al.*, 2003; Pauls and Esté, 2004). The target mRNA is cleaved at a single site in the centre of the duplex region between the guide siRNA and the target mRNA, 10-nt from the 5' end of the siRNA. Mismatches greater than 1-2bp within the 21- to 23-nt siRNA effectively disrupts proper degradation of the target mRNA (Dillin, 2003). The synthesis of the protein encoded by the mRNA targeted by the siRNAs is thus prevented and that protein is selectively depleted from the cell thus RNA interference. The RNAi process is shown in Figure 5.

In worms, interaction between the siRNA and mRNA can lead to immediate cleavage by Dicer, liberating a new siRNA, and the degradation of mRNA by endo- and exonucleases (Dillin, 2003). Alternatively, the siRNA can serve as a primer for an RNA-dependent RNA polymerase (RdRP), creating many more siRNAs. The action of RdRP may explain the catalytic mechanism of RNAi, because only a few dsRNA molecules are required to degrade a much larger population of mRNAs.

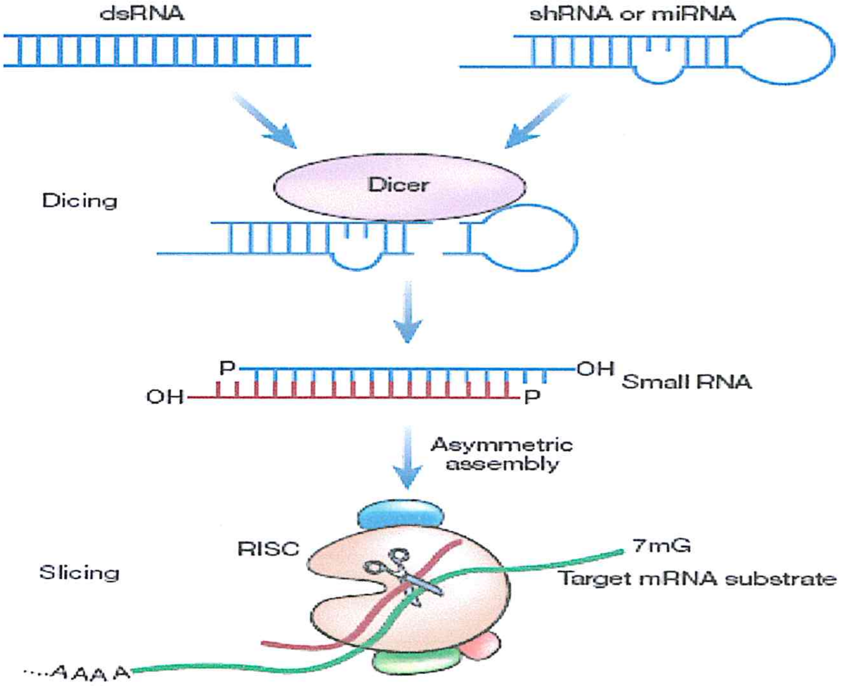


Figure 5. Overview of dsRNA-mediated mRNA degradation. Double stranded RNA is cleaved by dicer into 21- to 23-nt siRNAs, which is complexed with a large multiprotein complex, the RISC. RISC unwinds the siRNA to help target the appropriate mRNA (shown in green). The siRNA-mRNA hybrid is cleaved, releasing the siRNA, and the mRNA is degraded by endo and exonucleases.

Recent studies have identified the existence of endogenous siRNA molecules called microRNA (miRNA) (Jones *et al.*, 2004). These are transcribed in the nucleus as a large non-coding primary sequence that are processed by the RNase III enzyme Drosha to produce a hairpin RNA of ~70 nucleotides before being exported into the cytoplasm by exportin-5 (Jones *et al.*, 2004). Like siRNAi, pre-miRNA is then cleaved by the Dicer to produce a double stranded 21- to 23- nucleotide RNA duplex, which is incorporated into the RISC-like complex. However, in contrast to siRNA-mediated mRNA cleavage, miRNA is thought to block mRNA translation through imperfect complementary binding of the antisense sequence to the miRNA recognition elements (MREs) within the 3` untranslated region (UTR) (Jones *et al.*, 2004).

2.6.2 RNAi as a genetic tool

The discovery of the RNAi machinery provides a powerful tool for the study of gene function. Small fragments of double-stranded RNA (siRNA) can be introduced into cells to selectively degrade homologous endogenous mRNA enabling knock down of complementary genes. This technique has been used to investigate gene function in various organisms including *C. elegans* (Fire *et al.*, 1998), plants (Baulcombe, 1999), planaria (Sánchez-Alvarado and Mewmark, 1999), trypanosomes (Ngô *et al.*, 1998; Wirtz *et al.*, 1999; Shi *et al.*, 2000; LaCount *et al.*, 2002; Hendriks *et al.*, 2003; Alibu *et al.*, 2004), hydras (Lohmann *et al.*, 1999), *Drosophila* (Kinnerdell and Carthew, 1998; Misquitta *et al.*, 1999), mosquitoes (Caplen *et al.*, 2002) and mouse oocytes (Svoboda *et al.*, 2000).

2.6.3 RNAi and trypanosome genome studies

Characterisation of putative coding regions is a prerequisite for converting raw genomic sequence data into biologically relevant information (Gopal *et al.*, 2003). With the completion of trypanosome and leishmania genome projects (Berriman *et al.*, 2005; El-Sayed *et al*, 2005a; Ivens *et al.*, 2005), a wealth of information is available at the repository genome database at GeneDB. Currently, GeneDB provides access to 32 genomes of various stages of sequencing curation pipeline as well as complete genomes with extensive manual curation. This includes sequences and associated annotation of bacteria, fungi, protozoa and arthropod (Arnaud Kerhornou and The Sanger Institute Pathogen Sequencing Unit, 2002; Hertz-Fowler *et al.*, 2004). The genome information will provide insight on parasitism, virulence, epidemiology (Gull, 2000) and parasite biology among other aspects. To exploit this information and the rapid accumulation of genome sequence in databases, techniques such as microarrays and RNAi (Donelson, 2002) that allow gene function studies are important.

The technique of RNA interference is useful for knocking down the expression of a specific mRNA in African trypanosomes and other organisms for the purpose of examining the function of its gene (DaRocha *et al.*, 2004). The technique circumvents the necessity of generating double knockouts, which is a requirement for such studies in diploid organisms. It works well in *T. brucei* (Ngô *et al.*, 1998) due to the presence of two predicted proteins representing potential Dicer candidates (El-Sayed *et al.*, 2005b). It can therefore be used to examine the genes or a set of genes essential for trypanosome processes. Since this finding, more has been done on utilization of this technique in trypanosome gene functions studies (Wang *et al.*, 2000; Morris *et al.*, 2001; Hendriks *et al.*, 2003), designing siRNAs (Redmond *et al.*, 2003), development of cell lines (Chen *et al.*, 2002), construction of plasmid vectors for RNAi delivery (Wirtz *et al.*, 1999; Wang *et al.*, 2000; Gull *et al.*, 2002; Alibu *et al.*, 2004; DaRocha *et al.*, 2004; Motyka *et al.*, 2004), *in vivo* analysis (Tschudi *et al.*, 2003) and transfection methods. Redmond *et al.*, 2003 have developed a web-based tool for designing RNAi targets in *T. brucei*. This *T.brucei* functional genomics project (TrypanoFAN) aims to utilise the information from the *T.brucei* genome project to produce a research resource and systematic collection of mutants by targeted gene inactivation using RNAi. These developments are important in the utilization of genome sequence for valuable functional assays.

2.6.4 RNAi vectors for trypanosome gene research

The simplest method to achieve RNA interference in *T. brucei* is to use vectors with opposing bacteriophage T7 promoters, controlled by binding of the *tet* repressor to *tet* operators (Wang *et al.*, 2000; Morris *et al.*, 2001; LaCount *et al.*, 2002). The inducible RNAi vectors pZJM and p2T7^{Ti} are widely used for they allow for insertion of a sequence of interest between opposing T7 promoters under the control of tetracycline operator (Motyka *et al.*, 2004). This allows the transcription of the insert in both directions upon induction with tetracycline. Linearized vector allows integration by homologous recombination into ribosomal DNA (rDNA) spacer region, a transcriptionally inactive segment of the *T. brucei* genome (Morris *et al.*, 2001; LaCount *et al.*, 2002; Alibu *et al.*, 2004; Motyka *et al.*, 2004).

2.6.4.1 p2T7Ta blue vector

p2T7TA blue is a 6740bp vector derived from p2T7^{Ti}B/GFP (LaCount *et al.*, 2002) and is shown in Figure 6. It has two promoters for bacteriophage T7 polymerase (shown as the big arrows facing each other). Each promoter has two binding sites for the bacterial *tet* repressor. The vector also has a hygromycin resistance gene, transcription of which is driven by a separate promoter. This gene allows for the selection of recombinants. Parasites express the T7 polymerase and the repressor so that transcription of the piece of DNA between the T7 promoters only happens when tetracycline is added to the medium.

Digestion with *Eam*1105I generates 3' overhanging T-ends, which allow PCR products with 3' A-ends to be cloned directly. Once inserted, opposing T7 promoters with downstream operators flank RNAi targets. Recombinant vectors are linearised at the *Not*I site and targeted to the rDNA locus (Wirtz *et al.*, 1999) in cells that express T7 RNA polymerase, *tet*-repressor and the selectable marker. Recombinants are selected with hygromycin B. Upon integration, the construct lies between rDNA spacer sequences. The operators bind *tet*-repressor in the absence of tetracycline so dsRNA is only synthesised when tetracycline is added to the growth medium. T7 terminators prevent transcription from extending into rDNA.

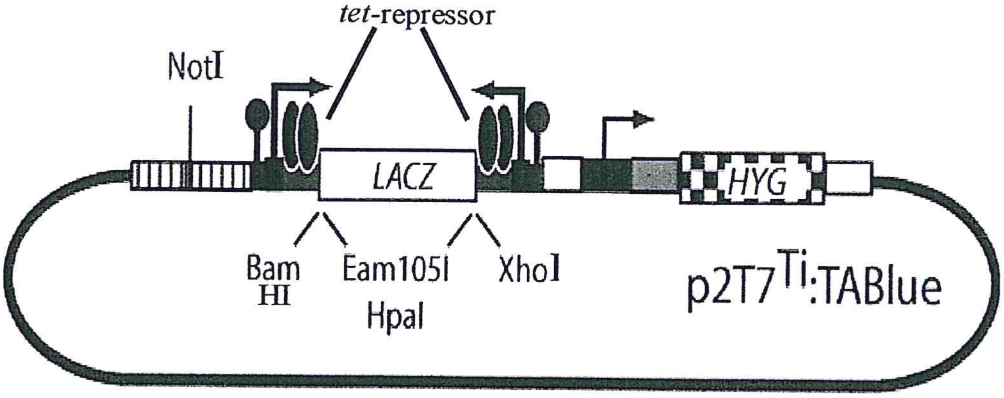


Figure 6. p2T7Ta blue vector. Two promoters for T7 polymerase are shown as big arrows facing each other, *Eam*11051 cloning site and hygromycin resistance gene for selection of transformants.

2.6.4.2 Stem-loop vectors

Stem-loop vectors such as pLEW79, pLEW100, pHD1146 and pHD1336 (Wirtz *et al.*, 1999; Estévez *et al.*, 2003) are also used (Appendix 3). They involve cloning of the insert twice in opposite direction on either sides of a 'stuffer' fragment. A representative map is shown in Figure 7. Within the transfected cells, a stem loop of complementary RNA is formed that can be `Diced` as shown in Figure 8.

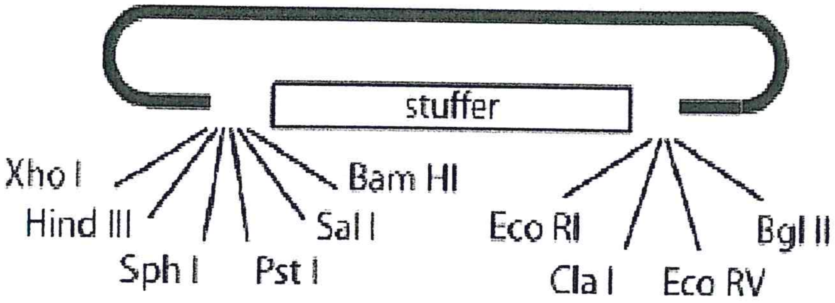


Figure 7. Vector for stem-loop cloning. The two multiple cloning sites for cloning of the insert are in opposite direction.

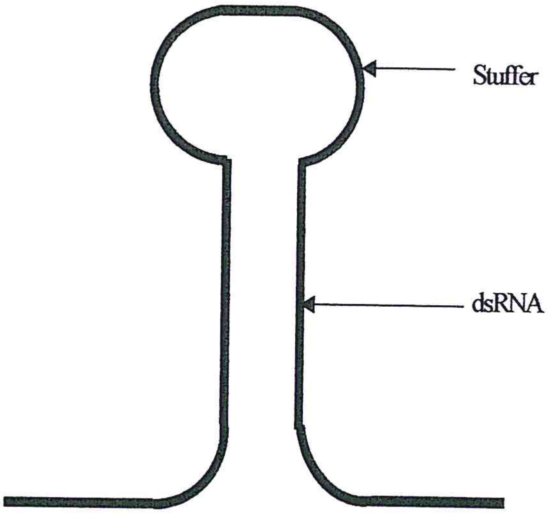


Figure 8. Double strand RNA (dsRNA) stem-loop. The dsRNA is formed in transfected cells. The siRNAs are formed from `diced` dsRNA.

These vectors allow the cloning of the insert, which then undergoes homologous recombination in rDNA loci when linearized with *Not*I. The transcription occurs on induction with tetracycline, hence producing mRNA homologous to target the gene from which siRNA is produced.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Data mining for *T. brucei* trans-splicing and polyadenylation homologs

In order to identify homologs of *T. brucei* genes for trans-splicing and polyadenylation, basic local alignment search tool (BLAST) searches (Altschul *et al.*, 1990; Altschul *et al.*, 1997) were done in GeneDB (Hertz-Fowler *et al.*, 2004), the repository of genome data for *T. brucei, T. cruzi* and *Leishmania major*. *T. cruzi*, human, and/or yeast splicing and polyadenylation factors were used to query the database. The accession numbers of nucleotide and protein sequences used in the search and *T. brucei* homologs generated are shown in Appendix 1. The search results were used to design primers, which were subsequently used to recover the genes from *T. brucei* by PCR.

3.2 Trypanosome strain

The trypanosome strain used in this study was KETRI 3741 (MHOM/UG/72/KETRI 3741). It is a *T. b. rhodesiense* derived from a clone, KETRI 3666. These were grown in mice. KETRI 3666 is a derivative of KETRI 2537 that was isolated from human in 1972 in Busoga, Uganda. It was isolated by inoculation of whole blood from patient into monkey.

3.3 DNA preparation

KETRI 3741 *T. b. rhodesiense* strain was grown in mice and purified using anion exchange chromatography (Lanham and Godfrey, 1970). DNA extraction was conducted according to the methods of Van der Ploeg *et al.*, (1982) with few modifications. Briefly, the column-purified trypanosomes were harvested by centrifugation (3000 rpm, 10 min, 4°C) in a refrigerated microcentrifuge (Heraeus fresco Biofuge, Kendro Laboratory Products GmbH, Germany) and washed once with phosphate buffered saline (PBS) (Appendix 2). The cells were resuspended in 200µl of cell lysis solution (Appendix 2), transferred into 1.5 ml microcentrifuge tube and incubated at 37°C overnight. The lysate was phenol extracted by adding an equal volume of phenol/chloroform/isoamylalcohol (25/25/1, v/v/v) for 5 minutes (Sambrook *et al.*, 1989). This was centrifuged (5000 rpm, 3 min, 25°C) in the refrigerated microcentrifuge and aqueous upper layer pipetted out leaving the interphase behind. The interphase was re-extracted and upper phases pooled. DNA was precipitated by adding 0.1 volume of 3M sodium acetate (pH 5.2) and 2 volumes of absolute ethanol. This was kept at -20°C overnight. The DNA was then pelleted (13,000 rpm, 10 min, 4°C) in the refrigerated microcentrifuge, washed with 1ml of 70% ethanol, air-dried and redissolved in 200µl of sterile water.

3.4 PCR amplification of genomic DNA

Individual amplifications were carried out in 25µl reaction volumes. The reaction mixture contained as final concentrations, 10mM Tris-HCl, pH 8.3, 1.5mM $MgCl_2$, 200µM each of the four deoxynucleoside triphosphates (dNTPs), 100ng of each primer, 1µl DNA template, 1 unit of *Taq* polymerase (MBI Fermentas, Lithuania) and 15.3µl of nuclease free water. The reaction mixture was placed in a thermocycler (PTC-100^{TM}, MJ Research, Inc. Watertown MA.) and incubated at 94°C for 1 min, the 35 cycles of 94°C for 30 sec (denaturation), 55°C for 30 sec (annealing), 72°C for 90 sec (extension) and a final extension at 72°C for 10 minutes. 10µl of the amplification product was electrophoresed through a 1% ethidium bromide stained agarose gel and viewed under ultraviolet (UV) illumination. The primers and amplified homologs are shown in Table 3.

Table 3. Primers used for genomic DNA amplification.

SPLICING FACTORS	PRIMER SEQUENCES (5'-3' direction)
P14	FOR: GCGAATGTCCGATGAGCC
	REV: ATCCTCGGCCTTGACCT
CPSF 30	FOR: GCATGTTTACTGACAACGCTGCC
	REV: CTGCCTTCCCGTTGCATCACCACG
AF 35	FOR: GCATGTATCARGAYCGYTGCA
	REV: TTTAAGGGGCATTCGCGAGAGATG
AF65	FOR: ATGGGGCGTGATAGTCGCGGACAC
	REV: GCTACTCCACAAAAACACGGGCT
CFII a1	FOR: GCATGTCTTCTAATTGTCGCAG
	REV: GCTGGGGCTTGTAAAGGTGG
U1 70k	FOR: GCATGGAGGCGTCCCACCGAGT
	REV: CGAGCATCTTCCCTTCTG
CST 50	FOR: GCATGTCTGGAGAAAACTTG
	REV: CGCCGTGGTCCATGCCGG
Zn 1	FOR: GCATGTGAGAGAGGGGAGGGA C
	REV: CAAGGAAAGAAACATATGCAG
Zn 2	FOR: GTCTGCCCGCAGCTTTCGCGC
	REV: CAACGAAAGAAACATATGCAG
SF3b145	FOR: GCTGCGGCCGCATGTATATACATCTTTTCTCCTAT
	REV: CCAGCGGCCGCCTATCAAAATTTTGTTGGTGCACG
SF3b125	FOR: CCTGCGGCCGCATGGAGGAGACGTACAGTCCCTT
	REV: GCAGCGGCCGCCTATCTCGCGTATTTCTTGCGGG
	FOR: CCTGCGGCCGCATGACAATCGCTGCACAGGGGG
Sf3b49	REV: GCAGCGGCCGCCTATTAAGCGCTTGAGGCGTGT
SF3b10	FOR: CCTGCGGCCGCATGGATGTGCCTGGTGAGCTTCTC
	REV: GCAGCGGCCGCCTATCAAACACACTCACTGGTCAT

3.5 Purification of amplified products.

The full-length amplification products were purified directly using QIAquick Gel Extraction Kit (Qiagen, GmbH Germany). Briefly, the DNA fragments were excised from the agarose gel using a clean sharp scalpel and weighed. Three volumes of yellow core buffer (QG) were added to 1 volume of the gel and incubated at 50°C until the gel was completely dissolved. One volume of isopropanol was then added to the sample and mixed. Sample was then applied to the QIAquick spin column and centrifuged at 13000 rpm for 1 minute. The flow through was discarded and QIAquick column placed back in the same collection tube. 0.5 ml of buffer QG was added to QIAquick column and centrifuged at 13000 rpm for 1 min. To wash, 0.75 ml of buffer PE was added to the QIAquick column, allowed to stand for 5 min and centrifuged at 1300 rpm for 1 min. The flow through was discarded and the QlAquick column centrifuged for an additional 1 min at 13000 rpm. The QlAquick column was then placed in a clean 1.5 ml centrifuge tube and 30µl of nuclease free water added to the center of the QIAquick membrane. Column was allowed to stand for 1 min and centrifuged at 13000 rpm for 1 min to elute DNA. The centrifugations were done in a table-top microcentrifuge (eppedorf centrifuge 5415 C, GmbH & Co. Bremen, Germany).

3.6 Cloning of PCR products

Purified PCR products were cloned in pGEM-T Easy vector (pGEM-T® EASY vector Systems kit, Promega Corp., Madison, WI, U.S.A; map shown in Appendix 3) using a shotgun cloning strategy. Ligations were performed in a total volume of 10µl. Purified PCR product (3µl) was mixed with 50ng (1µl) pGEM-T Easy vector, 5µl 2x rapid ligation buffer and 1 µl of T4 DNA ligase. The reactions were mixed by pipetting and incubated at 4°C overnight. Control reactions containing pGEM-T Easy vector alone were also carried out. Transformation of DH5α high efficiency competent cells was carried out using the ligation mixture. 50µl of cells were mixed with the ligation reaction and incubated on ice for 20 min. Uptake of the DNA was facilitated by 1 min heat shock at 42°C, followed by 5 min on ice. 950µl of room temperature SOC medium (Appendix 2) was added and incubated for 3 hrs at 37°C. 50µl of IPTG (Isopropyl-β-D-thiogalactopyranoside; 100mM) and 40µl of X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside; 50mg/ml) was spread on *Luria-Bertani* (LB)-ampicillin (125µg/ml) agar plates (Appendix 2). The transformed cells were pelleted by centrifugation at 3000 rpm for 10 min, resuspended in 200µl of SOC medium and 100µl plated on each of two plates. Plates were incubated overnight at 37°C. Recombinant clones were identified by blue: white screening of the colonies. Potential positives were screened by PCR using the primers used for genomic DNA amplification to check for insertion of the correct template sequence. A single colony of the positive clones was inoculated in 5ml LB medium (Appendix 2) containing ampicillin (125 µg/ml) in a falcon tube and incubated overnight at 37°C with shaking.

3.7 Purification of plasmid DNA

Plasmid DNA was purified using QlAprep® Spin Miniprep kit (Qiagen, GmbH Germany) according to manufacturers protocol. Briefly, overnight cultures were pelleted by centrifugation at 4000 rpm for 15 min on a refrigerated centrifuge (GEMCO refrigerated centrifuge, Sigma-Aldrich, St. Louis, MO). Pelleted bacterial cells were resuspended in 250µl buffer P1 containing RNase A, mixed and transferred into a 1.5ml microcentrifuge tube. Following this, 250µl of lysis buffer (P2) was added to the suspension and mixed gently by inverting the tube 4-6 times. The lysate was neutralized by addition of 350µl buffer N3 and mixed immediately by inversion until a cloudy precipitate was obtained. This was centrifuged for 10 min at 13000 rpm. The supernatant was applied to the QlAprep column by pippeting and centrifuged for 1 min at 13000 rpm. The flow through was discarded. The QlAprep spin column was washed by adding 0.5ml buffer PB, recentrifuged for 1 min and flow through discarded. Bound plasmid DNA was washed by adding 750µl of buffer PE and centrifuged for 1 min. The flow through was discarded and re-centrifuged for an additional 1 min to remove residual wash buffer. The QlAprep column was placed in a clean 1.5ml microcentrifuge tube and DNA eluted with 30µl of nuclease free water. DNA was stored at –20°C. The centrifugations were done in a table-top microcentrifuge (eppedorf centrifuge 5415 C, GmbH & Co. Bremen, Germany).

3.8 Sequencing

Automated sequencing was undertaken for the purified plasmid with PCR inserts using dideoxy chain termination method (Sanger *et al.*, 1977) in an ABI 3100 sequencer (Applied Biosystems, Foster City, CA, USA) using appropriate fluorescent labelled terminators.

