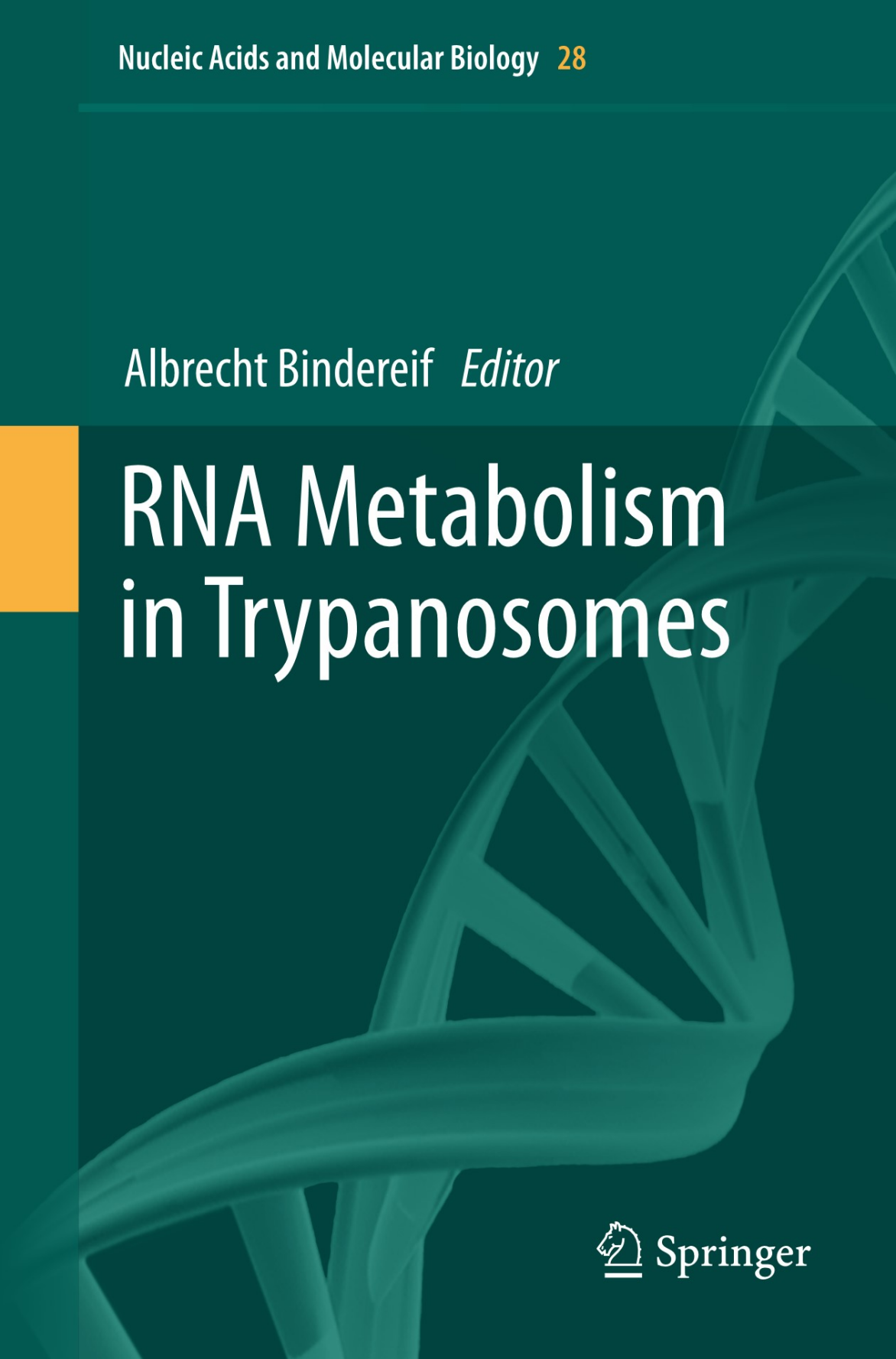
Nucleic Acids and Molecular Biology 28

Albrecht Bindereif *Editor*

RNA Metabolism in Trypanosomes



Springer

Nucleic Acids and Molecular Biology

Volume 28

Series Editor
Janusz M. Bujnicki
International Institute of Molecular
and Cell Biology
Laboratory of Bioinformatics and
Protein Engineering
Trojdena 4
02-109 Warsaw
Poland

For further volumes:
http://www.springer.com/series/881

Albrecht Bindereif
Editor

RNA Metabolism in Trypanosomes

Springer

Editor
Albrecht Bindereif
Department of Biology and Chemistry
Justus Liebig University of Giessen
Giessen
Germany

ISSN 0933-1891 ISSN 1869-2486 (electronic)
ISBN 978-3-642-28686-5 ISBN 978-3-642-28687-2 (eBook)
DOI 10.1007/978-3-642-28687-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012940659

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1 RNA Polymerases and Transcription Factors of Trypanosomes . . . 1
Arthur Günzl

2 SL RNA Biogenesis in Kinetoplastids: A Long and Winding Road . . . 29
Nancy R. Sturm, Jesse R. Zamudio, and David A. Campbell

3 Pre-mRNA Splicing in *Trypanosoma brucei*: Factors, Mechanisms, and Regulation . . . 49
Christian Preußer, Nicolas Jaé, Arthur Günzl, and Albrecht Bindereif

4 mRNA Turnover in Trypanosomes . . . 79
Christine Clayton

5 tRNA Biogenesis and Processing . . . 99
Jessica L. Spears, Mary Anne T. Rubio, Paul J. Sample, and Juan D. Alfonzo

6 rRNA Biogenesis in Trypanosomes . . . 123
Shulamit Michaeli

7 RNA Editing in African Trypanosomes: A U-ser's G-U-ide . . . 149
H. Ulrich Göringer

8 The RNA Interference Pathway in *Trypanosoma brucei* . . . 167
Elisabetta Ullu, Nikolay G. Kolev, Rebecca L. Barnes, and Christian Tschudi

9 Translation in Trypanosomatids . . . 187
Michal Shapira and Alexandra Zinoviev

10 Mitochondrial Translation in Trypanosomatids 215
Dmitri A. Maslov and Rajendra K. Agrawal

11 RNA-Seq Analysis of the Transcriptome of *Trypanosoma brucei* . . . 237
Jan Mani, Kapila Gunasekera, and Isabel Roditi

Index 267

Chapter 1
RNA Polymerases and Transcription Factors of Trypanosomes

Arthur Günzl

Abstract RNA synthesis in trypanosomatid organisms deviates substantially from what we see in model organisms. In these parasites, protein-coding genes are arranged in long tandem arrays that are polycistronically transcribed by an unresolved mode of transcription initiation. Moreover, the African *Trypanosoma brucei* has evolved a multifunctional RNA polymerase I system which it employs for pre-rRNA synthesis as well as for transcription of specific gene units encoding its major cell surface antigens. Additionally, the trypanosomatid RNA polymerase III system, by relying mostly on bidirectional tRNA gene promoters, exhibits clear differences to other eukaryotic systems. Interestingly, annotation of completed trypanosomatid genomes revealed only a small subset of basal transcription factors (BTFs), which suggested that trypanosomes have a simplified transcription machinery. Recent research, however, has demonstrated that trypanosomes possess extremely divergent orthologs of most BTFs and that these factors deviate in many aspects from their human counterparts; they are the focus of this chapter.

1.1 Introduction

Trypanosomatids represent a phylogenetically early diverged, protistan family of parasites which include the human-pathogenic *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major* ("Tritryp"). The transcriptional processes in these organisms differ substantially from those described in the model systems of human cells and of the budding yeast *Saccharomyces cerevisiae*. In the model systems, the

A. Günzl (✉)
Department of Genetics and Developmental Biology, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, CT 06030, USA

Department of Molecular, Microbial and Structural Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030, USA
e-mail: gunzl@uchc.edu

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_1,

three canonical nuclear RNA polymerases (pols) I, II, and III perform distinct tasks: RNA pol I transcribes exclusively the large ribosomal gene units (*RRNA*) in the nucleolus, RNA pol II all protein-coding genes and some small RNA genes, and RNA pol III genes encoding tRNAs, 5S rRNA, and remaining small RNAs. Transcription is typically monocistronic and each gene includes specific transcription initiation and termination sites. In each case, conserved basal transcription factors (BTFs) bind to core promoters, recruit RNA pols to specific transcription initiation sites (TISs), and facilitate correctly initiated transcription. Conversely, trypanosomatid protein-coding genes are organized in long tandem arrays, which are transcribed polycistronically with transcription initiation at these arrays apparently not depending on classical promoters and BTF assembly. Furthermore, the African trypanosome *T. brucei* has evolved a multifunctional RNA pol I system, which, in addition to rRNA synthesis, is utilized for the transcription of specific protein-coding gene units. Finally, small nuclear and cytoplasmic RNA genes are transcribed by RNA pol III from unique, bidirectional tRNA gene promoters.

Polycistronic as well as RNA pol I-mediated transcription of protein-coding genes in trypanosomes is based on an atypical mode of mRNA processing in which polycistronic precursor RNAs are resolved into individual mRNAs by spliced leader (SL) *trans* splicing and polyadenylation. In *trans* splicing, the capped 5′ terminal region of the small nuclear SL RNA, referred to as the SL or mini-exon, is transferred onto the 5′ end of each mRNA. SL *trans* splicing is therefore a posttranscriptional capping mechanism which is in contrast to the co-transcriptional capping mode in yeast and higher eukaryotes. In those systems, the mRNA capping enzyme physically interacts with the carboxy-terminal domain (CTD) of the largest RNA pol II subunit RPB1 (from here on referred to as "CTD") and cotranscriptionally caps pre-mRNA (Bentley 2005). This enzyme association has two consequences: pre-mRNA is exclusively synthesized by RNA pol II, because the capping enzyme does not associate with other RNA pols, and protein-coding genes are transcribed monocistronically because the capping enzyme requires a free 5′ end. However, in trypanosomatids, SL *trans* splicing uncouples capping from RNA pol II transcription thereby eliminating the requirement for a free RNA end as well as enabling RNA pol I to productively express functional mRNA. This difference between trypanosomatids and other eukaryotes becomes apparent when the *RRNA* promoter is used to express a reporter gene. In the mouse system, this experiment yielded virtually no reporter protein despite high transcription rates (Grummt and Skinner 1985), whereas in trypanosomes the *RRNA* promoter directed the expression of functional mRNA and very high reporter levels (Zomerdijk et al. 1991; Rudenko et al. 1991).

Transcriptional studies in trypanosomatids first focused on promoter characterizations. These studies were conducted more than a decade ago and have been reviewed several times since, most recently by Martinez-Calvillo et al. (2010), Günzl et al. (2007), and Palenchar and Bellofatto (2006). Promoter structures will, therefore, only be briefly discussed in this chapter. In contrast, our knowledge of RNA pol subunits and transcription factors stems predominantly from recent research. While the largest RNA pol subunits were identified early on (Köck et al.

1988; Jess et al. 1989; Evers et al. 1989; Smith et al. 1989a, b), the first transcription factor was only characterized in 2003 (Das and Bellofatto 2003). Since then, RNA pol complexes and several transcription factors have been characterized in detail; together they currently account for an impressive repertoire of 80 proteins (Tables 1.1 and 1.2). Finding such a large number of transcriptional proteins was unexpected because annotation of the completed TriTryp genomes (Berriman et al. 2005; El Sayed et al. 2005; Ivens et al. 2005) revealed only very few orthologs of known transcription factors. Now we know that these factor sequences are extremely divergent from their eukaryotic counterparts, which has made bioinformatic identification nearly impossible. Thus, most of the transcriptional proteins were identified biochemically as part of protein complexes. The advent of tandem affinity purification technology enabled efficient protein complex isolation at nearly physiological conditions and in amounts that could be analyzed by liquid chromatography–tandem mass spectrometry (LC/MS/MS) (Rigaut et al. 1999; Günzl and Schimanski 2009). While the original method proved to be effective for some transcription complexes, most transcription factors were purified by a modified approach, which depended on a new epitope combination, termed PTP, that was specifically developed for the isolation of complexes from transcriptionally active extracts (Schimanski et al. 2005b). Most of these studies were carried out on *T. brucei* and therefore the trypanosome transcription machinery is the focus of this chapter.

1.2 RNA pol II Transcription

Except for gene units encoding the major cell surface antigens variant surface glycoprotein (VSG) and procyclin in *T. brucei*, trypanosomatid RNA pol II transcribes protein-coding gene arrays, which also contain genes encoding small nucleolar RNAs and, as recently shown, a single tRNA gene (Aeby et al. 2010). The only other known RNA pol II transcription unit is the SL RNA gene (*SLRNA*).

1.2.1 RNA pol II

Across eukaryotes, RNA pol II consists of 12 highly conserved subunits termed RPB1-12. The two largest proteins RPB1 and RPB2 form the catalytic core, the subunits RPB3, 10, 11, and 12 represent an assembly platform of the enzyme, while the remaining subunits have specialized functions in the transcription process. The subunits RPB5, 6, 9, 10, and 12 are shared between all three eukaryotic RNA pols, whereas there are paralogous subunits of the other seven subunits in RNA pols I and III. RPB1-12 are highly conserved with orthologs present in archaea (Werner 2007), and therefore all of them could be identified in silico in the Tritryp genomes (Table 1.1) (Kelly et al. 2005). Interestingly, trypanosomatids harbor two divergent sets of the shared subunits RPB5, 6, and 10, and due to trypanosomatid-specific

Table 1.1 *T. brucei* RNA pol II subunits and transcription factors

RNA pol II	TRF4/SNAP$_c$/TFIIA	TFIIB	TFIIH[a]	Med-T	Others
RPB1	**TRF4**	**TFIIB**	**XPB**	**Med-T1**	**BDF3**
Tb927.4.5020	Tb927.10.15950	Tb09.160.4220	Tb927.3.5100	Tb11.01.7350	Tb11.01.1830
Tb927.8.7400					
RPB2	**SNAP50**		**XPD**	**Med-T2**	**FACT SSRP1**
Tb927.4.3810	Tb09.211.1510		Tb927.8.5980	Tb927.4.3030	Tb927.10.14390
RPB3	**SNAP2**		**'p62'[b]**	**Med-T3**	**FACT Spt16**
Tb927.3.5500	Tb927.5.3910		Tb11.01.1200	Tb927.3.5170	Tb927.3.5620
RPB4	**SNAP3**		**'p52'[b]**	**Med-T4**	**TFIIS1**
Tb927.3.5270	Tb927.10.7070		Tb10.70.1900	Tb927.8.2360	Tb11.02.2600
RPB5	**TFIIA-2**		**'p44'[b]**	**Med-T5**	**TFIIS2-1**
Tb927.10.13320	Tb927.10.4840		Tb927.8.6540	Tb11.03.0430	Tb927.2.3580
RPB6	**TFIIA-1/TAF ?**		**'p34'[b]**	**Med-T6**	**TFIIS2-2**
Tb927.4.3490	Tb927.10.15570		Tb11.01.7730	Tb11.02.4940	Tb927.2.3480
Tb927.4.3510					
RPB7			**TFB5**	**Med-T7**	
Tb11.01.6090			Tb10.61.2600	Tb09.211.2270	
RPB8			**TSP1**	**Med-T8**	
Tb11.02.5790			Tb927.1.1080	Tb927.4.2960	
RPB9			**TSP2**	**Med-T9[c]**	
Tb11.02.5180			Tb11.01.5700	XP_951672	
RPB10[d]					
"Tb11.02.1185"					
RPB11					
Tb11.57.0004					
RPB12					
Tb927.1.1170					

Proteins highlighted in *gray* have been experimentally identified/characterized. Nucleotide and amino acid sequences can be retrieved at http://www.GeneDB.org and http://TritrypDB.org with accession numbers cited. For each gene, these websites provide links to trypanosomatid orthologs

[a]Human nomenclature

[b]The apparent sizes of the trypanosome orthologs do not correspond to the sizes of the human proteins

[c]NCBI accession number; the gene has not been identified in the GeneDB/TritrypDB databases

[d]The *RPB10* gene has not been annotated in the *T. brucei* genome

Table 1.2 *T. brucei* RNA pol I/pol III subunits and transcription factors

RNA pol I[a]	CITFA	RNA pol III[a]	TFIIIB	TFIIIA/TFIIIC
RPA1	**CITFA-1**	**RPC1**	**BRF1**	?
Tb927.8.5090	Tb11.47.0010	Tb927.10.2780	Tb11.03.0670	
RPA2	**CITFA-2**	**RPC2**	**TRF4**	
Tb11.03.0450	Tb09.211.3440	Tb927.1.540	Tb927.10.15950	
RPC40	**CITFA-3**	**RPC40**	putative BDP1	
Tb927.10.15370	Tb11.47.0008	Tb927.10.15370	Tb927.10.7840	
RPA14[b]	**CITFA-4**	**putative RPC17[d]**		
?	Tb11.01.0240	Tb927.3.2700		
RPB5z	**CITFA-5a**	**RPB5**		
Tb927.10.13310	Tb927.8.4030	Tb927.10.13320		
	Tb927.8.4080			
RPB6z	**CITFA-5b**	**RPB6**		
Tb11.03.0935	Tb927.8.4130	Tb927.4.3490		
		Tb927.4.3510		
TbRPA31[d]	**CITFA-6**	RPC25		
Tb927.10.3540	Tb927.5.970	Tb11.01.4820		
RPB8	**CITFA-7**	**RPB8**		
Tb11.02.5790	Tb927.7.2600	Tb11.02.5790		
RPA12	**DYNLL1**	RPC11		
Tb11.01.2190	Tb11.50.0007	Tb927.10.15150		
	Tb11.0845			
RPB10z		RPB10[e]		
Tb927.3.1250		"Tb11.02.1185"		
RPC19		**RPC19**		
Tb11.01.0625		Tb11.01.0625		
RPB12		**RPB12**		
Tb927.1.1170		Tb927.1.1170		
RPA49		**putative 'RPC82'**		
?		Tb927.2.2990		
RPA34		**putative 'RPC53'**		
?		Tb09.211.2090		
		putative 'RPC37'		
		Tb11.02.0970		
		putative 'RPC34'		
		Tb927.3.3910		
		RPC31		
		?		

Proteins highlighted in *gray* have been experimentally identified/characterized. Nucleotide and amino acid sequences can be retrieved at http://www.GeneDB.org and http://TritrypDB.org with accession numbers cited. For each gene, these websites provide links to trypanosomatid orthologs

[a]Yeast nomenclature

[b]"?" indicates that a trypanosome ortholog has not been found

[c]"Putative" indicates weak sequence conservation

[d]TbRPA31 is potentially the ortholog of the yeast RPB7 paralog RPA43

[e]The *RPB10* gene has not been annotated in the *T. brucei* genome

sequence insertions in one set the corresponding subunits were termed RPB5z, 6z, and 10z (Kelly et al. 2005; Nguyen et al. 2006). RNA pol II was tandem affinity purified in *T. brucei* (Devaux et al. 2006; Das et al. 2006) and *L. major* (Martinez-Calvillo et al. 2007), and across these three studies LC/MS/MS identified all subunits as part of the enzyme. Due to RNA pol I and III characterizations (see below), it became clear that RNA pols II and III share RPB5, 6, and probably 10, whereas RPB5z, 6z, and 10z are specific subunits of RNA pol I (Martinez-Calvillo et al. 2007; Walgraffe et al. 2005; Nguyen et al. 2006).

Besides a high degree of sequence divergence in its subunits, trypanosome RNA pol II does not appear to differ from its human/yeast counterpart. The exception is the largest subunit RPB1. An early observation was that trypanosomes, in contrast to other eukaryotes, harbor two *RPB1* genes that are located on different chromosomes. Although the amino acid sequence encoded in the two genes differed only in four positions, it was hypothesized that trypanosomes harbor two RNA pol II enzymes with distinct essential functions (Evers et al. 1989; Smith et al. 1989b). This scenario proved to be wrong because three of the four *RPB1* alleles could be deleted without affecting parasite viability in culture (Ruan et al. 2004). A more striking difference is located in the CTD. Conserved from some protists to humans, the CTD contains 15–52 repeats of the heptad sequence YSPTSPS. Dephosphorylation of the serine residues in this motif directs recruitment of RNA pol II to core promoters, while their phosphorylation transforms RNA pol II into an elongation-competent enzyme (Buratowski 2009). However, the trypanosome CTD lacks this motif raising the possibility that trypanosome RNA pol II transcription does not depend on CTD (de-)phosphorylation. Accordingly, CTD phosphorylation in other eukaryotes leads to the binding of RNA processing machinery, including capping and splicing complexes, whereas trypanosome mRNA processing is not coupled to RNA pol II transcription (Stewart et al. 2010). On the other hand, trypanosome RPB1 including the CTD *is* phosphorylated (Chapman and Agabian 1994; Nett et al. 2009) and it was recently shown that deletion of the larger part of the CTD abolished RNA pol II transcription (Das and Bellofatto 2009).

1.2.2 SLRNA *Transcription*

SL RNA, the SL donor in the *trans* splicing process, is a key molecule in trypanosome gene expression because all mRNAs are *trans* spliced and SL RNA is consumed in the process. Therefore, trypanosomes need to continuously express large amounts of this small nuclear RNA to sustain viable levels of gene expression. This importance and the fact that SL RNA is a parasite-specific RNA without counterpart in mammalian and arthropod hosts of trypanosomes has made SL RNA biogenesis a research focus and *SLRNA* the best characterized transcription unit of the parasites.

SLRNAs are tandemly repeated on *T. brucei* chromosome 9 and there may be up to 100 *SLRNAs* per array according to hybridization experiments (de Lange et al. 1983;

Nelson et al. 1983) although only 28 repeats were annotated in the *T. brucei* genome (Berriman et al. 2005). Each *SLRNA* is transcribed in a monocistronic fashion by RNA pol II (Gilinger and Bellofatto 2001) from a concrete TIS and with a T-stretch as termination signal (Sturm et al. 1999). The *SLRNA* promoter is conserved among trypanosomatids and consists of a bipartite upstream sequence element (USE) and an initiator element at the TIS (Fig. 1.1a) (Hartree and Bellofatto 1995; Günzl et al. 1997; Nunes et al. 1997; Yu et al. 1998; Luo et al. 1999).

In eukaryotes, transcription initiation by RNA pol II invariably depends on a set of BTFs, which bind to a core promoter, recruit RNA pol II to a defined TIS, open up the DNA strands at the TIS, and enable the enzyme's escape from the promoter. The protein complex that assembles at the core promoter is known as the preinitiation complex (PIC) and is comprised of RNA pol II and the six BTFs TFIIA, TFIIB, TFIID/TBP, TFIIE, TFIIF, and TFIIH (recently reviewed by Thomas and Chiang 2006; Sikorski and Buratowski 2009). PIC assembly and transcription initiation have been meticulously studied in protein-coding genes and function as follows: The initial core promoter contact is made by TFIID, which consists of TBP and up to 15 TBP-associated factors (TAFs). TFIIA, which comprises two subunits, then binds to TBP stabilizing the DNA–protein interaction. TBP binding leads to strong DNA bending allowing TFIIB to enter the PIC. The single polypeptide TFIIB is of central importance to PIC formation because it directly interacts with RNA pol II and positions the enzyme at the correct TIS. TFIIB also binds to TBP and to the DNA on both sides of the TATA box forming a bridge between core promoter and enzyme. RNA pol II that interacts with TFIIB is associated with the bipartite TFIIF and is transcription elongation incompetent because the transcriptionally important CTD is dephoshorylated. To start up the polymerase, the interacting TFIIE and TFIIH are needed. TFIIH is a large complex consisting of a core of seven subunits, including the two helicases *Xeroderma pigmentosum* B (XPB) and XPD, and of an additional cyclin-activating kinase (CAK) subcomplex of three subunits. The helicase activity of TFIIH opens up the DNA strands at the TIS, and its kinase activity phosphorylates the CTD enabling RNA pol II to transform into an elongation-competent enzyme and escape from the promoter. TFIIE consists of two subunits and is important for TFIIH recruitment and function.

