Cloning and Analysis of Some Trans-Splicing Factors in *Trypanosoma brucei*

Owino, V.[1,2], Limoh, M. K.[2] and Masiga, D.[1]
[1]Molecular Biology and Biotechnology Unit, International Centre of Insect Physiology and Ecology. P. O. Box 30772, Nairobi
[2]Department of Biochemistry and Molecular Biology, Egerton University, P. O. Box 536, Egerton
Corresponding Author E-mail vowino@gmail.com

Abstract

RNA maturation is an important process in gene expression in all eukaryotes. In trypanosomes, mature mRNA is derived from independent pre-mRNA molecules by a protein complex, spliceosome, in a process known as 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 the present study, we identified *in silico* 13 proteins of trypanosome spliceosome. 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 *Leishmania major*, *Trypanosoma cruzi* 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. Significant variations of the kinetoplastid sequences from human suggest that some components of the trypanosome spliceosome are targets for the design of novel drugs.

Key Words: *Trypanosoma brucei,* Kinetoplastid, Trans-Splicing, Spliceosome.

Introduction

The process of mRNA maturation in trypanosomes differs from that of most eukaryotes. The protein-coding genes are transcribed into polycistronic

RNAs rather than monocistronic transcription units (Sather and Agabian, 1985; Mair *et al.*, 2000; Mandelboim *et al.*, 2002) since they lack introns (Lücke *et al.*, 1997; Denker *et al.*, 2002; Liang *et al.*, 2003). In addition, the 5' ends of all mature mRNAs are contributed by independent RNA transcript called splice leader (SL) in a process called trans-splicing (Sutton and Boothroyd, 1986; Denker *et al.*, 2002; Garcia-Blanco, 2003). The process involves interaction between 5' and 3' splice sites on separate transcripts and occurs in a variety of eukaryotic organisms including 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 in trypanosomes` mammalian hosts. 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 undertaking 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).

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). Nematodes, trematodes and euglenoids however, carry out both trans- and cis-splicing. Lack of introns in trypanosomes has led to a notion that trypanosomes lack a machinery to carry out cis-splicing. This almost two-decade-old tenets that trypanosomes exhibit only trans-splicing has been refuted by a surprising report of cis-splicing in poly (A) polymerase (PAP) genes in *T. brucei* and *T. cruzi* (Mair *et al.*, 2000). 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). In this study, we sought to clone, sequence and analyse some trans-spliceosome factor and compare them to those of human host. This could give insight on trans-spliceosome as a potential drug target.

Materials and Methods

Trypanosoma brucei trans-splicing homologs were generated by BLAST searches at GeneDB (Hertz-Fowler *et al.*, 2004), the repository of genome data for the kinetoplastids *T. brucei*, *T. cruzi* and *Leishmania major*. *T. cruzi*, human, and/or yeast splicing factors were used to query the database. Primers were designed from the search results and subsequently used to recover the genes from genomic DNA of a *T. brucei rhodesiense* strain KETRI 3741 (MHOM/UG/72/KETRI 3741) by PCR. The full-length amplification products were purified directly using QIAquick Gel Extraction Kit (Qiagen, GmbH Germany) and subsequently cloned in pGEMT-Easy vector (pGEM-T® EASY vector Systems kit, Promega Corp., Madison, WI, U.S.A). The purified plasmid constructs with PCR inserts were sequenced in an ABI 3100 sequencer (Applied Biosystems, Foster City, CA, USA) using appropriate fluorescent labelled terminators.

Nucleotide sequences of cloned inserts were translated to protein using the translation tool at Swiss Bioinformatics Institute website - Expasy (Bairoch, 1991). The generated amino acid sequences were used to query *T. brucei* database at GeneDB to determine P-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), 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), prosite (Falquet *et al.*, 2002; Hulo *et al.*, 2004), MotifScan (Falquet *et al.*, 2002) and Pfam (Sonnhammer *et al.*, 1998; Bateman *et al.*, 1999; Bateman *et al.*, 2004).