3.9 Expression and purification of recombinant protein

3.9.1 Determination of open reading frames (ORFs)

The nucleotide sequences generated from the cloned products were used to query the *T. brucei* database GeneDB at http://www.genedb.org to evaluate similarity and therefore infer homology with sequences in the database. The sequences were translated using translation tool at Expert Protein Analysis systems (Expasy -http://us.expasy.org) to determine the ORFs. ORF specific primers containing restriction sites for *Not*I were designed to allow their restriction ligation into pET-28a expression vector (Appendix 3). All the factors except Zn2, SF3b10, SF3b49, SF3b125 and SF3b145 were ligated into pET-28a vector.

3.9.2 Amplification and purification of cloned inserts

Amplifications were carried out in 25µl reaction volumes. The reaction mixture contained as final concentration 10mM Tris-HCl, pH 8.3, 1.5mM $MgCl_2$, 200µM each of the four dNTPs, 100ng of each primer, 0.1µl template, 1 unit of Taq polymerase (MBI Fermentas, Lithuania) and 16.3 µl of nuclease free water. The reaction mixture was placed in a thermocycler (PTC-100^{TM}, MJ Research, Inc. Watertown MA.) and incubated at 94°C for 2 min, then 35 cycles of 94°C for 30 sec (denaturation), 55°C for 30 sec (annealing) and 72°C for 120 sec (extension) and a final extension at 72°C for 10 minutes. 20µl of the amplification product were electrophoresed through a 1% ethidium bromide stained agarose gel and viewed under ultraviolet (UV) illumination. The amplification products were purified using QIAquick Gel Extraction Kit (Qiagen, GmbH Germany) as above.

3.9.3 Restriction digestion and ligation into pET-28a

A restriction digestion reaction in a volume of 20µl was prepared, with the reaction mixture containing 8.0µl of purified amplification product, 2.0µl of 10X buffer, 20 units of *Not*I (MBI Fermentas, Lithuania) and 9.8 µl of nuclease free water. The reaction mixture was incubated for 4 hrs at 37°C and was thereafter gel purified using QIAquick Gel Extraction Kit (Qiagen, GmbH Germany). A similar restriction digestion and purification reactions were carried out for 4.0µl of pET-28a vector using *Not*I (MBI Fermentas, Lithuania).

Ligation was conducted using Novagen pET-28 system (Biosciences, Inc, Darmstadt, Germany) according to manufacturers protocol. This vector allows the expression of a fusion protein tagged to six histidine amino acids (the His-tag) allowing easy purification on nickel columns. In brief, a reaction containing 4µl of *Not*I linearized pET-28a vector, 8µl of *Not*I digested insert, 2µl of 10X buffer and 6 units of T4 DNA ligase (MBI Fermentas, Lithuania) was set up and incubated overnight at 16°C. 5µl of the reaction mixture was used to transform DH5α high efficiency competent cells, plated on LB-kanamycin (50µg/ml) plates and incubated overnight at 37°C. Colonies were PCR screened using insert specific primers. Positive colonies were grown on LB-kanamycin liquid medium (50µg/ml) with overnight shaking and plasmid purified using QIAprep Spin Miniprep kit (Qiagen, GmbH Germany) as above.

3.9.4 Expression and induction

Transformation of BL21 bacterial cells was carried out using 1µl of purified recombinant plasmid as described before (section 3.6). This was platted on LB-kanamycin plates (50µg/ml) and incubated overnight at 37°C. A single positive colony was used to inoculate 10ml of LB-kanamycin medium and incubated at 37°C with shaking. 1ml of the culture was used to inoculate 100ml of LB-kanamycin medium and incubated at 37°C with shaking at 200 rpm. At OD_{600} of 0.5 (Beckman DU 6480B spectrophotometer, Beckman Coulter, Inc. Fullerton, CA.), the culture was split into two equal parts. To one part, IPTG was added to a final concentration of 1mM (induced) and the other served as an uninduced control. Both were further incubated for 4 hrs. The bacterial cells were harvested by centrifugation at 4000 rpm for 15 min (GEMCO refrigerated centrifuge, Sigma).

3.9.5 Extraction of total cell proteins

Harvested cells were resuspended in 2ml of 1X PBS buffer. 200µl of 4X SDS sample buffer (Appendix 2) was added and cells sonicated for 10 sec at 5-µm amplitude for 10 cycles using Soniprep 150 ultrasonic disintegrator (Sanoy, Integrated Services, TCP Inc.). Immediately the samples were heated at 85°C for 3 min. The prepared total cell protein was separated on 4 – 20% SDS PAGE gels.

3.9.6 Purification of recombinant proteins

The recombinant P14 and CPSF30 proteins expressed in BL21 bacterial cells were purified using Ni-NTA Spin Kit (Qiagen, GmbH Germany) according to the manufacturers protocol. Briefly, cells were harvested by centrifugation at 5000 rpm for 15 min (GEMCO refrigerated centrifuge, Sigma). The pellet was resuspended in 1 ml lysis buffer (Appendix 2) and incubated for 1 hr with agitation at room temperature. The lysate was thereafter centrifuged at 9500 rpm for 30 min at 25°C (GEMCO refrigerated centrifuge, Sigma). 20µl of the clear lysate was used as crude extract for SDS-PAGE analysis.

Ni-NTA spin column was equilibrated with 600µl of lysis buffer by centrifugation at 2000 rpm for 2 min in a microcentrifuge. Subsequent centrifugations were done at the same conditions. 300µl of lysate containing 6xHis-tagged recombinant protein was applied onto the Ni-NTA spin column, centrifuged and flow through collected. The column was washed three times with 600µl of wash buffer (Appendix 2) by centrifugation. The 6xHis-tagged recombinant was eluted twice with 150µl of elution buffer (Appendix 2) by centrifugation. The protein concentration was determined and protein used for antibody generation and SDS-PAGE analysis.

3.9.7 Determination of protein concentration

The purified protein was dialysed in 1X PBS buffer and concentrated. The protein concentration was determined using bicinchoninic acid (BCA*) Protein Assay Reagent kit (Pierce, Rockford, Ill. USA). Briefly, working reagent was prepared by addition of 1 part of reagent B to 50 parts of reagent A and mixed. 0.1 ml each of standard, concentrated protein (serially diluted in 1X PBS buffer) and a blank (diluent/ 1X PBS) were pipetted into test tubes. 2 ml of working reagent was added to each tube and incubated for 30 min at 37°C. This reaction was cooled to room temperature and absorbance read at 562 nm on a Beckman DU 6480B spectrophotometer (Beckman Coulter, Inc. Fullerton, CA.). Concentration was then determined from a standard curve.

3.10. Antibody generation

Ten millilitres of pre-immune rabbit blood was obtained by bleeding aseptically from the central ear artery using a hypodermic syringe and needle. 1000µg of the purified recombinant P14 protein in Freund`s complete adjuvant (Pierce and Warriner, Chester, UK) was injected intramuscularly into the right thigh muscle. After 2 weeks, a boost was given on the left thigh by injecting 700µg of the recombinant protein in Freund`s incomplete adjuvant (Pierce and Warriner, Chester, UK). The rabbit was bled two weeks later and the blood stored at 4°C for two days. This was centrifuged at 13000 rpm for 10 min and serum pippeted out. The serum was stored at -20°C.

3.11 Sequence analysis

Nucleotide sequences of cloned inserts were translated to protein using the translation tool at Swiss Bioinformatics Institute website - Expasy (http://us.expasy.org) (Bairoch, 1991). The generated amino acid sequences were used to query *T. brucei* database at GeneDB to determine E-values (the probability that the alignment is due to chance) at statistical significance threshold of 0.0001. The amino acid sequences were also compared with those of *Homo sapiens*, *Leishmania major* and *T. cruzi* via alignment with ClastalW (Thompson *et al.*, 1994; Altschul *et al.*, 1997) at http://www.ebi.ac.uk/clustalw/, biological sequence alignment editor – BioEdit (Tom Hall, *Ibis Therapeutics* Carlsbad CA.) and Needleman-Wunsch global alignment (NeedleN) (Needleman and Wunsch, 1970; Kruskal, 1983; Rice *et al.*, 2000). The amino acid sequences were used to query various public protein databases to identify conserved domains. These included the integrated resource of protein domains and functional sites, InterPro (Apweiler *et al.*, 2000; Mulder *et al.*, 2005; http://www.ebi.ac.uk/interpro/), prosite (Falquet *et al.*, 2002; Hulo *et al.*, 2004; http://au.expasy.org/tools/scanprosite/), MotifScan (Falquet *et al.*, 2002; http://myhits.isb-sib.ch/cgi-bin/motif_scan) and Pfam (Sonnhammer *et al.*, 1998; Bateman *et al.*, 1999; Bateman *et al.*, 2004; http://pfam.wustl.edu/hmmsearch.shtml).

3.12 RNA interference

3.12.1 Plasmid constructs for RNAi

A 30µl restriction digestion reaction was set up consisting of 1µl of p2T7Ta vector, 3 µl of 10X buffer, 2µl *Eam*1051 restriction enzyme (MBI Fermentas, Lithuania) and 24 µl of nuclease free water. This was incubated at 37°C for 4 hrs. All the reaction mixture was electrophoresed through a 1% ethidium bromide stained agarose gel and viewed under ultraviolet (UV) illumination. The band containing the linearized plasmid was purified using QIAquick Gel Extraction Kit (Qiagen, GmbH Germany) as above.

Primers (Table 4) were designed to amplify products 500bp or less. The amplification was carried out in a 25µl reaction volume. The reaction mixture was as in section 3.9.2 above. 20µl of the amplification product were electrophoresed through a 1% ethidium bromide stained agarose gel and purified using QIAquick Gel Extraction Kit (Qiagen, GmbH Germany). The linearized plasmid generates T-overhangs that allow T-A cloning of PCR products.

Purified amplification products were ligated into linearized p2T7Ta blue vector and incubated at 4°C overnight as described before (section 3.9.3). 5µl of the ligation reaction was used to transform competent DH5α cells and plated on LB-ampicillin plates (100µg/ml) containing 150µl of IPTG (isopropylthio-β-D-galactosidase; 100mM) and 40µl of X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactosidase; 50mg/ml) and incubated overnight at 37°C. White colonies were screened using insert specific primers. Positive clones were grown in LB-ampicillin medium at 37°C overnight. The plasmid was purified using QIAprep Spin Miniprep Kit (Qiagen, GmbH Germany) as described above.

A restriction digestion reaction in a volume of 50µl was set up. The reaction mixture contained 14µl of purified plasmid, 5µl of 10X buffer, 30µl of nuclease free water and 20 units of *Not*I (MBI Fermentas, Lithuania). The reaction mixture was incubated at 37°C for 4 hrs. Linearized plasmid DNA was precipitated by adding 0.1 volume of 3M sodium acetate (pH 5.2) and 2 volumes of absolute ethanol. This was kept at -20°C overnight. The DNA was then pelleted (13,000 rpm, 10 min, 4°C) in a refrigerated microcentrifuge (Heraeus fresco Biofuge, Kendro Laboratory Products GmbH, Germany), washed with 1 ml of 70% ethanol, air-dried and redissolved in 20µl of nuclease free water.

Table 4. Primers for construction of RNAi plasmids.

FACTORS	**PRIMER SEQUENCES** (5'-3' direction)
P14	FOR: GAGAAGATCTGCATGCCTTGTAACGGGGATTCCC
	REV: CGGAATTCGTCGACATCCTCGGCCTTGACCTCCG
CPSF30	FOR: GAGAAGATCTGCATGCTGCAAGCATTGGTTTCGCGG
	REV: GGAATTCGTCGACTTGCATACCCGGACAGTTAGGTG
U170k	FOR: GAGAAGATCTGCATGCACCGAGTACGATAAGGACAC
	REV: CGGAATTCGTCGACAGCTTCCGCGGCACGCTCATAG

3.12.2 Trypanosome culture

Procyclic forms of *T. brucei* 1313-1333 cell line (Alibu *et al.*, 2004) was cultured in SDM79 (Brun and Schönenburger, 1979) supplemented with 10% fetal calf serum (FCS) at 27°C in the continuous presence of 0.5 µg/ml and 15µg/ml of phleomycin and G418 respectively (Wang *et al.*, 2000; Allen *et al.*, 2003; Bakshi and Shapiro, 2004). The cells were cultured to a final density of between 5 x 10^5/ml and 5 x 10^6/ml with the cells number determined with a haematocrit counter.

3.12.3 Transfection

The trypanosome cultures were centrifuged at 3000 rpm for 10 min (Koolspin up centrifuge, Burkard scientific, Uxbridge, Middx, UB8 2RT, UK) and supernatant filtered and stored in a new sterile tube. The cells were washed once in 5ml of Zimmerman post-fusion medium (ZPFM) electroporation medium (Appendix 2) and resuspended in 2.5ml of the same medium. 50µg of *Not*I linearized plasmid was added to 0.5ml of the 2 x 10^7 cells in 4 mm BTX cuvette (Harvard Apparatus, Holliston, Ma.) and mixed. Transfection was carried in a Gene Pulser™ electroporator (Bio-Rad, Hercules, Ca.) at 1.5kV in a resistance-timing mode of R2 (24 ohms). Immediately after transfection, the cells were transferred into 5ml of SDM-79 medium supplemented with G418 and phleomycin and incubated overnight at 27°C. Transfections were done in pairs for each gene. Circular plasmid DNA was used as control.

3.12.4 Selection of transformants

The cells were diluted to 2 x 10^5/ml using conditioned medium that was saved the previous day and selective antibiotic (hygromycin) added to a concentration of 50µg/ml. The culture was plated on a 24 well plate in a four serial fold dilutions. The plates were put in a modular chamber and incubated. The transformants were cultured until a density of 2 x 10^6/ml was achieved, transferred into new wells and some stored in liquid nitrogen (-196°C). The cells were maintained in conditioned media with hygromycin.

3.12.5 Induction of transformants

Induction of dsRNA was done by culturing the clonal population in the same medium with 100ng/ml of tetracycline. The cultures were monitored to detect morphological changes and the growth rate.

3.13 Western blotting for P14 transformants

Mid-log phase cells (2 x 10^6) were harvested, washed in 1x PBS buffer (Appendix 2) then resuspended in 50µl SDS-polyacrylamide sample buffer (Appendix 2). Control and transformants were not counted because they tended to clump together making counting virtually impossible. These were incubated at 95°C for 10 minutes. The samples were electrophoresed on 4-20% SDS-PAGE gel and transferred to nitrocellulose membrane (Bio-Rad Laboratories, Hercules, CA.) by electroblotting in transfer buffer (Appendix 2) at 168 mA for 2hrs. Nonspecific sites on the membranes were blocked by incubation in blocking buffer (Appendix 2) for 2hrs. The membrane was then incubated overnight in TBS buffer (Appendix 2) containing primary antibody. This was washed thrice in wash buffer (Appendix 2) at intervals of 5 min followed by incubation for 1hr with alkaline phosphatase conjugated secondary antibody (goat anti-rabbit IgG, Sigma-Aldrich) diluted at 1:1000 (v/v). This was subsequently washed twice in wash buffer for 5 min each followed by a third wash in substrate buffer (Appendix 2) for 10 min. Bound antibodies were detected by incubation with Nitro Blue tetrazolium (NBT) (Sigma-Aldrich) and 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) (Sigma-Aldrich) in the dark. The reaction was stopped by adding distilled deionised water.

CHAPTER FOUR

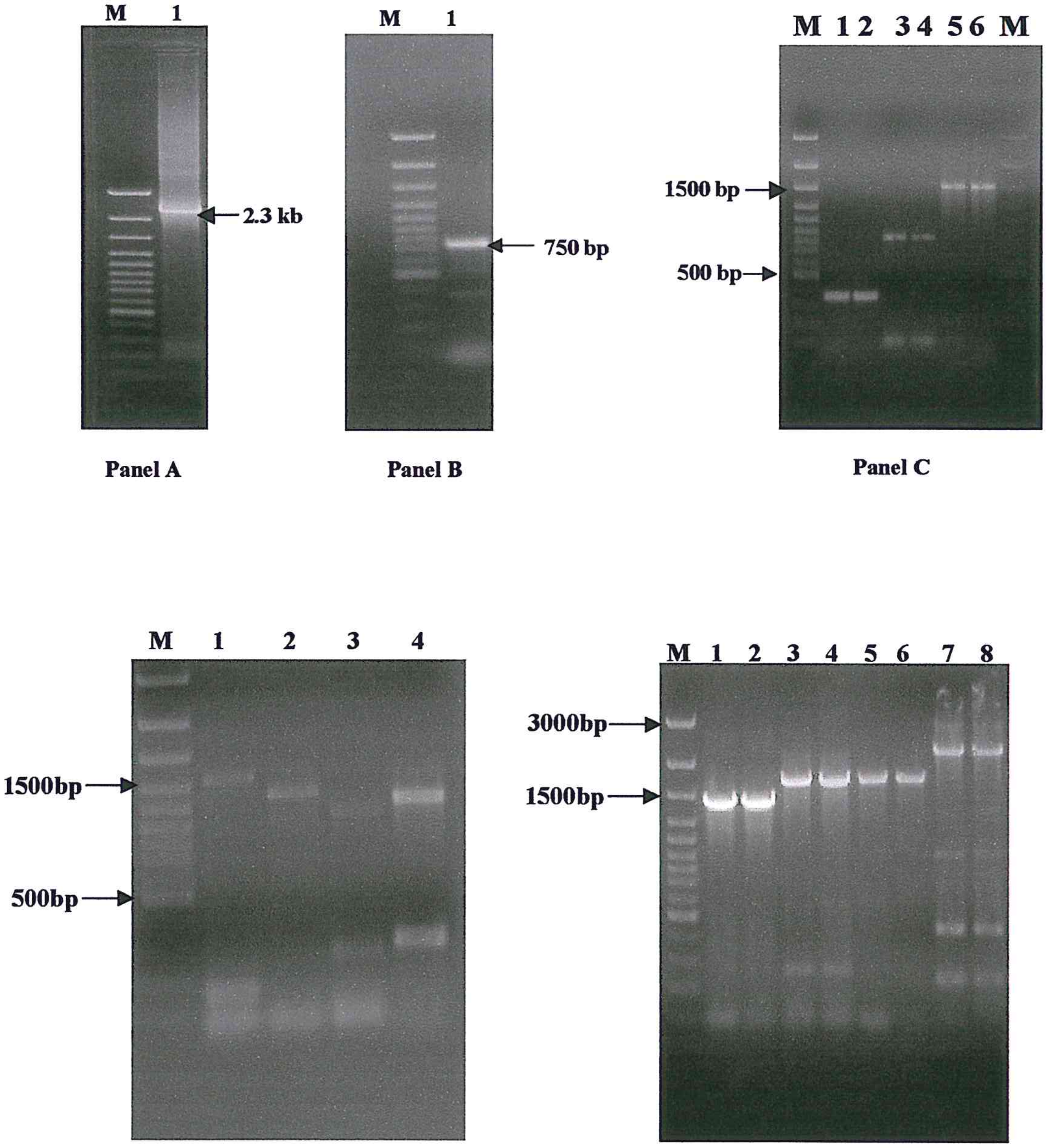
RESULTS AND DISCUSSION

4.1 Data mining

A clean dataset of *T. brucei* nucleotide and protein sequences related to spliceosome was generated from sequences available at GeneDB. The searches had motifs also identified in *Saccharomyces cerevisie*, *Caenorhabditis elegans*, *T. cruzi*, *L. major* and *H. sapiens*. Combination of methods such as reciprocal BLAST hits, presence of protein domains and motifs, and phylogenetic trees exhibited orthologous relationship among these genes and those of better studied organisms such as yeast and human. The accession numbers of the data mining results are shown in Appendix 1.

4.2 Amplification of genomic DNA

Amplification of auxiliary factors AF^{35} and AF^{65} gave products of approximately 2500 and 750 bp respectively (Figure 9 Panel A and B). P14, cleavage and polyadenylation splicing factor 30 (CPSF 30) and cleavage stimulating factor 50 (CstF 50) amplification gave products of approximately 350, 800 and 1500 bp respectively (Figure 9 Panel C). PCR amplification of cleavage factor II (CFII a1), U1-70k, zinc finger 1 (Zn 1) and zinc finger 2 (Zn 2) gave products of between 1200 and 1300 bp (Figure 9 Panel D). Amplification products for splicing factors (SF) SF3b 145, SF3b 49, SF3b 125 and SF3b 10 gave products of between 1200 and 3000 bp (Figure 9 Panel E).



Panel A

Panel B

Panel C

Panel D

Panel E

Figure 9: 1.0% ethidium bromide stained agarose gel showing amplification of splicing and polyadenylation factors from genomic DNA. Panel A: U2AF^{65}; Panel B: U2AF^{35}; Panel C: lanes 1-2, P14; lanes 3-4, CPSF 30; lanes 5-6, CstF 50; Panel D: lane 1, CFII-a1; lane 2, U1-70k; lane 3, Zn 1; lane 4, Zn 2; Panel E: lanes 1-2, SF3b 145; lanes 3-4, SF3b 49; lanes 5-6, SF3b 125; lane 7-8, SF3b 10. In all panels, M represents a 1 Kb ladder for panel A and E, and 100bp plus size ladder for B, C and D.

4.3 DNA sequencing

The genes associated with both E and A complex were sequenced. CFII-a1, Zn 1, Zn 2, U1-70k, AF^{35} and AF^{65} for the E complex while P14, SF3b 10, SF3b 49, SF3b 125 and SF3b 145 for the A complex. Cleavage and polyadenylation factors CFSF 30 and CstF 50 were also sequenced. The insert sizes from the sequencing and E-values from searches at geneDB using amino acid sequences from translated nucleotide sequences as query are shown in Table 5.

Table 5. Sequence from data mining, sequencing and E-values

FACTORS	EXPECTED SIZE (from GeneDB)	SEQUENCING (bp)	E - value (at 0.0001)
		E Complex	
CFII-a1	1272	1273	3.7e-227
Zn 1	1047	943	6.8e-148
U1-70k*	1131	1103	2.2e-129
AF^{35}	741	740	4.1e-111
AF^{65}	2352	2312	9.8e-275
Zn 2^{Ψ}	1047	1039	8.4e-215
		A Complex	
P14*	507	355	1.0e-41
Sf3b 145	1434	1433	1.2e-197
Sf3b 49	1701	1688	4.1e-217
Sf3b 125	1707	1699	3.0e-280
Sf3b 10	2199	2200	0. 00
		Cleavage and Polyadenylation Factor	
CPSF 30*	879	833	8.0e-161
CstF 50	1566	1550	2.7e-256

---

* Factors used for RNAi experiments

$^{\Psi}$ Nucleotide sequence was used instead of amino acid sequence due to short ORFs resulting from multiple stop codons, presumably resulting from sequencing errors

e = base 10

4.4 Sequence analysis

The E-values of amino acid sequences (and nucleotide sequence for Zn 2) from the cloned factors were very low at statistical significance threshold of 0.0001 (Table 5). Comparison of amino acid sequences from the cloned genes with those from *H. sapiens*, *T. cruzi* and *L. major* are shown in Tables 6, 7 and 8. Percentage identity and similarity between cloned genes and those of *H. sapiens* were in the range of 15.4 – 31.1 and 22.8 – 49.3 respectively (Table 6). Human and *T. brucei* SF3b 10 homologs were incomparable since TbSF3b 10 had 732 amino acid residues while hSF3b 10 had only 86 amino acid residues. *T. cruzi* and *T. brucei* orthologs had percentage identity and similarity ranging between 52.8 – 83.0 and 61.7 – 88.8 respectively (Table 7). Results obtained for *T. brucei* and *L. major* orthologs varied between 22.8 – 66.5 and 32.3 – 76.7 for percentage identity and similarity respectively (Table 8).