As mentioned earlier, the SL RNA is a small nuclear RNA. PIC formation at small RNA gene promoters was shown in the human system to be nucleated by a dedicated factor, termed small nuclear RNA-activating protein complex ($SNAP_c$) (Henry et al. 1995; Yoon et al. 1995). While human $SNAP_c$ consists of five subunits, only three of these factors are essential and conserved in other eukaryotes (Li et al. 2004). PIC formation at small RNA genes is less well understood than at protein-coding genes but an in vitro analysis in the human system demonstrated that TFIIA, TFIIB, TFIIE, and TFIIF are essential for transcription of this gene class (Kuhlman et al. 1999).

Annotation of the completed Tritryp genomes in 2005 revealed only the trypanosome TBP homolog, termed TBP-related factor 4 (TRF4), and the TFIIH helicases XPB and XPD (Ivens et al. 2005); the claimed identification of cyclin-dependent kinase 7 (CDK7) in the same study is most likely incorrect (see below).

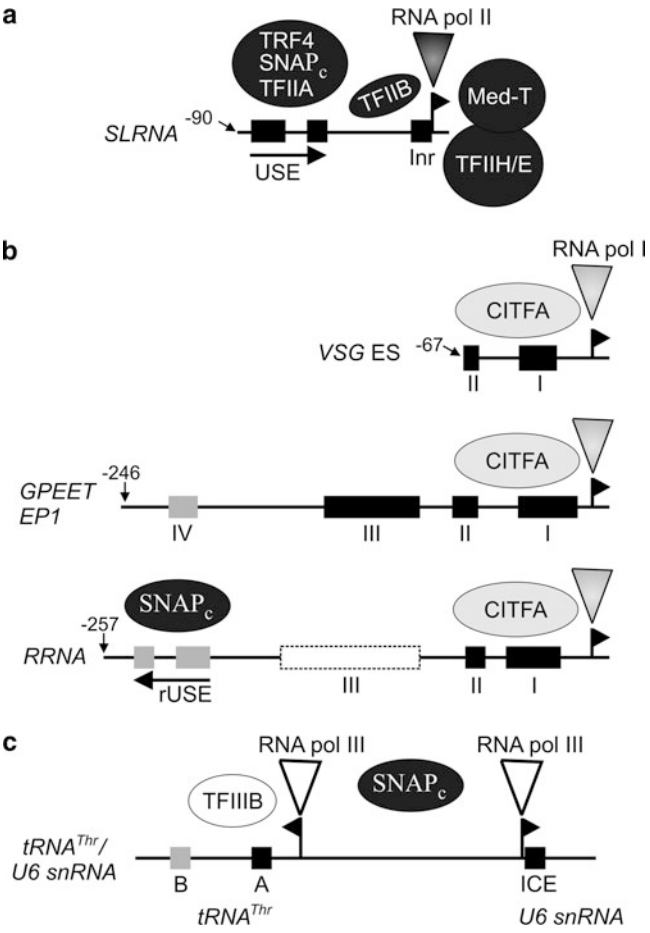


Fig. 1.1 Schematic depiction of *T. brucei* promoters and transcription factors. Transcription initiation sites (TISs) are represented by *flags* and essential promoter elements by *black* boxes. *Gray* boxes indicate promoter elements that are important only in vivo. Furthermore, RNA pols are drawn as *rectangles* and RNA pol I, II, and III transcription factors as *gray*, *black*, and *white ovals*, respectively. (**a**) *SLRNA* promoter consisting of a bipartite upstream sequence element (USE) and an initiator (Inr). The preinitiation complex assembling at the *SLRNA* promoter consists of RNA pol II, TRF4/$SNAP_c$/TFIIA, TFIIB, TFIIH/E, and mediator (Med-T). The *arrow* points to position −90 relative to the TIS which is the minimum promoter length required for full transcriptional activity. (**b**) RNA pol I promoters. The *VSG* ES promoter is short and consists of only two sequence domains whereas the two nearly identical procyclin promoters *GPEET* and *EP1* as well as the *RRNA* promoter have four distinct domains resembling the RRNA promoter of yeast. *T. brucei RRNA* promoter domain III has not been mapped and, as discussed in the text, *RRNA* promoter domain IV resembles the *SLRNA* USE in opposite direction indicated by the *arrow* (rUSE). Although there is no obvious sequence similarity among the three different promoters CITFA binds to all of them. (**c**) Depiction of the threonine tRNA gene ($tRNA^{Thr}$)/U6 snRNA gene association in *T. brucei*. The U6 promoter elements include the A and B boxes of the upstream tRNA gene as well as an intragenic control element (ICE) downstream of the TIS. TFIIIB binds to the A box, whereas it is unclear how $SNAP_c$ interacts with RNA pol III genes

This suggested that trypanosomatids either have a strongly reduced BTF repertoire or that BTF sequences are too divergent to identify in silico. In the latter case, an analysis of *SLRNA* transcription factors would depend on biochemical and genetic experiments. The pioneering work, conducted in the trypanosomatid *Leptomonas seymouri* by the research group of Dr. Vivian Bellofatto (New Jersey Medical School, USA), used conventional chromatography and an electrophoretic mobility shift assay to purify a tripartite factor that specifically interacted with the *SLRNA* USE. One of the two proteins that were identified in this study by mass spectrometry was a clear ortholog of human SNAP50, an essential $SNAP_c$ subunit, whereas the second protein appeared to be parasite specific in sequence (Das and Bellofatto 2003). TRF4 was the second factor found to play a role in *SLRNA* transcription, since *TRF4* expression silencing in *T. brucei* resulted in a decrease of SL RNA abundance (Ruan et al. 2004). Subsequently, tandem affinity purification of SNAP50 and TRF4 in *T. brucei* identified a larger, *SLRNA* USE binding complex that consisted of a tripartite $SNAP_c$, TRF4, a clear ortholog of the smaller TFIIA subunit (TFIIA-2), and a sixth protein whose sequence conservation was too weak for an unambiguous assignment as the larger TFIIA subunit and thus may equally well represent a TAF (Das et al. 2005; Schimanski et al. 2005a, b).

TRF4/$SNAP_c$/TFIIA has been implicated in transcriptional regulation. The complex binds to the *SLRNA* USE most likely through its SNAP2 subunit, also known as tSNAP42, because it contains divergent myb domains by virtue of which its putative human ortholog SNAP190 binds the proximal sequence element of small RNA genes (Das et al. 2005; Schimanski et al. 2005a; Hernandez 2001). Interestingly, under certain stress conditions such as prolonged interference with endoplasmic reticulum function, trypanosome $SNAP_c$ falls apart: SNAP2 dissociates from the *SLRNA* promoter and accumulates in the nucleus whereas SNAP50 and the third $SNAP_c$ subunit, SNAP3 or tSNAP26, become degraded. This response leads to effective and specific inhibition of *SLRNA* transcription and therefore was termed *SL RNA silencing* (Lustig et al. 2007). It was speculated that SL RNA silencing leads to programmed cell death to "eliminate unfit parasites from the population" (Goldshmidt et al. 2010). In this regard, it is interesting to note that in *T. brucei* the most distal domain of the *RRNA* promoter, in opposite direction, resembles the *SLRNA* USE, binds SNAP50 in vitro, and is in part functionally interchangeable with the *SLRNA* USE (Schimanski et al. 2004a). These findings suggest that $SNAP_c$ dissociation in SL RNA silencing would also affect *RRNA* transcription, but 28S rRNA amounts dropped only somewhat during the silencing period (Lustig et al. 2007).

The identification of a clear TFIIA subunit in the TRF4/$SNAP_c$/TFIIA complex was surprising because it suggested for the first time that trypanosomes do harbor BTFs for RNA pol II transcription initiation, and it triggered in silico searches for the key factor TFIIB. An extremely divergent TFIIB ortholog was finally identified in *T. brucei* and shown to be recruited to the PIC, to be essential for *SLRNA* transcription, and to interact with both RNA pol II and TRF4 (Palenchar et al. 2006; Schimanski et al. 2006). TFIIB has two functional domains: the N-terminal part consists of a zinc ribbon domain and a novel structure termed a B finger that, by

reaching into the active center, takes hold of RNA pol II (Bushnell et al. 2004). The C-terminal part harbors a tandem cyclin domain for the interaction with TRF4 and promoter DNA (Nikolov et al. 1995; Tsai and Sigler 2000). In trypanosome TFIIB, the zinc ribbon domain and a putative B finger are present, although there are clear differences at the amino acid sequence level. Unfortunately, the N-terminal domain did not fold appropriately in solution and therefore could not be structurally analyzed so far (Ibrahim et al. 2009). In contrast, the C-terminal TFIIB domain, which also deviates substantially from its eukaryotic counterparts, formed stable monomers and was amenable to crystallographic analysis. At 2.3Å this analysis produced the first high resolution structure obtained from a trypanosomatid transcription factor and revealed several unique structural elements that are not present in human TFIIB (Ibrahim et al. 2009). Importantly, this study demonstrated that the high degree of sequence divergence found in trypanosome BTFs confers structural differences.

The identification of TFIIA and TFIIB as essential *SLRNA* transcription factors suggested that a PIC is formed at the *SLRNA* promoter and that trypanosomes have a full complement of highly divergent BTFs. Accordingly, expression silencing of the TFIIH subunit XPD and a corresponding in vitro transcription analysis unambiguously demonstrated that TFIIH is essential for *SLRNA* transcription (Lee et al. 2007; Lecordier et al. 2007). The subsequent biochemical characterization of the TFIIH complex revealed a full core complex of seven subunits and two additional bona fide subunits which were trypanosomatid specific in sequence and therefore termed trypanosomatid-specific protein 1 (TSP1) and TSP2 (Lee et al. 2009). Although there is no direct evidence yet, it is likely that TSP1 and 2 are the orthologs of the two TFIIE subunits β and α, respectively, and that in trypanosomes, TFIIH and TFIIE form a single, stable complex because of the following reasons: (i) Like their putative counterparts in humans and yeast, TSP1 and 2 were indispensable for RNA pol II transcription (Lee et al. 2009); (ii) trypanosomatid TSP2s, like all known TFIIEα orthologs, harbor an invariant, internal C_2C_2 zinc finger; and (iii) the known interaction of TFIIE and XPB (Maxon et al. 1994) correlated with the finding that the TSPs become part of TFIIH together with XPB. In addition, solving the TFIIH structure by single particle electron microscopy (EM) and comparing 3D reconstructions of human and trypanosome TFIIH revealed extra protein densities in the trypanosome core domain which resembled the molecular structure of human TFIIE (Jawhari et al. 2006; Lee et al. 2009). Most interestingly, trypanosome TFIIH lacked a CAK subcomplex because its subunits do not belong to the well-characterized trypanosome kinome (Parsons et al. 2005) or harbor the invariant N-terminal C3HC4 RING finger domain of the CAK subunit MAT1 and because the TFIIH structure lacked the CAK-characteristic knob of human TFIIH (Schultz et al. 2000; Lee et al. 2009). This finding supported a previous comparative genomics analysis which predicted that early diverged protists, whose CTD does not contain the heptad motif, lack CDK7 (Guo and Stiller 2004). Hence, if CTD phosphorylation is important for transcription initiation in trypanosomes, as it is in all other eukaryotes analyzed so far, then it has to be conducted by a TFIIH-independent kinase. This would be a first mechanistic difference in RNA pol II transcription initiation between mammalian/yeast and trypanosome systems.

While trypanosome TFIIH is not associated with a kinase, it interacted with another protein complex of at least nine subunits. Again, this complex proved to be indispensable for *SLRNA* transcription in vivo and in vitro but this time the amino acid sequences were so divergent that they did not reveal the identity of the complex (Lee et al. 2010). However, the multisubunit nature of the complex, its association with TFIIH, and its transcriptional importance raised the possibility that it might be a trypanosome mediator complex. Mediator has been characterized in humans and yeast and consists of ~25 subunits with a combined mass of >1 MDa comprising four distinct domains, namely head, middle, tail, and CDK8 modules. Mediator has been predominantly characterized as a coactivator linking a DNA-bound activator via a protein bridge to the PIC (reviewed by Malik and Roeder 2010). Recently though, it has emerged as a seventh BTF because studies in yeast demonstrated that, in vivo, this complex has a basal function in transcription of the majority of genes (Takagi and Kornberg 2006; Takagi et al. 2006). Thus far, mediator was shown to interact with TFIIH, RNA pol II, and TBP, and to be important for TFIIB and TFIIH recruitment into the PIC (Baek et al. 2006; Esnault et al. 2008; Cai et al. 2010). The single-particle EM analysis of the trypanosome complex (Med-T) revealed a structure strikingly similar to the head module of yeast mediator which resembles pliers with a "handle" and a flexible "jaws" domain (Lee et al. 2010). Moreover, ChIP assays demonstrated that, as expected for mediator, Med-T was essential for the recruitment of TFIIH and TFIIB to the *SLRNA* promoter (Lee et al. 2010). Interestingly, identification of Med-T is based on structural and functional data only because Med-T sequences are so divergent that a similarity to yeast/human mediator head subunits could still not be established. Nevertheless, this study represented the first characterization of a protistan mediator complex, it showed that the basal transcription function of mediator has evolved very early in eukaryotic evolution, and it represented the first demonstration in any system that mediator is essential for the transcription of a small nuclear RNA gene.

Overall, the BTFs for *SLRNA* transcription now comprise 25 distinct proteins (Table 1.1) invalidating the notion that the basal transcription machinery of trypanosomatids is simplified (Ivens et al. 2005). Trypanosome BTF sequences are extremely divergent from their nontrypanosomatid orthologs, which is why they were not identified in standard genome annotation. This sequence divergence is most likely a direct consequence of a strongly reduced set of PIC formation sites in the trypanosome genome, diminishing evolutionary constraints. It should be kept in mind that the *SLRNA* promoter is the only characterized RNA pol II promoter in kinetoplastids thus far, and it currently is not known whether protein-coding gene transcription requires PIC formation.

1.2.3 RNA pol II Transcription of Protein-Coding Genes

One of the most contentious questions in trypanosome research has been how transcription initiation of RNA pol II is achieved for protein-coding genes. In contrast to other systems, it has not been possible to characterize a promoter for

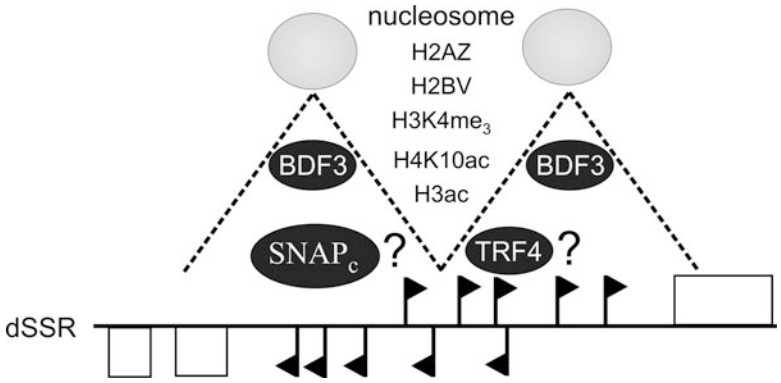


Fig. 1.2 Schematic drawing of a divergent strand switch region (dSSR). In ChIP-seq experiments, two broad peaks of open chromatin marks (*dashed lines*) are present within a dSSR, and there are multiple TISs for transcription in both directions over a broad region. Factor BDF3 appears to be linked to open chromatin marks. Binding of TRF4 and $SNAP_c$ was detected by ChIP-on-chip experiments in *L. major*, but binding sites and functional roles in protein-coding gene transcription remain to be established for those factors. Similarly, a dSSR sequence element involved in transcription has not been identified so far. *Open boxes* mark genes on both strands

coding gene transcription; the two putative promoters that have been described in *T. brucei*, one for an actin gene (Ben Amar et al. 1991) and one for a heat shock protein 70 gene (Lee 1996) did not reveal concrete TISs and the promoter functions could not be reproduced in another study (McAndrew et al. 1998). However, careful nuclear run-on analyses on microarrays in *L. major* have established that transcription in trypanosomatids is predominantly initiated in divergent strand switch regions (dSSRs; Fig. 1.2) in which tandem gene arrays are arranged head to head (Martinez-Calvillo et al. 2003, 2004). In *L. major*, dSSR sequences were partially successful in driving reporter gene expression although transcription appeared to initiate at several sites and promoter elements were not identified (Martinez-Calvillo et al. 2003). Interestingly, recent genome-wide mapping of histone variants and modifications established that there are typically two clear peaks of open chromatin marks within a dSSR most likely facilitating transcription initiation in both directions (Respuela et al. 2008; Thomas et al. 2009; Siegel et al. 2009; Wright et al. 2010). Single peaks also occur within polygenic arrays typically when tRNA genes disrupt protein-coding gene arrays. It is thought that tRNA genes block RNA pol II and necessitate re-initiation of RNA pol II transcription (Marchetti et al. 1998). The open chromatin marks comprise the two H2 histone variants H2AZ and H2BV and the modifications H3K4-trimethylation, H4K10 acetylation, and general histone 3 acetylation. The two open chromatin peaks per dSSR suggest that transcription is unidirectional and that this directionality is guided by a DNA-binding factor. However, no clear motif conservation could be found in dSSRs except for short, central GC-rich elements in *L. major* (Martinez-Calvillo et al. 2003) and longer G-runs in *T. brucei* (Siegel et al. 2009). It has been speculated that the G-rich regions, by adopting a special secondary structure known as G

quadruplexes, bind factors that direct transcription or alternatively, that the G-runs provide a transcriptional barrier ensuring unidirectional transcription toward the gene array (Siegel et al. 2009). Interestingly, in a recent transcriptome study, primary, unspliced transcripts were enriched and analyzed by RNA-seq allowing the determination of putative TISs in dSSRs (Kolev et al. 2010). The TISs were located predominantly in the regions of open chromatin marks corroborating the notion that transcription initiation requires open chromatin. Moreover, in most instances there were clear peaks of read alignments within dSSRs suggesting that transcription initiation is sequence determined and not random although transcription initiation was bidirectional and occurred over a broad region resembling transcription initiation from mammalian CpG promoters (Carninci et al. 2006; Kolev et al. 2010). One possibility is therefore that dSSRs do harbor degenerate core promoter sequences that so far have escaped bioinformatic detection.

Not much is known about the transcription factor requirement for protein-coding gene transcription because PIC formation at dSSRs has not been demonstrated. However, a ChIP-on-chip analysis in *L. major* provided evidence that transcription from dSSRs may depend at least on some BTFs because SNAP50 and TRF4 appeared to be enriched right upstream of acetylated histone H3 peaks within dSSRs (Thomas et al. 2009). In addition, it was shown in *T. brucei* that the essential bromodomain containing factor 3 (BDF3) co-localized with H4K10ac in the nucleus and cross-linked to open chromatin peaks (Siegel et al. 2009). Since BDFs, in general, function in nucleosome remodeling, histone modification, and transcription, it is likely that BDF3 has an important role in protein-coding gene transcription.

The polygenic arrays can span more than 1 Mb, and therefore, transcription of such long units depends on high processivity of RNA pol II. The latter is typically conferred by transcription elongation factors that enable RNA pol II to overcome transcription arrests. One such factor is TFIIS and trypanosomes have two divergent paralogs of this factor, termed TFIIS1 and TFIIS2 (Uzureau et al. 2008). While an analysis of a combined knockdown of genes encoding both factors indicated a synthetic lethal phenotype, it remains to be shown whether these elongation factors associate with RNA pol II and increase the enzyme's processivity. Another important factor for effective RNA pol II transcription in trypanosomatids appears to be the chromatin remodeler FACT ("*fa*cilitates *c*hromatin *t*ranscription"). Human FACT consists of the larger subunit SPT16 and the smaller subunit SSRP1 (Orphanides et al. 1999), and it can promote effective RNA pol II transcription elongation by dislocating histone H2A–H2B dimers (Belotserkovskaya et al. 2003; Belotserkovskaya and Reinberg 2004). Trypanosome orthologs of the two subunits have been identified and shown to be encoded by essential genes (Patrick et al. 2008). Interestingly, both subunits could be cross-linked to the retrotransposon SLAC (*s*pliced *l*eader-*a*ssociated *c*onserved sequence) which specifically invades *SLRNA*s (Aksoy et al. 1987; Carrington et al. 1987) and uses the *SLRNA* promoter for transcription. In contrast, FACT was not associated with *SLRNA*s indicating that the retrotransposon sequence modifies the *SLRNA* promoter such that RNA pol II binds FACT. This interaction may increase RNA pol II processivity because it is associated with the transformation from a T-stretch-terminating enzyme to one that

effectively elongates through the SLAC open reading frames (Patrick et al. 2008). This role of FACT in RNA pol II processivity appears not to be restricted to SLACs because in *L. major* isolation of RNA pol II led to the mass-spectrometric identification of the larger FACT subunit, suggesting that a major fraction of RNA pol II complexes in trypanosomatids comprises FACT (Martinez-Calvillo et al. 2007).

Besides *SLRNA* and protein-coding genes, RNA pol II transcribes a single tRNA gene (Aeby et al. 2010) and arrays of genes encoding small nucleolar RNAs (Dunbar et al. 2000; Liang et al. 2005). These small RNA genes are not transcribed from dedicated promoters like *SLRNA*s, rather they are part of the protein-coding gene arrays. Accordingly, their primary transcripts appear to be processed equivalently to pre-mRNA by *trans* splicing and polyadenylation (Kolev et al. 2010).

1.3 The Multifunctional RNA pol I System of *T. brucei*

1.3.1 Transcription Units and Promoter Structures

T. brucei lives freely in the blood of its mammalian host and uses antigenic variation of its dense VSG cell surface coat as its main strategy to evade the immune system. The coat, consisting of ~10^7 identical molecules, shields invariant membrane proteins and appears to be the only protein recognized by the humoral immune system (Ziegelbauer and Overath 1993; Schwede et al. 2010). The parasite harbors more than one thousand *VSG* genes but to ensure a homogenous coat, all VSG is expressed from a single gene in a monoallelic fashion. A switch to the expression of a different *VSG* gene then results in a glycoprotein coat of different antigenicity. The single *VSG* gene is expressed from one of about 15 telomeric expression sites (ESs), which are polycistronic transcription units of 45–60 kb, containing several ES-associated genes and a single *VSG* as the last gene of the array (Hertz-Fowler et al. 2008).

Early on, it was noticed in bloodstream trypanosomes that *VSG* transcription was resistant to the RNA pol II inhibitor α amanitin, whereas transcription of other protein-coding genes was α amanitin-sensitive, suggesting that the parasite employed RNA pol I for *VSG* transcription (Kooter and Borst 1984). Later, this notion was confirmed when it was demonstrated that RNA pol I depletion in vivo by RNA interference or in vitro by immunoprecipitation specifically and clearly affected *RRNA* and *VSG* transcription but not RNA pol II transcription of protein-coding genes or of *SLRNA*s (Günzl et al. 2003).

In *T. brucei*, RNA pol I-mediated transcription of protein-coding genes is not restricted to *VSG* ESs. During the parasite's life cycle it is also used to express procyclin in procyclic, insect-stage trypanosomes and VSG in metacyclic trypanosomes, the mammalian-infective forms in the tsetse fly. Procyclin, the major cell surface antigen in procyclics, is essential for efficient fly infection (Ruepp et al. 1997) and thought to protect against digestive enzymes in the fly

midgut (McConville and Ferguson 1993). There are two types of procyclin characterized by internal dipeptide (EP procyclin) or pentapeptide (GPEET procyclin) repeats. Procyclin genes reside in two chromosome-internal loci as parts of polycistronic transcription units (Roditi and Clayton 1999). In metacyclic trypanosomes, VSG is monoallelically expressed from special metacyclic (m)VSG ESs that are monocistronic (Donelson 2003).