Results

A clean dataset (not shown) of 13 *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*. 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 P-values from searches at geneDB using amino acid sequences showed successful cloning of the targets. Comparison of amino acid sequences from the cloned genes with those from *H. sapiens*, *T. cruzi* and *L. major* are shown in Table 1.

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. 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. 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. The tryptophan residue in $U2AF^{35}$ that interacts with the "groove" in $U2AF^{65}$ has been replaced with a lysine residue in all the kinetoplastids.

Table 1. Comparison of the amino acid sequences of cloned *T. brucei* factors and *H. sapiens*, *T.cruzi* and *L. major* sequences

	% Identity			% Similarity		
Factors	Hs	Tc	Lm	Hs	Tc	Lm
SF3b125	22.3	69.3	61.2	31.7	78.9	72.2
SF3b49	19.9	71.7	66.5	29.5	78.6	76.7
SF3b145	15.4	59.1	44.8	22.8	70.9	59.5
SF3b10	-	53.9	22.8	-	68.5	32.3
P14	29.1	79.2	40	46.3	87.5	50.3
AF35	31.1	80.6	63.7	49.3	88.7	72.2
AF65	18.3	52.8	30.8	27.7	61.7	43.4
CstF50	22.4	67.6	47.3	38.3	77.3	58.7
CFII-a1	26.6	57.5	36.3	40.9	75	54.4
CPSF30	30.6	83	47.4	39	88.8	59.4

Domains

The protein domains from the sequenced factors are shown in Table 2. Domain signatures were detected in WD, DEAD box, PSP, MutS and RRM domains. The signatures in RRM were RMM1, RRM2 and RMM3 in U2AF^{35} and U2AF^{65}.

Table 2. Protein domain in cloned trypanosome trans-spliceosome factors

Factor	Domains
AF^{35}	RRM, Zn Finger, Signalp
AF^{65}	RRM, Arginine –rich region
CstF50	WD domain, Signalp
CFII-a1	Pre-mRNA cleavage complex II protein Clpi, Signalp, GTPase, P-loop.
SF3b 125	DEAD box
SF3b 145	PSP proline rich, Signalp
SF3b 10	MutS domain III, MutS domain V, DNA binding domain for DNA mismatch repair, ATPase domain for DNA mismatch repair, Signalp.
SF3b 49	RRM, PABPh, PABP, Signalp.
P14	RRM, Signalp.
CPSF 30	Zn Finger.

RRM, RNA Recognition Motif; PABP, Poly (A) binding protein.

Discussion

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 P–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 represent 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 a 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).

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 (Ito *et al.*, 1999; Kielkopf *et al.*, 2001; Selenko *et al.*, 2003) 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 prote ins 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).

Conclusions and Recommendations

The analysis of some 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. 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, we 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.

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.

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 Research 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.

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

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 Research 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 Research (Database issue) 32:D138-D141.

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.

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.

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.

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., 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. 23:2478-2487.

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.

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.

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.

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 Research 30:235-238.

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.

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

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.

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.

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.

Hendriks, E., Abdulrazak, 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 Research 32:D339-D343.

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 Research 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.

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.

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

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.

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.

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. USA. 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.

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.

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.

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.

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

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.

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.

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.

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

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.

Sather, S. and Agabian, N. (1985). A 5' spliced leader is added in trans to both α- and ß -tubulin transcripts in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. U.S.A. 82:5695-5699.

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. USA. 101:17628-17633.

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 Research 26:320-322.

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.

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

Tanner N. and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol. Cell. 2:251-262.

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 Research 22:4673-4680.

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.

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.

Will, C. L., Urlaub, H., Achsel, T., Gentzel1, 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.

Xu, Y., Newnham, M., Kameoka, S., Huang, T., Konarska 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.

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.