Table 6. Comparison of the amino acid sequences of cloned *T. brucei* factors and *H. sapiens* sequences.

FACTORS	% IDENTITY	% SIMILARITY
SF3b125	22.3	31.7
SF3b49	19.9	29.5
SF3b145	15.4	22.8
P14	29.1	46.3
AF35	31.1	49.3
AF65	18.3	27.7
CstF50	22.4	38.3
CFII-a1	26.6	40.9
CPSF30	30.6	39.0

Table 7. Comparison of the amino acid sequences of cloned *T. brucei* factors and *T. cruzi* sequences.

FACTORS	%IDENTITY	%SIMILARITY
SF3b125	69.3	78.9
SF3b49	71.7	78.6
SF3b145	59.1	70.9
SF3b10	53.9	68.5
P14	79.2	87.5
AF35	80.6	88.7
AF65	52.8	61.7
CstF50	67.6	77.3
CFII-a1	57.5	75.0
CPSF30	83.0	88.8

Table 8. Comparison of the amino acid sequences of cloned *T. brucei* factors and *L. major* sequences.

FACTORS	%IDENTITY	%SIMILARITY
SF3b125	61.2	72.2
SF3b49	66.5	76.7
SF3b145	44.8	59.5
SF3b10	22.8	32.3
P14	40.0	50.3
AF35	63.7	72.2
AF65	30.8	43.4
CstF50	47.3	58.7
CFII-a1	36.3	54.4
CPSF30	47.4	59.4

4.5 Protein domains

4.5.1 U2 auxiliary factor 35 (AF^{35})

Genes homologous to *T. brucei* U2AF^{35} for *T. cruzi*, *L. major* and *H. sapiens* had a central RNA recognition motif (RRM) flanked by two zinc finger motifs (Cx8Cx5Cx3H/CCCH). However, the human U2AF^{35}, hU2AF^{35} had an arginine-serine (RS) domain and a glycine tract at the C-terminus that were absent in the three kinetoplastids. Instead, *T. brucei* and *T. cruzi* accommodated a third and different zinc knuckles (CCHC/Cx2Cx4Hx4C) that is absent in *L. major*. The C-terminal of the kinetoplastids U2AF^{35} had arginine and serine residues interrupted with other residues. Signalp domain was present in *T. brucei* and *T. cruzi* (Figure 10A). The amino acid alignment of the sequences exhibits significant similarity in the kinetoplatids (Figure 10B). The RRM domain characterized by two conserved sequence elements, RNP1 and RNP2 was conserved in the kinetoplastids and had helices. The tryptophan residue that interacts with the "groove" in U2AF^{65} has been replaced with a lysine residue (green asterisk in Figure 10A) in all the kinetoplastids.

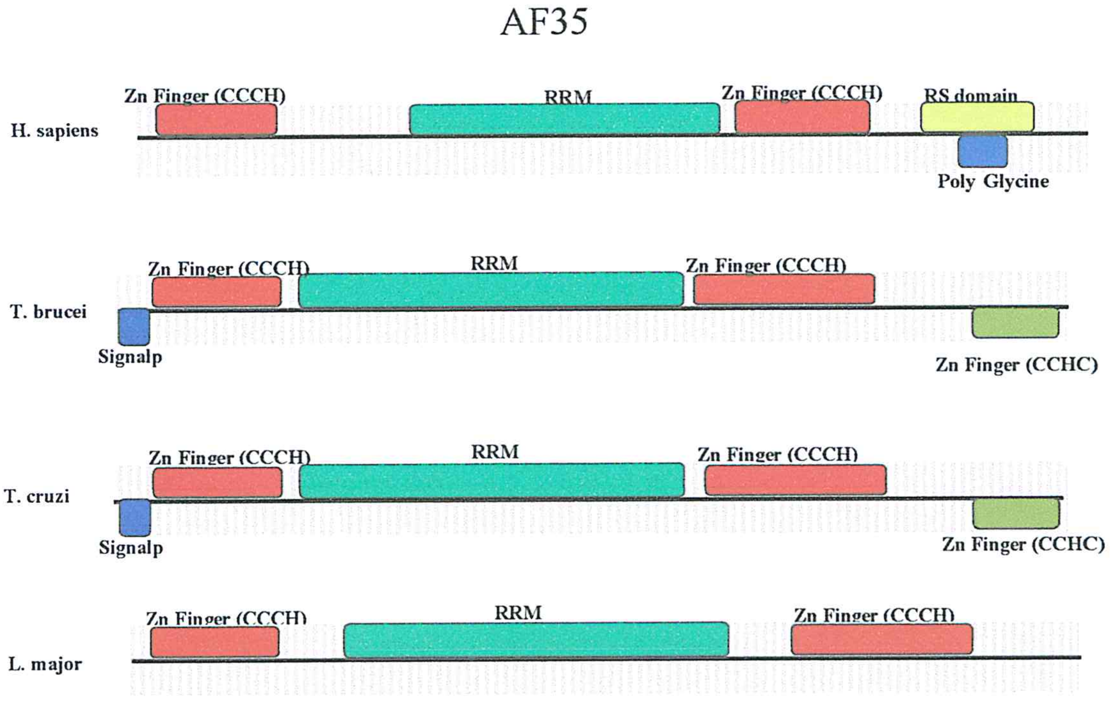
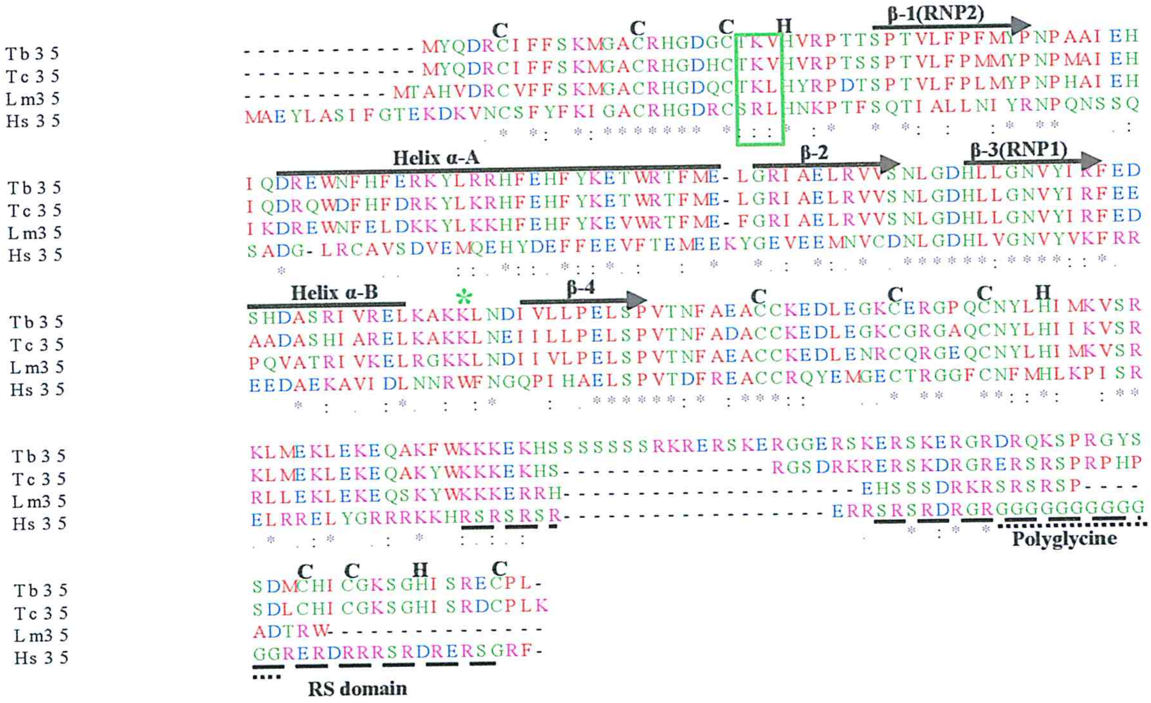


Figure 10A: U2AF^{35} protein domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*

AF35



```
Tb35  ------------MYQDRCIFFSKMGACRHGDGCTKVHVRPTTSPTVLFPFMYPNPAAIEH
Tc35  ------------MYQDRCIFFSKMGACRHGDHCTKVHVRPTSSPTVLFPMMYPNPMAIEH
Lm35  ----------MTAHVDRCVFFSKMGACRHGDQCTKLHYRPDTSPTVLFPLMYPNPHAIEH
Hs35  MAEYLASIFGTEKDKVNCSFYFKIGACRHGDRCSRLHNKPTFSQTIALLNIYRNPQNSSQ

Tb35  IQDREWNFHFERKYLRRHFEHFYKETWRTFME-LGRIAELRVVSNLGDHLLGNVYIKFED
Tc35  IQDRQWDFHFDRKYLKRHFEHFYKETWRTFME-LGRIAELRVVSNLGDHLLGNVYIRFEE
Lm35  IKDREWNFELDKKYLKKHFEHFYKEVWRTFME-FGRIAELRVVSNLGDHLLGNVYIRFED
Hs35  SADG-LRCAVSDVEMQEHYDEFFEEVFTEMEEKYGEVEEMNVCDNLGDHLVGNVYVKFRR

Tb35  SHDASRIVRELKAKKLNDIVLLPELSPVTNFAEACCKEDLEGKCERGPQCNYLHIMKVSR
Tc35  AADASHIARELKAKKLNEIILLPELSPVTNFADACCKEDLEGKCGRGAQCNYLHIIKVSR
Lm35  PQVATRIVKELRGKKLNDIIVLPELSPVTNFAEACCKEDLENRCQRGEQCNYLHIMKVSR
Hs35  EEDAEKAVIDLNNRWFNGQPIHAELSPVTDFREACCRQYEMGECTRGGFCNFMHLKPISR

Tb35  KLMEKLEKEQAKFWKKKEKHSSSSSSSRKRERSKERGGERSKERSKERGRDRQKSPRGYS
Tc35  KLMEKLEKEQAKYWKKKEKHS--------------RGSDRKRERSKDRGRERSRSPRPHP
Lm35  RLLEKLEKEQSKYWKKKERRH--------------------EHSSSDRKRSRSRSP----
Hs35  ELRRELYGRRRKKHRSRSRSR-------------------ERRSRSRDRGRGGGGGGGGG

Tb35  SDMCHICGKSGHISRECPL-
Tc35  SDLCHICGKSGHISRDCPLK
Lm35  ADTRW---------------
Hs35  GGRERDRRRSRDRERSGRF-
```

*= Conserved amino acid; green box represent conserved RNA recognition amino acids

* Lys (lysine) substitution for Trp (tryptophan).

Figure 10B. U2AF^{35} amino acid alignment of *T. brucei, T. cruzi, L. major* and *H. sapiens*

4.5.2 U2 auxiliary factor 65 (AF^{65})

The three kinetoplastids lacked the N-terminal RS domain that is present in the hU2AF^{65} (Figure 11A). *T. brucei* however, had an arginine rich region at the N-terminal. hU2AF^{65} had three RRMs as compared to one in *T. brucei* and *L. major*. *T. cruzi* has two RRMs that are close. The amino acid alignment in Figure 11B is based on the signatures for RMM1, RRM2 and RMM3 from human U2AF^{65}.

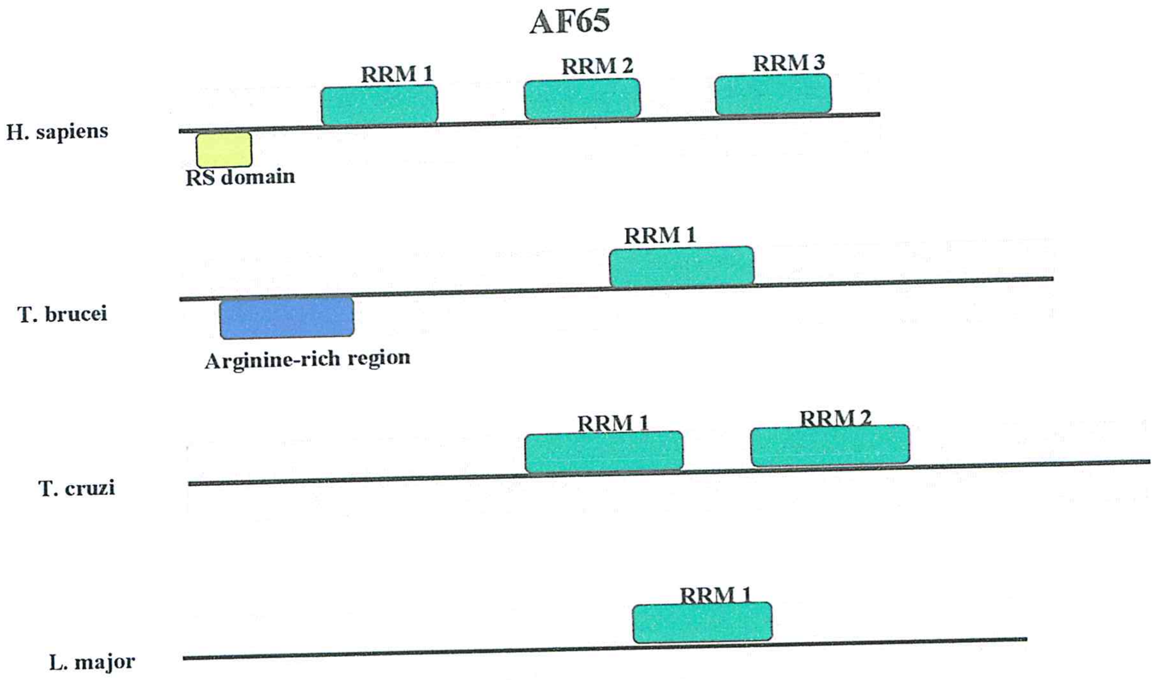
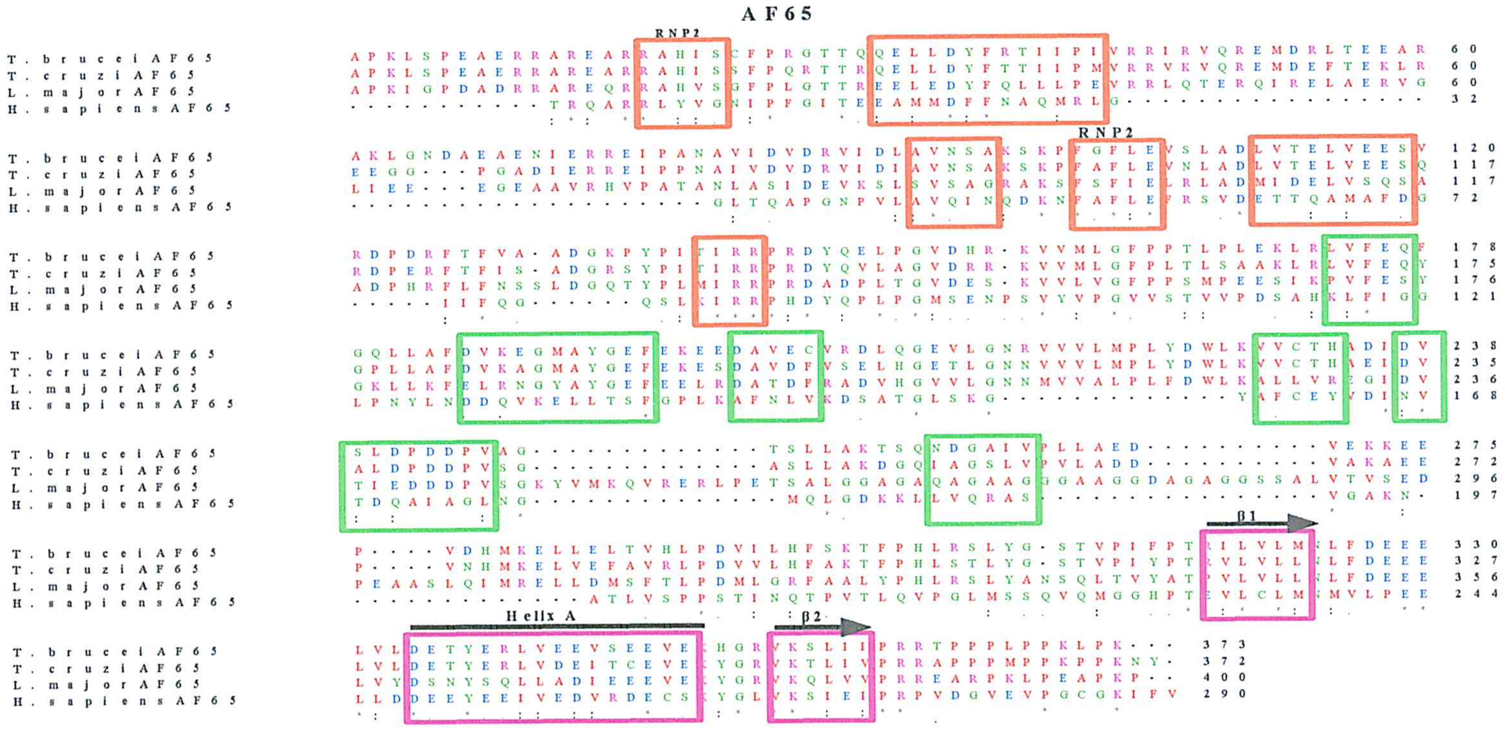


Figure 11A: U2AF^{65} protein domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*



```
T.bruceiAF65     APKLSPEAERRAREARRAHISCFPRGTTQQELLDYFRTIIPIVRRIRVQREMDRLTEEAR   60
T.cruziAF65      APKLSPEAERRAREARRAHISSFPQRTTRQELLDYFTTIIPMVRRVKVQREMDEFTEKLR   60
L.majorAF65      APKIGPDADRRAREQRRAHVSGFPLGTTREELEDYFQLLLPEVRRLQTERQIRELAERVG   60
H.sapiensAF65    ...........TRQARRLYVGNIPFGITEEAMMDFFNAQMRLG.................   32

T.bruceiAF65     AKLGNDAEAENIERREIPANAVIDVDRVIDLAVNSAKSKPFGFLEVSLADLVTELVEESV  120
T.cruziAF65      EEGG...PGADIERREIPPNAIVDVDRVIDIAVNSAKSKPFAFLEVNLADLVTELVEESQ  117
L.majorAF65      LIEE...EGEAAVRHVPATANLASIDEVKSLSVSAGRAKSFSFIELRLADMIDELVSQSA  117
H.sapiensAF65    ....................GLTQAPGNPVLAVQINQDKNFAFLEFRSVDETTQAMAFDG   72

T.bruceiAF65     RDPDRFTFVA.ADGKPYPITIRRPRDYQELPGVDHR.KVVMLGFPPTLPLEKLRLVFEQF  178
T.cruziAF65      RDPERFTFIS.ADGRSYPITIRRPRDYQVLAGVDRR.KVVMLGFPLTLSAAKLRLVFEQY  175
L.majorAF65      ADPHRFLFNSSLDGQTYPLMIRRPRDADPLTGVDES.KVVLVGFPPSMPEESIKPVFESY  176
H.sapiensAF65    ......IIFQG.....QSLKIRRPHDYQPLPGMSENPSVYVPGVVSTVVPDSAHKLFIGG  121

T.bruceiAF65     GQLLAFDVKEGMAYGEFEKEEDAVECVRDLQGEVLGNRVVVLMPLYDWLKVVCTHADIDV  238
T.cruziAF65      GPLLAFDVKAGMAYGEFEKESDAVDFVSELHGETLGNNVVVLMPLYDWLKVVCTHAEIDV  235
L.majorAF65      GKLLKFELRNGYAYGEFEELRDATDFRADVHGVVLGNNMVVALPLFDWLKALLVREGIDV  236
H.sapiensAF65    LPNYLNDDQVKELLTSFGPLKAFNLVKDSATGLSKG.............YAFCEYVDINV  168

T.bruceiAF65     SLDPDDPVAG.............TSLLAKTSQNDGAIVPLLAED..........VEKKEE  275
T.cruziAF65      ALDPDDPVSG.............ASLLAKDGQIAGSLVPVLADD..........VAKAEE  272
L.majorAF65      TIEDDDPVSGKYVMKQVRERLPETSALGGAGAQAGAAGGGAAGGDAGAGGSSALVTVSED  296
H.sapiensAF65    TDQAIAGLNG..............MQLGDKKLLVQRAS................VGAKN.  197

T.bruceiAF65     P....VDHMKELLELTVHLPDVILHFSKTFPHLRSLYG.STVPIFPTRILVLMNLFDEEE  330
T.cruziAF65      P....VNHMKELVEFAVRLPDVVLHFAKTFPHLSTLYG.STVPIYPTRVLVLLNLFDEEE  327
L.majorAF65      PEAASLQIMRELLDMSFTLPDMLGRFAALYPHLRSLYANSQLTVYATPVLVLLNLFDEEE  356
H.sapiensAF65    .............ATLVSPPSTINQTPVTLQVPGLMSSQVQMGGHPTEVLCLMNMVLPEE  244

T.bruceiAF65     LVLDETYERLVEEVSEEVEKHGRVKSLIIPRRTPPPLPPKLPK...  373
T.cruziAF65      LVLDETYERLVDEITCEVEKYGRVKTLIVPRRAPPPMPPKPPKNY.  372
L.majorAF65      LVYDSNYSQLLADIEEEVEKYGRVKQLVVPRREARPKLPEAPKP..  400
H.sapiensAF65    LLDDEEYEEIVEDVRDECSKYGLVKSIEIPRPVDGVEVPGCGKIFV  290
```

*= Conserved amino acid, Orange= RRM1 green= RRM2 and violet= RRM3 boxes, ⟶ = β sheets, ——— = Helix

Figure 11B. Structure-based multiple sequence alignment of U2AF^{65} amino acid sequences of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*. RNP1 and RNP2 signature sequences in RRMs are indicated on top.

4.5.3 Cleavage stimulating factor 50 (CstF 50)

Comparison of CstF 50 homologs showed that they have a tryptophan-aspartic acid (WD) and signalp domains except in *T. cruzi* where signalp domain was absent (Figure 12A). The WD domain signature is conserved among the species (Figure 12B). However, in the human homolog, aspartic acid (D) residue is replaced with glutamic acid (E) residue.