The promoters of these RNA pol I transcription units have been characterized in detail; they are structurally different and share no obvious sequence homology. Nevertheless, they can be divided into two structural classes (Fig. 1.1b): the *RRNA* and the two nearly identical *GPEET* and *EP1* procyclin promoters form one class by extending to ~250 bp upstream of the TIS and consisting of four distinct domains (Sherman et al. 1991; Brown et al. 1992; Janz and Clayton 1994) similar to the well-characterized yeast *RRNA* promoter. In contrast, the *VSG* and *mVSG* ES promoters are very short and possess only two upstream elements residing within 67 bp of the TIS. Interestingly, these elements are different in size and located at different positions, and only in the *mVSG* ES promoter does the sequence around the TIS seem critical for transcription, suggesting that *VSG* and *mVSG* ES promoters function differently (Vanhamme et al. 1995; Pham et al. 1996; Ginger et al. 2002). There is also a clear difference in domain IV of *RRNA* and procyclin promoters (Fig. 1.1b): transcription competition experiments unexpectedly revealed that adding an excess of *RRNA* promoter DNA to extract effectively inhibited RNA pol II-mediated *SLRNA* transcription (Laufer and Günzl 2001). The competition result was specific for the *RRNA* promoter due to its domain IV which was found to resemble the USE of the *SLRNA* promoter in opposite orientation. And indeed, the *SLRNA* and *RRNA* USEs were to some extent functionally interchangeable and the *RRNA* USE bound, according to SNAP50 immunodetection, the SNAP complex in vitro (Schimanski et al. 2004a). Although this interaction remains to be verified in vivo, it suggests that the *RRNA* promoter directs RNA pol II transcription in the opposite direction to *RRNA* transcription and that the $SNAP_c$-based silencing effect on *SLRNA* transcription under certain stress conditions (see above) concomitantly affects *RRNA* transcription. This is an attractive possibility because it would enable the parasite to regulate global gene expression simultaneously at the level of translation (ribosome biogenesis) and mRNA maturation (SL RNA synthesis). The USE sequence is specific to the *RRNA* promoter and not found in the procyclin promoters.

1.3.2 RNA pol I and the Class I Transcription Factor A

Transcription of the active *VSG* ES takes place in a DNase-resistant, extranucleolar compartment termed expression site body or ESB (Navarro and Gull 2001). Therefore, during the *T. brucei* life cycle, RNA pol I is not only recruited to four structurally different promoters but, in the bloodstream form, it is also sequestered into two different nuclear compartments, the nucleolus and the ESB. It has been

speculated that this unprecedented versatility of the *T. brucei* RNA pol I system requires specific pol subunits, factors, or protein domains (Nguyen et al. 2006) but all deviations found in the *T. brucei* RNA pol I machinery so far, including the specific set of subunits RPB5z, 6z, and 10z (see above), are conserved among all trypanosomatids (Table 1.2). In an initial partial purification of RNA pol I, it was found that the second largest subunit, RPA2, was unusually large. The subsequent cloning and sequencing of this subunit revealed a unique ~28 kDa-large N-terminal extension domain which is conserved only among trypanosomatids and not present in RPA2s of other eukaryotes or in trypanosomatid RPB2 and RPC2 paralogs (Schimanski et al. 2003). The functional role of this domain is not known and besides a serine-rich region, the sequence did not exhibit known motifs. A better characterization of RNA pol I was subsequently achieved by tandem affinity purification which revealed enzyme complexes that overall contained eight subunits but were not active, presumably because they did not contain the essential subunit RPB6z (Walgraffe et al. 2005; Nguyen et al. 2006). Accordingly, tagging the RPB6z subunit resulted in the purification of RNA pol I, which was active in both nonspecific and promoter-dependent transcription assays (Nguyen et al. 2007). This complex contained the eight previously identified subunits, RPB6z, and a tenth subunit, termed RPA31, that is trypanosomatid specific in sequence and essential for RNA pol I transcription in vivo and in vitro (Nguyen et al. 2007). Of the twelve subunits which are either paralogs of RNA pol II subunits or shared with RNA pol II, nine were biochemically characterized, whereas a tenth subunit, RPB12, has only been identified in silico so far (Kelly et al. 2005). The two missing subunits are the paralogs of the RNA pol II subunit doublet RPB4/RPB7, and it has been hypothesized that trypanosome RNA pol I utilizes these subunits instead of RNA pol I-specific paralogs (Kelly et al. 2005). Indeed, this appeared to be the case because a small percentage of RNA pol I in extract co-precipitated RPB7, and results from *RPB7* expression silencing suggested that RPB7 is essential for RNA pol I transcription (Penate et al. 2009). However, RPB7 was not detected in active RNA pol I preparations, it did not cross-link to the *VSG* ES promoter or within the 18S rRNA coding region, and RPB7 immunodepletion from extract only affected *SLRNA* transcription, but not *VSG* ES or *RRNA* promoter transcription (Park et al. 2011), suggesting that the published RNA pol I defects upon *RPB7* silencing were secondary in nature. Instead, RPA31 may be the RNA pol I-specific paralog of RPB7 because, like its putative yeast counterpart RPA43, it is essential for transcription and its recruitment into the enzyme complex depends on the presence of RPB6z (Nguyen et al. 2007). Besides the twelve core subunits, yeast RNA pol I contains two additional specific subunits, termed RPA49 and RPA34. While orthologs of these subunits have not been found in trypanosomes, purification of active RNA pol I revealed protein bands of 27 and 29 kDa which are yet to be identified (Nguyen et al. 2007).

The identification of RNA pol I transcription factors has been difficult in any system because these “class I” factors have dramatically diverged in eukaryotes due to the fact that they interact only with a single type of promoter. Hence, it was no surprise that orthologs of yeast and mammalian class I transcription factors could not

be identified in trypanosomatid genomes. It took conventional multistep chromatography to purify a *VSG* ES promoter-binding activity as well as a systematic tagging approach to analyze proteins of the purified fraction to finally identify a protein that was involved in RNA pol I transcription. Subsequent tagging and tandem affinity purification of this protein isolated a complex of at least eight subunits that was termed class I transcription factor A or CITFA (Table 1.2) (Brandenburg et al. 2007). Seven of the proteins are only conserved among trypanosomatids whereas the eighth subunit surprisingly turned out to be the dynein light chain DYNLL1 (Brandenburg et al. 2007). In vivo and in vitro analyses showed that CITFA is a basal factor for all RNA pol I transcription, and competition of a *VSG* ES promoter gel shift indicated that CITFA binds to *VSG* ES, *RRNA,* and procyclin promoters. Interestingly, these data also suggested that the procyclin promoter binds purified CITFA with much lower affinity than *VSG* ES and *RRNA* promoters (Brandenburg et al. 2007). Since reporter gene assays previously showed that the procyclin promoter is ~5-fold less active in the bloodstream stage than in the procyclic stage (Biebinger et al. 1996) and since the relative activity of the procyclin promoter compared to that of the *RRNA* promoter was ~4-fold reduced in extract of bloodstream trypanosomes as compared to extract of procyclics (Park et al. 2012), it is likely that procyclics possess a specific transactivator that assists CITFA binding to the procyclin promoter.

1.3.3 Monoallelic VSG *Expression*

While *RRNA* promoter transcription appears to be constitutive and procyclin promoter transcription regulated during the life cycle, monoallelic *VSG* expression in bloodstream trypanosomes is achieved by regulating individual VSG ESs on the transcriptional level in the same cells. Telomeric silencing is a key mechanism in restricting *VSG* transcription to a single ES because knockdowns of the genes encoding the telomere-binding protein RAP1 (Yang et al. 2009), the histone methyltransferase DOT1B (disruptor of telomeric silencing 1B) (Figueiredo et al. 2008), and the Sir2 homolog TbSIR21-rp1 (Alsford et al. 2007) resulted in clear derepression of silent ESs and, in case of RAP1 loss, in additional extranucleolar RNA pol I foci. In these studies, however, derepression of ESs did not affect the expression level of the active ES, which remained magnitudes higher than that of derepressed ESs. These results, therefore, suggested that telomeric silencing primarily prevents transcription, initiating at silent ESs at a low level (Vanhamme et al. 2000), to reach the *VSG* gene and that additional mechanisms affecting transcription initiation rates operate in the ES promoter region. One likely mechanism is the formation of repressive chromatin at silent promoters because silent ESs, in contrast to the active ES, are highly occupied by nucleosomes (Figueiredo and Cross 2010; Stanne and Rudenko 2010) and because gene silencing of the chromatin remodelers TbISWI (Hughes et al. 2007) and FACT (Denninger et al. 2010) as well as of the histone modifiers HAT1 (Kawahara et al. 2008) and DAC3

(Wang et al. 2010) led to partial derepression of silent ESs. However, the rather modest derepression levels in these studies and the fact that the active *VSG* is transcribed at a rate that is ~50-fold higher than that of a β-tubulin gene (Ehlers et al. 1987) suggest that a dedicated activation mechanism operates on the promoter of the active ES. Accordingly, we could recently show for the subunit CITFA-7 that this protein is concentrated in the nucleolus and the ESB and predominantly occupies the promoter of the active ES. These results suggest a model in which sequestration of CITFA confines productive RNA pol I transcription to these two subnuclear compartments. In this context, the presence of the dynein motor subunit in CITFA is intriguing because in other systems dynein light chains have been implicated in sequestration and modulation of transcription factors (Kaiser et al. 2003; Yeh et al. 2005).

1.4 RNA pol III-Mediated Transcription of Small RNA Genes

The classic RNA pol III-transcribed genes are tRNA genes, the 5S rRNA gene, and the U6 snRNA gene. These three gene types have structurally different promoters and assemble different protein complexes. The tRNA gene promoter consists of two sequence elements downstream of the TIS termed A and B box, the 5S rRNA gene promoter of a larger intragenic control region, and the U6 snRNA gene of three USEs termed distal sequence element, proximal sequence element, and TATA box. While the tripartite TFIIIB is essential for all three types of promoters, its recruitment into the RNA pol III PIC requires the multisubunit factor TFIIIC in tRNA genes, TFIIIC and the single polypeptide TFIIIA in the 5S rRNA gene, and $SNAP_c$ in the U6 snRNA gene (reviewed by Schramm and Hernandez 2002).

With the exception of *SLRNA*s and snoRNA genes embedded in the polygenic protein-coding gene arrays, trypanosome RNA pol III appears to transcribe all small RNA genes including the spliceosomal U1, U2, U4, and U5 snRNA genes which in humans and yeast are transcribed by RNA pol II. Trypanosome tRNA genes, like their eukaryotic counterparts, possess A and B boxes that are required for transcription (Nakaar et al. 1994). Interestingly, the U snRNA genes are always arranged head to head with a functional tRNA gene or with a tRNA pseudogene in which the A and B boxes are intact (Fig. 1.1c) (Mottram et al. 1991). Moreover, the distance between A box and the TIS of the downstream snRNA gene was found to vary only by 4 bp, being 104–107 bp in length across different snRNA genes and trypanosomatid organisms (Günzl 2003). These findings suggested a functional association between tRNA and U snRNA genes and, indeed, functional analyses demonstrated that the A and B boxes of the upstream tRNA gene were indispensable promoter elements for the transcription of the downstream snRNA gene (Fantoni et al. 1994; Nakaar et al. 1994, 1997; Djikeng et al. 2001). In case of the U1 snRNA gene, the tRNA A and B box elements are sufficient to drive snRNA gene transcription (Djikeng et al. 2001), whereas in U2 and U6 snRNA genes, an internal control region just downstream of the TIS represents a third essential

promoter element (Fantoni et al. 1994; Nakaar et al. 1997). These snRNA gene promoter structures are unique and it is unclear how the tRNA elements drive transcription in the opposite direction. An attractive model states that TFIIIB-induced DNA bending aligns the TIS of the snRNA gene with the PIC assembled at the upstream A and B boxes, enabling bidirectional transcription from the tRNA gene promoter (Nakaar et al. 1995).

Trypanosome RNA pol III subunits have been identified bioinformatically (Kelly et al. 2005), and a biochemical analysis of tandem affinity-purified enzyme was conducted in *L. major* (Martinez-Calvillo et al. 2007). The complex purification revealed eight of the twelve core subunits as well as four subunits, which are likely to be orthologs of the yeast RNA pol III-specific subunits although their sequences are not well conserved (Table 1.2). Since bioinformatics identified the missing four core subunits, only the yeast RPC31 subunit appears to be without ortholog in trypanosomatids. RNA pol III transcription factors have not been investigated in trypanosomes so far. However, TFIIIB appears to be conserved in trypanosomes. Eukaryotic TFIIIB consists of the TFII*B*-*r*elated *f*actor *1* (BRF1), TBP, and *B d*ouble *p*rime *1* (BDP1), previously also known as B″ (Schramm and Hernandez 2002). When the TRF4/SNAP$_c$/TFIIA complex was isolated in *T. brucei*, the BRF1 protein was found to co-purify with TRF4 indicating that TFIIIB is conserved in trypanosomes (Schimanski et al. 2005a). Accordingly, a protein which shares a conserved SANT domain with eukaryotic BDP1 could be retrieved from trypanosomatid genomes (Günzl et al. 2007). In addition and as discussed earlier, trypanosome SNAP$_c$ has been characterized. Although the trypanosome U snRNA gene promoters have no resemblance to the eukaryotic U6 snRNA gene promoter, ChIP experiments have indicated that SNAP$_c$ occupies U snRNA genes (Gilinger et al. 2004; Thomas et al. 2009). However, the interaction of SNAP$_c$ with these RNA pol III-transcribed genes and with the *SLRNA* promoter must be different because promoter pull-down assays and transcription competition experiments detected SNAP50 only at *SLRNA* and *RRNA* promoters as discussed earlier (Schimanski et al. 2004b). The trypanosome 5S rRNA gene promoter has not been characterized and TFIIIA and TFIIIC have not been identified in trypanosomes yet. However, these factors are present in all eukaryotes and it is most likely that the prevalent sequence divergence of trypanosome transcription factors has prevented their identification thus far.

1.5 Conclusion

Research in the past 5 years established that trypanosomes harbor many proteins that function in transcription. Nearly all of these proteins have orthologs in humans and yeast but they are extremely divergent in sequence. This was unexpected because other gene expression factors such as spliceosomal proteins are better conserved (Günzl 2010). Possibly, the divergence level is a direct result of fewer PIC formation sites in the genome. The identification of so many transcriptional proteins demonstrates that the basal transcription machinery of trypanosomes is not

simplified as in archaea (Werner 2007), but as complex as in higher eukaryotes. Thus, it can be anticipated that yet unidentified factors such as TFIIF, TFIIIA, or TFIIIC are still hiding somewhere in the trypanosome genome. The trypanosome transcription machineries are not only divergent, but the findings thus far suggest clear mechanistic differences between trypanosomes and humans. These include TFIIH-independent phosphorylation of RNA pol II CTD, the presence of dedicated factors interacting with the structurally different RNA pol I promoters, the sequestration of RNA pol I into two different subnuclear compartments in bloodstream *T. brucei*, and the genetic associations of tRNA and small RNA genes. Hopefully, these differences will eventually be exploitable in an antiparasitic strategy. This is not a remote possibility because with the exception of the functionally redundant TFIIS paralogs, all transcriptional proteins investigated thus far were absolutely essential for trypanosome viability.

Acknowledgments I thank Tu N. Nguyen and Justin K. Kirkham for critical reading of the manuscript. This work was supported by National Institutes of Health grants AI059377 and AI073300 to A.G.

References

Aeby E, Ullu E, Yepiskoposyan H, Schimanski B, Roditi I, Mühlemann O, Schneider A (2010) tRNASec is transcribed by RNA polymerase II in *Trypanosoma brucei* but not in humans. Nucleic Acids Res 38:5833–5843

Aksoy S, Lalor TM, Martin J, Van der Ploeg LH, Richards FF (1987) Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. EMBO J 6:3819–3826

Alsford S, Kawahara T, Isamah C, Horn D (2007) A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. Mol Microbiol 63:724–736

Baek HJ, Kang YK, Roeder RG (2006) Human mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. J Biol Chem 281:15172–15181

Belotserkovskaya R, Reinberg D (2004) Facts about FACT and transcript elongation through chromatin. Curr Opin Genet Dev 14:139–146

Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D (2003) FACT facilitates transcription-dependent nucleosome alteration. Science 301:1090–1093

Ben Amar MF, Jefferies D, Pays A, Bakalara N, Kendall G, Pays E (1991) The actin gene promoter of *Trypanosoma brucei*. Nucleic Acids Res 19:5857–5862

Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol 17:251–256

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Böhme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, MacLeod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G,

Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El Sayed NM (2005) The genome of the African Trypanosome *Trypanosoma brucei*. Science 309:416–422

Biebinger S, Rettenmaier S, Flaspohler J, Hartmann C, Pena Diaz J, Wirtz LE, Hotz HR, Barry JD, Clayton C (1996) The PARP promoter of *Trypanosoma brucei* is developmentally regulated in a chromosomal context. Nucleic Acids Res 24:1202–1211

Brandenburg J, Schimanski B, Nogoceke E, Nguyen TN, Padovan JC, Chait BT, Cross GA, Günzl A (2007) Multifunctional class I transcription in *Trypanosoma brucei* depends on a novel protein complex. EMBO J 26:4856–4866

Brown SD, Huang J, Van der Ploeg LH (1992) The promoter for the procyclic acidic repetitive protein (PARP) genes of *Trypanosoma brucei* shares features with RNA polymerase I promoters. Mol Cell Biol 12:2644–2652

Buratowski S (2009) Progression through the RNA polymerase II CTD cycle. Mol Cell 36:541–546

Bushnell DA, Westover KD, Davis RE, Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science 303:983–988

Cai G, Imasaki T, Yamada K, Cardelli F, Takagi Y, Asturias FJ (2010) Mediator head module structure and functional interactions. Nat Struct Mol Biol 17:273–279

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38:626–635

Carrington M, Roditi I, Williams RO (1987) The structure and transcription of an element interspersed between tandem arrays of mini-exon donor RNA genes in *Trypanosoma brucei*. Nucleic Acids Res 15:10179–10198

Chapman AB, Agabian N (1994) *Trypanosoma brucei* RNA polymerase II is phosphorylated in the absence of carboxyl-terminal domain heptapeptide repeats. J Biol Chem 269:4754–4760

Das A, Bellofatto V (2003) RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. Proc Natl Acad Sci USA 100:80–85

Das A, Bellofatto V (2009) The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in *Trypanosoma brucei*. PLoS One 4:e6959

Das A, Zhang Q, Palenchar JB, Chatterjee B, Cross GA, Bellofatto V (2005) Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. Mol Cell Biol 25:7314–7322

Das A, Li H, Liu T, Bellofatto V (2006) Biochemical characterization of *Trypanosoma brucei* RNA polymerase II. Mol Biochem Parasitol 150:201–210

de Lange T, Liu AY, Van der Ploeg LH, Borst P, Tromp MC, Van Boom JH (1983) Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell 34:891–900

Denninger V, Fullbrook A, Bessat M, Ersfeld K, Rudenko G (2010) The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in *Trypanosoma brucei*. Mol Microbiol 78:459–474

Devaux S, Lecordier L, Uzureau P, Walgraffe D, Dierick JF, Poelvoorde P, Pays E, Vanhamme L (2006) Characterization of RNA polymerase II subunits of *Trypanosoma brucei*. Mol Biochem Parasitol 148:60–68

Djikeng A, Ferreira L, D'Angelo M, Dolezal P, Lamb T, Murta S, Triggs V, Ulbert S, Villarino A, Renzi S, Ullu E, Tschudi C (2001) Characterization of a candidate *Trypanosoma brucei* U1 small nuclear RNA gene. Mol Biochem Parasitol 113:109–115

Donelson JE (2003) Antigenic variation and the African trypanosome genome. Acta Trop 85:391–404

Dunbar DA, Chen AA, Wormsley S, Baserga SJ (2000) The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. Nucleic Acids Res 28:2855–2861

Ehlers B, Czichos J, Overath P (1987) RNA turnover in *Trypanosoma brucei*. Mol Cell Biol 7:1242–1249

El Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309:409–415

Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M (2008) Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Mol Cell 31:337–346

Evers R, Hammer A, Köck J, Jess W, Borst P, Memet S, Cornelissen AW (1989) *Trypanosoma brucei* contains two RNA polymerase II largest subunit genes with an altered C-terminal domain [published erratum appears in Cell 58:419]. Cell 56:585–597

Fantoni A, Dare AO, Tschudi C (1994) RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. Mol Cell Biol 14:2021–2028

Figueiredo LM, Cross GA (2010) Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. Eukaryot Cell 9:148–154

Figueiredo LM, Janzen CJ, Cross GA (2008) A histone methyltransferase modulates antigenic variation in African trypanosomes. PLoS Biol 6:e161

Gilinger G, Bellofatto V (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Res 29:1556–1564

Gilinger G, Luo H, Bellofatto V (2004) *In vivo* transcription analysis utilizing chromatin immunoprecipitation reveals a role for trypanosome transcription factor PBP-1 in RNA polymerase III-dependent transcription. Mol Biochem Parasitol 134:169–173

Ginger ML, Blundell PA, Lewis AM, Browitt A, Günzl A, Barry JD (2002) *Ex vivo* and *in vitro* identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. Eukaryot Cell 1:1000–1009

Goldshmidt H, Matas D, Kabi A, Carmi S, Hope R, Michaeli S (2010) Persistent ER stress induces the spliced leader RNA silencing pathway (SLS), leading to programmed cell death in *Trypanosoma brucei*. PLoS Pathog 6:e1000731

Grummt I, Skinner JA (1985) Efficient transcription of a protein-coding gene from the RNA polymerase I promoter in transfected cells. Proc Natl Acad Sci USA 82:722–726

Günzl A (2003) Transcription. In: Marr JJ, Nielsen TW, Komuniecki RW (eds) Molecular medical parasitology. Academic Press, London, pp 47–65

Günzl A (2010) The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell 9:1159–1170

Günzl A, Schimanski B (2009) Tandem affinity purification of proteins. Curr Protoc Protein Sci Chapter 19:Unit 19.19

Günzl A, Ullu E, Dörner M, Fragoso SP, Hoffmann KF, Milner JD, Morita Y, Nguu EK, Vanacova S, Wünsch S, Dare AO, Kwon H, Tschudi C (1997) Transcription of the *Trypanosoma brucei* spliced leader RNA gene is dependent only on the presence of upstream regulatory elements. Mol Biochem Parasitol 85:67–76

Günzl A, Bruderer T, Laufer G, Schimanski B, Tu LC, Chung HM, Lee PT, Lee MG (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. Eukaryot Cell 2:542–551

Günzl A, Vanhamme L, Myler PJ (2007) Transcription in trypanosomes: a different means to the end. In: Barry JD, Mottram JC, McCulloch R, Acosta-Serrano A (eds) Trypanosomes—after the genome. Horizon Press, Pittsburgh, pp 177–208

Guo Z, Stiller JW (2004) Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. BMC Genomics 5:69

Hartree D, Bellofatto V (1995) Essential components of the mini-exon gene promoter in the trypanosomatid *Leptomonas seymouri*. Mol Biochem Parasitol 71:27–39

Henry RW, Sadowski CL, Kobayashi R, Hernandez N (1995) A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. Nature 374:653–656

Hernandez N (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. J Biol Chem 276:26733–26736

Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, Brooks K, Churcher C, Fahkro S, Goodhead I, Heath P, Kartvelishvili M, Mungall K, Harris D, Hauser H, Sanders M, Saunders D, Seeger K, Sharp S, Taylor JE, Walker D, White B, Young R, Cross GA, Rudenko G, Barry JD, Louis EJ, Berriman M (2008) Telomeric expression sites are highly conserved in *Trypanosoma brucei*. PLoS One 3:e3527

Hughes K, Wand M, Foulston L, Young R, Harley K, Terry S, Ersfeld K, Rudenko G (2007) A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. EMBO J 26:2400–2410

Ibrahim BS, Kanneganti N, Rieckhof GE, Das A, Laurents DV, Palenchar JB, Bellofatto V, Wah DA (2009) Structure of the C-terminal domain of transcription factor IIB from *Trypanosoma brucei*. Proc Natl Acad Sci USA 106:13242–13247