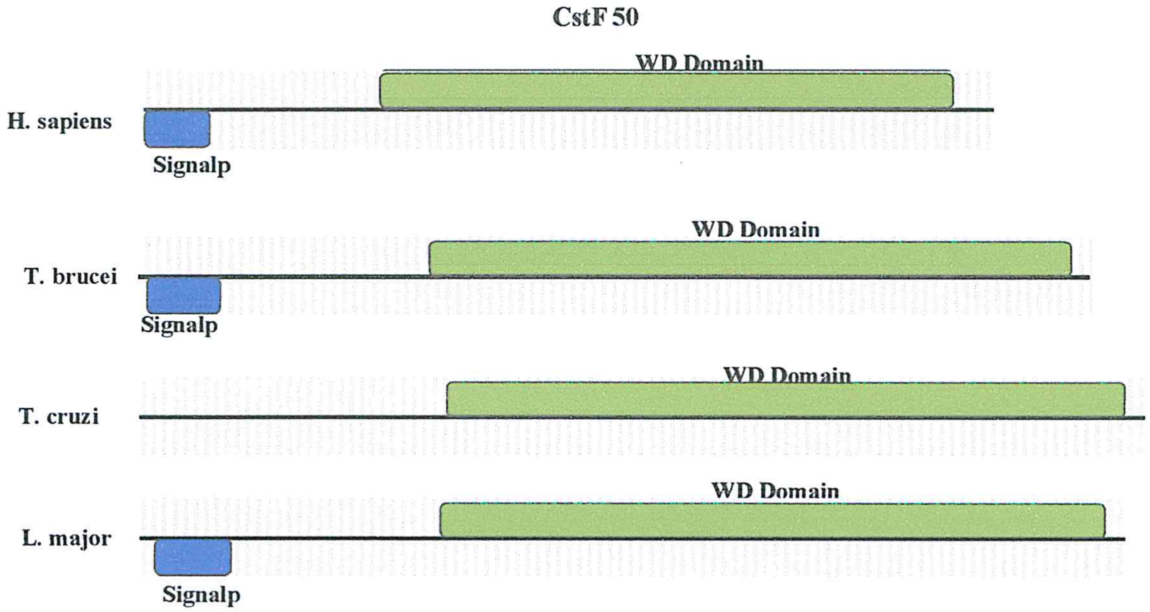


Figure 12A: Cleavage stimulating factor 50 (CstF 50) protein domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*

```
                                    Trp-Asp (WD) signature
T.bruceiCstF50        --SVKFSRTGNILLTAGMDSVARLWDLR--R- 27
L.majorCstF50         VTSVVYSRTGNVVLTAGMDSTARLWDLR--RL 30
T.cruziCstF50         --SVKFSRTGNFILSAGMDSVARLWDLR--R- 27
H.sapiensCstF50       --SAIFSKNSKYILSSGKDSVAKLWEISTGRT 30
                        *. :*:..: :*::* **.*:**::   *
```

Figure 12B. CstF 50 tryptophan-aspartic acid (WD) signature

4.5.4 Cleavage factor II (CFII–al)

All the homologs of CFII-a1 had a pre-mRNA cleavage complex II protein Clpl domain. GTPase domain was absent only in *L. major*. However, *L. major* had a signalp domain, which was also present in *T. brucei*. ATP_bind_1 domain was only present in human protein and may be similar to AAA ATPase family domain in *T. cruzi*. P-loop domain was present only in *T. brucei* protein sequence (Figure 13).

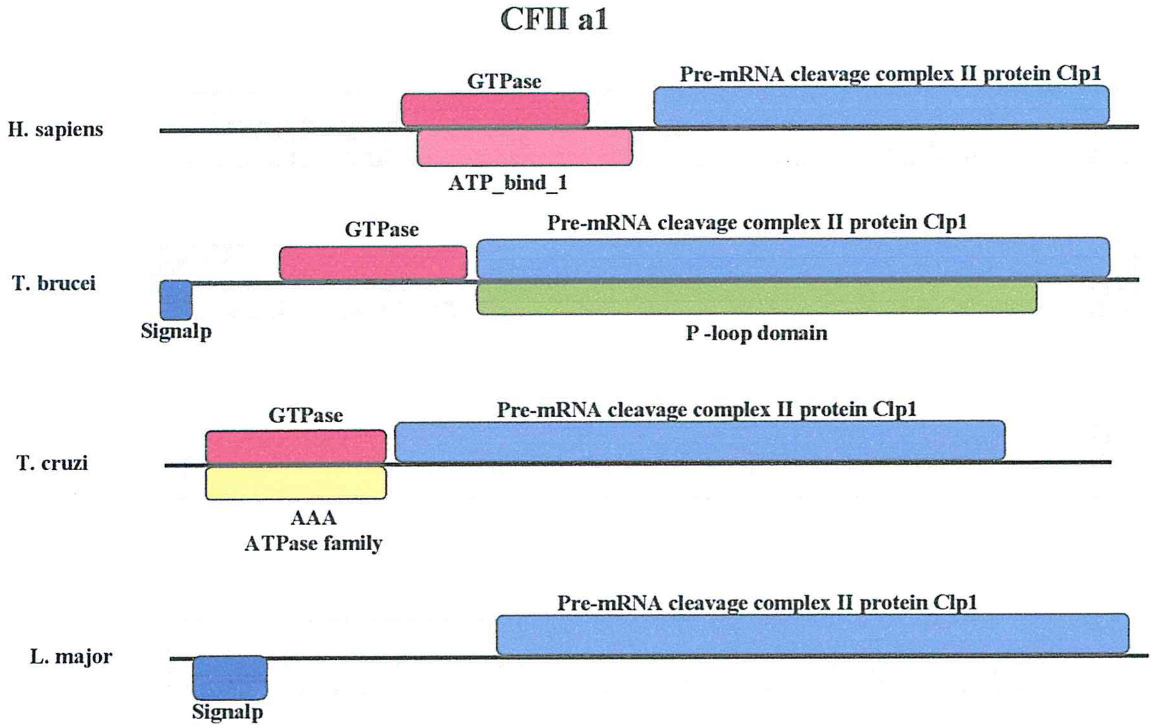


Figure 13: Cleavage factor II (CFII a1) protein domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*.

4.5.5 Splicing factor 125 (SF3b 125)

The four sequences had centrally located DEAD/D-box domain (Figure 14A), which is a conserved feature among the homologs (Figure 14B).

Figure 14A: SF3b 125 DEAD box domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*

```
T.bruceiSF3b125      NLHRVTYLVLDEADRMLDMGFEPQ 24
T.cruziSF3b125       NFF RVTYLVLDEADRMLDMGFEPQ 24
L.majorSF3b125       NLL RVTYLVMDEADRMLDMGFEPQ 24
H.sapiensSF3b125     NLQRVSYLVFDEADRMFDMGFEYQ 24
                     *:   ** : *** : ******  : *****   *
```

* = Conserved amino acids

Figure 14B. SF3b 125 DEAD-box signature from cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*.

4.5.6 Splicing factor 145 (SF3b 145)

The four species (*H. sapiens, T. brucei, T. cruzi* and *L. major)* had a proline rich (PSP) domain. However, the kinetoplastids had a signalp domain at the N-terminal while *H. sapiens* had a DNA binding splicing associated protein (SAP) domain in similar position. The human SF3b 145 had a DUF382 domain observably absent among the kinetoplastids (Figure 15A). The PSP domain signature was conserved amongst the four (Figure 15B).

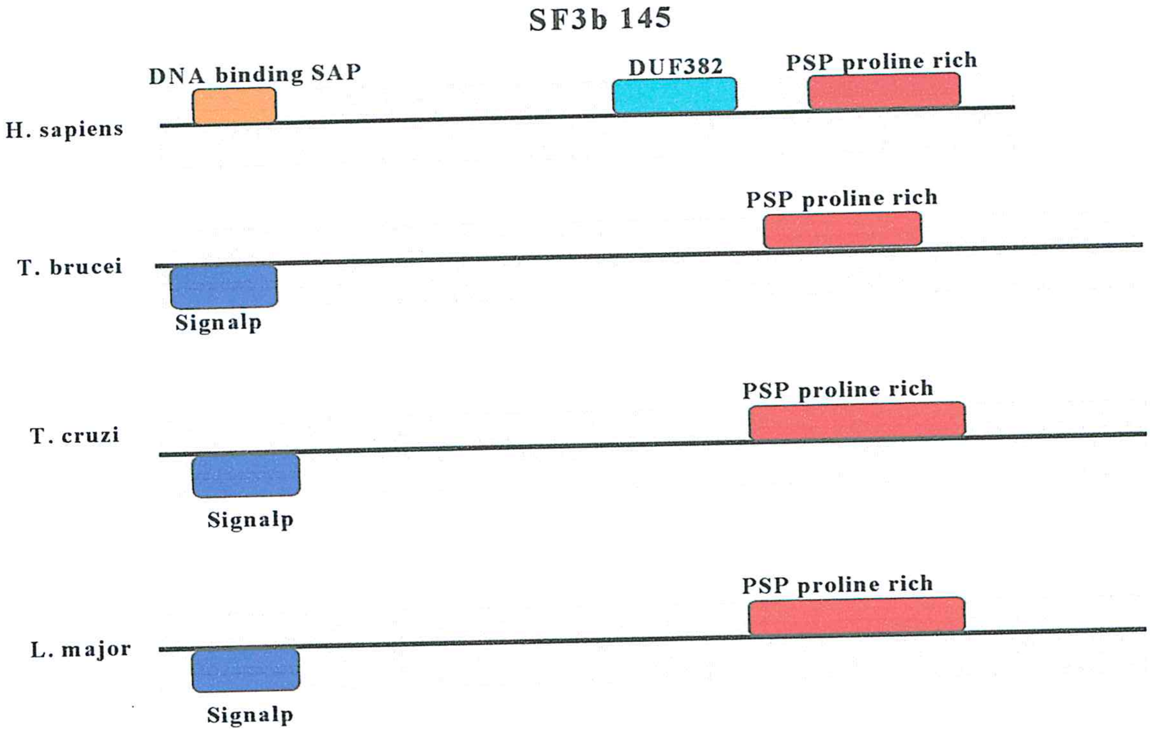
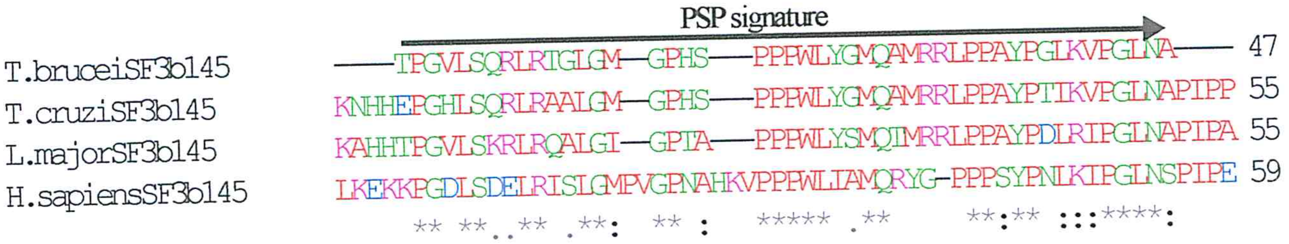


Figure 15A: Splicing factor (SF) 3b 145 domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*.



```
T.bruceiSF3b145     ----TPGVLSQRLRTGLGM--GPHS---PPPWLYGMQAMRRLPPAYPGLKVPGLNA---- 47
T.cruziSF3b145      KNHHEPGHLSQRLRAALGM--GPHS---PPPWLYGMQAMRRLPPAYPTIKVPGLNAPIPP 55
L.majorSF3b145      KAHHTPGVLSKRLRQALGI--GPTA---PPPWLYSMQIMRRLPPAYPDLRIPGLNAPIPA 55
H.sapiensSF3b145    LKEKKPGDLSDELRISLGMPVGPNAHKVPPPWLIAMQRYG-PPPSYPNLKIPGLNSPIPE 59
                         ** **..** .**:  ** :   ***** .**     **:** :::****:
```

* = Conserved amino acids

Fig 15B. Proline rich region (PSP) signature of SF3b 145

4.5.7 Splicing factor 10 (SF3b 10)

The hSF3b 10 sequence is small (86 amino acid residues) thus incomparable to the kinetoplastid homologs. The kinetoplastids had MutS III and V domains, DNA binding domain for DNA mismatch repair and ATPase domain for DNA mismatch repair (Figure 16A). Both *T. brucei* and *T. cruzi* had a signalp domain while *T. cruzi* and *L. major* had a P-loop containing nucleoside triphosphate hydrolases. The DNA mismatch repair MutS family signature was present in the three kinetoplastid sequences (Figure 16B).

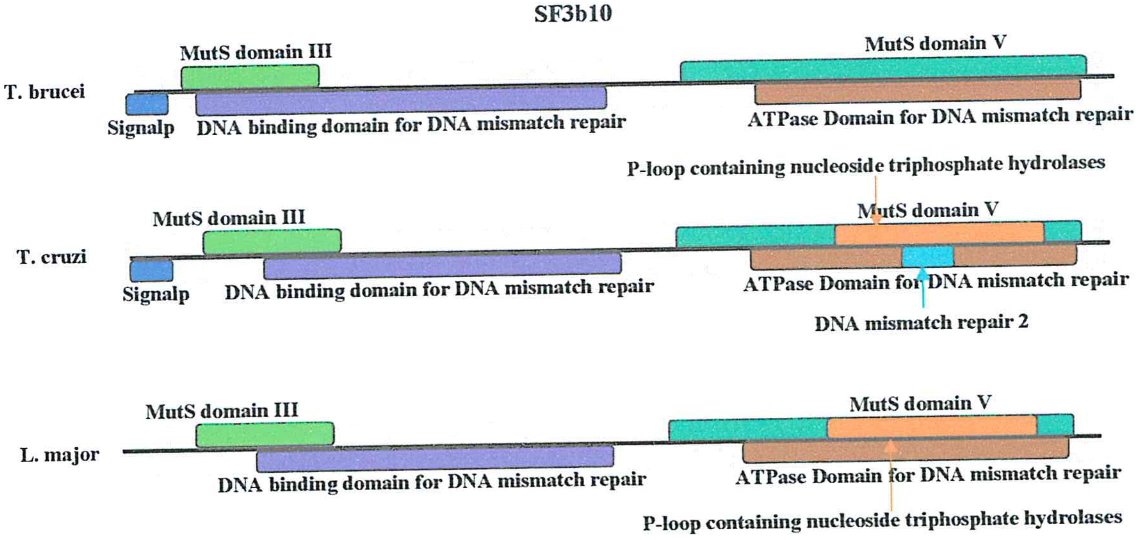
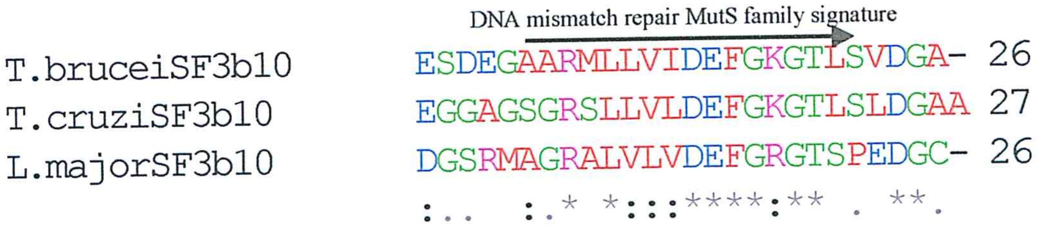


Figure 16A: SF3b 10 domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*



```
T.bruceiSF3b10    ESDEGAARMLLVIDEFGKGTLSVDGA- 26
T.cruziSF3b10     EGGAGSGRSLLVLDEFGKGTLSLDGAA 27
L.majorSF3b10     DGSRMAGRALVLVDEFGRGTSPEDGC- 26
                  :..  :.* *:::****:** . **.
```

* = Conserved amino acids

Figure 16B: DNA mismatch repair MutS family signature

4.5.8 Splicing factor 49 (SF3b 49)

Homo sapiens SF3b 49 had two RRMs as compared to four RRMs in the kinetoplastids. N-terminal signalp domain as well as poly (A)-binding protein, h (PABP, h) and PABP domains were only present in the kinetoplatids (Figure 17).

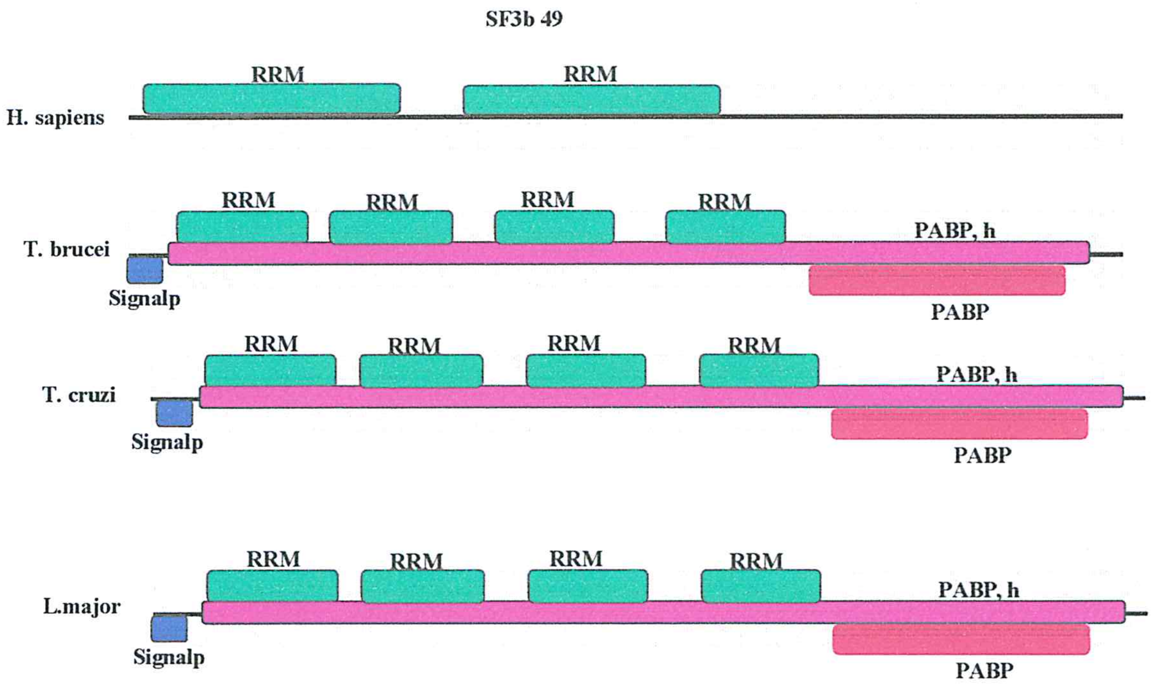


Figure 17: SF3b 49 domains of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*

4.5.9 P14

All the sequences (*H. sapiens, T. brucei, T. cruzi* and *L. major*) had a single RRM (Figure 18). *T. brucei* and *L. major* also had a signalp domain that was absent in *H. sapiens* and *T. cruzi*.

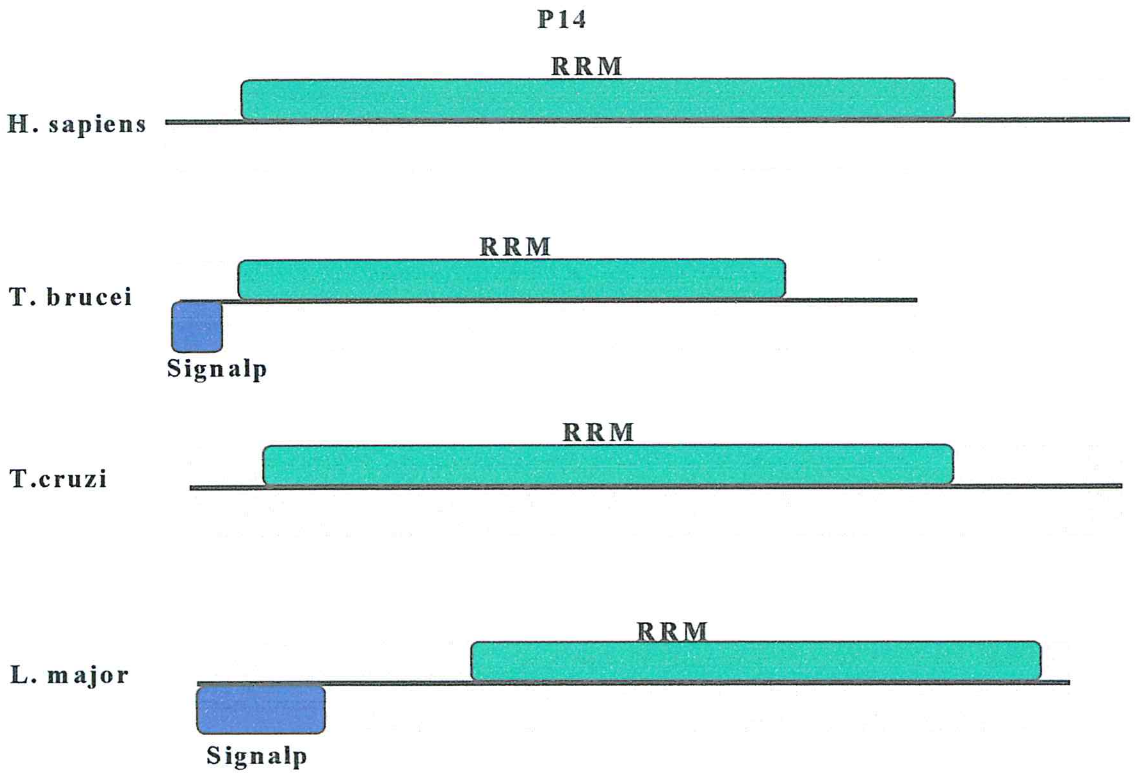
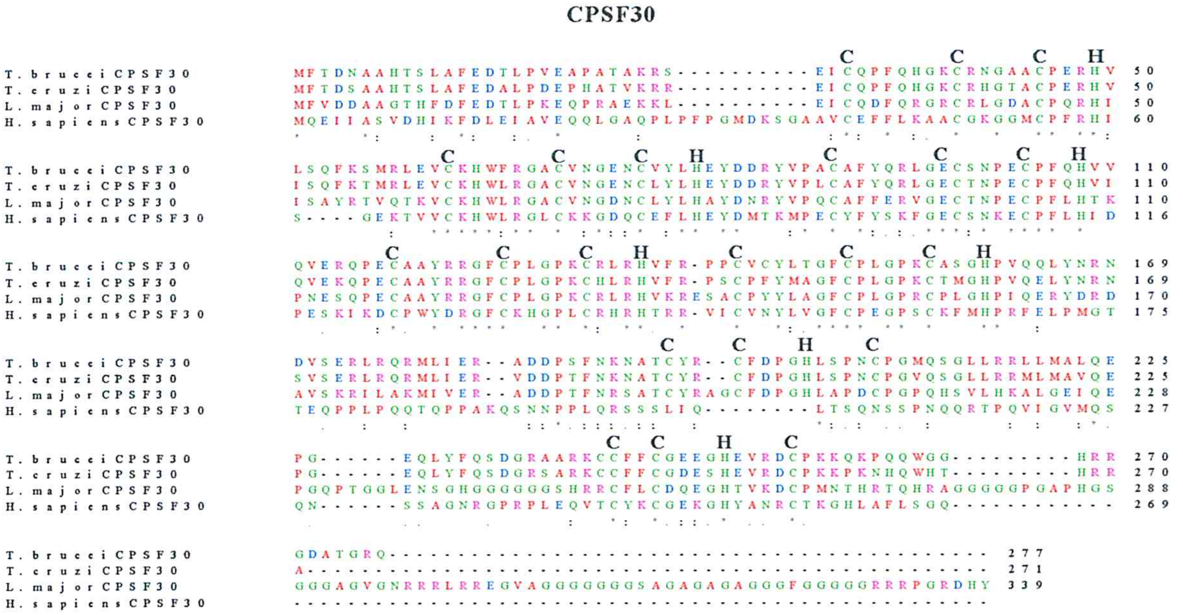


Figure 18: RRM domains in P14 of cloned *T. brucei*, *T. cruzi*, *L. major* and *H. sapiens*

4.5.10 Cleavage and polyadenylation factor 30 (CPSF 30)

CPSF 30 had zinc fingers of type CCCH and CCHC (zinc knuckle). CPSF 30 of *T. brucei*, *L. major* and *T. cruzi* had five CCCH zinc fingers and two zinc knuckles while *H. sapiens* had four CCCH zinc fingers and one zinc knuckle (Figure 19). Various deletions and additions are noticeable in the homologs. The *L. major* CPSF 30 is longer with several glycine residues at the C-terminal.



```
T.bruceiCPSF30     MFTDNAAHTSLAFEDTLPVEAPATAKRS----------EICQPFQHGKCRNGAACPERHV 50
T.cruziCPSF30      MFTDSAAHTSLAFEDALPDEPHATVKRR----------EICQPFQHGKCRHGTACPERHV 50
L.majorCPSF30      MFVDDAAGTHFDFEDTLPKEQPRAEKKL----------EICQDFQRGRCRLGDACPQRHI 50
H.sapiensCPSF30    MQEIIASVDHIKFDLEIAVEQQLGAQPLPFPGMDKSGAAVCEFFLKAACGKGGMCPFRHI 60

T.bruceiCPSF30     LSQFKSMRLEVCKHWFRGACVNGENCVYLHEYDDRYVPACAFYQRLGECSNPECPFQHVV 110
T.cruziCPSF30      ISQFKTMRLEVCKHWLRGACVNGENCLYLHEYDDRYVPLCAFYQRLGECTNPECPFQHVI 110
L.majorCPSF30      ISAYRTVQTKVCKHWLRGACVNGDNCLYLHAYDNRYVPQCAFFERVGECTNPECPFLHTK 110
H.sapiensCPSF30    S----GEKTVVCKHWLRGLCKKGDQCEFLHEYDMTKMPECYFYSKFGECSNKECPFLHID 116

T.bruceiCPSF30     QVERQPECAAYRRGFCPLGPKCRLRHVFR-PPCVCYLTGFCPLGPKCASGHPVQQLYNRN 169
T.cruziCPSF30      QVEKQPECAAYRRGFCPLGPKCHLRHVFR-PSCPFYMAGFCPLGPKCTMGHPVQELYNRN 169
L.majorCPSF30      PNESQPECAAYRRGFCPLGPKCRLRHVKRESACPYYLAGFCPLGPRCPLGHPIQERYDRD 170
H.sapiensCPSF30    PESKIKDCPWYDRGFCKHGPLCRHRHTRR-VICVNYLVGFCPEGPSCKFMHPRFELPMGT 175

T.bruceiCPSF30     DVSERLRQRMLIER--ADDPSFNKNATCYR--CFDPGHLSPNCPGMQSGLLRRLLMALQE 225
T.cruziCPSF30      SVSERLRQRMLIER--VDDPTFNKNATCYR--CFDPGHLSPNCPGVQSGLLRRMLMAVQE 225
L.majorCPSF30      AVSKRILAKMIVER--ADDPTFNRSATCYRAGCFDPGHLAPDCPGPQHSVLHKALGEIQE 228
H.sapiensCPSF30    TEQPPLPQQTQPPAKQSNNPPLQRSSSLIQ--------LTSQNSSPNQQRTPQVIGVMQS 227

T.bruceiCPSF30     PG------EQLYFQSDGRAARKCCFFCGEEGHEVRDCPKKQKPQQWGG---------HRR 270
T.cruziCPSF30      PG------EQLYFQSDGRSARKCCFFCGDESHEVRDCPKKPKNHQWHT---------HRR 270
L.majorCPSF30      PGQPTGGLENSGHGGGGGGSHRRCFLCDQEGHTVKDCPMNTHRTQHRAGGGGGPGAPHGS 288
H.sapiensCPSF30    QN------SSAGNRGPRPLEQVTCYKCGEKGHYANRCTKGHLAFLSGQ------------ 269

T.bruceiCPSF30     GDATGRQ-------------------------------------------- 277
T.cruziCPSF30      A-------------------------------------------------- 271
L.majorCPSF30      GGGAGVGNRRRLRREGVAGGGGGGGSAGAGAGAGGGFGGGGGRRRPGRDHY 339
H.sapiensCPSF30    ---------------------------------------------------
```

* = Conserved amino acids

Figure 19: Zinc fingers of CPSF 30.