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble 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 JC, Müller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schäfer M, Schein J, Schwartz DC, 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 DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ (2005) The genome of the kinetoplastid parasite, *Leishmania major*. Science 309:436–442

Janz L, Clayton C (1994) The PARP and rRNA promoters of *Trypanosoma brucei* are composed of dissimilar sequence elements that are functionally interchangeable. Mol Cell Biol 14:5804–5811

Jawhari A, Uhring M, De Carlo S, Crucifix C, Tocchini-Valentini G, Moras D, Schultz P, Poterszman A (2006) Structure and oligomeric state of human transcription factor TFIIE. EMBO Rep 7:500–505

Jess W, Hammer A, Cornelissen AW (1989) Complete sequence of the gene encoding the largest subunit of RNA polymerase I of *Trypanosoma brucei*. FEBS Lett 249:123–128

Kaiser FJ, Tavassoli K, Van den Bemd GJ, Chang GT, Horsthemke B, Moroy T, Lüdecke HJ (2003) Nuclear interaction of the dynein light chain LC8a with the TRPS1 transcription factor suppresses the transcriptional repression activity of TRPS1. Hum Mol Genet 12:1349–1358

Kawahara T, Siegel TN, Ingram AK, Alsford S, Cross GA, Horn D (2008) Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. Mol Microbiol 69:1054–1068

Kelly S, Wickstead B, Gull K (2005) An *in silico* analysis of trypanosomatid RNA polymerases: insights into their unusual transcription. Biochem Soc Trans 33:1435–1437

Köck J, Evers R, Cornelissen AW (1988) Structure and sequence of the gene for the largest subunit of trypanosomal RNA polymerase III. Nucleic Acids Res 16:8753–8772

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 6:e1001090

Kooter JM, Borst P (1984) Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucleic Acids Res 12:9457–9472

Kuhlman TC, Cho H, Reinberg D, Hernandez N (1999) The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Mol Cell Biol 19:2130–2141

Laufer G, Günzl A (2001) *In-vitro* competition analysis of procyclin gene and variant surface glycoprotein gene expression site transcription in *Trypanosoma brucei*. Mol Biochem Parasitol 113:55–65

Lecordier L, Devaux S, Uzureau P, Dierick JF, Walgraffe D, Poelvoorde P, Pays E, Vanhamme L (2007) Characterization of a TFIIH homologue from *Trypanosoma brucei*. Mol Microbiol 64:1164–1181

Lee MG (1996) An RNA polymerase II promoter in the hsp70 locus of *Trypanosoma brucei*. Mol Cell Biol 16:1220–1230

Lee JH, Nguyen TN, Schimanski B, Günzl A (2007) Spliced leader RNA gene transcription in *Trypanosoma brucei* requires transcription factor TFIIH. Eukaryot Cell 6:641–649

Lee JH, Jung HS, Günzl A (2009) Transcriptionally active TFIIH of the early-diverged eukaryote *Trypanosoma brucei* harbors two novel core subunits but not a cyclin-activating kinase complex. Nucleic Acids Res 37:3811–3820

Lee JH, Cai G, Panigrahi AK, Dunham-Ems S, Nguyen TN, Radolf JD, Asturias FJ, Günzl A (2010) A TFIIH-associated mediator head is a basal factor of small nuclear spliced leader RNA gene transcription in early-diverged trypanosomes. Mol Cell Biol 30:5502–5513

Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumph WE (2004) Architectural arrangement of cloned proximal sequence element-binding protein subunits on *Drosophila* U1 and U6 snRNA gene promoters. Mol Cell Biol 24:1897–1906

Liang XH, Uliel S, Hury A, Barth S, Doniger T, Unger R, Michaeli S (2005) A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. RNA 11:619–645

Luo H, Gilinger G, Mukherjee D, Bellofatto V (1999) Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two DNA-binding proteins and a tripartite architecture that includes an initiator element. J Biol Chem 274:31947–31954

Lustig Y, Sheiner L, Vagima Y, Goldshmidt H, Das A, Bellofatto V, Michaeli S (2007) Spliced-leader RNA silencing: a novel stress-induced mechanism in *Trypanosoma brucei*. EMBO Rep 8:408–413

Malik S, Roeder RG (2010) The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11:761–772

Marchetti MA, Tschudi C, Silva E, Ullu E (1998) Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersionof RNA polymerase II- and III-transcribed genes. Nucleic Acids Res 26:3591–3598

Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. Mol Cell 11:1291–1299

Martinez-Calvillo S, Nguyen D, Stuart K, Myler PJ (2004) Transcription initiation and termination on *Leishmania major* chromosome 3. Eukaryot Cell 3:506–517

Martinez-Calvillo S, Saxena A, Green A, Leland A, Myler PJ (2007) Characterization of the RNA polymerase II and III complexes in *Leishmania major*. Int J Parasitol 37:491–502

Martinez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martinez LE, Manning-Cela RG, Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010:525241

Maxon ME, Goodrich JA, Tjian R (1994) Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev 8:515–524
McAndrew M, Graham S, Hartmann C, Clayton C (1998) Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. Exp Parasitol 90:65–76
McConville MJ, Ferguson MA (1993) The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. Biochem J 294:305–324
Mottram JC, Bell SD, Nelson RG, Barry JD (1991) tRNAs of *Trypanosoma brucei*. Unusual gene organization and mitochondrial importation. J Biol Chem 266:18313–18317
Nakaar V, Dare AO, Hong D, Ullu E, Tschudi C (1994) Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. Mol Cell Biol 14:6736–6742
Nakaar V, Tschudi C, Ullu E (1995) An unusual liaison: small nuclear and cytoplasmic RNA genes team up with tRNA genes in trypanosomatid protozoa. Parasitol Today 11:225–228
Nakaar V, Günzl A, Ullu E, Tschudi C (1997) Structure of the *Trypanosoma brucei* U6 snRNA gene promoter. Mol Biochem Parasitol 88:13–23
Navarro M, Gull K (2001) A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. Nature 414:759–763
Nelson RG, Parsons M, Barr PJ, Stuart K, Selkirk M, Agabian N (1983) Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in *Trypanosoma brucei*. Cell 34:901–909
Nett IR, Martin DM, Miranda-Saavedra D, Lamont D, Barber JD, Mehlert A, Ferguson MA (2009) The phosphoproteome of bloodstream form *Trypanosoma brucei*, causative agent of African sleeping sickness. Mol Cell Proteomics 8:1527–1538
Nguyen TN, Schimanski B, Zahn A, Klumpp B, Günzl A (2006) Purification of an eight subunit RNA polymerase I complex in *Trypanosoma brucei*. Mol Biochem Parasitol 149:27–37
Nguyen TN, Schimanski B, Günzl A (2007) Active RNA polymerase I of *Trypanosoma brucei* harbors a novel subunit essential for transcription. Mol Cell Biol 27:6254–6263
Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG, Burley SK (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature 377:119–128
Nunes LR, Carvalho MR, Shakarian AM, Buck GA (1997) The transcription promoter of the spliced leader gene from *Trypanosoma cruzi*. Gene 188:157–168
Orphanides G, Wu WH, Lane WS, Hampsey M, Reinberg D (1999) The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature 400:284–288
Palenchar JB, Bellofatto V (2006) Gene transcription in trypanosomes. Mol Biochem Parasitol 146:135–141
Palenchar JB, Liu W, Palenchar PM, Bellofatto V (2006) A divergent transcription factor TFIIB in trypanosomes is required for RNA polymerase II-dependent spliced leader RNA transcription and cell viability. Eukaryot Cell 5:293–300
Park S, Nguyen TN, Kirkham JK, Lee J, Günzl A (2011) Transcription by the multifunctional RNA polymerase I in *Trypanosoma brucei* functions independently of RPB7. Mol Biochem Parasitol 180:35–42
Park S, Nguyen TN, Günzl A (2012) Development of an efficient in vitro transcription system for bloodstream form *Trypanosoma brucei* reveals life cycle-independent functionality of class I transcription factor A. Mol Biochem Parasitol 181:29–36
Parsons M, Worthey EA, Ward PN, Mottram JC (2005) Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. BMC Genomics 6:127
Patrick KL, Luz PM, Ruan JP, Shi H, Ullu E, Tschudi C (2008) Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference-deficient *Trypanosoma brucei*. Mol Microbiol 67:435–447
Penate X, Lopez-Farfan D, Landeira D, Wentland A, Vidal I, Navarro M (2009) RNA pol II subunit RPB7 is required for RNA pol I-mediated transcription in *Trypanosoma brucei*. EMBO Rep 10:252–257

Pham VP, Qi CC, Gottesdiener KM (1996) A detailed mutational analysis of the VSG gene expression site promoter. Mol Biochem Parasitol 75:241–254

Respuela P, Ferella M, Rada-Iglesias A, Aslund L (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. J Biol Chem 283:15884–15892

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 17:1030–1032

Roditi I, Clayton C (1999) An unambiguous nomenclature for the major surface glycoproteins of the procyclic form of *Trypanosoma brucei*. Mol Biochem Parasitol 103:99–100

Ruan JP, Arhin GK, Ullu E, Tschudi C (2004) Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. Mol Cell Biol 24:9610–9618

Rudenko G, Chung HM, Pham VP, Van der Ploeg LH (1991) RNA polymerase I can mediate expression of CAT and neo protein- coding genes in *Trypanosoma brucei*. EMBO J 10:3387–3397

Ruepp S, Furger A, Kurath U, Renggli CK, Hemphill A, Brun R, Roditi I (1997) Survival of *Trypanosoma brucei* in the tsetse fly is enhanced by the expression of specific forms of procyclin. J Cell Biol 137:1369–1379

Schimanski B, Klumpp B, Laufer G, Marhöfer RJ, Selzer PM, Günzl A (2003) The second largest subunit of *Trypanosoma brucei's* multifunctional RNA polymerase I has a unique N-terminal extension domain. Mol Biochem Parasitol 126:193–200

Schimanski B, Laufer G, Gontcharova L, Günzl A (2004a) The *Trypanosoma brucei* spliced leader RNA and rRNA gene promoters have interchangeable TbSNAP50-binding elements. Nucleic Acids Res 32:700–709

Schimanski B, Ma L, Günzl A (2004b) Failure to detect binding of *Trypanosoma brucei* SNAPc to U2 and U6 snRNA gene sequences by in vitro transcription competition and pull-down assays. Mol Biochem Parasitol 137:293–296

Schimanski B, Nguyen TN, Günzl A (2005a) Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. Mol Cell Biol 25:7303–7313

Schimanski B, Nguyen TN, Günzl A (2005b) Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. Eukaryot Cell 4:1942–1950

Schimanski B, Brandenburg J, Nguyen TN, Caimano MJ, Günzl A (2006) A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. Nucleic Acids Res 34:1676–1684

Schramm L, Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. Genes Dev 16:2593–2620

Schultz P, Fribourg S, Poterszman A, Mallouh V, Moras D, Egly JM (2000) Molecular structure of human TFIIH. Cell 102:599–607

Schwede A, Jones N, Engstler M, Carrington M (2010) The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Mol Biochem Parasitol 175:201–204

Sherman DR, Janz L, Hug M, Clayton C (1991) Anatomy of the PARP gene promoter of *Trypanosoma brucei*. EMBO J 10:3379–3386

Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GA (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. Genes Dev 23:1063–1076

Sikorski TW, Buratowski S (2009) The basal initiation machinery: beyond the general transcription factors. Curr Opin Cell Biol 21:344–351

Smith JL, Levin JR, Agabian N (1989a) Molecular characterization of the *Trypanosoma brucei* RNA polymerase I and III largest subunit genes. J Biol Chem 264:18091–18099

Smith JL, Levin JR, Ingles CJ, Agabian N (1989b) In trypanosomes the homolog of the largest subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain structure. Cell 56:815–827

Stanne TM, Rudenko G (2010) Active VSG expression sites in *Trypanosoma brucei* are depleted of nucleosomes. Eukaryot Cell 9:136–147

Stewart M, Haile S, Jha BA, Cristodero M, Li CH, Clayton C (2010) Processing of a phosphoglycerate kinase reporter mRNA in *Trypanosoma brucei* is not coupled to transcription by RNA polymerase II. Mol Biochem Parasitol 172:99–106

Sturm NR, Yu MC, Campbell DA (1999) Transcription termination and 3′-End processing of the spliced leader RNA in kinetoplastids. Mol Cell Biol 19:1595–1604

Takagi Y, Kornberg RD (2006) Mediator as a general transcription factor. J Biol Chem 281:80–89

Takagi Y, Calero G, Komori H, Brown JA, Ehrensberger AH, Hudmon A, Asturias F, Kornberg RD (2006) Head module control of mediator interactions. Mol Cell 23:355–364

Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41:105–178

Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ (2009) Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. BMC Genomics 10:e152

Tsai FT, Sigler PB (2000) Structural basis of preinitiation complex assembly on human pol II promoters. EMBO J 19:25–36

Uzureau P, Daniels JP, Walgraffe D, Wickstead B, Pays E, Gull K, Vanhamme L (2008) Identification and characterization of two trypanosome TFIIS proteins exhibiting particular domain architectures and differential nuclear localizations. Mol Microbiol 69:1121–1136

Vanhamme L, Pays A, Tebabi P, Alexandre S, Pays E (1995) Specific binding of proteins to the noncoding strand of a crucial element of the variant surface glycoprotein, procyclin, and ribosomal promoters of *Trypanosoma brucei*. Mol Cell Biol 15:5598–5606

Vanhamme L, Poelvoorde P, Pays A, Tebabi P, Xong HV, Pays E (2000) Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. Mol Microbiol 36:328–340

Walgraffe D, Devaux S, Lecordier L, Dierick JF, Dieu M, Van Den AJ, Pays E, Vanhamme L (2005) Characterization of subunits of the RNA polymerase I complex in *Trypanosoma brucei*. Mol Biochem Parasitol 139:249–260

Wang QP, Kawahara T, Horn D (2010) Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Mol Microbiol 77:1237–1245

Werner F (2007) Structure and function of archaeal RNA polymerases. Mol Microbiol 65:1395–1404

Wright JR, Siegel TN, Cross GA (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. Mol Biochem Parasitol 172:141–144

Yang X, Figueiredo LM, Espinal A, Okubo E, Li B (2009) RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*. Cell 137:99–109

Yeh TY, Chuang JZ, Sung CH (2005) Dynein light chain rp3 acts as a nuclear matrix-associated transcriptional modulator in a dynein-independent pathway. J Cell Sci 118:3431–3443

Yoon JB, Murphy S, Bai L, Wang Z, Roeder RG (1995) Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Mol Cell Biol 15:2019–2027

Yu MC, Sturm NR, Saito RM, Roberts TG, Campbell DA (1998) Single nucleotide resolution of promoter activity and protein binding for the *Leishmania tarentolae* spliced leader RNA gene. Mol Biochem Parasitol 94:265–281

Ziegelbauer K, Overath P (1993) Organization of two invariant surface glycoproteins in the surface coat of *Trypanosoma brucei*. Infect Immun 61:4540–4545

Zomerdijk JC, Kieft R, Borst P (1991) Efficient production of functional mRNA mediated by RNA polymerase I in *Trypanosoma brucei*. Nature 353:772–775

Chapter 2
SL RNA Biogenesis in Kinetoplastids: A Long and Winding Road

Nancy R. Sturm, Jesse R. Zamudio, and David A. Campbell

Abstract The spliced leader (SL) RNA is a defining element in the gene expression of kinetoplastids. The first 39 nt of this small RNA are *trans*-spliced onto every nuclear message, providing a unique hypermethylated cap and sequence elements required for stability and translation. Transcribed from a large tandem array, the journey that each primary SL transcript takes en route to splicing is marked by molecular modifications. Methylation, pseudouridylation, and 3′-end nuclease processing contribute to the mature product. The consequences of this elaborate pathway are not understood fully, but may reveal distinctions that will make these oft-parasitic organisms yield their foothold in the vertebrate host.

2.1 Introduction

The spliced leader (SL) RNA defines the 5′ end of every nuclear mRNA in kinetoplastids, accounting for a huge chunk of transcriptional activity in the cell (Haanstra et al. 2008). In a multistep process, these small transcripts are extensively modified, including hypermethylation of the 5′-cap structure $m^7Gpppm^{6,6}AmpAmpCmpm^3Um$ referred to as cap 4, pseudouridylation at nucleotide 28, and removal of a 3′-extended tail (Bangs et al. 1992; Liang et al. 2002; Sturm et al. 1999).

N.R. Sturm (✉) • D.A. Campbell
Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA
e-mail: nsturm.ucla@gmail.com; dc@ucla.edu

J.R. Zamudio
Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

Koch Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: jzamudio@mit.edu

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_2,

The SL RNA is required for kinetoplastid gene expression because of the strategy these organisms adopted for the generation of their mRNA. Transcription of protein-coding genes is polycistronic, with processing to essentially typical eukaryotic monocistronic mRNAs occurring through coupled *trans*-splicing and polyadenylation. *Trans*-splicing allows the use of RNA polymerase I in addition to RNA polymerase II for the synthesis of mRNA, as the mRNA cap structure is provided to both by the capped SL RNA transcript. Under normal circumstances, gene regulation at the level of transcription initiation is minimal, if at all, in these organisms (Clayton 2002). The predominant absence of transcriptional control sets the kinetoplastids apart from their host organisms and shunts the control of gene expression further downstream in the chain of events. As such, mRNA processing and stability, translation, and posttranslational signals functionally regulate the levels of any particular protein (Lahav et al. 2011).

In this review of SL RNA biogenesis, we will follow the path of the SL RNA from the gene to the mRNA. Although several steps in the process have been characterized in recent years, much remains to be understood about the underlying mechanisms.

2.2 Genomic Organization

The SL RNA genes in kinetoplastids are found in extensive head-to-tail tandem arrays (Fig. 2.1) that are sometimes interspersed with 5S rRNA genes (Santana et al. 2001). Gene copy number is in the range of 100–200 (Aksoy et al. 1992), making SL an excellent marker for kinetoplastid taxonomy (Murthy et al. 1992; Westenberger et al. 2004; Maslov et al. 2007). The arrays are located in one to two distinct nucleoplasmic locations proximal to the nucleolus (Dossin and Schenkman 2005; Hitchcock et al. 2007; Uzureau et al. 2008). Among different genera and/or species, the transcribed and nontranscribed regions of the array show substantial sequence variation, except for nucleotides 1–6 and 21–39 of the SL. Within a taxon, the active use of copy correction in the maintenance of gene and array integrity is indicated by the prevalence of specific alterations throughout given arrays, particularly in the intergenic regions, but also in the SL RNA gene itself (Thomas et al. 2005). The presence of a self-cleaving ribozyme in a transcript encompassing the SL RNA and 5S rRNA sequences of *Diplonema papillatum* hints at further RNA function and activity derived from these multicopy repeats (Webb et al. 2009).

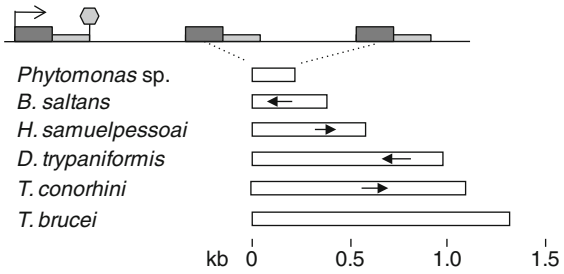


Fig. 2.1 Organization of the kinetoplastid SL RNA gene array. SL RNA genes are found as tandem head-to-tail arrays in all kinetoplastid genomes. Conserved SL sequences are shown as *dark gray boxes* and variable intron sequences are shown as *light gray boxes*. Transcription start sites and termination sites are indicated by a line *arrowhead* and a *hexagon*, respectively. The *single line* represents a variable length nontranscribed spacer that may contain a 5S ribosomal RNA gene (*solid arrowhead* within *rectangle*) transcribed by RNA polymerase III in either orientation relative to the SL RNA gene. SL RNA gene repeats representing the extremes of length and all combinations of SL RNA and 5S rRNA are shown for the trypanosomatids *Herpetomonas samuelpessoai*, *Trypanosoma brucei*, *T. conorhini*, the plant trypanosome *Phytomonas* (phloem-restricted group), and the bodonids *Bodo saltans* and *Dimastigella trypaniformis*

2.3 Transcription and Cotranscriptional Processing

2.3.1 Promoter

Two upstream elements drive SL RNA transcription by RNA polymerase II (Yu et al. 1998; Campbell et al. 2003; Günzl 2003; Palenchar and Bellofatto 2006; Martínez-Calvillo et al. 2010). A third element referred to as an initiator was defined in *Leptomonas seymouri* (Das et al. 2008), however it does not function independently of other upstream elements in bona fide initiator fashion; the equivalent sequences in *L. tarentolae* affect transcript stability, but not transcription (Saito et al. 1994). The basal transcription factors, core small nuclear (sn) RNA gene transcription factors, and the mediator complex (Lee et al. 2010) associated with SL RNA gene transcription have been identified (Das et al. 2008; Günzl 2010b; see Chap. 1 by Günzl). The promoter and coding regions are free of nucleosomes (Hitchcock et al. 2007), as anticipated for highly transcribed genes. Under stress conditions, transcription of the SL RNA genes can be shut down, most likely through the degradation of transcription factors, via a process termed spliced leader silencing or SLS (Goldshmidt et al. 2010).

2.3.2 m^7G (Cap 0) Formation

SL RNA is modified early in its synthesis by a trio of activities starting with RNA triphosphatase TbCet1 (Ho and Shuman 2001), followed by RNA

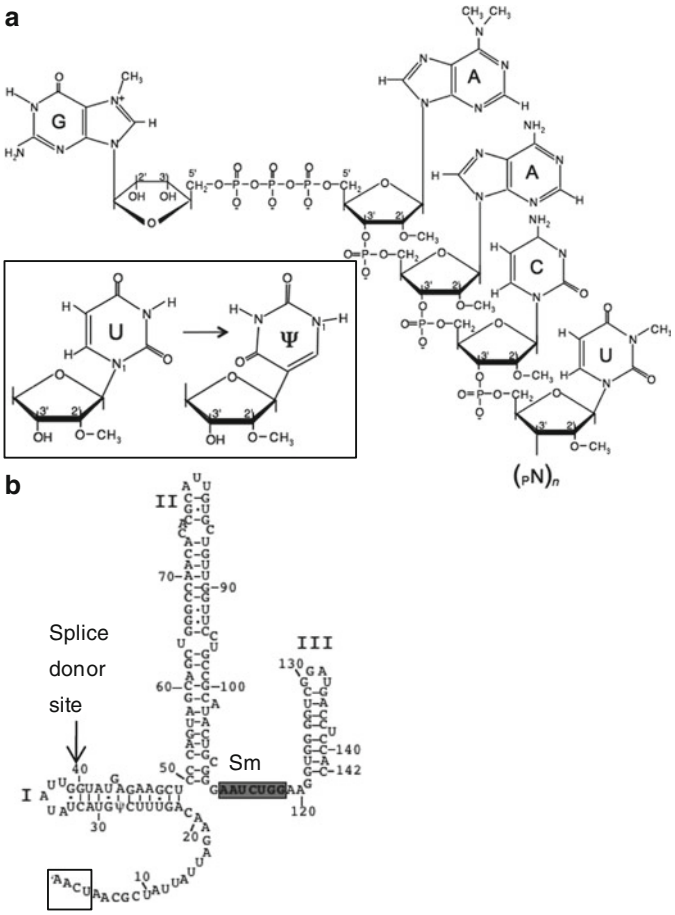


Fig. 2.2 RNA modifications and secondary structure of the SL RNA. (**a**) Chemical composition of the hypermethylated cap 4, which includes the following modifications: m^7G in the inverted cap 0 structure; 2′-*O*-ribose methylation of the first four transcribed nucleotides; ^{6,6}A-dimethylation of the first base; and ^{3}U methylation of the fourth base. The *inset* shows rotation of the uracil ring in Ψ_{28} formation from N1 to C5 relative to the C1 of ribose. (**b**) Secondary structure model of the *T. brucei* SL RNA. Stem loops I, II, and III are indicated, as are the splice-donor site and the single-stranded Sm-binding site. Cap 4 nucleotides are *boxed*

guanylyltransferase/(guanine N-7) methyltransferase, provided by the bifunctional protein TbCgm1 specialized for SL RNA processing (Ruan et al. 2007; Takagi et al. 2007). Chromatin immunoprecipitation studies support cotranscriptional recruitment resulting in the formation of the cap 0 structure on nascent SL. The precise cellular localization and functional significance of what follows is the subject of interest and debate, as studies of cap 4, the most extensive cap structure described in any system (Fig. 2.2a), and other unique aspects of SL RNA maturation, are critical for efficient function. Cotranscriptional cap 4 formation is championed by some

(Mair et al. 2000; Hury et al. 2009), while others argue for dissociation of the downstream methylations from RNA polymerase II activity (Zeiner et al. 2003a, 2004a; Zamudio et al. 2007).