4.6 *T. brucei* U2 auxiliary factor 65 (TbU2AF^{65})

Analysis of TbU2AF^{65} sequence and the well characterized hU2AF^{65} showed conservation of some of the amino acid residues that could be associated with "groove" formation. Some were not conserved as shown in Figure 20 and Table 9 below. Of particular interest are two proline residues with six amino acid spacer. These are important in "groove" formation in hU2AF^{65} hence highly conserved (Kielkopf *et al.,* 2001). Sequence alignment between hU2AF^{65} and TbU2AF^{65} gave two proline residues at positions TbU2AF^{65}Pro162 and TbU2AF^{65}Pro169 with a six amino acid residue spacer. The other residues important in "groove" formation were substituted with amino acids of similar properties except in TbU2AF^{65}Pro160 (Table 9). Sequence comparison between hU2AF^{65} and the two *Trypanosoma* species homologs (TbU2AF^{65} and TcU2AF^{65}) and later between hU2AF^{65} and the three kinetoplastid homologs (TbU2AF^{65}, TcU2AF^{65} and LmU2AF^{65}) gave almost similar results. In both cases, the two proline residues with six amino acid spacer were found at positions TbU2AF^{65}Pro213 and TbU2AF^{65}Pro220. The representatives of hU2AF^{65}Pro96 were proline residues which occurred at different positions, TbU2AF^{65}Pro129 in the first comparison and TbU2AF^{65}Pro174 in the second comparison. 83 residue spacer was noted in the alignment between the three kinetoplastids and human U2AF^{65} homologs (TbU2AF^{65} Pro129 and TbU2AF^{65}Pro213), while 38 residue spacer was noted in the alignment between the *Trypanosoma* species and human U2AF^{65} homologs (TbU2AF^{65}Pro174 and TbU2AF^{65} Pro213). No spacer was present in the case of alignment between human U2AF^{65} and *T. brucei* U2AF^{65} (Figure 20).

```
                               * *      *
TbU2AF65      147 QQQQQQHAGLLSDPLPAAVAIPPVETTPMISVGTVA 182
hU2AF65        82 HEKKKKVRKYWDVPPPGFEHITPMQYKAMQAAGQIP 117
                  ::::::     . * *.   *.*:: ..* :.* :.
```

* = Conserved proline residues

Figure 20. Alignment of amino acid residues of *H. sapiens* and *T. brucei* U2AF^{65} for "groove" formation.

Table 9. Residues in *H. sapiens* and *T. brucei* U2AF^{65}. The amino acid residues are implicated in "groove" formation based on sequence alignment between the *T. cruzi* (Tc), *T. brucei* (Tb), *L. major* (Lm) and *H. sapiens* (Hs). The columns under TbU2AF^{65} represent *T. brucei* U2AF^{65} amino acid residues at corresponding positions to amino acid residues of human U2AF^{65}.

hU2AF^{65}	**TbU2AF^{65}**		
	Tb / Hs	**Tb / Tc / Hs**	**Tb / Tc / Lm / Hs**
Trp92	Leu157	Val170	Pro125
Pro95	Pro160	Thr173	Asp128
Pro96	Leu161	Pro174	Pro129
Pro 97	Pro162	Pro213	Pro213
Phe99	Ala164	Val215	Val215
Pro 104	Pro169	Pro220	Pro220
Try107	Thr172	Pro223	Pro223

4.7 Purification of recombinant proteins in bacterial cells

4.7.1 CPSF 30 recombinant protein

Transformed BL21 *E. coli* cells expressed the CPSF 30 recombinant protein when induced with IPTG. The recombinant CPSF 30 protein purified on a Ni-NTA Spin Kit (Qiagen, GmbH Germany) gave a product of approximately 24 kDa when resolved on a 4 - 20% SDS-PAGE gel (Figure 21).

Purification of recombinant CPSF 30

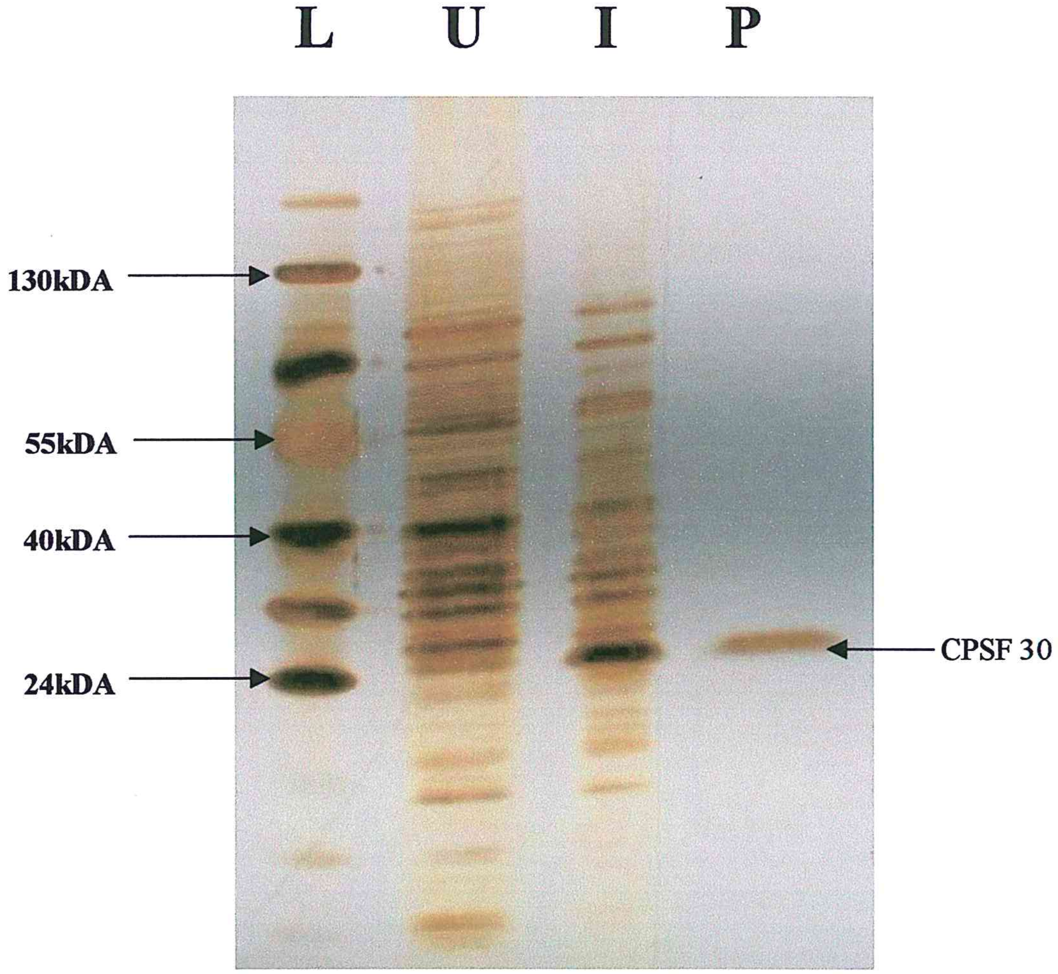


Figure 21. 4 - 20% SDS-PAGE gel showing purification of recombinant CPSF 30 protein. **Lanes** L: Protein ladder, U: uninduced recombinant, I: induced recombinant and P: purified recombinant CPSF 30.

4.7.2 P14 recombinant protein

BL21 *E. coli* cells transformed with pET28a plasmid with P14 insert expressed the recombinant P14 protein when induced with IPTG. Purification of the recombinant P14 protein on a Ni-NTA Spin Kit (Qiagen, GmbH Germany) gave a product of approximately 20 kDa on separation on a 4 - 20% SDS-PAGE gel (Figure 22).

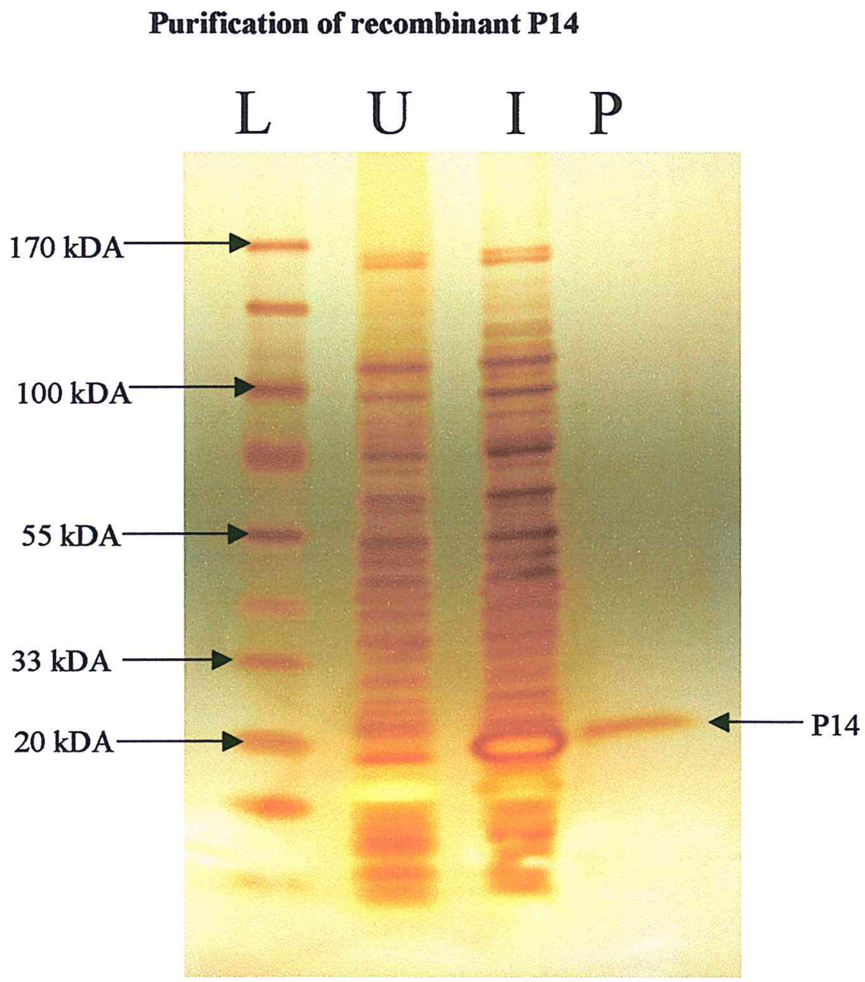

Purification of recombinant P14

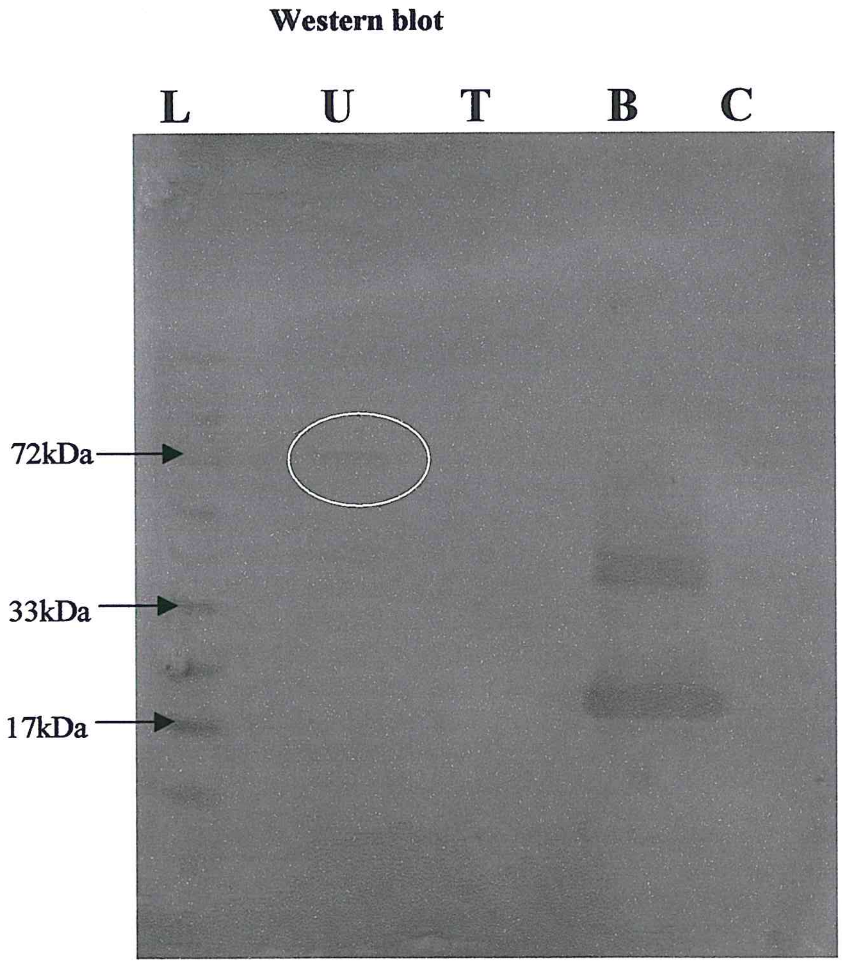
Figure 22. 4 - 20% SDS-PAGE gel showing purification of recombinant P14 protein. Lanes L: Ladder, U: uninduced recombinant, I: induced recombinant and P: purified recombinant P14.

4.8 Gene silencing of P14, U1-70k and CPSF 30 by RNAi

Difficulties were experienced in generation of clonal populations after induction due to clumping of the cells. However, growth arrest was evident among transfected cells in all the genes investigated (P14, U1-70k and CPSF 30). Unique morphological features were noticed in transgenic P14 and U1-70k cells but not in transgenic CPSF 30 cells (data not shown). The cells were abnormally large in size. In P14 transformants, the cells could not undergo complete division and remained held at the head in a group of about six cells just before death. Spherical cells like FAT cell stage (Ngô *et al.*, 1998) were noticed on induction for all the three factors.

4.9 Western blot

Western blot on proteins derived from transfected, untransfected and control trypanosomes is shown in Figure 23. Detection of P14 protein in the untransfected trypanosome 1313-1333 cell line indicated that the gene encoding P14 was expressed. Similar detection was expected for control, however its absence could be due to "leakage" in the system (Alibu *et al.*, 2004). Depletion of P14 was observed in the transfected recombinants a phenomenon attributed to knock down of the P14 gene.



Western blot

Figure 23. Results of western blot analysis of a trypanosome cell line transformed with P14 RNAi construct. L, protein ladder; U, 1313-1333 lysate; T, P14 RNAi transfected 1313-1333 cell; B, recombinant total bacterial lysate and C, control. The circle shows detection of P14 in untransfected 1313-1333 cell line lysate.

4.10 DISCUSSION

4.10.1 Sequence comparison

The amplification and sequencing results showed that *T. brucei* CFII-a1, AF^{35}, SF3b 145 and SF3b 10 were of similar sizes as those sequences from data mining at GeneDB, while U1-70k, AF^{65}, CstF 50, Zn 1, Zn 2, CPSF 30, P14, SF3b 49 and SF3b 125 were smaller in size. Comparison of the nucleotide and amino acid sequences of the cloned genes and those from data mining showed that the correct genes were recovered by the degenerate PCR amplification approach. This observation was further supported by the extremely low E–values (between 0.000 - 3.0e-280) at the stringent threshold limit of 0.0001.

Comparison of the *T. brucei* and human homologs showed the lowest percentage identity and similarity. This is because they are evolutionarily distinct and represents one of the earliest branches in eukaryotic lineage (Bringaud *et al.*, 1998; Stevens *et al.*, 1998; Verlinde *et al.*, 2001). Small U2 auxiliary factor ($U2AF^{35}$) had high percentage identity and similarity across the four species in comparison to other factors. Moreover, the RNP1 and RNP2 motifs of RRM are conserved. This could be due to conserved intimate heterodimeric interaction of auxiliary factor (AF^{35} and AF^{65}) in eukaryotes (Vázquez *et al.*, 2003). This could further be attributed to phylogenetic conservation of AF^{35} and residues of the $U2AF^{65}$ peptide that are critical for $U2AF^{35}$ binding (Kielkopf *et al.*, 2001). Similarly, CPSF 30 had an appreciably high percentage identity and similarity, presumably due to the conservation of the overall zinc finger motif structure and function (Hendriks *et al.*, 2003).

T. cruzi and *L. major* orthologs showed the highest percentage identity and similarity to *T. brucei*. Similar closeness was observed by El-Sayed *et al.* (2005b) at whole genome level. However, *T. cruzi* orthologs are more closely related to *T. brucei* than *L. major*. This closeness is in agreement with amino acid sequence alignment of a large sample of three-way cluster of orthologous genes (COGs) earlier observed by Haag *et al.* (1998) and El-Sayed *et al.* (2005b). The alignment revealed an identity of 57% between *T. brucei* and *T. cruzi* and 44% between *L. major* and the two other trypanosomes, reflecting phylogenetic relationships. Similarly, analysis of glucose transporter gene cluster (Bringaud *et al.*, 1998) showed close evolutionary relationship between *T. brucei* and *T. cruzi*; members of the same genus. The difference between

T. brucei and *T. cruzi* is supported by the suggestion that among the monophyletic trypanosomatids, the Salivarian trypanosomes (also called African trypanosomes: subgenus *Trypanozoon* or *Trypanosoma brucei* group, *T. congolense* and *T. vivax*) emerged before *T. cruzi* (Bringaud *et al.*, 1998). This variation could also be due to varied acquisition of an accelerated rate of evolutionary substitutions in *Trypanosoma* (Lake *et al.*, 1988) and different rates of evolution (Stevens *et al.*, 1998).

4.10.2 Domains / motifs

The thirteen *T. brucei* trans-spliceosome genes studied showed domains that suggest their involvement in RNA splicing. The TbU2AF^{35}, TbU2AF^{65}, TbP14 and TbSF3b 49 have RRM domains involved in RNA recognition; a fundamental process in precise splice site and branch point recognition during RNA maturation. The TcU2AF^{35} RRM domain has conserved residues (Thr 45, Leu 47 and Tyr 114) known to be directly involved in RNA recognition (Vázquez *et al.*, 2003). These residues are also conserved in TbU2AF^{35}. However, Trp 134, the hallmark of the U2AF^{35} RRM domain of eukaryotes (Vázquez *et al.*, 2003) is absent in TbU2AF^{35}. This residue, which is necessary for the reciprocal "tongue in groove" heterodimerization with U2AF^{65} is changed to Lys in the *T. brucei* ortholog. The *T. cruzi* ortholog has the same substitution (Vázquez *et al.*, 2003). This is a fundamental difference with the human homolog suggesting that the trypanosome gene products of U2AF^{35} and U2AF^{65} interact differently during spliceosome assembly. The third and different zinc knuckles (CCHC/Cx2Cx4Hx4C) in TbU2AF^{35} and TcU2AF^{35} is similar to the zinc finger domain found in a protein that binds the universal minicircle sequence of trypanomastids and is indicative of kinetoplastid DNA/RNA single strand binding protein (Tzfati *et al.*, 1995; Abu-Elneel *et al.*, 1999).

U2 auxiliary factor large subunit U2AF^{65} interacts directly with the pyrimidine (Py) tract and branch point (BP) by the C-terminal RRM (Kielkopf *et al.*, 2001) and RS domain (Förch *et al.*, 2003) respectively. Its RRM also interacts with SF3b 155, a component of U2 snRNP (Shepard *et al.*, 2002). The protein is therefore thought to be involved in stabilization of the interaction of U2 snRNP with the BP through base-pairing interactions (Gozani *et al.*, 1998; Förch *et al.*, 2003). The hU2AF^{65} N-terminal RS domain is missing in the three kinetoplastids. TbU2AF^{65} however, has an arginine rich region at the N-terminal, which could be involved in direct interaction with the BP and stabilization of the interaction of U2 snRNP with the BP as in hU2AF^{65}. The RMM domains could be involved in interaction with pyrimidine (Py) tract and splicing factor 1/branch point binding protein (SF1/BBP) as suggested by Varani and Ramos (2003). TcU2AF^{65} has two RRMs. This is suspected to be a split RRM when compared with the hU2AF^{65}.