2.3.3 Termination

A run of T nucleotides is the common downstream limit of the SL RNA gene, along with many other small RNA genes. This T tract leads to attenuated termination of transcription in *L. tarentolae* (Sturm et al. 1999) resulting in a primary transcript with a 3′-polyuridylate tail. The poly(U)-binding protein La has minimal or no effect on the maturation of the SL RNA (Foldynová-Trantírková et al. 2005; Arhin et al. 2005). The unpaired nucleotides at the 3′ end of the nascent SL RNA are trimmed off later in SL RNA biogenesis.

In different kinetoplastids, the mature SL RNA transcript varies in size between 96 and 142 nt (Hitchcock et al. 2004) and folds into a secondary structure consisting of three stem loops (Fig. 2.2b) that is common to SL RNAs from nematodes and many other *trans*-splicing eukaryotes (Bruzik et al. 1988); exceptions to this structural rule include the chordates *Ciona* and *Oikopleura* (Vandenberghe et al. 2001; Ganot et al. 2004) and the *Perkinsus*/dinoflagellate lineage (Zhang et al. 2007, 2011). The majority of the exon is single stranded (Harris et al. 1995), with the loop of the first stem loop adjacent to the splice donor site. The second and third stem loops, which can vary considerably in length and structure, flank a consensus single-stranded Sm-binding site that is required for SL RNA maturation and function.

Following cap 0 addition, the SL RNA embarks on a quest to obtain the additional modifications required for its maturation. How the molecule is directed through this pathway is largely undefined. The nonessential nature of the individual cap 4 modifications was surprising (Arhin et al. 2006a, b; Zamudio et al. 2006, 2007); however, the cumulative effects of these methylations are subtle but potentially vital in the natural setting (Zamudio et al. 2009a). An early step in maturation involves an SL RNA processing complex specialized for Sm-protein independent modifications in biogenesis.

2.4 Cap 1/Pseudouridylation

The activities of two SL RNA modification enzymes are linked in a molecular complex (Zamudio et al. 2009b) that resides in or near the nucleolus.

2.4.1 Pseudouridylation

Pseudouridine (Ψ) is a common RNA nucleotide modification resulting from the rotation of a uridine base relative to its ribose ring (Fig. 2.2a inset). These modifications are ubiquitous on tRNAs, snRNAs, and rRNAs, and optimize structural base-pairing interactions. The pseudouridylation at nucleotide 28 (Ψ_{28}) of SL RNA occurs prior to *trans*-splicing on the SL RNA substrate molecule, resulting in the presence of this modification on every mRNA. When and where this modification plays its biological role is unclear, but RNA–RNA interactions at both the substrate and mRNA levels leave many options.

SL RNA pseudouridylation is directed by the small nucleolar (sno) RNA SLA1 (Watkins et al. 1994; Liang et al. 2002; Hury et al. 2009). SLA1 snoRNA is transcribed with a group of three other small RNAs by RNA polymerase II (Roberts et al. 1996, 1998; Xu et al. 2001), and possesses an unusual, but undetermined 5′-cap structure (Watkins et al. 1994). Ablation of the pseudouridine synthase TbCBF5 eliminates Ψ_{28} formation on SL RNA and results in incomplete cap 4 formation (Barth et al. 2005); attempts at snoRNA silencing of the SLA1 RNA were not successful (Liang et al. 2003b). Mutagenesis of the exon at position 28 yielded no catastrophic phenotype (Sturm et al. 1998; Liang et al. 2002), however minor effects cannot be ruled out. Colocalization of SLA1 RNA with the transcription factor tSNAP42 is consistent with an early step in SL processing (Hury et al. 2009).

2.4.2 Cap 1 2′-O-Ribose Methylation

The enzyme catalyzing the cap 1 2′-*O*-ribose methylation in *T. brucei*, TbMTr1 (Zamudio et al. 2007), is a member of the Rossman Fold family of methyltransferases (Feder et al. 2003) that possesses a cap-binding domain (Mittra et al. 2008) and is the first of a series of enzymes responsible for formation of the hypermethylated cap 4. TbMTr1 methyltransferase activity is optimized for the SL RNA primary sequence (Mittra et al. 2008), which is shared by the U1 snRNA substrate at six of the first seven nucleotides. A related cap 1 methyltransferase was confirmed subsequently in humans (Bélanger et al. 2010; Werner et al. 2011). Although *T. brucei* cells can survive in culture without TbMTr1 (Zamudio et al. 2007), the maturation profile of SL RNA shifts in its absence with the accumulation of cap 0 and 3′-extended intermediates, indicating that movement through the biogenesis pathway is hindered. The mRNA population in TbMTr1$^{-/-}$ cells is methylated at the remaining cap 4 positions, demonstrating that downstream capping activities are efficient once the substrate processing delay is overcome. Accumulation of underprocessed SL RNA is not observed with knockout of the downstream Sm-protein-dependent cap methyltransferases, supporting distinct steps and localizations for maturation. Neither TbMTr1$^{-/-}$ mRNA stability nor translatability is affected in TbMTr1$^{-/-}$ background (Zamudio et al. 2009a).

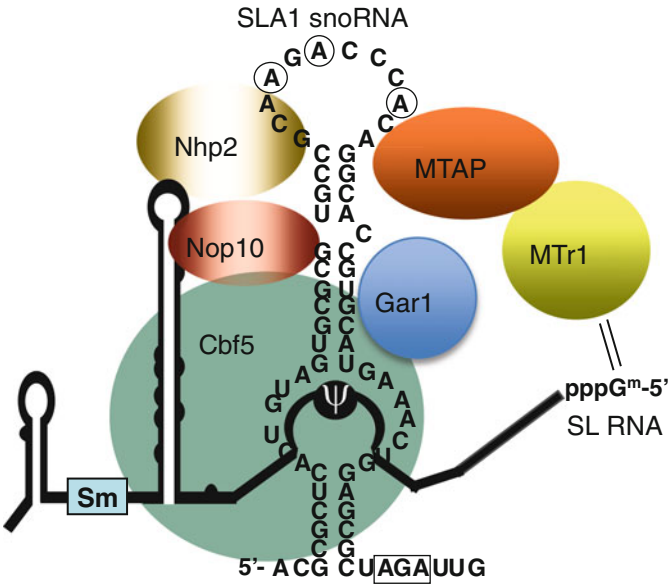


Fig. 2.3 Early modification of the SL RNA is performed by a bifunctional complex. Composition of the ribose-cap 1 methyltransferase/pseudouridylation-SLA1 snoRNA complex. The conserved pseudouridylation complex consists of four proteins (Cbf5, containing the enzymatic activity; Nop10; Gar1; and Nhp2) and a boxH/ACA small nucleolar RNA (snoRNA) scaffold (Hamma and Ferre-D'Amare 2006). The SLA1 snoRNA provides specificity for the Ψ_{28} modification by base pairing with the SL RNA (Liang et al. 2002). *Circled* nucleotides in the SLA1 RNA indicate the conserved box H; boxed nucleotides represent the AGA motif found in kinetoplastids rather than the canonical ACA (Myslyuk et al. 2008). The cap 1 methyltransferase, TbMTr1, is thought to associate with the pseudouridylation complex via the MTAP protein. Cap-binding activity of the TbMTr1 is indicated by the *double lines*. Thus far, the SL RNA has not copurified with the complex. Figure reproduced with permission from Zamudio et al. (2009b)

Furthermore, the remaining cap 4 2′-*O*-ribose methyltransferases cannot be deleted from TbMTr1$^{-/-}$ cells (Zamudio et al. 2009a), suggesting a lack of redundancy in function.

Purification of complexes containing TbMTr1 revealed an association with the pseudouridylation machinery (Fig. 2.3) that contains the SLA1 snoRNA as a specific component (Zamudio et al. 2009b). Localization of a TbMTr1-GFP fusion shows nuclear accumulation (Zamudio et al. 2007), but TbMTr1 tagged with the HA epitope results in a discrete spot within the nucleus that is likely to represent the nucleolus (Zamudio, Sturm, and Campbell, unpublished data), consistent with the typical pseudouridylation site within the cell, and possibly representing an SL RNA processing center. Characterization of the TbMTr1/pseudouridylation complex identified a component designated TbMTAP for *m*ethyl*t*ransferase *a*ssociated *p*rotein that is required for the physical link between the two enzymatic activities. This complex is specific to SL RNA processing as the SLA1 snoRNA, but not other snoRNAs, is associated with the purified complex, and elimination of TbMTAP

reduces both SL RNA Ψ_{28} and cap 1 formation, but does not affect U1 snRNA cap 1 levels (Zamudio et al. 2009b). Initially, the TbMTr1/pseudouridylation association was thought to be yet another kinetoplastid-specific development to accommodate SL RNA processing; however, a homologous WD40-domain protein has been linked to Cajal body localization of small Cajal body RNAs or scaRNAs in humans (Tycowski et al. 2009).

Acquisition of early modifications involves the transporter exportin 1, or Xpo1 (Zeiner et al. 2003a; Biton et al. 2006). TbXpo1 inactivation by the toxin leptomycin B results in nuclear accumulation of SL RNA that is trapped in a cap 0-modified and 3′-end extended state, and cell death. Given the localization of TbMTr1 and the SL RNA cap 0 phenotype in leptomycin B-treated cells, TbXpo1 interaction with the SL RNA appears to occur before the cap 1 methylation, and suggests a function of TbXpo1 in the transport of the nascent transcript into the nucleolus for the first set of posttranscriptional modifications, similar to its role in U3 snoRNA trafficking (Boulon et al. 2004) rather than the proposed nuclear export function that predicted a cytosolic localization of the cap 1 enzyme (Zeiner et al. 2004b). In mammals, Xpo1 binds to neither the 5′ cap nor the nuclear cap-binding protein CBP20 directly (Ohno et al. 2000), implying that other players (such as PHAX in humans) remain to be identified. Whether a putative cap-binding protein is temporarily displaced by TbMTr1 or replaced by another chaperone for subsequent maturation steps is unclear.

2.5 Sm Association

The assembly of the heptameric Sm protein complex on the SL RNA represents a distinct stage in the maturation of substrate SL without which the molecule is stalled at a cap 1, 3′-end extended form that does not participate in *trans*-splicing. The composition of the SL RNP has been inferred by tandem-affinity tagging (Luz Ambrosio et al. 2009; Palfi et al. 2009; Tkacz et al. 2010). The seven canonical Sm proteins are associated with the SL RNA (Table 2.1). In contrast, variant SmD3 proteins Ssm4 and 16.5 K/Ssm2-2 are associated with the U4 and U2 snRNA, respectively, the variant SmB protein 15 K/Ssm2-1 with the U2 snRNA (Wang et al. 2006; Tkacz et al. 2007; Günzl 2010a; Jaé et al. 2010). The question of where this Sm-complex assembly occurs is controversial, with support on both the nuclear and cytosolic sides of the argument.

The cytosolic trafficking model invokes the nuclear egress of substrate SL RNA at the cap 1/3′-extended form to acquire the Sm-protein complex that then signals reimport of the SL RNP into the nucleus for completion of SL maturation. Mutation of the Sm-binding site (Sturm and Campbell 1999; Sturm et al. 1999) or knockdown of an Sm subunit (Mandelboim et al. 2003; Zeiner et al. 2004a) yielded similar immature/nonfunctional cytosolic SL RNA species, supporting the transit of SL RNA through the cytosol to acquire the Sm-protein complex in a manner analogous to the pathway in humans (Neuenkirchen et al. 2008). Copurification of importin

Table 2.1 Proteins involved in SL maturation

Sm-independent		Sm assembly		Sm-dependent	
Cap 0		SMN	Tb11.01.6640	**Cap 2**	
TbCet1	Tb927.3.2190	Gemin2	Tb927.10.5640	TbMTr2/TbCom1/ TbMT48	Tb11.02.2500
TbCgm1	Tb927.7.2080	SmB	Tb927.2.4540	TbMTr3/TbMT57	Tb09.211.3130
Cap 1		SmD3	Tb927.4.890	**Cap 4**	
TbMTr1	Tb927.10.7940	SmD1	Tb927.7.3120	[TbMTr3?][a]	
[TbMT^{66}A]		SmD2	Tb927.2.5850	[TbMT^{3}U]	
Ψ_{28}		SmF	Tb09.211.1695	**3′ end formation**	
TbCBF5	Tb10.100.0060	SmE	Tb927.6.2700	[poly(U) removal]	
TbGar1	Tb927.4.470	SmG	Tb11.01.5915	TbSNIP	Tb927.8.3710
TbNhp2	Tb927.4.750	PRP19	Tb927.2.5240		
TbNop10	Tb10.70.2465				
TbMTAP	Tb11.01.8210				
Other					
TbXpo1	Tb11.01.5940				

Square brackets [ ] represent proteins/activities that have yet to be identified
TbMTr3 may have dual activity including cap 4 ribose methylation

α with PTP-tagged SmB (Palfi et al. 2009) and SmD3 (Tkacz et al. 2010; Martinez, Zamudio, Sturm and Campbell, unpublished) may be a footprint of a trafficking pathway.

The nuclear Sm-assembly model was hypothesized as part of a cotranscriptional processing scenario (Mair et al. 2000) and is supported by colocalizations within a proposed SL RNP factory (Tkacz et al. 2007; Hury et al. 2009). The Sm-assembly factor SMN localizes to the nucleus also (Palfi et al. 2009; Jaé et al. 2011), consistent with nuclear Sm-complex formation. SMN specificity for assembly of the SL RNP, but not the U2 snRNP, has been demonstrated (Palfi et al. 2009) and is likely due to interaction of SMN with the canonical SmB/SmD3, but not the corresponding Sm variants found in the U2 snRNP. Additionally, the trypanosome SmB possesses a nuclear localization signal that is functional in human cells (Girard et al. 2004). Under this scenario, assembly of the Sm complex within the nucleus, likely after cap 1 2′-*O*-ribose methylation, would be required to transport the immature substrate to the next stage of processing. In this model, the cytosolic accumulation of SL RNA would be part of a turnover pathway for defective molecules termed spliced leader discard or SLD (Lustig et al. 2007; Goldshmidt et al. 2010). Conditional knockdown of the assembly factor SMN and the general splicing factor PRP19 also results in incomplete cap 4 formation on the SL RNA, but not cytosolic accumulation (Tkacz et al. 2010).

2.6 Cap2/Cap3/Cap4/3′ Processing

2.6.1 Sm-Dependent Cap 2′-O-Ribose Methylation

Two related enzymes supply the cap 2′-*O*-ribose methylations at positions 2 and 3, called TbMTr2 (aka TbCom1 and TbMT48) and TbMTr3 (aka TbMT57), respectively (Arhin et al. 2006a, b; Hall and Ho 2006; Zamudio et al. 2006). One possibility is that TbMTr3 may catalyze the position 4 methylation as well (Arhin et al. 2006a; Zamudio et al. 2006) as cap 3 is not seen in the absence of cap 4 in vivo; however, this remains to be demonstrated experimentally. In contrast to TbMTr1, phylogenetic analysis suggests a common origin with the Vaccinia virus cap 1 2′-*O*-ribose methyltransferase with subsequent specialization for modification of downstream nucleotides (Feder et al. 2003).

Generally diffuse, nucleoplasmic localizations for both enzymes are indicated, segregating their activities from that of the nucleolar TbMTr1/pseudouridylation steps. The arguments for cotranscriptional cap 4 formation include the finding that the position 2, 3, and 4 2′-*O*-ribose methylations did not appear on tethered, nascent transcripts until the SL RNA length reached position 117 and beyond in *T. brucei* (Mair et al. 2000). That position corresponds to the 3′-end of the Sm-binding site, so if these truncated transcripts were released from the transcription apparatus, the ability for the Sm-protein complex to form may have been the discriminating factor for the subsequent methylation events. The spatial separation of the cap 1 TbMTr1 in the nucleolus and TbMTr2 and TbMTr3 in the nucleoplasm as well as different processing defects upon cap methyltransferase depletion supports the original observation that cap 4 formation occurs posttranscriptionally (Ullu and Tschudi 1995).

Procyclic *T. brucei* survive the loss of either or both TbMTr2 and TbMTr3 in culture, and mRNA cap analysis determined that undermethylated substrate SL RNA are comparable substrates for *trans*-splicing suggesting downstream function. Consistently, translation rates decrease in proportion to the loss of cap methylation, an effect further exacerbated when the cells are under stress (Zamudio et al. 2009a). The basis for this reduction is unclear, but as stability of the SL RNA is unaffected, the answer could lie in reduced levels of cap recognition by the multitude of trypanosome cap-binding proteins (Li and Tschudi 2005; Dhalia et al. 2005; Yoffe et al. 2006) producing downstream effects on nuclear to cytoplasmic transport and/or translation initiation. The consequences of reduced translation rates are tolerated in the isolation of culture conditions; however, a cell population at such a disadvantage would be eliminated in short order by their brethren with fully capped SL.

SL RNA in *L. tarentolae*, *T. brucei*, and *L. collosoma* mutated in the central portion of the exon present intriguing combinations of cap 4 phenotypes (Lücke et al. 1996; Sturm et al. 1998; Xu et al. 2000) in light of the *T. brucei* capping-enzyme deletion mutants. Transcripts with sequence alterations in the exon region of nucleotides 10–29 are not recognized by TbMTr2 or TbMTr3. These transcripts

are *trans*-spliced (Sturm et al. 1998), but are not loaded into polysomes (Zeiner et al. 2003b). Since the cap 1 mRNA is translatable in *T. brucei*, the reason for the absence of polysome formation on the mutated *L. tarentolae* SL may not be a function of cap recognition, but of base complementarity with the translation machinery. Still, undermethylation suggests a delay in processing, possibly due to defective recognition by the cap 2, 3, and 4 2′-*O*-ribose methyltransferases. Another interesting possibility is the presence of a physical block on the cap 1 structure in the *L. tarentolae* cells that cannot be removed to allow further cap processing or translation in the exon-mutated lines. The presence of this block could be a useful, and likely essential, element in the cytosolic trafficking scenario, in which substrate SL could efficiently compete for mature mRNA, to the detriment of the translation rate. Removal of the 5′ block upon re-entry into the nucleus would expose the 5′ end to further cap 4 maturation and allow subsequent access by translation initiation factors. Hence, exon mutagenesis disrupted recognition by the same machinery that interacts with the central exon region, and the 5′ block is not effectively removed. In the nuclear Sm assembly scenario, shielding of the substrate SL RNA cap structure would equally be a necessary precaution to prevent premature export of the unspliced molecule.

2.6.2 3′ Processing

Maturation of the SL RNA 3′ end involves the removal of the last 4–6 transcribed nucleotides and is performed by a minimum of two nucleases. The identity of the first of these enzymes is unknown. The final product is created by a general nuclease, named SNIP for *sn*RNA *i*ncomplete *p*rocessing (Zeiner et al. 2004b), that trims off the final nucleotide or two from a variety of snRNAs including the SL RNA, where the mature end is defined by the base of stem loop III. Knockdown of TbSNIP is lethal; however, the direct cause of this effect is unclear as multiple substrates are affected. As 3′ processing does not initiate until after Sm-complex formation, either some physical barrier exists to prevent association of the immature substrate SL and the 3′ nucleases, or the physical binding by the Sm proteins is required for the stable formation of stem loop III and definition of the mature 3′ end. If a spatial separation takes place within the nuclear environment or via partitioning by the nuclear envelope remains to be determined.

2.6.3 Base Methylation

In addition to the 4 cap 2′-*O*-ribose methylations, the SL receives three additional methylations, two on the first base ($m^{6,6}A$), and one on the fourth that is unique to kinetoplastids (m^3U). The identity of the methyltransferases performing these modifications is unknown, but their activity appears to follow a 5′-to-3′ polarity

(Mair et al. 2000) and is reduced in the absence of the 2′-*O*-ribose methylations (Zamudio et al. 2009a). Two candidate $m^{6,6}A$ dimethyltransferases are present in kinetoplastid genomes that are thought to act upon the ribosomal RNAs of the cytosol and mitochondrion. If the nuclear enzyme also catalyzes base methylation on the SL RNA remains to be determined. The likely nucleolar localization of this enzyme for both SL RNA processing in tandem with cap 1 2′-*O*-ribose methylation and ribosomal modification is consistent with a dual substrate function. The distinctive features that the substrates present to the enzyme, with this conserved ribosomal modification found near the 3′ end of the small subunit versus the extreme 5′-end position of the SL RNA site, cast some doubt on a combined action. No specific candidates have come to light for the m^3U methylation, but other methyltransferases with undetermined specificities remain to be explored experimentally.

2.6.4 Cap 4 Binding by CBP20

In mammalian cells the CBP20 component of the nuclear cap-binding complex CBC is involved in transport of snRNAs to the cytosol. In trypanosomes, the CBP20 homolog preferentially binds cap 4 over cap 0, suggesting its interaction with the SL RNA is late in the biogenesis pathway (Li and Tschudi 2005). RNAi inhibition of TbCBP20 results in the accumulation of the SL RNA, suggesting it functions in bringing the mature SL RNP to the site of *trans*-splicing (Li and Tschudi 2005). By analogy with the mammalian homolog, TbCBC may accompany *trans*-spliced mRNA to the cytosol (Li and Tschudi 2005). The interaction of TbCBP20 with substrate SL, mRNA, or both SL forms remains to be demonstrated in vivo.

2.7 *Trans*-Splicing

From the coding region perspective, the requirements for *trans*-splicing are defined loosely as the presence of a GU splice-donor site in the SL RNA and two elements in the pre-mRNA: an AG splice-acceptor site and an upstream polypyrimidine tract (Matthews et al. 1994). Transcriptome analysis documents differential usage of alternative splice acceptor sites for a large number of genes (Siegel et al. 2010; Nilsson et al. 2010; Kolev et al. 2010), while experimental and computational approaches should elucidate the mechanistic rules for the selection of specific sites (Benz et al. 2005; Gopal et al. 2005; Siegel et al. 2005). These splicing events have the potential to impact transcript translation through control of stability, localization, and translatability based on the sequence present or absent from the 5′ untranslated region of the mRNA. The much-anticipated development of an

in vitro *trans*-splicing system (Shaked et al. 2010) will enable functional testing of the *trans*-splicing mechanism.

Mechanistically, *trans*-splicing is similar to *cis*-splicing (Liang et al. 2003a; see Chap. 3 by Jaé et al.) except that the SL RNA itself appears to replace the U1 snRNA as a component of the spliceosome. Any truly catalytic residues of SL RNA have not been determined, but some of the intron residues in the vicinity of stem loop II have been implicated. Exon-tagged SL that carry mutations in the 40–50 nt region yield molecules that are processed accurately, but not *trans*-spliced in *L. tarentolae* (Sturm and Campbell 1999). Various proteins copurify with tagged Sm proteins representing spliceosomal snRNPs and general splicing factors (Luz Ambrosio et al. 2009; Palfi et al. 2009; Tkacz et al. 2010). Unlike for the nematode SL RNP (Denker et al. 2002; MacMorris et al. 2007), no proteins unique to the kinetoplastid SL RNP have been identified so far.