The cleavage stimulating factor 50 (CstF 50) sequences have WD domain (WD or beta-transducin repeats) with a terminating Trp-Asp (W-D) dipeptide characteristic of the domain. The WD domain proteins form a large family with a high degree of diversity in sequence, multidomains and cellular functions (Yu *et al.*, 2000). The sequence diversity occurs primarily in the two variable regions within the WD-repeat itself (Yu *et al.*, 2000) thus substitution of aspartic acid with glutamic acid in human CstF 50. The TbCstF 50 could be involved in directing spliceosome complex assembly in which interactions between several proteins are involved. This is because the underlying common function of the domain is to coordinate multi-protein complex assemblies in signal transduction, transcription initiation complex assembly, chromatin assembly, RNA splicing, vascular trafficking, cell cycle control and apoptosis (Smith *et al.*, 1999; Madrona and Wilson, 2004). The motif also provides an interface for protein–protein interactions (Zhao *et al.*, 1999; Li and Roberts, 2001) either with other members of the WD family or with proteins carrying different motifs; most of the known proteins being members of multiprotein complexes (Yehuda *et al.*, 1998). These interactions can occur simultaneously with several different proteins and their specificity is determined by sequences outside the repeats (Li and Roberts, 2001). Its interaction with RNA polymerase II and CPSF to stabilize the cleavage complex (Zhao *et al.*, 1999) could be through the WD domain that interfaces for protein-protein interaction among different proteins (Zhao *et al.*, 1999; Li and Roberts, 2001).

TbCFII-a1 has pre-mRNA cleavage complex II protein Clp1 domain that may be involved in ATP/GTP binding, P-loop (phosphate binding loop) with nucleoside triphosphate hydrolases and GTPase domain. These domains may be involved in nucleoside hydrolysis during cleavage and are similar to members of AAA+ family of ATPases. This family of ATPases forms dynamic oligomeric rings, and carries out diverse and important cellular functions, including those of helicases, unfoldases and ATPases (Orlova and Saibil, 2004), which have been recorded in the splicing process. The function of the P-loop is to correctly position the triphosphate moiety of a bound nucleotide (Caruthers and McKay, 2002; Leipe *et al.*, 2002).

SF3b 125 proteins have a DEAD box signature. The DEAD box represents the one letter code for the tetrapeptide, Asp-Glu-Ala-Asp and is a helicase domain characteristic of members of DExH/D box family domain (Will *et al.*, 2002; Shi *et al.*, 2004). The helicase 'superfamily' of proteins (RNA unwindases/ RNPases/ helicases) is characterized by a common general function of an ATP-dependent nucleic acid unwinding (de la Cruz *et al.*, 1999). The 'superfamily' has been implicated in various aspects of RNA metabolism which include nuclear transcription, pre-mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression (de la Cruz *et al.*, 1999; Tanner and Linder, 2001; Cordin *et al.*, 2004). TbSF3b 125 may therefore be involved in the ATP dependent A complex formation in which base pairing and ATP hydrolysis are involved. The domain could be specifically implicated in directing precise base pairing and correcting mismatch in the recruitment of U2 snRNP to the degenerate branch point. This process could occur through displacement of splicing factor 1/branch point binding protein (SF 1/BBP) as suggested by Fleckner *et al.* (1997) on the role of two DEAD box proteins, Prp5p and UAP56. The nucleic acid unwinding ability is very important in structural rearrangements and conformational changes during spliceosome assembly and correction of mismatches. The motif is specific to proteins that couple ATP-binding/hydrolysis and structural rearrangement (Fleckner *et al.*, 1997; Xu *et al.*, 2004) fitting well with formation of A complex, an ATP dependent process.

TbSF3b 145 exhibited a proline-rich domain (PSP) similar to homologs from *H. sapiens*, *T. cruzi* and *L. major*. It probably interacts with TbSF49 via its proline-rich domain since this domain is dispensable for the protein–protein interaction between human SF3b 145 and SF3b 49 (Igel *et al.*, 1998). *H. sapiens* SF3b 145 has a DNA binding SAP or SF found in ATP-dependent DNA helicase.

Interaction between TbSF3b 49 and TbSF3b 145 could be through RRM of TbSF3b 49 as observed in yeast homologs (Igel *et al.*, 1998). TbSF3b 49 may also bind the pre-mRNA via the RRMs. These interactions are for the stable recruitment of U2 snRNP to the degenerate BP, a process that involves base pairing (Gozani *et al.*, 1998; Igel *et al.*, 1998). These inferences are supported by the facts that SF3b 49 can cross-link efficiently to RNA substrates in complexes A and B and also to bind both U2 snRNP and the pre-mRNA (Chiara *et al.*, 1996). The poly adenylate binding protein (PABP) and PABPh domains in TbSF3b 49 are thought to recognize the poly-A tail of mRNA and may be involved in the linkage of cleavage and polyadenylation.

The domains in TbSF3b 10 implicate the protein in the energy dependent mismatch repair during A complex formation. It has MutS (III and IV domains); a key protein of the *Escherichia coli* DNA mismatches repair system that recognizes mispaired and unpaired bases and has intrinsic ATPase activity (Lamers *et al.*, 2004). The ATPase domain in TbSF3b 10 could be associated with ATP binding activity that induces a state in which MutS slides away from the mismatch to allow new molecules to bind the mismatch (Lamers *et al.*, 2004) or discrimination between homoduplex and heteroduplex DNA (Schofield *et al.*, 2001). Alternatively, MutS domain can act as a motor protein that uses the ATPase activity to translocate along the DNA in search of a signal for strand discrimination (Blackwell *et al.*, 1998). In *T. cruzi* and *L. major*, a P-loop domain with nucleoside triphosphate hydrolase could be associated with ATP binding and hydrolysis. The binding and hydrolysis cause ATP-dependent conformational change that allows recruitment of other proteins (Alani *et al.*, 2003). These features concur with the structural rearrangements and energy consumption associated with spliceosome assembly (Schwer and Guthrie, 1992; Chan *et al.*, 2003; Xu *et al.*, 2004).

TbCPSF 30 has five zinc finger (type CCCH) motifs and two zinc knuckles (CCHC) as described by Hendriks and colleagues, 2003. TbCPSF 30 may be involved in both cleavage and polyadenylation. These involve RNA binding and protein-protein interactions through the motifs. These motifs typically function as interaction modules and bind to a wide variety of compounds such as nucleic acids (Hendriks *et al.*, 2003), proteins and small molecules (Krishna *et al.*, 2003). They are also structurally diverse and are present among proteins that perform a broad range of functions in various cellular processes, such as replication and repair, transcription and translation, metabolism and signalling, cell proliferation and apoptosis (Krishna *et al.*, 2003). Zhao *et al.* (1999) reported that CPSF as well as poly (A) polymerase (PAP) remains bound to the cleaved RNA and elongate the poly A tail in the presence of poly (A)-binding protein II (PAB II). Therefore, TbCPSF 30 could be involved in the transcription.

Protein synthesis occurs in the cytoplasm, but many proteins are required in the nucleus and have to be imported. The splicing process which occurs in the nucleus requires recruitment of spliceosome complex proteins. Marchetti *et al.* (2000) demonstrated the presence of an energy dependent nuclear import system in trypanosomes. The signalp domain found in most of the trans-splicing factors could be involved in directing the transportation of these proteins across the trypanosome nucleus membrane. The nuclear import process depends on nuclear localization signals (NLS) present only in nuclear proteins and can be either signal sequences or signal patches (Görlich, 1998; Moore, 1998). The signalp domain could therefore be a signal sequence or patch that directs importation into the nucleus, by nuclear import receptors. Each type of receptor protein is specialized for the transport of a group of nuclear proteins sharing structurally similar nuclear localization signals (Smith and Raikhel, 1999).

4.10.3 *T. brucei* U2 auxiliary factor heterodimer (TbU2AF)

The U2AF heterodimer consists of 65-kDa (U2AF^{65}) and 35-kDa (U2AF^{35}) subunits (Ito *et al.*, 1999; Shepard *et al.*, 2002) that form an intimate heterodimeric complex in which the RRM of U2AF^{35} and a central polyproline segment of U2AF^{65} interact via reciprocal "tongue in groove" tryptophan residues (Vàzquez *et al.*, 2003). Three-dimensional structure of this interaction is mallet-shaped (Figure 24). In the "tongue in groove" interaction, the U2AF^{65} peptide forms a type II proline helix that wraps tightly around the globular U2AF^{35} domain at the head, while one long U2AF^{35} α helix escapes the molecular embrace to form the handle (Kielkopf *et al.*, 2001). The proline residues (forming the "groove") enclose the hU2AF^{35} tryptophan 134 (hU2AF^{35} Trp134) in a tight molecular embrace of aromatic/aliphatic interactions (Kielkopf *et al.*, 2001). In *T. brucei,* the non-polar and aromatic U2AF^{35} Trp134 observed in human is replaced with a polar and basic U2AF^{35} Lys122. However, the residues Thr 45, Leu 47 and Tyr 114 of U2AF^{35} RRM known to be directly involved in RNA recognition are conserved.

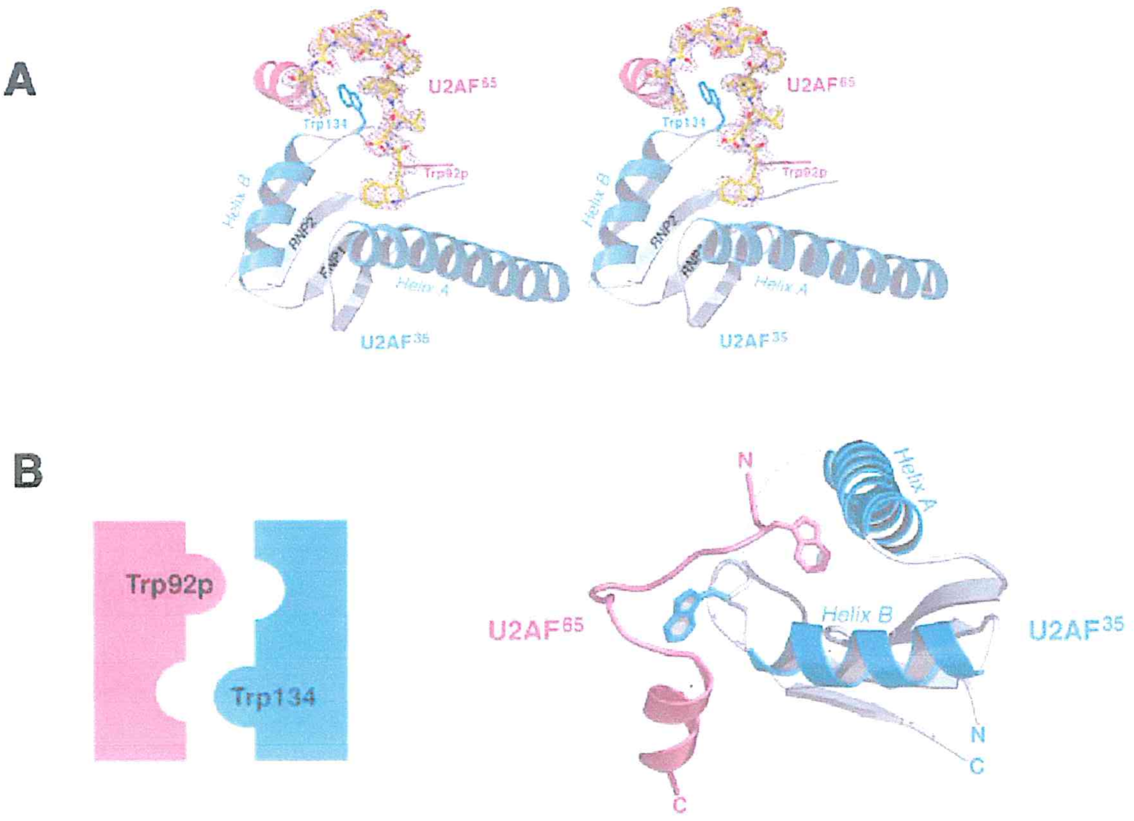


Figure 24. Schematic representation of human U2AF heterodimer. **(A)** Stereoview of the hU2AF^{65} proline-rich loop enveloping hU2AF^{35}Trp134. The hU2AF^{65} peptide is shown in pink, or color-coded for atom type (yellow, carbon; red, oxygen; and blue, nitrogen) and the U2AF^{35} domain is shown in light blue. **(B)** A schematic representation of the reciprocal "tongue in groove" and tryptophan binding sites. (Adopted from Kielkopf *et al.*, 2001)

The residues important in "groove" formation in hU2AF^{65} are Trp92, Pro95, Pro96, Pro97, Phe99, Pro104, and Try107 with a mandatory six residue spacer between the conserved hU2AF^{65}Pro97 and hU2AF^{65}Pro104 (Kielkopf *et al.*, 2001). Comparison of hU2AF^{65} and TbU2AF^{65} showed two sets of proline residue pairs with a six residue spacer at TbU2AF^{65}Pro162 and TbU2AF^{65}Pro169, and TbU2AF^{65}Pro213 and TbU2AF^{65}Pro220. Conservation of these proline residues is suspected to be high. The role played by hU2AF^{65}Pro97 and hU2AF^{65}Pro104 might be undertaken by either TbU2AF^{65}Pro162 and TbU2AF^{65}Pro169 or TbU2AF^{65}Pro213 and TbU2AF^{65}Pro220 in TbU2AF^{65}. In the sequence alignment between all the kinetoplastids and human U2AF^{65} homologs, the amino acid residue at similar position to hU2AF^{65}Pro95 in *T. brucei* is an aspartic acid residue (TbU2AF^{65}Asp128). Interestingly, at position 127 of TbU2AF^{65}, there is a proline residue (TbU2AF^{65}Pro127). The TbU2AF^{65}Pro127 could be implicated in a similar role played by hU2AF^{65} Pro95. No inference could be drawn on the 83 and 38 residue spacers unless the proline residues implicated in interaction with TbU2AF^{35} are determined. Despite these variations, the residues in TbU2AF^{65} could be involved in formation of the TbU2AF^{65} "groove" that encloses TbU2AF^{35}Lys122 in a tight molecular aliphatic interaction similar to hU2AF heterodimer interaction suggested by Kielkopf *et al.* (2001).

The variations between hU2AF^{65} and TbU2AF^{65} could be due to the fact that the U2AF^{35} residue actively involved in interaction with residues of U2AF^{65} polyproline "groove" has switched from hU2AF^{35}Trp134 to TbU2AF^{35}Lys122. These changes may be implicated in positioning amino acid residues in appropriate positions allowing the formation of a "groove" and stable interactions with the TbU2AF^{35}Lys122. If these variations are experimentally proven by mutational analysis to assess the residues that are important in molecular recognition hence interaction, it can demonstrate an example of host and parasite specific protein-protein interaction in the spliceosome complex. Such fundamental differences could be exploited as drug targets by designing specific disruptors.

4.10.4 Gene knock down

Cell death observed in cell lines expressing recombinant CPSF 30, U1-70k and P14 suggests the importance of these factors in trypanosome viability. TbCPSF 30 depletion was lethal and no morphological changes were evident similar to findings by Hendriks *et al.* (2003). However, Hendriks and co-workers recorded complete cell death after 5-6 days as compared to 10 days observed in this study. This could be attributed to the fact that a different cell line was used.

TbP14 depletion resulting into incomplete division in cells and death infers the importance of this factor in viability. This may be due to the fact that P14 is positioned within the inner cage of the SF3b structure (Golas *et al.*, 2003). The depletion could result into significant disruption of SF3b consequently affecting the trans-spliceosome complex and trans-splicing process. The abnormal morphological shapes in transgenic P14 and U1-70k cells could be due to disruption of expression of genes that are involved in maintenance of shape. Abnormal shapes have been documented in suppression of clathrin heavy chain (Allen *et al.*, 2003) in which enlargement of the flagella pocket resulted into a `BigEye` phenotype. The appearance of FAT cells (Ngô *et al.*, 1998; Inoue *et al.*, 2002) suggests effects on expression of α-tubulin. This can be due to disruption of processing of α-tubulin pre-mRNA. There were difficulties in monitoring growth kinetics since the cells formed clumps. The clumping is characteristic of dying trypanosomes. However, the degree of clumping was unusual. The reasons for the extensive clumping were unknown.

The RNAi experiments highlighted the importance of trans-spliceosome and polyadenylation factors in trypanosomes. However, the levels of RNA and proteins should be determined to conclusively determine the effects of specific silencing. This was not achieved in this study because clumping of cells precluded accurate monitoring of cell numbers. Investigations on protein-protein interactions in trypanosome trans-spliceosome could be done using tandem affinity purification (TAP) tagging (Rigaut *et al.*, 1999; Puig *et al.*, 2001; Knuesel *et al.*, 2003; Gould *et al.*, 2004). This could offer insights into the nature of interacting pairs of proteins. Small molecules that disrupt such interactions are potential drugs.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

Trypanosomosis is a widespread constraint on human health, livestock production and mixed farming in tropical Africa. The disease has denied Africa her agricultural potential and causes considerable mortality and morbidity in endemic foci. The main control strategies applied (vector control and chemotherapy) have varied limitations. In recent years, increased incidents of resistance to available drugs have limited their value. Continued use of the few available drugs may result in multiple resistance that could make chemotherapy increasingly ineffectual, though it remains the mainstay for control of parasitic diseases (Verlinde *et al.,* 2001). It currently protects more cattle than all other techniques combined (FAO, 2002), can be used under any production system, in any ecological zone and allows farmers to individually apply trypanocides to their cattle. The limited number of available drugs is a consequence of market economy principles: since people most at risk from tropical diseases are among the poorest in the world, there is currently little perceived financial incentive for pharmaceutical companies to invest in development of new drugs. Trypanosomosis has therefore remained among the most neglected tropical diseases (Morel, 2003) in terms of drug development. Re-emergence of the disease (Legros *et al.,* 2002) will worsen the delicate situation. This justifies the investigation of trypanosome specific process such as regulation of gene expression as a potential chemotherapeutic target.

Gene regulation is vital to cell viability, growth and development in all eukaryotic organisms. One aspect of gene regulation is the requirement for pre-mRNA processing through splicing into mature RNA species. This process is particularly important in regulation of gene expression in trypanosomes whose genes are organized in polycistronic transcription units. This occurs via trans-splicing, specific to trypanosomes but absent in human and mammalian hosts.

The identification and functional analysis of the trans-splicing and polyadenylation factors in *T. brucei* is an important contribution to understanding the trans-spliceosome as a potential drug target. The low percentage identity and similarity between the *T. brucei* trans-splicing and polyadenylation factors and those of human and suspected difference in protein-protein interactions, defines the variations in the process of RNA maturation. Additionally, the long evolutionary distance between trypanosomatids and their mammalian hosts (Verlinde *et al.*, 2001) endows the trans-splicing and polyadenylation factors with distinct properties. For optimum utilization of these findings, I would recommend further studies aimed at generating exhaustive information that would be exploited in development of disruptors specific to parasite trans-splicing process. RNAi technology could be used as a tool to analyse the genes for validation as potential drug targets during such studies. Interacting factors whose silencing is lethal to the parasite should be adequately characterized and amino acid residues involved in molecular recognition determined. This should include auxiliary factor (AF^{35} and AF^{65}), SF3b 145, SF3b 49, CPSF 30, U170k and P14 as well as other factors that could be important in viability. Successful undertaking of the above recommendations would improve chemotherapeutic control not only to trypanosomosis, but also to other diseases caused by parasites that exhibit trans-splicing such as leishmania, *T. cruzi*, trematode infections caused by schistosomes and nematode infections caused by filaria. This would enhance realization of Africa's optimum agricultural potential that would in turn support her economically disadvantaged inhabitants.

REFERENCES

Abu-Elneel, K., Kapeller, I. and Shlomai, J. (1999) Universal minicircle sequence binding protein, a sequence-specific DNA binding protein that recognizes the two replication origins of the kinetoplast DNA minicircle. *J. Biol. Chem.* **274**, 13419-13429.

Aksoy, S. (2003) Control of tsetse flies and trypanosomes using molecular genetics. *Vet. Parasitol.* **115**, 125-145.

Aksoy, S., Maudlin, I., Dale, C., Robinson, A. and O`Neill, S. (2001) Prospects for control of African trypanosomosis by tsetse vector manipulation. *Trends Parasitol.* **17**, 29-35.

Alani, E., Lee, Y. J., Schofield, J. M., Kijas, W. A., Hsieh, P. and Yang, W. (2003) Crystal structure and biochemical analysis of the MutS·ADP·beryllium fluoride complex suggests a conserved mechanism for ATP interactions in mismatch repair. *J. Biol. Chem.* **278**, 16088-16094.

Alibu, V., Storm, L., Haile, S., Clayton, C. and Horn, D. (2005) A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **139**, 75-82.

Allen, C., Goulding, D. and Field, M. (2003) Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. *EMBO J.* **22**, 4991-5002.

Allsopp, R. (2001) Options for vector control against trypanosomosis in Africa. *Trends Parasitol.* **17**, 15-19.

Altschul, S., Gish, W., Miller, W., Myers, W. and Lipman, J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Apweiler, R., Attwood K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N., Oinn, M., Pagni, M., Servant, F., Sigrist, J., Zdobnov, M. and InterPro Consortium (2000) InterPro- an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* **12**, 1145-1150.

Arnaud, K. and The Sanger Institute Pathogen Sequencing Unit. (2002) GeneDB: a repository for genome resource. *JOBIM* 185.

Ayadi, L., Callebaut, I., Saguez, C., Villa, T., Mornon, J. and Banroques, J. (1998) Functional and structural characterization of the Prp3 binding domain of the yeast Prp4 splicing factor. *J. Mol. Biochem.* **284**, 673-687.

Bairoch, A. (1991) SEQANALREF: a sequence analysis bibliographic reference databank. *Comput. Appl. Biosci.* **2**, 286.

Bakshi, R. and Shapiro, T. (2004) RNA interference of *Trypanosoma brucei* topoisomerase IB: both subunits are essential. *Mol. Biochem. Parasitol.* **136**, 249-255.

Balaban, N., Waithaka, H. K., Njogu, A. R. and Goldman, R. (1995) Intracellular antigens (microtubule-associated protein copurified with glycosomal enzymes) - possible vaccines against trypanosomiasis. *J. Infect. Dis.* **172**, 845-850.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D. and Sonnhammer, E. L. (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27**, 260-262.

Bateman, A., Coin, L., Durbin, R., Finn, R., Hollich, V., Griffith-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E., Studholme, D., Yeats, C. and Eddy, S. (2004) The Pfam protein families database. *Nucleic Acid Res.* (Database issue) **32**, D138-D141.

Baulcombe, D. (1999) Gene silencing: RNA makes RNA makes no protein. *Curr. Biol.* **9**, R599-R601.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, C. D., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T., Churcher, C., Clark, N. L., Corton, H. C. I., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., MacLeod, A., Mooney, P. J., Moule, S., Martin, D. M. A., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M. R., Tait, A., Tivey, A. R., Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E. and El-Sayed, N. M. (2005) The Genome of the African trypanosome *Trypanosoma brucei. Science* **309**, 416-422.

Blackwell, J. L., Martik, D., P. Bjornson, K., Bjornson, S. E. and Modrich, P. (1998) Nucleotide-promoted release of hMutSα from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. *J. Biol. Chem.* **273**, 32055-32062.

Breitbart, E., Andreadis, A. and Nadal-Ginard, B. (1987) Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* **56**, 467-495.

Bringaud, F., Vedrenne, C., Cuvillier, A., Parzy, D., Baltz, D., Tetaud, E., Pays, E., Venegas, J., Merlin, G. and Baltz, T. (1998) Conserved organization of genes in trypanosomatids. *Mol. Biochem. Parasitol.* **94**, 249–264.

Brown, J., Simpson, C., Thaw, G., Clark, G., Jennings, S., Medina-Escobar, N., Haupt, S., Chapman, S. and Oparka, K. (2002) Splicing signals and factors in plant intron removal. *Biochem. Soc. Transact.* **30**, 146-149.

Brun, R. and Schönenburger, M. (1979). Cultivation and *in vitro* cloning of procyclic forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Trop.* **36**, 289–292.

Caplen, N., Zheng, Z., Falgout, B. and Morgan, R. (2002) Inhibition of viral gene expression and replication in mosquito cells by ds-RNA triggered RNA interference. *Mol. Ther.* **6**, 243-251.

Caruthers, J. and McKay, D. (2002) Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123-133.