2.8 Translation

The eukaryotic translation process begins with an interaction between the mRNA cap and cap binding protein eIF4E, which displaces the CBC in higher eukaryotes. Kinetoplastids follow this convention, but possess a wider array of eIF4E family members with a variety of cap binding affinities for cap 0 through cap 4 (Clayton and Shapira 2007; see Chapt. 9 by Shapira). Likewise, other proteins associated with ribosomal loading have a multiplicity of homologs, potentially resulting in a level of nuanced translational control not found in other eukaryotes. The other participants in the eIF4F complex, eIF4G (Dhalia et al. 2005; Yoffe et al. 2009) and eIF4A (Dhalia et al. 2006), have six and two homologs, respectively. This shift in gene expression control may be a functional consequence of the absence of individual gene promoters in these organisms. Differential assembly of these factors must use mRNA information beyond the SL, such as that contained in the 5′- and 3′-untranslated regions.

2.9 Conclusion

Production of the SL RNA is a major undertaking for kinetoplastids and has evolved into an efficient, specialized process. From the hypermethylated 5′ end through Ψ_{28} formation, protein assembly, and the trimming of the 3′ end, each molecule is shuttled through a chain of events that each play an important role in SL substrate or mRNA function. The sheer quantity of RNA flowing through this pathway makes the existence of devoted SL processing centers likely. As a strategy in gene expression for eukaryotes, the *trans*-splicing of polycistronically transcribed transcripts has either been retained or has emerged multiple times throughout the tree of life with a remarkably low level of variation on the SL RNA theme

(Nilsen 2001; Hastings 2005). Thus, kinetoplastids have not simply maintained an archaic mode of genetic control, but have maximized the use of a viable pathway and the downstream processes that characterize this group of widespread and highly successful organisms.

Acknowledgements This work was funded by NIH award AI056034 to DAC and NRS; JRZ was supported by USPHS National Research Service Award GM07104.

References

Aksoy S, Shay GL, Villanueva MS, Beard CB, Richards FF (1992) Spliced leader RNA sequences of *Trypanosoma rangeli* are organized within the 5S rRNA-encoding genes. Gene 113:239–243

Arhin GK, Shen S, Perez IF, Tschudi C, Ullu E (2005) Downregulation of the essential *Trypanosoma brucei* La protein affects accumulation of elongator methionyl-tRNA. Mol Biochem Parasitol 144:104–108

Arhin GK, Li H, Ullu E, Tschudi C (2006a) A protein related to the vaccinia virus cap-specific methyltransferase VP39 is involved in cap 4 modification in *Trypanosoma brucei*. RNA 12:53–62

Arhin GK, Ullu E, Tschudi C (2006b) 2′-*O*-Methylation of position 2 of the trypanosome spliced leader cap 4 is mediated by a 48 kDa protein related to vaccinia virus VP39. Mol Biochem Parasitol 147:137–139

Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC (1992) Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J Biol Chem 267:9805–9815

Barth S, Hury A, Liang XH, Michaeli S (2005) Elucidating the role of H/ACA-like RNAs in *trans*-splicing and rRNA processing via RNA interference silencing of the *Trypanosoma brucei* CBF5 pseudouridine synthase. J Biol Chem 280:34558–34568

Bélanger F, Stepinski J, Darzynkiewicz E, Pelletier J (2010) Characterization of hMTr1, a human cap1 2′O-ribose methyltransferase. J Biol Chem 285:33037–33044

Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL (2005) Messenger RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol 143:125–134

Biton M, Mandelboim M, Arvatz G, Michaeli S (2006) RNAi interference of XPO1 and Sm genes and their effect on the spliced leader RNA in *Trypanosoma brucei*. Mol Biochem Parasitol 150:132–143

Boulon S, Verheggen C, Jady BE, Girard C, Pescia C, Paul C, Ospina JK, Kiss T, Matera AG, Bordonné R, Bertrand E (2004) PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. Mol Cell 16:777–787

Bruzik JP, Van Doren K, Hirsh D, Steitz JA (1988) *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature 335:559–562

Campbell DA, Thomas S, Sturm NR (2003) Transcription in the kinetoplastid protozoa: why be normal? Microbes Infect 5:1231–1240

Clayton C (2002) Life without transcriptional control? From fly to man and back again. EMBO J 21:1881–1888

Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156:93–101

Das A, Banday M, Bellofatto V (2008) RNA polymerase transcription machinery in trypanosomes. Eukaryot Cell 7:429–434

Denker JA, Zuckerman DM, Maroney PA, Nilsen TW (2002) New components of the spliced leader RNP required for nematode *trans*-splicing. Nature 417:667–670

Dhalia R, Reis CR, Freire ER, Rocha PO, Katz R, Muniz JR, Standart N, de Melo Neto OP (2005) Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues. Mol Biochem Parasitol 140:23–41

Dhalia R, Marinsek N, Reis CR, Katz R, Muniz JR, Standart N, Carrington M, de Melo Neto OP (2006) The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct. Nucleic Acids Res 34:2495–2507

Dossin FdM, Schenkman S (2005) Actively transcribing RNA polymerase II concentrates on spliced leader genes in the nucleus of *Trypanosoma cruzi*. Eukaryot Cell 4:960–970

Feder M, Pas J, Wyrwicz LS, Bujnicki JM (2003) Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2′-*O*-methyltransferases. Gene 302:129–138

Foldynová-Trantírková S, Paris Z, Sturm NR, Campbell DA, Lukeš J (2005) The *Trypanosoma brucei* La protein is a candidate poly(U) shield that impacts spliced leader RNA maturation and tRNA intron removal. Int J Parasitol 35:359–366

Ganot P, Kallesoe T, Reinhardt R, Chourrout D, Thompson EM (2004) Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. Mol Cell Biol 24:7795–7805

Girard C, Mouaikel J, Neel H, Bertrand E, Bordonne R (2004) Nuclear localization properties of a conserved protuberance in the Sm core complex. Exp Cell Res 299:199–208

Goldshmidt H, Matas D, Kabi A, Carmi S, Hope R, Michaeli S (2010) Persistent ER stress induces the spliced leader RNA silencing pathway (SLS), leading to programmed cell death in *Trypanosoma brucei*. PLoS Pathog 6(1):e1000731. doi:10.1371/journal.ppat.1000731

Gopal S, Awadalla S, Gaasterland T, Cross GA (2005) A computational investigation of kinetoplastid *trans*-splicing. Genome Biol 6:R95

Günzl A (2003) Transcription. In: Marr JJ, Nilson TW, Komuniecki RW (eds) Molecular medical parasitology. Academic, London, pp 47–65

Günzl A (2010a) The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell 9:1159–1170

Günzl A (2010b) RNA polymerases in trypanosomes. In: Bindereif A (ed) RNA metabolism in trypanosomes. Springer, Heidelberg

Haanstra JR, Stewart M, Luu VD, van Tuijl A, Westerhoff HV, Clayton C, Bakker BM (2008) Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form *Trypanosoma brucei*. J Biol Chem 283:2495–2507

Hall MP, Ho CK (2006) Functional characterization of a 48 kDa *Trypanosoma brucei* cap 2 RNA methyltransferase. Nucleic Acids Res 34:5594–5602

Hamma T, Ferre-D'Amare AR (2006) Pseudouridine synthases. Chem Biol 13:1125–1135

Harris KA Jr, Crothers DM, Ullu E (1995) In vivo structural analysis of spliced leader RNAs in *Trypanosoma brucei* and *Leptomonas collosoma*: a flexible structure that is independent of cap4 methylations. RNA 1:351–362

Hastings KE (2005) SL trans-splicing: easy come or easy go? Trends Genet 21:240–247

Hitchcock RA, Zeiner GM, Sturm NR, Campbell DA (2004) The 3′-termini of small RNAs in *Trypanosoma brucei*. FEMS Microbiol Lett 236:73–78

Hitchcock RA, Thomas S, Campbell DA, Sturm NR (2007) The promoter and transcribed regions of the *Leishmania tarentolae* spliced leader RNA gene array are devoid of nucleosomes. BMC Microbiol 7:44

Ho CK, Shuman S (2001) *Trypanosoma brucei* RNA triphosphatase. J Biol Chem 276:46182–46186

Hury A, Goldshmidt H, Tkacz ID, Michaeli S (2009) Trypanosome spliced-leader-associated RNA (SLA1) localization and implications for spliced-leader RNA biogenesis. Eukaryot Cell 8:56–68

Jaé N, Wang P, Gu T, Hühn M, Palfi Z, Urlaub H, Bindereif A (2010) Essential role of a trypanosome U4-specific Sm core protein in small nuclear ribonucleoprotein assembly and splicing. Eukaryot Cell 9:379–386

Jaé N, Preußer C, Krüger T, Tkacz ID, Engstler M, Michaeli S, Bindereif A (2011) snRNA-specific role of SMN in trypanosome snRNP biogenesis in vivo. RNA Biol 8:90–100

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 6: e1001090

Lahav T, Sivam D, Volpin H, Ronen M, Tsigankov P, Green A, Holland N, Kuzyk M, Borchers C, Zilberstein D, Myler PJ (2011) Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. FASEB J 25:515–525

Lee JH, Cai G, Panigrahi AK, Dunham-Ems S, Nguyen TN, Radolf JD, Asturias FJ, Gunzl A (2010) A TFIIH-associated mediator head is a basal factor of small nuclear spliced leader RNA gene transcription in early-diverged trypanosomes. Mol Cell Biol 30:5502–5513

Li H, Tschudi C (2005) Novel and essential subunits in the 300-kilodalton nuclear cap binding complex of *Trypanosoma brucei*. Mol Cell Biol 25:2216–2226

Liang X-h, Xu Y-x, Michaeli S (2002) The spliced leader-associated RNA is a trypanosome-specific sn(o) RNA that has the potential to guide pseudouridine formation on the SL RNA. RNA 8:237–246

Liang X-h, Haritan A, Uliel S, Michaeli S (2003a) *trans* and *cis* splicing in trypanosomatids: mechanisms, factors, and regulation. Eukaryot Cell 2:830–840

Liang X-h, Liu Q, Michaeli S (2003b) Small nucleolar RNA interference induced by antisense or double-stranded RNA in trypanosomatids. Proc Natl Acad Sci USA 100:7521–7526

Lücke S, Xu GL, Palfi Z, Cross M, Bellofatto V, Bindereif A (1996) Spliced leader RNA of trypanosomes: *in vivo* mutational analysis reveals extensive and distinct requirements for *trans* splicing and cap 4 formation. EMBO J 15:4380–4391

Lustig Y, Sheiner L, Vagima Y, Goldshmidt H, Das A, Bellofatto V, Michaeli S (2007) Spliced-leader RNA silencing: a novel stress-induced mechanism in *Trypanosoma brucei*. EMBO Rep 8:408–413

Luz Ambrosio D, Lee JH, Panigrahi AK, Nguyen TN, Cicarelli RM, Gunzl A (2009) Spliceosomal proteomics in *Trypanosoma brucei* revealed new RNA splicing factors. Eukaryot Cell 8:990–1000

MacMorris M, Kumar M, Lasda E, Larsen A, Kraemer B, Blumenthal T (2007) A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. RNA 13:511–520

Mair G, Ullu E, Tschudi C (2000) Cotranscriptional cap 4 formation on the *Trypanosoma brucei* spliced leader RNA. J Biol Chem 275:28994–28999

Mandelboim M, Barth S, Biton M, X-h L, Michaeli S (2003) Silencing of Sm proteins in *Trypanosoma brucei* by RNAi captured a novel cytoplasmic intermediate in SL RNA biogenesis. J Biol Chem 278:51469–51478

Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010:525241. doi:10.1155/2010/525241

Maslov DA, Westenberger SJ, Xu X, Campbell DA, Sturm NR (2007) Discovery and barcoding by analysis of spliced leader RNA gene sequences of new isolates of Trypanosomatidae from Heteroptera in Costa Rica and Ecuador. J Eukaryot Microbiol 54:57–65

Matthews KR, Tschudi C, 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

Mittra B, Zamudio JR, Bujnicki JM, Stepinski J, Darzynkiewicz E, Campbell DA, Sturm NR (2008) The TbMTr1 spliced leader RNA cap 1 2′-*O*-ribose methyltransferase from *Trypanosoma brucei* acts with substrate specificity. J Biol Chem 283:3161–3172

Murthy VK, Dibbern KM, Campbell DA (1992) PCR amplification of mini-exon genes differentiates *Trypanosoma cruzi* from *Trypanosoma rangeli*. Mol Cell Probes 6:237–243

Myslyuk I, Doniger T, Horesh Y, Hury A, Hoffer R, Ziporen Y, Michaeli S, Unger R (2008) Psiscan: a computational approach to identify H/ACA-like and AGA-like non-coding RNA in trypanosomatid genomes. BMC Bioinformatics 9:471

Neuenkirchen N, Chari A, Fischer U (2008) Deciphering the assembly pathway of Sm-class U snRNPs. FEBS Lett 582:1997–2003

Nilsen TW (2001) Evolutionary origin of SL-addition *trans*-splicing: still an enigma. Trends Genet 17:678–680

Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. PLoS Pathog 6:e1001037

Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell 101:187–198

Palenchar JB, Bellofatto V (2006) Gene transcription in trypanosomes. Mol Biochem Parasitol 146:135–141

Palfi Z, Jae N, Preusser C, Kaminska KH, Bujnicki JM, Lee JH, Gunzl A, Kambach C, Urlaub H, Bindereif A (2009) SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes. Genes Dev 23:1650–1664

Roberts TG, Dungan JM, Watkins KP, 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

Roberts TG, Sturm NR, Yee BK, Yu MC, Hartshorne T, Agabian N, Campbell DA (1998) Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans. Mol Cell Biol 18:4409–4417

Ruan J-p, Shen S, Ullu E, Tschudi C (2007) Evidence for a capping enzyme with specificity for the trypanosome spliced leader RNA. Mol Biochem Parasitol 156:246–254

Saito RM, Elgort MG, Campbell DA (1994) A conserved upstream element is essential for transcription of the *Leishmania tarentolae* mini-exon gene. EMBO J 13:5460–5469

Santana DM, Lukes J, Sturm NR, Campbell DA (2001) Two sequence classes of kinetoplastid 5S ribosomal RNA gene revealed among bodonid spliced leader RNA gene arrays. FEMS Microbiol Lett 204:233–237

Shaked H, Wachtel C, Tulinski P, Yahia NH, Barda O, Darzynkiewicz E, Nilsen TW, Michaeli S (2010) Establishment of an *in vitro* trans-splicing system in *Trypanosoma brucei* that requires endogenous spliced leader RNA. Nucleic Acids Res 38(10):e114

Siegel TN, Tan KS, Cross GA (2005) Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. Mol Cell Biol 25:9586–9594

Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GA (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res 38:4946–4957

Sturm NR, Campbell DA (1999) The role of intron structures in *trans*-splicing and cap 4 formation for the *Leishmania* spliced leader RNA. J Biol Chem 274:19361–19367

Sturm NR, Fleischmann J, Campbell DA (1998) Efficient *trans*-splicing of mutated spliced leader exons in *Leishmania tarentolae*. J Biol Chem 273:18689–18692

Sturm NR, Yu MC, Campbell DA (1999) Transcription termination and 3′-end processing of the spliced leader RNA in kinetoplastids. Mol Cell Biol 19:1595–1604

Takagi Y, Sindkar S, Ekonomidis D, Hall MP, Ho CK (2007) *Trypanosoma brucei* encodes a bifunctional capping enzyme essential for cap 4 formation on the spliced leader RNA. J Biol Chem 282:15995–16005

Thomas S, Westenberger SJ, Campbell DA, Sturm NR (2005) Intragenomic spliced leader RNA array analysis of kinetoplastids reveals unexpected transcribed region diversity in *Trypanosoma cruzi*. Gene 352:100–108

Tkacz ID, Lustig Y, Stern MZ, Biton M, Salmon-Divon M, Das A, Bellofatto V, Michaeli S (2007) Identification of novel snRNA-specific Sm proteins that bind selectively to U2 and U4 snRNAs in *Trypanosoma brucei*. RNA 13:30–43

Tkacz ID, Gupta SK, Volkov V, Romano M, Haham T, Tulinski P, Lebenthal I, Michaeli S (2010) Analysis of spliceosomal proteins in Trypanosomatids reveals novel functions in mRNA processing. J Biol Chem 285:27982–27999

Tycowski KT, Shu MD, Kukoyi A, Steitz JA (2009) A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. Mol Cell 34:47–57

Ullu E, Tschudi C (1995) Accurate modification of the trypanosome spliced leader cap structure in a homologous cell-free system. J Biol Chem 270:20365–20369

Uzureau P, Daniels JP, Walgraffe D, Wickstead B, Pays E, Gull K, Vanhamme L (2008) Identification and characterization of two trypanosome TFIIS proteins exhibiting particular domain architectures and differential nuclear localizations. Mol Microbiol 69:1121–1136

Vandenberghe AE, Meedel TH, Hastings KEM (2001) mRNA 5′-leader *trans*-splicing in the chordates. Genes Dev 15:294–303

Wang P, Palfi Z, Preusser C, Lücke S, Lane WS, Kambach C, Bindereif A (2006) Sm core variation in spliceosomal small nuclear ribonucleoproteins from *Trypanosoma brucei*. EMBO J 25:4513–4523

Watkins KP, Dungan JM, Agabian N (1994) Identification of a small RNA that interacts with the 5′ splice site of the *Trypanosoma brucei* spliced leader RNA *in vivo*. Cell 76:171–182

Webb C-HT, Riccitelli NJ, Ruminski DJ, Lupták A (2009) Widespread occurrence of self-cleaving ribozymes. Science 236:953

Werner M, Purta E, Kaminska KH, Cymerman IA, Campbell DA, Mittra B, Zamudio JR, Sturm NR, Jaworski J, Bujnicki JM (2011) 2′-*O*-ribose methylation of cap 2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res 39:4756–4768

Westenberger SJ, Sturm NR, Yanega D, Podlipaev SA, Zeledón R, Campbell DA, Maslov DA (2004) Trypanosomatid biodiversity in Costa Rica: genotyping of parasites from Heteroptera using the spliced leader RNA gene. Parasitology 129:537–547

Xu Y, Liu L, Michaeli S (2000) Functional analyses of positions across the 5′ splice site of the trypanosomatid spliced leader RNA—implications for base-pair interaction with U5 and U6 snRNAs. J Biol Chem 275:27883–27892

Xu Y, Liu L, Lopez-Estraño C, Michaeli S (2001) Expression studies on clustered trypanosomatid box C/D small nucleolar RNAs. J Biol Chem 276:14289–14298

Yoffe Y, Zuberek J, Lerer A, Lewdorowicz M, Stepinski J, Altmann M, Darzynkiewicz E, Shapira M (2006) Binding specificities and potential roles of isoforms of eukaryotic initiation factor eIF4E in *Leishmania*. Eukaryot Cell 5:1969–1979

Yoffe Y, Léger M, Zinoviev A, Zuberek J, Darzynkiewicz E, Wagner G, Shapira M (2009) Evolutionary changes in the *Leishmania* eIF4F complex involve variations in the eIF4E-eIF4G interactions. Nucleic Acids Res 37(10):3243–3253

Yu MC, Sturm NR, Saito RM, Roberts TG, Campbell DA (1998) Single nucleotide resolution of promoter activity and protein binding for the *Leishmania tarentolae* spliced leader RNA gene. Mol Biochem Parasitol 94:265–281

Zamudio JR, Mittra B, Zeiner GM, Feder M, Bujnicki JM, Sturm NR, Campbell DA (2006) Complete cap 4 formation is not required for viability in *Trypanosoma brucei*. Eukaryot Cell 5:905–915

Zamudio JR, Mittra B, Foldynová-Trantírková S, Zeiner GM, Lukeš J, Bujnicki JM, Sturm NR, Campbell DA (2007) The 2′-*O*-ribose methyltransferase for cap 1 of spliced leader RNA and U1 small nuclear RNA in *Trypanosoma brucei*. Mol Cell Biol 27:6084–6092

Zamudio JR, Mittra B, Campbell DA, Sturm NR (2009a) Hypermethylated cap 4 maximizes *Trypanosoma brucei* translation. Mol Microbiol 72:1100–1110

Zamudio JR, Mittra B, Chattopadhyay A, Wohlschlegel JA, Sturm NR, Campbell DA (2009b) *Trypanosoma brucei* spliced leader RNA maturation by the cap 1 2′-*O*-ribose methyltransferase and SLA1 H/ACA snoRNA pseudouridine synthase complex. Mol Cell Biol 29:1202–1211

Zeiner GM, Sturm NR, Campbell DA (2003a) Exportin 1 mediates nuclear export of the kinetoplastid spliced leader RNA. Eukaryot Cell 2:222–230

Zeiner GM, Sturm NR, Campbell DA (2003b) The *Leishmania tarentolae* spliced leader contains determinants for association with polysomes. J Biol Chem 278:38269–38275

Zeiner GM, Foldynová S, Sturm NR, Lukeš J, Campbell DA (2004a) SmD1 is required for spliced leader RNA biogenesis. Eukaryot Cell 3:241–244

Zeiner GM, Hitchcock RA, Sturm NR, Campbell DA (2004b) 3′-end polishing of the kinetoplastid spliced leader RNA is performed by SNIP, a $3' \rightarrow 5'$ exonuclease with a motley assortment of small RNA substrates. Mol Cell Biol 24:10390–10396

Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, Gaasterland T, Lin S (2007) Spliced leader RNA trans-splicing in dinoflagellates. Proc Natl Acad Sci USA 104:4618–4623

Zhang H, Campbell DA, Sturm NR, Dungan CF, Lin S (2011) Spliced leader RNAs, mtDNA reading frameshifts and multiprotein phylogeny expand support for the genus *Perkinsus* as a unique group of Alveolates. PLoS One 5:e19933

Chapter 3
Pre-mRNA Splicing in *Trypanosoma brucei*: Factors, Mechanisms, and Regulation

Christian Preußer, Nicolas Jaé, Arthur Günzl, and Albrecht Bindereif

Abstract Processing of polycistronic pre-mRNAs in trypanosomes requires *trans* splicing and polyadenylation of the primary transcripts to generate mature 5′ and 3′ ends of the mRNAs, respectively. Splicing is catalyzed by the spliceosome, which contains the spliceosomal small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4/U6, and U5, as well as the SL RNP, which is essential for *trans* splicing. As in other eukaryotes, the trypanosomal snRNPs share a common set of seven Sm polypeptides, which assemble in a ring-like structure on the Sm binding site of the snRNAs, a process that is mediated by the survival of motor neurons (SMN) protein. Whereas studies in the last decade disclosed numerous spliceosomal components, little is known about splicing regulation in trypanosomes, even though recent studies revealed evidence for alternative *trans* splicing. Here we summarize the current state of research on mRNA splicing in *Trypanosoma brucei* which differs in several aspects from the well-studied mechanisms in other eukaryotes.

The authors Christian Preußer and Nicolas Jaé have contributed equally.

C. Preußer • N. Jaé • A. Bindereif (✉)
Institute of Biochemistry, Justus Liebig University of Giessen, Heinrich-Buff-Ring 58, 35392 Giessen, Germany
e-mail: albrecht.bindereif@chemie.bio.uni-giessen.de

A. Günzl
Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030-6403, USA

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_3,

3.1 Gene Expression in Trypanosomes and the Discovery of *Trans* Splicing

Trypanosomatids are unicellular, vector-borne parasites of invertebrates, vertebrates, and even plants, which differ in various cellular and biochemical aspects from other eukaryotes. Exemplary for this is the mechanism of gene expression and their unusual way of pre-mRNA processing.

In the general view of eukaryotic gene expression, a precursor mRNA (pre-mRNA) is transcribed whose protein-coding exonic sequences are disrupted by introns. In order to generate mature mRNAs, those introns have to be precisely removed and the exons joined together. In most eukaryotes, this important mRNA maturation step takes place within a single pre-mRNA transcript, therefore referred to as *cis* splicing (reviewed by Burge et al. 1999). The splicing reaction is catalyzed by a macromolecular complex, the spliceosome, composed of five small nuclear ribonucleoproteins (U1, U2, U4, U5, and U6 snRNPs) and many additional non-snRNP proteins. Spliceosome assembly and splicing catalysis occur in an ordered multistep process, which includes numerous conformational rearrangements (reviewed by Brow 2002; Will and Lührmann 2006; Wahl et al. 2009).

The correct discrimination between exonic and intronic sequences is a prerequisite for precise splicing and subsequent protein expression. In this context, four conserved pre-mRNA sequence elements were found to play essential roles by interacting with components of the splicing machinery during assembly and catalysis stages. These are the 5′- and 3′-splice site consensus sequences, which define the exon–intron boundaries and function as splicing donor and acceptor sites, respectively, as well as the branch point and a polypyrimidine tract of variable length (reviewed by Black 2003). Studies over the last decades on spliceosome structure, function, and dynamics revealed that besides a high degree of conservation in many basic aspects, there are numerous species-specific variations and characteristics. In this context, trypanosomes are particularly remarkable: Here, gene expression is initiated by the transcription of large polycistronic gene arrays, which are interrupted by tRNA and snRNA genes, suggesting that the polycistronic transcription of protein-coding genes by RNA polymerase II is terminated by RNA polymerase III-transcribed genes (reviewed by Martínez-Calvillo et al. 2010). Back in 1982, studies concerning the expression of variant surface glycoproteins (VSGs) in *T. brucei*, which play an essential role in evading the mammalian immune response, revealed that different VSG transcripts carry a short, identical leader sequence at their 5′ ends, which was missing in the gene sequences (Van der Ploeg et al. 1982; Boothroyd and Cross 1982). At the same time, in vitro splicing assays in HeLa nuclear extract showed that exon segments of separate RNA molecules can be joined together in a *trans* splicing reaction (Solnick 1985; Konarska et al. 1985). This stood in clear contrast to classical *cis* splicing and suggested that *trans* splicing can be used to distribute a single exon to different mRNAs. In the trypanosomal system, further analysis revealed that the addition of the same 5′ leader sequence was not restricted to *VSG* genes (De Lange et al. 1983, 1984; Parsons et al. 1984).

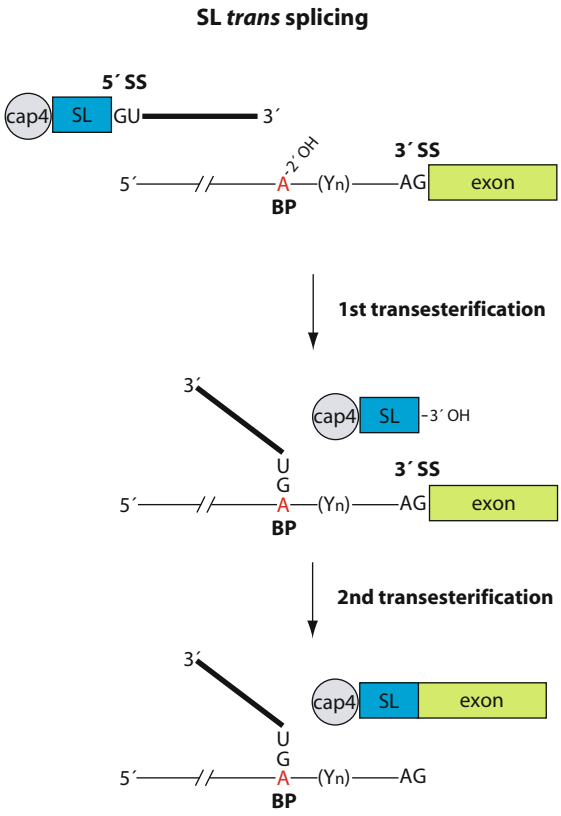


Fig. 3.1 *Trans* splicing mechanism. Schematic representation of the *trans* splicing mechanism. Exons are shown as *boxes*, introns as *solid lines*. The 5′ and 3′ splice sites, defined by their conserved dinucleotides (GU and AG, respectively), as well as the branch point (BP) adenosine (A), located upstream of the polypyrimidine tract (Y_n) are indicated. Like *cis* splicing, *trans* splicing proceeds as a two-step transesterification reaction; however, in *trans* splicing exons from two independently transcribed precursor RNAs are joined together. During the first step of *trans* splicing, the hydroxyl group of the adenosine residue performs a nucleophilic attack at the 5′ splice site, resulting in the release of the 39-nt spliced leader sequence and the formation of a Y-branched intermediate, characterized by its 2′–5′ phosphodiester bond. In the second step, the 3′ hydroxyl group of the released spliced leader sequence performs a nucleophilic attack at the 3′ splice site, thus joining the exons together; the Y-branched intron is degraded

Strikingly, a separate small nuclear RNA named SL RNA or mini-exon-derived RNA was identified as the donor of this 39-nucleotide long spliced leader (SL) sequence (Campbell et al. 1984; Kooter et al. 1984; Milhausen et al. 1984). At this time, different possibilities were proposed to explain the biochemical mechanism of *trans* splicing. The final discovery of a Y-branched splicing intermediate, instead of the *cis* splicing-specific lariat structure, indicated that *trans* splicing proceeds, in analogy to *cis* splicing, through a two-step transesterification reaction, although using two independently transcribed precursor RNAs (Fig. 3.1; Murphy et al. 1986;

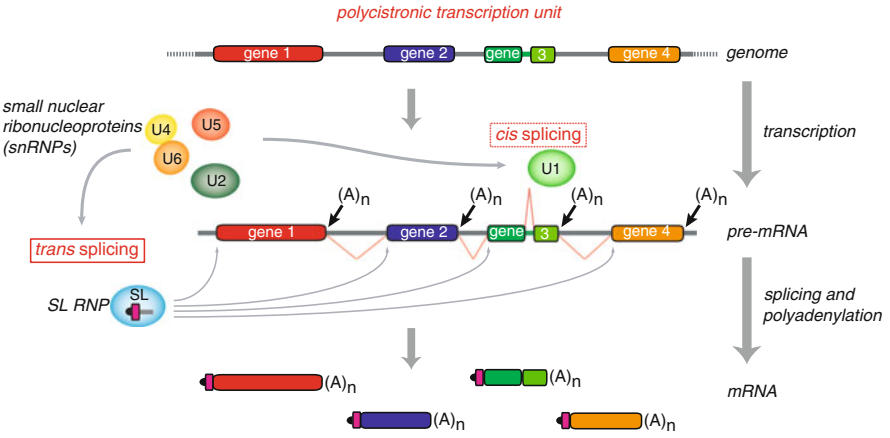


Fig. 3.2 mRNA processing in trypanosomatids. Schematic representation of the major mRNA processing steps in trypanosomatids. After transcription of polycistronic transcription units (*colored boxes*), which are separated by strand switch regions (SSR), the pre-mRNAs undergo *trans* splicing and polyadenylation [$(A)_n$]. Each of the protein-coding genes (shown in different colors) is processed through *trans* splicing, whereby the SL RNA with its cap 4 structure provides the 5′ exon (miniexon, in *red*); in addition, very few genes require *cis* splicing of an internal intron (see *green boxes*). Both *cis* and *trans* splicing proceed through two transesterification steps and require the U2, U4/U6, and U5 small nuclear ribonucleoproteins (snRNPs) as well as many additional protein factors. In addition, *cis* splicing requires the U1 snRNP, *trans* splicing the SL RNP as specific factors

Sutton and Boothroyd 1986). The finding that site-specific cleavage of spliceosomal snRNAs blocks *trans* splicing supported this concept (Tschudi and Ullu 1990). Further investigations revealed that the process of *trans* splicing is not restricted to kinetoplastids (reviewed by Mayer and Floeter-Winter 2005). It was subsequently discovered in a multitude of different organisms, including additional members of the *Euglenozoa* phylum as well as in some metazoans, e.g., in the *Nematoda* and *Platyhelminthes*, in *Hydra*, and even in two members of the *Urochordata* subphylum (for a graphical overview of mRNA processing in trypanosomes, see Fig. 3.2).

3.2 Genome Structure of *Trypanosoma brucei* and mRNA Splicing

The genome of *T. brucei* is bipartite and comprises the nuclear genes as well as an unconventional mitochondrial genome that is organized in the kinetoplast. With respect to the nuclear genome, ~9,068 protein-coding genes (including ~1,700 *T. brucei*-specific genes) are distributed among 11, so-called megabase chromosomes (Berriman et al. 2005). An additional characteristic of kinetoplastid genomes is that both strands of the megabase chromosomes contain nonoverlapping directional gene

clusters (DGCs) in which all genes share the same transcriptional orientation (Hall et al. 2003; El-Sayed et al. 2003; Berriman et al. 2005). Based on nuclear run-on assays in *L. major*, these DGCs are thought to be transcribed into polycistronic pre-mRNAs, and single genes are subsequently released from these primary transcripts by the coupled processes of *trans* splicing and polyadenylation.

In contrast to the RNA polymerase II-transcribed polycistronic protein-coding genes, the small nuclear RNAs (U1, U2, U4, U5, and U6 snRNAs) are transcribed by RNA polymerase III, and the corresponding genes are scattered over four chromosomes. Notably, these snRNA genes are arranged in close proximity to tRNA or tRNA-like sequences in several genomic loci (Mottram et al. 1991; Fantoni et al. 1994; Nakaar et al. 1994, 1997), which often reside between convergently transcribed protein-coding genes (Ivens et al. 2005; Berriman et al. 2005; El-Sayed et al. 2005). In contrast to the snRNAs, the spliced leader RNA (SL RNA), as a key molecule in and substrate of *trans* splicing, is continuously consumed during the process. Hence, trypanosomatids depend on a very high SL RNA transcription rate (Laird et al. 1987; Ullu and Tschudi 1991), in which 139 nt-long SL RNA precursors are produced that are processed through multiple steps before the mature SL RNA is committed to *trans* splicing.

3.3 Spliceosomal snRNPs and Their Core Proteins: Canonical Versus snRNP-Specific Sm Proteins

3.3.1 Spliceosome Assembly and Sm Core Proteins

As noted previously, the process of pre-mRNA splicing is catalyzed by the spliceosome, a macromolecular complex composed of five small nuclear ribonucleoproteins (U1, U2, U4, U5, and U6 snRNPs) and numerous additional non-snRNP splicing factors (for the mammalian system, reviewed by Will and Lührmann 2001; Makarov et al. 2002). Note that our classical concept on dynamics of spliceosome assembly relies primarily on studies in mammalian and yeast cells. As outlined in the following, the assembly of the spliceosome on the pre-mRNA follows an ordered multistep pathway and involves the following key points (1) initial recognition of the 5′ splice site and the branch point by the U1 and U2 snRNP, respectively; (2) joining of the preassembled U4/U6*U5 tri-snRNP; (3) dissociation of the U1 and U4 snRNP from the complex along with rearrangements of the pre-mRNA/snRNA network; (4) two-step splicing catalysis; and (5) final disassembly of the spliceosome (reviewed by Brow 2002; Will and Lührmann 2006; Wahl et al. 2009). Despite their different functions during the splicing process, the spliceosomal snRNPs share several basic structural features. They all assemble from proteins and uridine-rich small nuclear RNAs (U snRNAs); their protein components can be further grouped into snRNP-specific and common proteins (for a detailed summary of trypanosomal snRNP-specific proteins, see below and Table 3.1). The common proteins are found in all spliceosomal snRNPs

Table 3.1 Spliceosomal proteins

Annotation[1]	M [kDa][2]	Accession #[3]	TAP[4,5]	References
Sm/LSm proteins				
SmB	12.3	Tb927.2.4540	a, b, c, d, e	Palfi et al. (2000)
SmD3	12.4	Tb927.4.890	a ,b, c, d	Palfi et al. (2000)
SmD1	11.7	Tb927.7.3120	a, b, e	Palfi et al. (2000)
SmD2	12.5	Tb927.2.5850	a, b, e	Palfi et al. (2000)
SmE	9.6	Tb927.6.2700	a, b, e	Palfi et al. (2000)
SmF	8.4	Tb09.211.1695	a, b, e	Palfi et al. (2000)
SmG	8.9	Tb11.01.5915	a, b, e	Palfi et al. (2000)
Sm15K (SSm2-1*)	12.8	Tb927.6.4340	a, b	Wang et al. (2006) and Tkacz et al. (2007)
Sm16.5K (SSm2-2*)	14.7	Tb927.10.4950	a, b	Wang et al. (2006) and Tkacz et al. (2007)
SSm4*	23.2	Tb927.7.6380	a, b, e	Tkacz et al. (2007) and Jaé et al. (2010)
LSm2	13.2	Tb927.8.5180	a, b	Tkacz et al. (2008) and Luz Ambrósio et al. (2009)
LSm3	10.1	Tb927.7.7380	–	Liu et al. (2004)
LSm4	14.2	Tb11.01.5535	a, b	Liu et al. (2004)
LSm5	12.0	not assigned	–	Tkacz et al. (2008)
LSm6	9.1	Tb09.160.2150	–	Liu et al. (2004)
LSm7	10.2	Tb927.5.4030	a, b	Liu et al. (2004)
LSm8	14.0	Tb927.3.1780	a, b	Liu et al. (2004)
U1 snRNP				
U1-70K	31.7	Tb927.8.4830	a, b	Palfi et al. (2005)
U1C	21.7	Tb927.10.2120	a, b	Palfi et al. (2005)
U1A	18.0	Tb927.10.8280/ 8300	a	Luz Ambrósio et al. (2009)
U1-24K*	24.2	Tb927.3.1090	a, b	Palfi et al. (2005)
U2 snRNP				
U2 A′ (U2-40K)	36.5	Tb927.10.14360	a, b	Cross et al. (1993)
U2 B”	13.6	Tb927.3.3480	a, b	Preußer et al. (2009)
Sf3b 14	13.3	Tb927.10.7470	–	Avila et al. (2007) and Manful et al. (2009)
Sf3b(SAP)130	195.0	Tb927.7.6980	a	Manful et al. (2009)
Sf3b(SAP)145	52.5	Tb927.6.2000	–	Manful et al. (2009)
Sf3b(SAP)155	122.0	Tb11.01.3690	–	Avila et al. (2007) and Manful et al. (2009)
Sf3b(SAP)49	29.8	Tb927.3.5280	–	Manful et al. (2009)
Sf3a60	61.5	Tb927.6.3160*	–	Günzl (2010)
Sf3b(SAP)14b	12.7	Tb927.10.7390*	–	Manful et al. (2009)
Sf3b(SAP)10	10.4	Tb09.211.2205*	–	Manful et al. (2009)
U4/U6 snRNP				
PRP3	63.2	Tb09.160.2900*	a, b	Luz Ambrósio et al. (2009) and Palfi et al. (2009)

(continued)

Table 3.1 (continued)

Annotation[1]	M [kDa][2]	Accession #[3]	TAP[4,5]	References
PRP4	65.5	Tb927.10.960	a, b	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
SNU13	13.6	Tb09.160.3670	–	Barth et al. (2008)
U5 snRNP				
PRP8	277.0	Tb09.211.2420	a, b, d	Lücke et al. (1997)
U5-40K	35.0	Tb11.01.2940	a, b	Luz Ambrósio et al. (2009)
U5-Cwc21*	16.2	Tb09.160.2110	a, b	Luz Ambrósio et al. (2009)
U5-15K	17.7	Tb927.8.2560†	a	Luz Ambrósio et al. (2009)
U5-116K	105.4	Tb11.01.7080†	a, b	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
U5-102K (PRP6)	111.0	Tb11.01.7330†	a	Luz Ambrósio et al. (2009)
U5-200K	249.3	Tb927.5.2290†	a	Luz Ambrósio et al. (2009) and Günzl (2010)
SMN/Gemin2 and associated proteins				
SMN	17.0	Tb11.01.6640	a, b, c, d	Palfi et al. (2009)
Gemin2	55.4	Tb927.10.5640†	a, b, c, d	Palfi et al. (2009)
Coatomer (alpha)	132.0	Tb927.4.450	c, d	Maier et al. (2001)
Coatomer (beta)	110.0	Tb927.1.2570	c, d	Maier et al. (2001)
Coatomer (beta′)	93.9	Tb927.2.6050	c, d	Maier et al. (2001)
Coatomer (gamma)	97.5	Tb11.01.3740	c	Maier et al. (2001)
Coatomer (delta)	57.3	Tb927.8.5250	c, d	Maier et al. (2001)
Coatomer (epsilon)	34.8	Tb11.01.6530	c, d	Maier et al. (2001)
Coatomer (zeta)	20.5	Tb927.10.4270	c	Maier et al. (2001)
PRP19 complex				
CDC5	80.1	Tb927.5.2060†	–	Günzl (2010)
PRP19	54.3	Tb927.2.5240†	a, b	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
CRN	87.7	Tb927.10.9660†	a	Luz Ambrósio et al. (2009)
SYF1	92.2	Tb927.5.1340†	a	Luz Ambrósio et al. (2009)
ISY1	31.7	Tb927.8.1930†	a	Luz Ambrósio et al. (2009)
KIAA1604/Cwc22	66.8	Tb11.01.2520†	b	Palfi et al. (2009)
PRP46	48.7	Tb927.10.10170	f, g	Tkacz et al. (2010)
Miscellaneous splicing factors				
U2AF35	29.1	Tb927.10.3200	–	Vázquez et al. (2003)
U2AF65	96.6	Tb927.10.3500	–	Vázquez et al. (2003)
SF1	31.6	Tb927.10.9400	–	Vázquez et al. (2003)
PRP31	39.7	Tb927.10.10700	–	Liang et al. (2006)
PRP43	82.9	Tb927.5.1150	–	Liang et al. (2006)
PTB1	37.0	Tb09.211.0560	–	Stern et al. (2009)
PTB2	54.7	Tb11.01.5690	–	Stern et al. (2009)
TSR1	37.5	Tb927.8.900	–	Ismaïli et al. (1999)
RRM1	50.0	Tb927.2.4710	–	Manger and Boothroyd (1998)
PRP17	52.8	Tb927.3.1930†	b	Palfi et al. (2009)

(continued)

Table 3.1 (continued)

Annotation[1]	M [kDa][2]	Accession #[3]	TAP[4,5]	References
Un-annotated proteins that co-purified in spliceosomal complexes				
c.h. (putative cyclophilin)	27.1	Tb927.8.6280†	a, b	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
c.h.	12.1	Tb11.02.0465†	a	Luz Ambrósio et al. (2009)
c.h.	31.0	Tb927.7.1890†	a	Luz Ambrósio et al. (2009)
c.h.	42.0	Tb927.2.3400†	a	Luz Ambrósio et al. (2009)
c.h.	20.0	Tb927.5.2910†	a	Luz Ambrósio et al. (2009)
c.h.	26.0	Tb927.8.4790†	a	Luz Ambrósio et al. (2009)
c.h. (putative Cwc15)		Tb927.10.11950†	b	
c.h. (WD40)	22.3	Tb927.7.3560†	b	Palfi et al. (2009)
c.h.	199.2	Tb927.7.2570†	f	Tkacz et al. (2010)
c.h.	52.5	Tb927.10.1490†	f	Tkacz et al. (2010)
c.h.	134.8	Tb927.8.2230†	f, h	Tkacz et al. (2010)
c.h.	41.6	Tb11.01.7210†	f, g	Tkacz et al. (2010)
c.h.	71.0	Tb11.03.0530†	f, g	Tkacz et al. (2010)
c.h.	31.2	Tb11.01.3480†	f, g	Tkacz et al. (2010)
c.h.	64.5	Tb927.2.2650†	f	Tkacz et al. (2010)
c.h.	367.6	Tb927.3.2700†	f	Tkacz et al. (2010)
c.h.	42.6	Tb11.01.8590†	g	Tkacz et al. (2010)
	24.4		h	Tkacz et al. (2010)
Proteins without known splicing functions				
La protein	37.7	Tb10.70.5360	a, b, c	Marchetti et al. (2000) and Westermann and Weber (2000)
HSP70	75.4	Tb11.01.3110	a, b, c	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
eEF-1α	49.1	Tb927.10.2100	c	Palfi et al. (2009)
Importin α subunit	58.0	Tb927.6.2640	b	Li and Tschudi (2005)
Poly(A)-binding protein 2	62.1	Tb09.211.2150	a, b	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
Poly(A)-binding protein 1	63.0	Tb09.211.0930	f, h	Tkacz et al. (2010)
TRYP1	22.4	Tb09.160.4250/80	a	Luz Ambrósio et al. (2009)
NORF1	93.3	Tb927.5.2140	a	Luz Ambrósio et al. (2009) and Palfi et al. (2009)
DHH1	46.4	Tb927.10.3990	f, g, h	Holetz et al. (2007)
PPIase	21.6	Tb927.8.2090	b	Palfi et al. (2009)
Putative spliceosomal helicases				
	114.0	Tb927.6.4600†	–	Günzl (2010)
	98.1	Tb927.10.5280†	–	Günzl (2010)
	121.2	Tb927.10.7280†	–	Günzl (2010)
	76.7	Tb927.10.9130†	–	Günzl (2010)
	116.6	Tb11.02.3460†	–	Günzl (2010)
	108.8	Tb927.7.7300†	–	Günzl (2010)
	248.4	Tb11.02.1930†	–	Günzl (2010)
	71.3	Tb927.10.14550†	f, g, h	Tkacz et al. (2010)

(continued)

Table 3.1 (continued)

Annotation[1]	M [kDa][2]	Accession #[3]	TAP[4,5]	References
	82.7	Tb09.211.3510[†]	g, h	Tkacz et al. (2010)
	49.2	Tb927.10.540[†]	h	Tkacz et al. (2010)
	62.4	Tb927.8.1510[†]	f, g, h	Tkacz et al. (2010)

c.h. conserved hypothetical protein
[1]Annotation according to the human system
[2]Molecular mass in kilodaltons (kDa)
[3]Accession numbers of the TriTryp data base (http://www.tritrypdb.org/tritrypdb/)
[4]Proteins were tandem-affinity co-purified with SmD1 (a), SmB (b), SMN (c), Gemin2 (d), or SSm4 (e)
[5]Homologous proteins from *Trypanosoma brucei,* co-purified with *Leishmania tarantolae* SmD3 (f), LSm3 (g), or U1A (h)
[*]Trypanosome-specific proteins
[†]Annotated in the TriTryp data base only as hypothetical or putative proteins; therefore note that some of the assignments based on sequence similarity may be putative

and comprise seven polypeptides (SmD1, SmD2, SmD3, SmB, SmE, SmF, and SmG in the U1, U2, U4, and U5 snRNP; LSm2-8 in the U6 snRNP), which assemble the so-called Sm core structure around a conserved sequence within the snRNAs (Kambach et al. 1999). This important single-stranded RNA region, termed the Sm site, with the consensus derived from the mammalian sequence -PuA$(U)_{3-6}$GPu- is often flanked by stem-loops (Branlant et al. 1982). In trypanosomatid snRNAs, this consensus sequence is less stringent (Fig. 3.3). Importantly, the Sm proteins do not assemble as single subunits on the Sm sites. Rather they preassemble into the heteromeric SmD1/D2, SmE/F/G, and SmD3/B subcomplexes, which associate with the snRNAs in at least two steps (Raker et al. 1996). Characteristic for all Sm proteins is the presence of two conserved amino acid sequence motifs, Sm1 and Sm2, which are separated by a spacer of variable length and are important for Sm protein–protein recognition (Hermann et al. 1995; Séraphin 1995). Based on crystal structures from two different human Sm subcomplexes (SmD3/B and SmD1/D2), Kambach et al. (1999) proposed a model for the full Sm core heptamer, in which the Sm proteins form a seven-membered ring around the Sm site of the snRNAs. Sm binding to this RNA motif is facilitated by the positively charged central hole of the Sm core. This heteroheptameric ring model is consistent with earlier electron microscopic studies (Kastner et al. 1990; Plessel et al. 1997) and was later confirmed by the crystal structure of the human U1 snRNP (Pomeranz Krummel et al. 2009; Leung et al. 2011). In addition, UV-crosslinking assays indicated that the nucleotides of the Sm site interact with the inner surface of the Sm ring, with the most efficient crosslinks observed for the proteins SmG and SmB/B′ (Urlaub et al. 2001, reviewed by Khusial et al. 2005).