Chan, S., Kao, D., Tsai, W. and Cheng, S. (2003) The Prp19p-associated complex in spliceosome activation. *Science* **302**, 279-282.

Chen, Y., Hung, C., Burderer, T. and Lee, G. (2002) Development of RNA interference revertants in *Trypanosoma brucei* cell lines generated with a double stranded DNA expression construct driven by two opposing promoters. *Mol. Biochem. Parasitol.* **126**, 275-279.

Chi, J., Chang, H., Wang, N., Chang, D., Dunphy, N. and Brown, P. (2003) Genomewide view of gene silencing by small interfering RNAs. *Proc. Natl. Acad. Sci.* U.S.A. **100**, 6343-6346.

Chiara, M.D., Gozani, O., Bennett, M., Champion, A.P., Palandjian, L. and Reed, R. (1996) Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol. Cell Biol.* **16**, 3317-3326.

Clayton E. C. (2002) Life without transcriptional control? From fly to man and back again. *EMBO J.* **21**, 1881-1888.

Cordin, O., Tanner, N. K., Doere, M., Linder, P. and Banroques, J. (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.* **23**, 2478-87.

DaRocha, W., Otsu, K., Teixeira, S. and Donelson, J. (2004) Test of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **133**, 175-186.

de la Cruz, J., Kressler, D. and Linder, P. (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem. Sci.* **24**, 192-198.

Denker, A., Zuckerman, M., Moroney, A. and Nilsen, N. (2002) New components of the spliced leader RNP required for nematode trans-splicing. *Nature* **417**, 667-67.

Department For International Development (DFID) (2001) Trypanosomosis, Tsetse and Africa. The year 2001 report.

Dillin, E. (2003) The specifics of small interfering RNA specificity. *Proc. Natl. Acad. Sci.* U.S.A. **100**, 6289-6291.

Donelson, J. (2002) Antigenic variation and the African trypanosome genome. *Acta Trop.* **85**, 391-404.

Dykxhoorn, D., Novina, C. and Sharp, P. (2003) Killing the messenger: short RNAs that silence gene expression. *Nature* **4**, 457-467.

EANETT (2003) Proceedings 5th Annual EANETT workshop. <www.eanett.org> (Accessed 12.6.2003)

El-Sayed, N. M., Myler, J. P., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Åslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. and Andersson, B. (2005a) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409–415.

El-Sayed, N. M., Myler P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D. and Hall, N. (2005b) Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**,404-409.

Estévez, A., Lehner, B., Sanderson, C., Ruppert, T. and Clayton, C. (2003) The roles of intersubunit interactions in exosome stability. *J. Biol. Chem.* **37**, 34943-34951.

Evans, D., Perez, I., MacMorris, M., Leake, D., Wilusz, C. and Blumenthal, T. (2001) A complex containing CstF-64 and the SL2 snRNP connects mRNA 3' end formation and trans-splicing in *C. elegans* operons. *Genes Dev.* **15**, 2562-2571.

Fairlamb, A. H. and Bowman, I. B. (1980) Uptake of the trypanocidal drug suramin by bloodstream forms of *Trypanosoma brucei* and its effect on respiration and growth rate *in vivo*. *Mol. Biochem. Parasitol.* **1**, 315-333.

Fairlamb, A.H. (2003) Chemotherapy of human African trypanosomosis: current and future prospects. *Trends Parasitol.* **19**, 488-494.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.* **30**, 235-238.

FAO (2002) The state of food and Agriculture. Agriculture and global public goods ten years after the Earth Summit.

Fire, A., Xu, S., Montgomery, M., Kostas, S., Driver, S. and Mello, C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.

Fleckner, J., Zhang, M., Valcarcel, J. and Green M., R. (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**, 1864-1872.

Förch, P., Merendino, L., Martínez, C. and Valcárcel, J. (2003) U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF^{65}, can promote U1 snRNP recruitment to 5′ splice sites. *Biochem. J.* **372**, 235–240.

Furuyama, S. and Bruzik, J. (2002) Multiple roles for SR proteins in trans-splicing. *Mol. Cell Biol.* **22**, 5337-5346.

Garcia-Blanco, A. M. (2003) Messenger RNA reprogramming by spliceosome-mediated RNA trans-splicing. *J. Clin. Invest.* **112**, 474-480.

Gascard, P., Lee, G., Coulombel, L., Auffray, I., Lum, M., Parra, M., Conboy, J.G., Mohandas, N., and Chasis, J.A. (1998) Characterization of multiple isoforms of protein 4.1R expressed during erythroid terminal differentiation. *Blood* **92**, 4404-4414.

Golas, M. M., Sander, B., Will, L. C., Lührmann, R. and Stark, H. (2003) Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**, 981-984.

Goncharov, I., Palfi, Z., Binderreif, A. and Michaeli, S. (1999) Purification of spliced leader ribonucleoprotein particle from *Leptomonas collosoma* revealed the existence of an Sm protein in trypanosomes. *J. Biol. Chem.* **274**, 12217-12221.

Gopal, S., Cross, G. and Gaasterland, T. (2003) An organism-specific method to rank predicted coding regions in *Trypanosoma brucei*. *Nucleic Acid Res.* **20**, 5877-5885.

Gould, K. L., Ren, L., Feoktistova, A. S., Jennings, J. L. and Link, A. J. (2004) Tandem affinity purification and identification of protein complex components. *Methods* **3**, 239-244.

Gordon, P., Sontheimer, E. and Piccirilli, J. (2000) Metal ion catalysis during exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA* **6**, 199-205.

Görlich, D. (1998) Transport into and out of the cell nucleus. *EMBO J.* **17,** 2721–2727.

Gozani, O., Potashkin, J. and Reed, R. (1998) A Potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell Biol.* **18**, 4752-4760.

Graveley, B. (2000) Sorting out the complexities of SR protein functions. *RNA* **6**, 1197-1211.

Gull, K. (2000) The biology of kinetoplastid parasites: insights and challenges from genomics to post-genomics. *Mol. Biochem. Parasitol.* **31**, 443-452.

Gull, K. (2002) The cell biology of parasitism in *Trypanosoma brucei*: insights and drug targets from genomic approaches? *Curr. Pharm. Des.* **4**, 241-256.

Gull, K., Ersfeld, K. and Wickstead, B. (2002) Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei. Mol. Biochem. Parasitol.* **125**, 211-216.

Guth, S. and Valcárcel, J. (2000) Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. *J. Biol. Chem.* **275**, 38059-38066.

Haag, J., O'hUigin, C. and Overath, P. (1998) The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. *Mol. Biochem. Parasitol.* **91**, 37-49.

Hastings, L. M. and Krainer, R. A. (2001) Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13,** 302–309

Hendriks, E., Abdul-Razak, A. and Matthews, K. (2003) tbCPSF30 Depletion by RNA interference disrupts polycistronic RNA processing in *Trypanosoma brucei. J. Biol. Chem.* **278**, 26870-26878.

Hertz-Fowler, C., Peacock, C., Wood, V., Aslett, M., Kerhornou, A., Mooney, P., Tivey, A., Berriman, M., Hall, N., Rutherford, K., Parkhill, J., Ivens, A., Rajandream, M. and Barrell, B. (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.* **32,** D339-D343.

Hug, M., Hotz, R., Hartmann, C. and Clayton, C. (1994) Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. *Mol. Cell Biol.* **11**, 7428-7435.

Hulo, N., Sigrist, A., Le Saux, V., Langendijk-Genevaux, S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P. and Bairoch, A. (2004) Recent improvements to the PROSITE database. *Nucleic Acids Res.* **32**, D134 -D137.

Igel, H., Wells, S., Perriman, R., and Ares, M. (1998) Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA.* **4**, 1–10.

Inoue, N., Otsu, K., Ferraro, D. and Donelson, J. (2002) Tetracycline-regulated RNA interference in *Trypanosoma congolense. Mol. Biochem. Parasitol.* **120**, 309-313.

Ismaili, N., Perez-Morga, D., Walsh, P., Cadogan, M., Pays, A., Tebabi, P. and Pays, E. (2000) Characterization of a *Trypanosoma brucei* SR domain-containing protein bearing homology to cis-spliceosomal U170kDa proteins. *Mol. Biochem. Parasitol.* **106**,109-120.

Ito, T., Muto, Y., Green, M. and Yokoyama, S. (1999) Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF^{65}). *EMBO J.* **18**, 4523-4534.

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

Jones, S., de Souza, P. and Lindsay, M. (2004) siRNA for gene silencing: a route to drug target discovery. *Curr. Opin. Pharm.* **4**, 522-527.

Jurica, S. and Moore, J. M. (2003) Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell* **12**, 5-14.

Kabayo, J. (2002) Aiming to eliminate tsetse from Africa. *Trends Parasitol.* **11**, 473-475.

Kennerdell, R. and Carthew, W. (1998). Use of dsRNA mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95,** 1017–1026.

Kielkopf, C., Rodionova, N., Green, M. and Burley, S. (2001) A novel peptide recognition mode revealed by the x-ray structure of a core U2AF^{35}/U2AF^{65} heterodimer. *Cell* **106**, 595-606.

Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W. and Liu, X. (2003) Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. *Mol. Cell Proteomics* **11**,1225-1233

Krishna, S., Majumdar, I., and Grishin, N. (2003) Structural classification of zinc fingers: Survey and Summary. *Nucleic Acids Res.* **31**, 532-550.

Kruskal, J. B. (1983) An overview of sequence comparison In D. Sankoff and J. B. Kruskal, (ed.), Time warps, string edits and macromolecules: the theory and practice of sequence comparison, pp. 1-44 Addison Wesley.

Kubata, K., Nagamune, K., Murakami, N., Merkel, P., Kabututu, Z., Martin, S., Kalulu, T., Mustakuk, H., Yoshida, M., Ohnishi-Kameyama, M., Kinoshita, T., Duszenko, M. and Urade, Y. (2005) *Kola acuminata* proanthocyanidins: a class of anti-trypanosomal compounds effective against *Trypanosoma brucei*. *Int. J. Parasitol.* **35**, 91–103.

Kioy, D. and Mattock, N. (2005) Control of sleeping sickness–time to integrate approaches. *The Lancet.* **366**, 695-696.

LaCount, D., Barrett, B. and Donelson, E. (2002) *Trypanosoma brucei FLA1* is required for flagellum attachment and cytokinesis. *J. Biol. Chem.* **277**, 17580-17588.

Lake, J., De La Cruz, V., Ferreira, P., Morels, C. and Larry Simpson, L. (1988) Evolution of parasitism: Kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. *Proc. Natl. Acad. Sci.* U.S.A. **85**, 4779-4783.

Lamers, M. H., Georgijevic, D., Lebbink, J. H., Winterwerp, H. H., Agianian, B., De Wind, N. and Sixma, T. K. (2004) ATP increases the affinity between MutS ATPase domains; implications for ATP hydrolysis and conformational changes. *J. Biol. Chem.* **279**, 43879-43885.

Landfear, S. (2003) Trypanomastid transcription factors: Waiting for Godot. *Proc. Natl. Acad. Sci.* U.S.A. **100**, 7-9.

Lanham, S. M and Godfrey, D. G. (1970) Isolation of Salivarian trypanosomes from man and other mammals using DEAE-Cellulose. *Exper. Parasitol.* **28**, 521-534.

LeBowitz, J., Smith, H., Rusche, L. and Beverley, S. (1993) Coupling of poly (A) site selection and trans-splicing in *Leishmania. Genes Dev.* **7**, 996-1007.

Lee, C. and Irizarry, K. (2003) Alternative splicing in the nervous system: an emerging source of diversity and regulation. *Biol. Psychiatry* **54**, 771-776.

Legros, D., Ollivier, G., Gastellu-Etchegorry, M., Paquet, C., Burri, C., Jannin, J. and Büscher, P. (2002) Treatment of human African trypanosomiasis-present situation and needs for research and development. *Lancet Infect Dis* **2**, 437–440.

Leipe, D. Wolf, Y., Koonin, E. and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biochem.* **317**, 41-72.

Li, D. and Roberts, R. (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell Mol. Life Sci.* **58**, 2085–2097.

Li, L., Otakes, L., Xu, Y. and Michael, S. (2000) The trans-spliceosomal U4 RNA from the monogenetic trypanosomastid *Leptomonas collosoma. J. Biol. Chem.* **275**, 2259-2264.

Liang, X., Haritan, A., Uleil, S. and Michaeli, S. (2003) trans and cis splicing in trypanosomatids: Mechanism, factors and regulation. *Eukary. Cell* **2**, 830-840.

Lohmann, J., Endl, I. and Bosch, T. (1999) Silencing of developmental genes in *Hydra. Dev. Biol.* **214**, 211–214.

Lubega, G., Byarugaba, D. and Prichard, R. (2002) Immunization with a tubulin-rich preparation from *Trypanosoma brucei* confers broad protection against African trypanosomosis. *Exper. Parasitol.* **102**, 9-22.

Lücke, S., Klöcker, T., Palfi, Z., Boshart, M. and Bindereif, A. (1997) Trans mRNA splicing in trypanosomes: cloning and analysis of a PRP8-homologous gene from *Trypanosoma brucei* provides evidence for a U5 analogous RNP. *EMBO J.* **16**, 4433-4440.

Madrona, Y. and Wilson, D. (2004) The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure. *Protein Science* **13**, 1557-1565.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, O., Bishop, J., Falcone, F., Gavrilescu, C., Montgomery, L., Santori, M., Stern, S., Wang, Z., Ullu, E. and Tschudi, C. (2000) A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* **6**, 163-169.

Mandelboim, M., Estraño C., Tschudi, C., Ullu, E. and Michaeli, S. (2002) On the role of exon and intron sequence in trans-splicing utilization and cap 4 modification of the Trypanosomastid *Leptomonas collosoma* SL RNA. *J. Biol. Chem.* **277**, 35210-35218.

Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 239-243.

Marchetti, A. M., Tschudi, C., Kwon, H., Wolin, L., S. and Ullu, E. (2000) Import of proteins into the trypanosome nucleus and their distribution at karyokinesis. *J. Cell Sci.* **113**, 899-906.

Matovu, E., Seebeck, T., Enyaru, J. and Kaminsky, R. (2001) Drug resistance in *Trypanosoma brucei spp.*, the causative agents of sleeping sickness on man and nagana in cattle. *Microbes Infect.* **3**, 763-770.

Matthews, K. R., Tschudi, C. and Ullu, E. (1994) A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* **8**, 491-501.

Milhavet, O., Gary, D. and Mattson, M. (2003) RNA interference in biology and medicine. *Pharm. Rev.* **55**, 629-648.

Misquitta, L. and Paterson, B. (1999) Targeted disruption of gene function in *Drosophila* by RNA interference (RNAi): a role for nautilus in embryonic somatic muscle formation. *Proc. Natl. Acad. Sci.* U.S.A. **96**, 1451-1456.

Moore, M. S. (1998) Ran and nuclear transport. *J. Biol. Chem.* **273**, 22857-22860.

Morel, C. (2003) Neglected diseases: under-funded research and inadequate health interventions. *EMBO reports* **4**, S35-S38.

Morris, J., Wang, Z., Drew, M. E., Paul, K. S. and Englund, P. T. (2001) Inhibition of bloodstream form *Trypanosoma brucei* gene expression by RNA interference using the pZJM dual T7 vector. *Mol. Biochem. Parasitol.* **117**, 111-113.

Motyka, S., Zhao, Z., Gull, K. and Englund, P. (2004) Intergration of pZJM library plasmids into unexpected locations in the *Trypanosoma brucei* genome. *Mol. Biochem. Parasitol.* **134**, 163-167.

Mulder, N., Apweiler, R., Attwood, K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchel, A., Nikolskaya, N., Orchard, S., Pagni, M., Ponting, P., Quevillon, E., Selengut, J., Sigrist, J., Silventoinen, V., Studholme, J., Vaughan, R. and Wu, C. (2005) InterPro, progress and status in 2005. *Nucleic Acid Res.* **33** (Database issue) D201-205.

Mwangi, E., Stevenson, P., Gettinby, G., Reid, S. and Murray, M. (1998) Susceptibility to trypanosomosis of the three *Bos indicus* cattle in areas of differing tsetse fly challenge. *Vet. Parasitol.* **79**, 1-17.

Needleman, S. B. and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443-453.

Ngô, H., Tschudi, C., Gull, K. and Ullu, E. (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci.* U.S.A. **95**, 14687-14692.

Nilsen, T. (2000) The case for an RNA enzyme. *Nature* **408**, 782-783.

Orlova, E. and Saibil, H. (2004) Structure determination of macromolecular assemblies by single-particle analysis of cryo-electron micrographs *Cur. Opin. Struct. Biol.* **14**, 584-590.

Padgett, R., Grabowski, P., Konarska, M., Seiler, S. and Sharp, P. (1986) Splicing of messenger RNA precursors. *Ann. Rev. Biochem.* **55**, 1119-1150.

Palfi, Z., Lücke, S., Lahm, H., Lane, W., Kruft, V., Bragado-Nilsson, E., Séphin, B. and Bindereif, A. (2000) The spliceosomal snRNP core complex of *Trypanosoma brucei*: cloning and functional analysis reveals seven Sm protein constituents. *Proc. Natl. Acad. Sci.* U.S.A. **97**, 8967-8972.

Palfi, Z., Lane, W. S. and Bindereif, A. (2002) Biochemical and functional characterization of the cis-spliceosomal U1 small nuclear RNP from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **121**, 233-243.

Pauls, E. and Esté, J. (2004). RNA interference as a tool for target validation. *Drug Disco. Today*: *Tech.* **3**, 1-6.

Philipps, D., Celotto, A., Wang, Q., Tarng, R. and Graveley, B. (2003) Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acid Res.* **31**, 6502-6508.

Pitula, J., Ruyechan, W. and Williams, N. (1998) *Trypanosoma brucei*: identification and purification of a poly (A) binding protein. *Exper. Parasitol.* **88**, 157-160.

Portugal, J. (1994) Berenil acts as a poison of eukaryotic topoisomerase II. *FEBS Let.* **344**, 136 - 138.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Séraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **3**, 218-29.

Reddy, A. (2004) Plant serine/arginine-rich proteins and the role in pre-mRNA splicing. *Trends Plant Sci.* **11**, 541-547.

Redmond, S., Vadivelu, J. and Field, M. (2003) RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **128**, 115-118.

Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the european molecular biology software suite. *Trends in Genetics* **16**, 276-277.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Séraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **10**, 1030-1032.

Roberts, G. and Smith, C. (2002) Alternative splicing, combinatorial output from the genome. *Curr. Opin. Chem. Biol.* **6**, 375-383.

Roberts, T., Dungan, J., Watkins, K. and Agabian, N. (1996) The SLA RNA gene of *Trypanosoma brucei* is organized in a tandem array which encodes several small RNAs. *Mol. Biochem. Parasitol.* **83**, 163-174.

Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sánchez-Alvarado, A and Mewmark, P. (1999) Double stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl. Acad. Sci.* U.S.A. **96**, 5049-5054.

Sanger, F., Nicklen, S. and Coulson, R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* U.S.A. **94**, 5463-5467.

Schofield, C. and Maudlin, I. (2001) Trypanosomiasis control. *Int. J. Parasitol.* **31**, 615-620.

Schofield, J. M., Nayak, S., Scott, H. T., Du, C. and Hsieh, P. (2001) Interaction of *Escherichia coli* MutS and MutL at a DNA Mismatch. *J. Biol. Chem.* **276**, 28291-28299

Schwer, B. and Guthrie, C. (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis . *EMBO J.* **11**, 5033-5039.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors AF^{65} and SF1/mBBP. *Mol. Cell* **11**, 965-976.

Shepard, J., Reick, M., Olson, S. and Graveley, B. (2002) Characterization of U2AF26, a splicing factor related to U2AF35. *Amer. Soc. Microbiol.* **22**, 223-230.

Shi, H., Cordin, O., Minder, C. M., Linder, P. and Xu, R. (2004) Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc. Natl. Acad. Sci.* U.S.A. **101**, 17628-17633.

Shi, H., Djikeng, A., Mark, T., Wirtz, E. Tschudi, C. and Ullu, E. (2000) Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *RNA* **6**, 1069-1076.

Smith, C. and Valcárcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *TIBS* **25**, 381-388.

Smith, M.S. H. and Raikhel, V. N. (1999) Protein targeting to the nuclear pore. What can we learn from plants? *Plant Physiol.* **119**,1157-1164.

Smith, T. F., Gaitatzes, C. G., Saxena, K. and Neer, E. J. (1999) The WD-repeat: A common architecture for diverse functions. *TIBS* **24**,181–185.

Sonnhammer, E., Eddy, S., Birney, E., Bateman, A. and Durbin, R. (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acid Res.* **26**, 320-322.

Sontheimer, E., Sun, S. and Piccirilli, J. A. (1997) Metal ions catalysis during splicing of pre-messenger RNA. *Nature* **388**, 801-805.

Stevens, J., Noyes, H. and Gibson, W. (1998) The evolution of trypanosomes infecting humans and primates. *Mem Inst. Oswaldo Cruz*, Rio de Janeiro **93**, 669-676.

Sturm, N. and Campbell, D. (1999) The role of intron structure in trans-splicing and cap 4 formation for *Leishmania* spliced leader RNA. *J. Biol. Chem.* **274**, 19361-19367.

Sutton, R. and Boothroyd, J. C. (1986) Evidence of trans splicing in trypanosomes. *Cell* **47**, 527-535.

Svoboda, P., Stein, P., Hayashi, H. and Schultz, R. (2000) Selective reduction of dormant maternal mRNAs in mouse oocyte by RNA interference. *Development* **127**, 4147-4156.

Swallow M., B. (1999). Impacts of trypanosomiasis on African agriculture. PAAT information resources.

Tanner N. K. and Linder, P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell.* **2**,251-62.

Thompson, D., Higgins, G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tschudi, C., Djikeng, A., Shi, H. and Ullu, E. (2003) *In vivo* analysis of the RNA interference mechanism in *Trypanosoma brucei*. *Methods* **30**, 304-312.

Tzfati, Y., Abeliovich, H., Avrahami, D. and Shlomai, J. (1995) Universal minicircle sequence binding protein, a CCHC-type zinc finger protein that binds the universal minicircle sequence of trypanomastids, purification and characterization. *J. Biol. Chem.* **270**, 21339-21345.

Van der Ploeg, L., Bernards, A., Rijsewijk, F. and Borst, P. (1982) Characterisation of the DNA duplication-transposition that control the expression of two genes for variant surface glyco- proteins in *Trypanosoma brucei*. *Nucleic Acids Res.* **10**, 593–609.

Varani, G. and Ramos, A. (2003) Splicing factor 1 in the pocket. *Structure* **11**, 481-487.

Vázquez, M., Atorrasagasti, C., Bercovich, N., Volcovich, R. and Mariano, J. (2003) Unique features of the *Trypanosoma cruzi* U2AF35 splicing factor. *Mol. Biochem. Parasitol.* **128**, 77-81.

Verlinde, C., Hannaert, V., Blonski, C., Willson, M., Périé,J., Fothergill-Gilmore, L., Opperdoes, F., Gelb, M., Hol, W. and Michel, M. (2001) Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resist. Updat.* **4**, 1–14.

Wang, Z., Morris, J., Drew, M. and Englund, P. (2000) Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an intergratable vector with opposing T7 promoters. *J. Biol. Chem.* **275**, 40174-40179.

Weits, B. (1970) Hosts of *Glossina*. The African Trypanosomiases (Ed. H. W. Mulligan), 317-326. Allen and Unwin, London.

Wells, C., Wilkes, J. and Peregrine, A.S. (1995) Proceedings of the 23rd ISCTRC meeting, p. 527. Banjul.

WHO (2000) World Health Report 2000 Health Systems Improving Performance Geneva. <http://www.who.int/whr/previous/en/index.html> (Accessed 20.6.2003)

WHO (2002) 55th World Health Assembly: Pan African Tsetse and Trypanosomosis Eradication. <http://www.who.int/whr/previous/en/index.html> (Accessed 20.6.2003)

WHO (2004) Strategic review of traps and targets for tsetse and African trypanosomosis control. <http://www.who.int/whr/previous/en/index.html> (Accessed 20.6.2003)

Will, C. L., Urlaub, H., Achsel, T., Gentzell, M., Wilm, M. and Lührmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.* **21**, 4978-4988.

Wirtz, E., Leal, S., Ochatt C. and Cross, G. (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.

Xu, Y., Newnham, M. C., Kameoka, S., Huang, T., Konarska M. M. and Query, C. (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J.* **23**, 376-385.

Yehuda, S., Dix, I., Russell, C., Levy, S., Beggs, J. and Kupiec, M. (1998) Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40yeast gene involved in splicing and cell cycle control. *RNA* **4**, 1304 – 1312.

Yu, L., Gaitatzes, C., Neer, E. and Smith, T. (2000) Thirty-plus functional families from a single motif. *Protein* **9**, 2470–2476.

Zhang, J. and Williams, N. (1997) Purification, cloning and expression of two closely related *Trypanosoma brucei* nucleic acid binding proteins. *Mol. Biochem. Parasitol.* **87**,145-158.

Zhao, J., Hyman, L. and Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **2**, 405-445.

APPENDIX 1

Accession numbers of spliceosome factors sequences

Spliceosome nucleotide and proteins were used and generated in data mining. The accession numbers are given for *T. brucei, L. major, T. cruzi* and *Homo sapiens* in that order. Those for *Saccharomyces cerevisiae* and *Caenorhabditis elegans* are given the initials sc and ce. (The accession numbers are temporary)

Data Mining

Factors	Accession number
	E complex
Zn1	TB05.28F8.690
	LmjF35.1040
	Tc00.1047053507305.40
Zn2	TB05.28F8.690
	LmjF35.1040
	Tc00.1047053507305.40
U170k	Tb10.70.4700
	LmjF34.0495
	Tc00.1047053504105.160
CFII-a1	Tc00.1047053507027.59
	LmjF30.2410
	Tb06.4F7.820
	Q92989
	sc YOR250C
AF35	Tb10.70.4300
	Tc00.1047053510943.60
	LmjF03.0190
	AAH01923
AF65	Tc00.1047053510265.40
	LmjF03.0520
	Tb10.70.3950
	CAA45409
	ce AAM44400
	sc AAA64215

Factors	Accession number
	A complex
P14	AAH15463
	Tb10.6k15.3190
	LmjF36.3040
	Tc00.1047053506885.70
	sc YIR005W
SF3b10	AAH00198
	Tb03.26J7.490
	LmjF29.1710
	Tc00.1047053509617.30
SF3b49	AAH04273
	Tb09.211.0930
	LmjF35.5040
	Tc00.1047053506885.70
	sc YOR319W
SF3b125	Tb10.70.0140
	LmjF36.2130
	Tc00.1047053510187.290
	AAH08208
SF3b145	Tb06.4M18.90
	LmjF30.0570
	Tc00.1047053504071.50
	AAH14125
	sc YMR240C
	Cleavage and polyadenylation factors
CstF50	Tb10.61.0570
	LmjF33.0280
	Tc00.1047053508899.30
	AAP36871
CPSF30	Tb11.01.4600
	LmjF09.0720
	Tc00.1047053510219.30
	EAL23878
	sc YPR107C

APPENIX 2

Buffers and solutions

DNA extraction

Cell lysis solution

Tris-Cl (pH8.0)	10mM
NaCl	100mM
EDTA	100mM
SDS	0.01%
Proteinase K	100ng/ml

Phosphate buffered saline (PBS) pH 7.3

NaCl	137 mM
KCl	2.7 mM
$Na_2HPO_4. 7H_2O$	4.3 mM
KH_2PO_4	1.4 mM

Bacterial culture media

SOC medium (per litre)

Tryptone	20g
Yeast extract	5g
NaCl	0.5g
Dissolve, then add:	
250mM KCl	10ml
2M $MgCl_2$	5ml
Autoclave,cool, then add:	
1M Glucose†	20ml

LB* medium

Tryptone	10g
Yeast extract	5g
NaCl	10g

LB* agar

Tryptone	10g
Yeast extract`	5g
NaCl	10g
Agar	15g

[†] Filter sterilize through a 0.2 µm filter

* Autoclave

Protein extraction and purification buffers

4X SDS Sample Buffer

Tris-HCl, (pH6.8)	250mM
Sodium dodecyl sulphate (SDS)	8 %
β-mercaptoethanol	4 %
Glycerol	40 %
Bromophenol blue	0.02 %

Lysis buffer (pH 8.0)

Urea	8M
NaH_2PO_4	0.1M
Tris-Cl	0.01M
NaCl	0.3M

Wash buffer (pH 6.3)

Urea	8M
NaH_2PO_4	0.1M
Tris-Cl	0.01M

Elution buffer (pH 4.5)

Urea	8M
NaH_2PO_4	0.1M
Tris-Cl	0.01M

Transfection medium

Zimmerman post-fusion medium (pH 7.0)

Nacl	132 mM
KCl	8 mM
Na_2HPO_4 (anhyd)	8 mM
KH_2PO_4 (anhyd)	1.5 mM
$MgOAC.4H_2O$	1.5 mM
$CaCl_2$	90µM

Western blot

Transfer Buffer (pH 8.3)

48mM Tris-Cl

29mM Glycine

20% methanol

0.037% SDS

Washing Buffer (pH 7.4)

25mM Tris-Cl

137mM NaCl

3mM KCl

0.3% Tween-20

Blocking Buffer (pH 7.4)

25mM Tris-Cl

137mM NaCl

3mM KCl

1% BSA /1% Blocking reagent (Boehringer Mannheim, GmbH, Germany)

Nitro Blue tetrazolium (NBT) (Sigma) Stock Solution

1tablet of NBT dissolved in 1ml distilled deionized water

5-Bromo-4-chloro-3-indolyl phosphate (BCIP) (Sigma) stock solution

1 tablet of BCIP in 100% N, N- dimethylformamide (Sigma).

Substrate Buffer (pH 9.5)

100mM Tris-Cl

100mM NaCl

5mM $MgCl_2$

Substrate solution

330μl NBT

33μlBCIP

10 ml of substrate buffer (pH 9.5)

TBS Buffer (pH 7.4)

25mM Tris-Cl

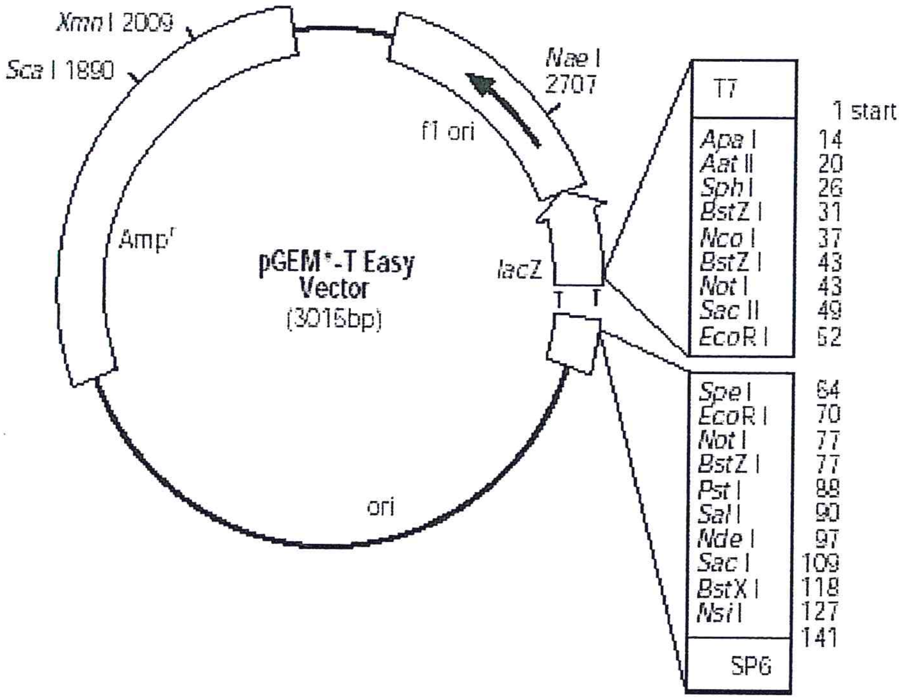
137mM NaCl

3mM KCl

APPENDIX 3

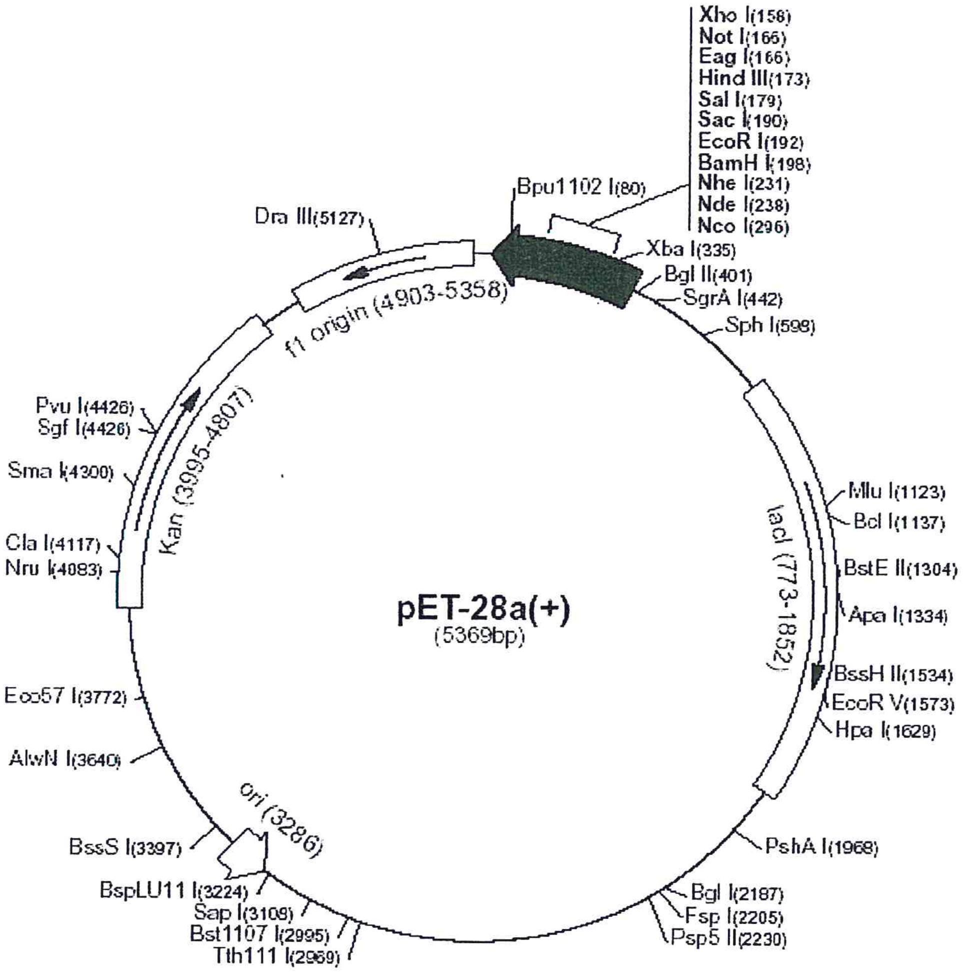
Vector maps

pGEM-T Easy Sequencing Vector



pGEM-T Easy Vector circle map (Promega Biotechnology)

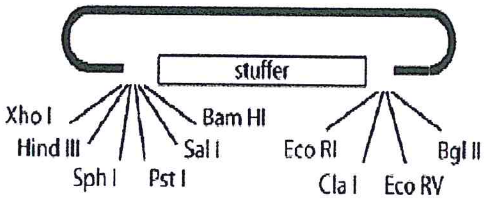
pET-28a(+) Expression Vectors



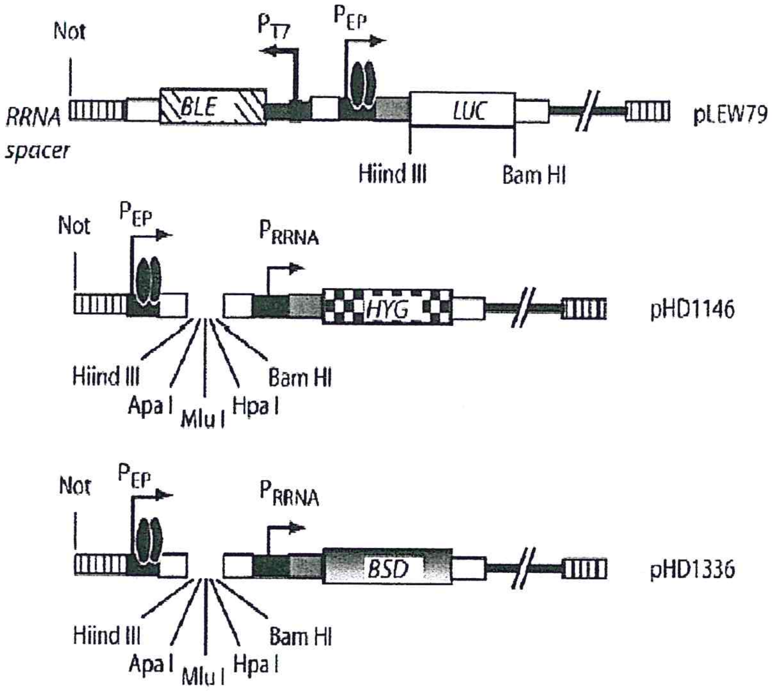
The pET-28a (+) vector map. It carries an N-terminal His•Tag/thrombin/T7•Tag configuration plus an optional C-terminal His•Tag sequence. Unique sites are shown on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown above.

RNAi VECTORS

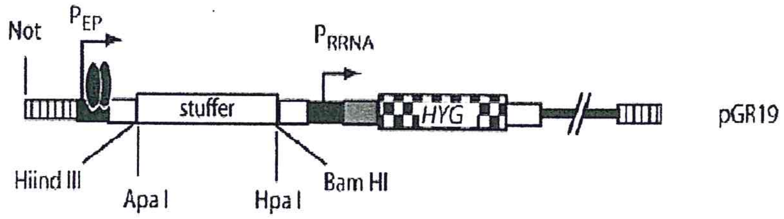
A. Vector for stem-loop cloning



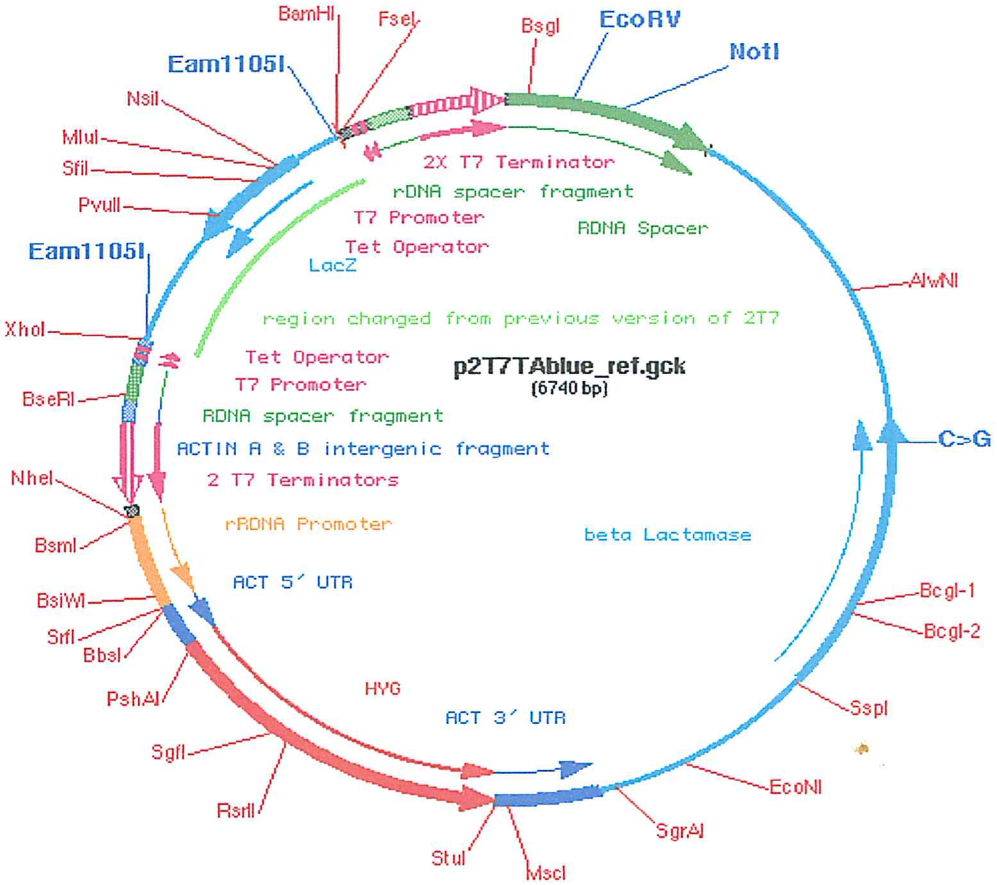
B. Inducible polymerase I vectors for insertion of ready-made stem-loops



C. Inducible polymerase I vector for direct cloning of stem-loops

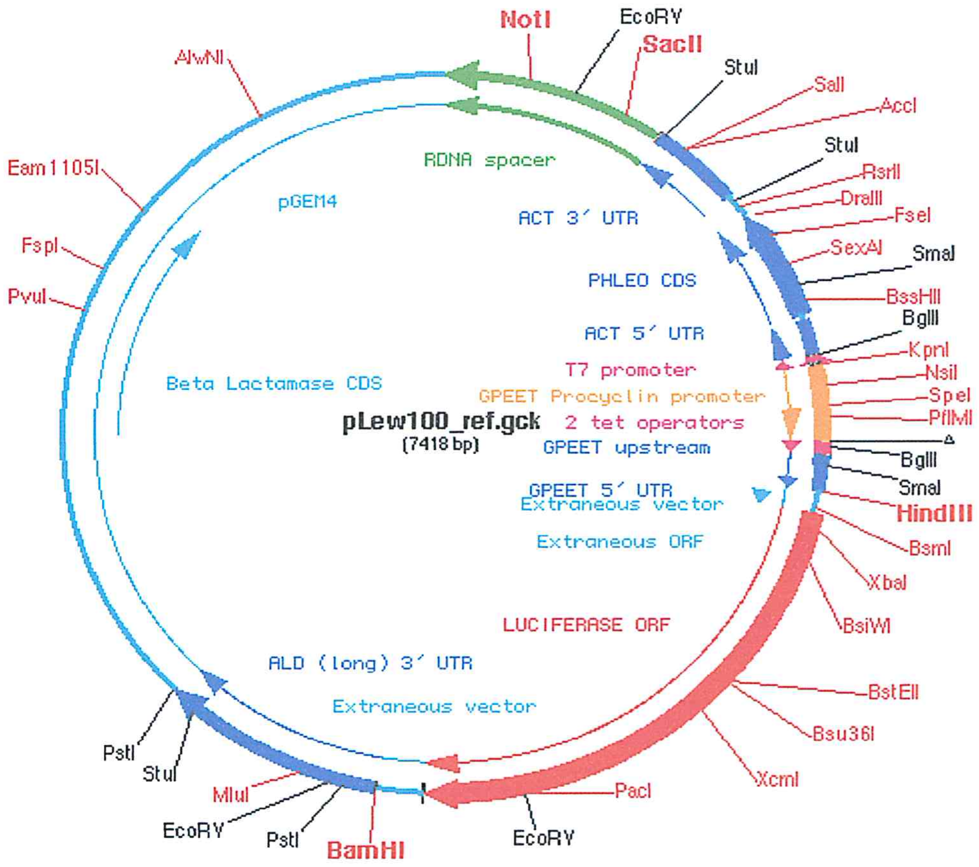


p2T7TAblue RNAi vector



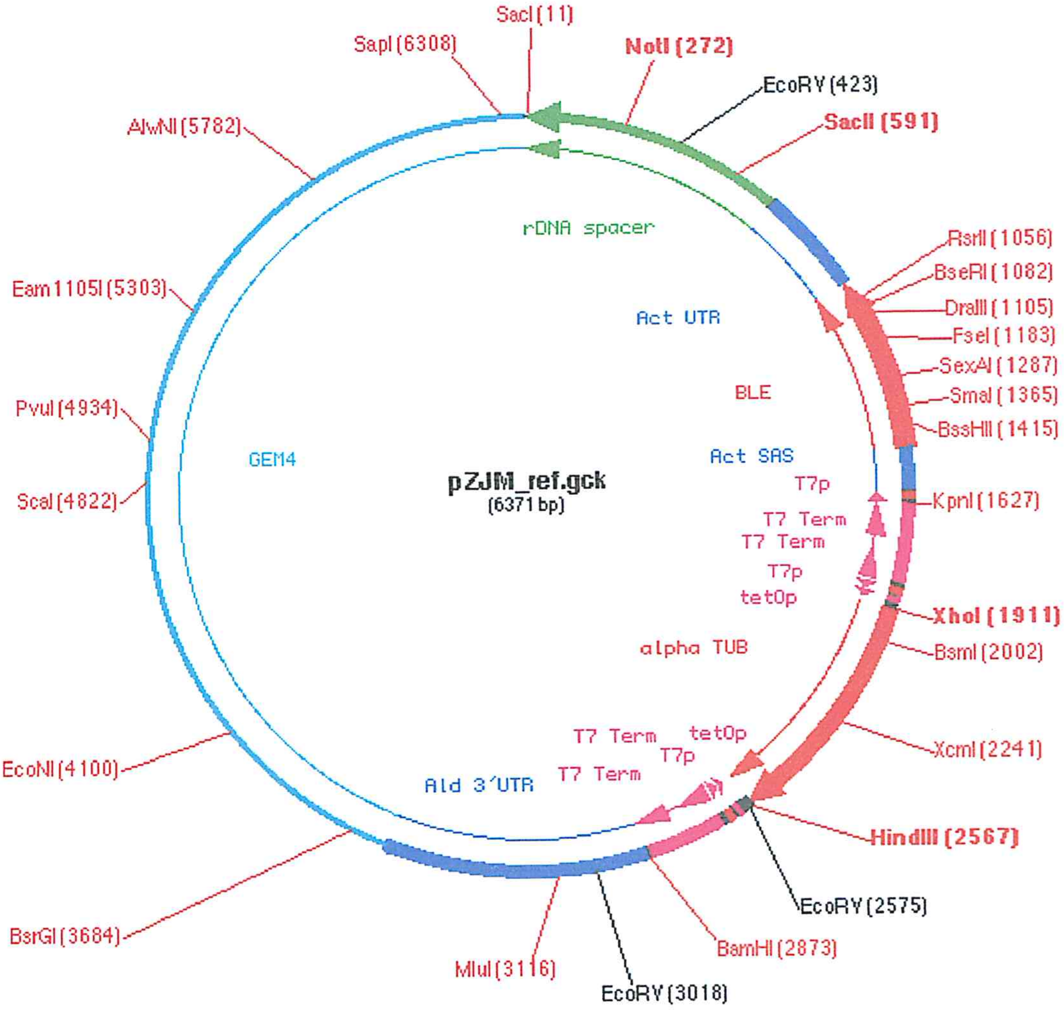
Source: http://tryps.rockefeller.edu/Seqs/Graphics/construct_images.pdf

pLEW 100



Source: http://tryps.rockefeller.edu/Seqs/Graphics/construct_images.pdf; Wirtz *et al.*, 1999

pZJM



Source: http://tryps.rockefeller.edu/Seqs/Graphics/construct_images.pdf