Almost 20 years ago, affinity purification and in silico analyses led to the identification and characterization of the trypanosomatid Sm protein homologs (Palfi et al. 1991, 2000). It was shown that these proteins bind to a conserved region on the snRNAs, which resembles the Sm binding site sequence of *cis*-spliceosomal

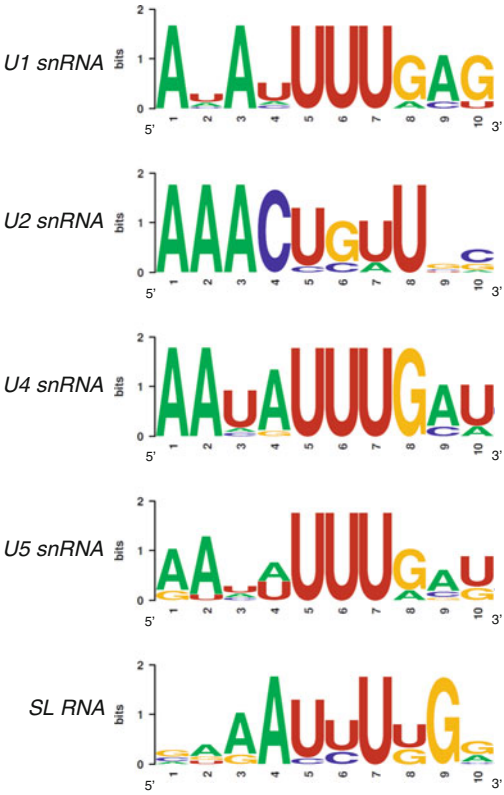


Fig. 3.3 Consensus sequences of Sm sites in snRNAs of different trypanosomatid species. Sm site sequence motifs of U1, U2, U4, U5 snRNAs as well as the SL RNA were aligned and depicted using Web logo (http://weblogo.berkeley.edu/). The following trypanosomatid species were used: *Trypanosoma brucei, T. congolense, T. cruzi, T. evansi, T. rangeli, T. vivax, Leishmania amazonensis, L. braziliensis, L. donovani, L. enrietti, L. infantum, L. major, L. mexicana, Leptomonas collosoma, L. tarentolae, L. seymouri, Crithidia fasciculata*. For the SL RNA alignment, the *T. cruzi* sequence was not included, because of an unusually degenerate Sm site. The *heights* of the letters indicate the degree of conservation

snRNPs (Günzl et al. 1992). Moreover, Palfi et al. (2000) revealed a network of specific heterodimeric and trimeric in vitro protein–protein interactions, supporting a heptameric model of the trypanosomal Sm ring consistent with its mammalian counterpart in protein composition and arrangement (Kambach et al. 1999).

Sequence alignment of the trypanosome Sm proteins with those from other systems revealed that the trypanosomal Sm motifs 1 and 2 are relatively degenerate and contain only a few highly conserved amino acids (Palfi et al. 2000). Moreover, in the mammalian system, SmB, SmD1, and SmD3 undergo symmetric arginine dimethylation (sDMA) at multiple RG dipeptides within their C-termini. This creates an epitope recognized by the monoclonal Y12 antibody (Brahms et al. 2000, 2001). Interestingly, trypanosome Sm proteins are not recognized by this antibody (Michaeli et al. 1990; Palfi and Bindereif 1992), consistent with the lack of C-terminal extensions with RG dipeptides beyond the Sm motif (Palfi et al. 2009).

Notably, several studies revealed unusual features of the spliceosomal Sm cores in the trypanosome system, in that they are not identical for all snRNPs; this was first discovered for the U2 snRNP and termed Sm core variation (Wang et al. 2006), later also for the U4 snRNP (Tkacz et al. 2007; Jaé et al. 2010; for details, see below).

3.3.2 *SL RNP: Unique Component of the* Trans *Splicing Machinery*

The spliced leader RNA, which is the most abundant spliceosomal RNA in trypanosomes, is transcribed by RNA polymerase II (Gilinger and Bellofatto 2001). During the last decade, transcription initiation of the SL RNA gene has been studied extensively. The trypanosome genome comprises a 1.4 kb large tandem array on chromosome 9, which contains 100–200 copies of the SL RNA gene that are transcribed in a monocistronic fashion (Gilinger and Bellofatto 2001). This reflects the high demand of SL RNA, which is stoichiometrically consumed during each *trans* splicing process and turned over at a very high rate (with a half life of ~4 min; Laird et al. 1987; Ullu and Tschudi 1991). As a result, SL RNA constitutes nearly 6% of all transcribed cellular RNA in *T. brucei* (Campbell et al. 2000).

One of the prominent characteristics of the SL RNA is the complex 5′ cap structure. In contrast to most eukaryotic pre-mRNAs, where an m^7G cap is acquired early during transcription by RNA polymerase II, a hypermethylated cap 4 structure is found on the SL RNA. Analysis of the 5′ end from *T. brucei* and *Crithidia fasciculata* revealed that the cap 4 structure is composed of a terminal m^7G residue followed by four 2′-*O*-methylated nucleotides (AACU) with additional base methylations on the first and fourth positions, resulting in the structure: $m^7G(5')ppp(5')m_2{}^6AmpAmpCmpm^3Ump$ (Bangs et al. 1992). These modifications are introduced co-transcriptionally in a 5′ to 3′ direction and were initially considered to be independent of core ribonucleoprotein formation (Mair et al. 2000b). However, recent studies revealed that full cap 4 modification requires correct Sm core assembly (Mandelboim et al. 2003; Jaé et al. 2011). In higher eukaryotes, the formation of the 2,2,7-trimethylguanosine cap (TMG) found on the snRNAs (except U6 snRNA) is coupled to Sm core assembly, and both constitute a bipartite nuclear import signal (Mattaj 1986). In trypanosomes, the m^7G is formed on the SL RNA by the bifunctional enzyme TbCet1 that carries structural and biochemical properties of a guanylyltransferase and a methyltransferase (MTase) (Ruan et al. 2007; Takagi et al. 2007). The functional role of the m^7G cap in the mammalian system was shown for several posttranscriptional processes, like mRNA stability and nuclear export (Lewis and Izaurralde 1997). In kinetoplastids, the modifications of the first four cap nucleotides are performed by methyltransferases that are named according to the positions modified.

Whether SL RNP biogenesis involves intracellular trafficking, where and in which order SL RNP core assembly and SL RNA modification reactions take place, have remained controversial issues until today. However, some of the modification reactions including the responsible enzymes have been localized. Since cap 4 modification and SLA1-guided internal pseudouridylation of the SL RNA occur co-transcriptionally, these modifications are linked to the nucleus (Mair et al. 2000b; Hury et al. 2009). In contrast, the process of 3′ end formation appears to involve both initial cytoplasmic and final nuclear 3′-trimming (Zeiner et al. 2004).

However, recent studies suggest that Sm core assembly is a nuclear process: RNAi-mediated knockdown of one of the Sm subunits (SmD1) led to an

accumulation of SL RNA inside the nucleus (Biton et al. 2006); in addition, SMN, which chaperones the Sm core assembly, was localized in the trypanosome nucleus (Palfi et al. 2009), suggesting that Sm core assembly takes place within this cellular compartment. Moreover, the SL RNA gene loci as well as the SL RNA transcript clearly co-localize with the SMN protein in distinct areas inside the nucleus (Jaé et al. 2011). For a detailed view on SL RNA biogenesis, see Chap. 2.

3.3.3 U2 snRNP

The trypanosomal U2 snRNA was initially identified by immunoprecipitation with antitrimethyl-guanosine (TMG) antibodies (Mottram et al. 1989), and phylogenetic sequence analysis revealed several unusual differences from the highly conserved counterparts in other eukaryotes: The U2 snRNA of *T. brucei* diverges in three major regions from other U2 snRNAs. First, there is no stem-loop III, eliminating the potential formation of a pseudoknot; second, the conserved branch-point recognition sequence GUAGUA, which is invariant in metazoan and yeast U2 snRNAs (Moore et al. 1993), is missing (Tschudi et al. 1990; Hartshorne and Agabian 1990). Only the adenosine at the branch point within the pre-mRNA is conserved, suggesting the possibility of a different mechanism for branch-point recognition in trypanosomatids (Lücke et al. 2005). Finally, another alteration in the U2 snRNA occurs in its Sm site, where the pyrimidine stretch is interrupted by an unusual purine nucleotide. Note that Urlaub et al. (2001) mapped a contact between human SmB/B′ and an Sm site RNA oligonucleotide to the third uridine position of the pyrimidine stretch, the position, at which in trypanosome U2 snRNA the exceptional guanosine resides (Fig. 3.3). Moreover, tandem affinity purification demonstrated that the identity of the Sm protein constituents varies between the spliceosomal snRNPs in *T. brucei*. In detail, the Sm subcomplex SmB/D3 is replaced by two novel U2-specific Sm proteins, Sm15K (SSm2-1) and Sm16.5K (SSm2-2). Subsequent in vitro reconstitution assays with recombinant Sm proteins and synthetic Sm site RNAs established that the unusual guanosine residue in the U2 Sm site provides an important element for discriminating between the U2-specific and the canonical Sm core (see Fig. 3.3; Wang et al. 2006; Tkacz et al. 2007).

What is the functional significance of Sm core variation in the trypanosome U2 snRNP? A recent study demonstrated that the unusual U2 Sm site is indirectly responsible for the accurate assembly of other U2 snRNP-specific proteins. Sequence comparisons had identified the trypanosomal U2A′ ortholog called U2-40K (Cross et al. 1993). The U2-specific heterodimer Sm15K/16.5K interacts with the U2A′ protein, which in turn contacts the loop IV-binding protein U2B″, another U2-specific protein. In contrast, the human ortholog of U2A′ does not bind the Sm subunits, perhaps due to their spatial separation by the additional stem-loop III; the trypanosome-specific U2A′/Sm protein interaction correlates with the lack of stem-loop III in the trypanosomatid U2 snRNA (Preußer et al. 2009). In sum, the variant

Sm core of the U2 snRNP does not only mediate specific binding to the special Sm binding site in U2, but also enables U2-specific protein–protein interactions.

3.3.4 U4/U6 di-snRNP and Singular U6 snRNP

Like the U2 snRNA, the trypanosomal U4 snRNA was initially identified by trimethyl cap-immunoprecipitation from *T. brucei* total RNA (Mottram et al. 1989). A subsequent sequence alignment revealed a considerable homology with its metazoan counterparts; based on extensive complementarity to the trypanosome U6 snRNA sequence, a secondary structure for the U4-U6 RNA duplex was proposed, which contains two main regions of U4-U6 base pairing, separated by a U4 hairpin loop. This U4-U6 base pairing interaction was experimentally validated by co-selection of trypanosomal U4 and U6 snRNAs (Palfi et al. 1991), as well as by UV light-induced RNA–RNA in vivo crosslinks (Watkins and Agabian 1991).

Similarly as described for the U2 snRNP, the U4 snRNP of *T. brucei* carries a variant Sm core, in which SmD3 is replaced by the U4-specific protein SSm4. Based on sequence homology, SSm4 had initially been considered to be part of the trypanosomal U6-specific LSm ring and therefore designated LSm2 (Liu et al. 2004). The correct assignment of SSm4 as U4 snRNP-specific Sm core protein was primarily based on a U4 snRNA-specific destabilization after SSm4 knock-down (Tkacz et al. 2007). Later on, mass-spectrometric analysis of an affinity-purified, partial U4-snRNP complex containing the 3′-terminal Sm domain provided direct biochemical evidence that only a single subunit, SSm4, replaces specifically the canonical SmD3 subunit in the heptameric trypanosome U4 Sm core (see Fig. 3.4; Jaé et al. 2010). In addition, this study ruled out that SSm4 might be part of the U6-specific LSm ring. Further analysis showed that SSm4 is required for the in vivo integrity of the U4 snRNA as well as the U4/U6 di-snRNP and that SSm4 specifies the assembly of U4 Sm cores during in vitro reconstitution (Jaé et al. 2010).

Regarding U4-specific protein components, a recent study suggested a candidate for the trypanosome ortholog of the U4/U6 di-snRNP-specific protein PRP4, based on a conserved PRP4 domain and seven C-terminal WD40 repeats (Luz Ambrósio et al. 2009). Further support for this assignment came from the efficient co-precipitation of U4 and U6 snRNAs. Surprisingly, U5 snRNA was not coprecipitated, in contrast to human PRP4, which stably associates with the U4/U6*U5 tri-snRNP. With PRP31, in contrast, a real tri-snRNP protein was identified for *T. brucei*; PRP31 is essential for *cis* and *trans* splicing and contributes to U4/U6*U5 tri-snRNP stability (Liang et al. 2006). However, no comprehensive protein identification of the trypanosomal tri-snRNP has been achieved so far.

With respect to the trypanosomal U6 snRNP, the U6 snRNA had originally been discovered by anti-TMG co-immunoprecipitation (Mottram et al. 1989). In analogy to other eukaryotes, the trypanosomal U6 snRNA is transcribed by RNA

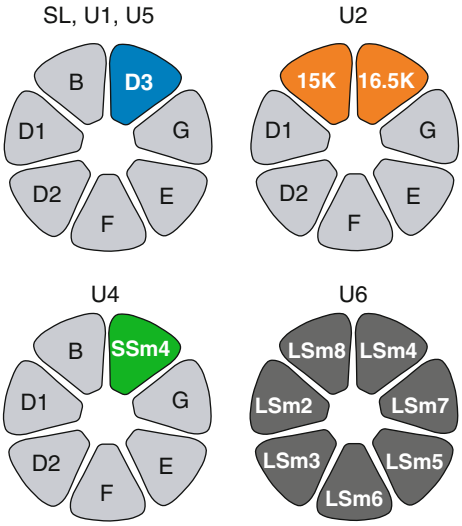


Fig. 3.4 Sm core variation The arrangement of Sm proteins in the canonical Sm core is according to Palfi et al. (2000). This composition has been experimentally verified for the SL RNP, the U1, and U5 snRNPs from *T. brucei*. The U2 snRNP, however, contains a variant Sm core, in which the canonical SmB and SmD3 are replaced by Sm15K and Sm16.5K, respectively (Wang et al. 2006). A second core variation was identified in the U4 snRNP, in which the canonical SmD3 is replaced by SSm4 (Tkacz et al. 2007; Jaé et al. 2010). The order of the U6-specific LSm subunits was defined based on protein–protein interaction data and in analogy to the canonical Sm ring (Liu et al. 2004; Khusial et al. 2005)

polymerase III and carries a γ-monomethylphosphate cap structure. Finally, biochemical studies demonstrated two distinct complexes for the U6 snRNA: singular U6 snRNPs and U4/U6 di-snRNPs (Cross et al. 1991).

As established in the yeast and mammalian systems, the U6 snRNA—in contrast to the other spliceosomal snRNAs—contains no Sm site; instead, it associates with seven LSm (like-Sm) proteins, which share structural similarity with the Sm proteins and assemble as a heteroheptameric ring (LSm2-8) around the 3′-terminal uridine stretch of the U6 snRNA (Fig. 3.4). This nuclear complex remains associated with the U6 snRNA until formation of the U4/U6 di- or the U4/U6*U5 tri-snRNP (Achsel et al. 1999; Mayes et al. 1999; Gottschalk et al. 1999; Stevens and Abelson 1999). A second complex, composed of the proteins LSm1-7, is found in the cytoplasm rather than in the nucleus, where it binds to the 3′ untranslated regions (UTR) of mRNAs destined to be degraded (Bouveret et al. 2000; Ingelfinger et al. 2002).

For the trypanosomal system, initial database homology searches identified seven putative *T. brucei* LSm proteins, LSm2p to LSm8p (Liu et al. 2004). In the same study, the function and identity of two of those candidate proteins (LSm3p and LSm8p) was assessed by RNA interference: Both proteins were found to be essential for mRNA decay, U6 snRNA stability, as well as U4/U6 di-snRNP and U4/U6*U5 tri-snRNP formation. Interestingly, a trypanosomal homolog of the LSm1 protein, which specifically participates in the process of mRNA decay, was

not identified, suggesting that in trypanosomes only a single LSm complex exists. The subsequent analysis of the remaining five LSm proteins by RNAi silencing and snRNA coprecipitation after tandem affinity purification revealed that two of these proteins were actually part of the variant U2- and U4-specific Sm cores (Tkacz et al. 2007, compare also Wang et al. 2006; Jaé et al. 2010). The remaining LSm subunits were finally completed to a heptameric system by the identification of the two missing LSm proteins (Tkacz et al. 2008).

3.3.5 U5 snRNP

The trypanosomal U5 snRNA was discovered as the last of the *trans*-spliceosomal snRNAs, because the initial screening for trypanosome snRNAs by anti-TMG immunoprecipitation failed to identify a U5 snRNA candidate (Mottram et al. 1989). Subsequently, the spliced leader-associated (SLA1) RNA was considered to be the U5 snRNA homolog, based on in vivo psoralen crosslinking to the SL RNA and a short stretch of homology (Watkins et al. 1994). Finally, Dungan et al. (1996) identified the real *T. brucei* U5 snRNA by psoralen crosslinks with the SL RNA and free 5′ exon splicing intermediates and found it to form an RNP complex with the U4 and U6 snRNAs. Further support for an appropriate identification came from detailed analysis of the corresponding U5 snRNAs from *Leptomonas collosoma* (Xu et al. 1997) and *Leptomonas seymouri* (Bell and Bindereif 1999). A remarkable characteristic of the trypanosomal U5 snRNA is the 5′ phosphate terminus instead of the usual TMG cap (Xu et al. 1997).

Besides the canonical set of seven Sm proteins, the homolog of the yeast U5-specific protein PRP8 (p220 in human) was found to be associated with the *T. brucei* U5 snRNA, conclusively confirming the existence of an U5-analogous RNP in trypanosomes (Lücke et al. 1997). PRP8, the largest known spliceosomal protein, is evolutionary highly conserved and thought to participate in the catalytic center of the spliceosome (reviewed by Grainger and Beggs 2005).

In addition, trypanosome orthologs of the U5-specific proteins U5-Cwc21 and U5-40K were recently identified (Luz Ambrósio et al. 2009). The trypanosomal U5-Cwc21 shares a highly conserved N-terminal region with the human SRm300/SRRM2 and the yeast Cwc21p proteins. In contrast to the yeast counterpart, U5-Cwc21 from *T. brucei* is essential for splicing and associates predominantly with the U5 snRNA, but not with U2, suggesting a unique function for trypanosomal U5-Cwc21 (Luz Ambrósio et al. 2009).

3.3.6 U1 snRNP

Regarding its structure and functional role during spliceosomal assembly, the mammalian U1 snRNP is one of the spliceosomal snRNPs that is characterized

best (Stark et al. 2001; Pomeranz Krummel et al. 2009). In addition to the Sm core, the U1 snRNP consists of three specific polypeptides: U1-70K, U1A, and U1C. Both U1-70K and U1A contain an RNA recognition motif, whereby they associate specifically with conserved loop positions in stem-loops I and II of the U1 snRNA, respectively (Query et al. 1989; Scherly et al. 1989). The U1 snRNP functions early in spliceosome assembly, specifically in 5′ splice site recognition, by base pairing with the 5′ splice site. Furthermore, it was demonstrated that U1C protein is able to recognize the 5′ splice site sequence in the absence of the U1 snRNP (Rossi et al. 1996; Du and Rosbash 2002; Lund and Kjems 2002). However, mammalian U1C does not bind to free U1 snRNA and integrates in the U1 snRNP via protein–protein interactions with U1-70K and the Sm core complex (Nelissen et al. 1994).

As the other spliceosomal snRNAs, the trypanosomatid U1 snRNA is much smaller than its known orthologs, lacking stem-loops II and III (75 nt versus 164 nt for the mammalian and 568 nt for the yeast U1; Schnare and Gray 1999; Djikeng et al. 2001; Palfi et al. 2002). The trypanosome U1 snRNA contains an Sm site, on which the seven canonical Sm proteins assemble. Notably, the U1 snRNP most probably acts only in *cis*, not *trans* splicing, as demonstrated by in vitro and in vivo studies in the nematode and trypanosomal systems, respectively (Denker et al. 1996; Tkacz et al. 2010). In trypanosomes, it was shown that the poly(A) polymerase (PAP) pre-mRNA, which undergoes *cis* splicing, can extensively base pair at its 5′ splice site with the U1 snRNA (Djikeng et al. 2001). That *cis* splicing function indeed requires 5′ splice site base pairing was supported by point mutations in the poly(A) polymerase (PAP) 5′ splice site, which were predicted to disrupt base pairing and did abolish splicing (Mair et al. 2000a).

The constituents of the trypanosomal U1 snRNP differ from its human counterparts: By tandem affinity purification (TAP), the orthologs of U1-70K and U1C could be identified, although both proteins are only distantly related to their known counterparts from other eukaryotes. In particular, the *T. brucei* U1-70K homolog represents a minimal version, and in contrast to the mammalian U1 snRNP, U1C from *T. brucei* is able to interact directly with the 5′ terminal region of U1 snRNA (Palfi et al. 2005). Interestingly, in an initial mass-spectrometric analysis, U1A could not be found, instead a novel trypanosomatid-specific U1 snRNP-specific protein, U1-24K, was identified. In contrast to U1A, U1-24K lacks any recognizable RNA-binding motif and integrates into the U1 snRNP by interaction with U1-70K (Palfi et al. 2005).

Luz Ambrósio et al. (2009) were successful in detecting a U1A ortholog in *T. brucei*; this was a surprising finding, since the U1 snRNA lacks the stem-loop II structure, where in other orthologs the U1A binding site resides. The U1 snRNP expression levels and its unusual protein constituents suggest that the trypanosome U1 snRNP may function beyond *cis* splicing. Recently, a follow-up study provided initial support for this, demonstrating that U1A is associated with the polyadenylation factor CPSF73 and also participates in polyadenylation (Tkacz et al. 2010), possibly linking 3′ end processing and splicing. Additional supportive evidence comes from the mammalian system, where splicing-independent functions of the U1 snRNP have already been reported. The U1A protein, for

instance, directly regulates the 3′ end maturation of its own mRNA (Boelens et al. 1993). Furthermore, a recent genome-wide study revealed that the U1 snRNP can suppress cryptic polyadenylation signals in introns (Kaida et al. 2010).

3.4 Trypanosome SMN, a Minimal Assembly Factor for Canonical Sm Cores

A key step in the biogenesis of the spliceosomal snRNPs is the association of the Sm proteins with the snRNAs. Several studies in the mammalian system have demonstrated that the so-called survival of motor neurons (SMN) complex plays a key role in this process (Meister et al. 2001; Pellizzoni et al. 2002). The SMN complex consists of the eponymous SMN protein and at least seven associated proteins, named Gemins 2–8, and collaborates with additional factors, such as the methyltransferase PRMT5 (for review, see Neuenkirchen et al. 2008). Interestingly, the human SMN protein was identified by searching for spinal muscular atrophy (SMA) determining genes. SMA is a common recessive disease, characterized by the degeneration of motor neurons in the spinal cord and caused by mutation in the *SMN1* gene (Lefebvre et al. 1995).

The mammalian SMN complex is predominantly localized in the cytoplasm, except for the nuclear Cajal bodies and Gems, and participates in these major snRNP maturation events: First, SMN functions as a chaperone during assembly of the Sm ring around the Sm site; second, SMN plays a crucial role during hypermethylation of the monomethyl cap after the assembly of the Sm ring, in which the trimethyl guanosine synthase 1 (Tgs1) is associated with the SMN complex (Mouaikel et al. 2003; Plessel et al. 1994); third, SMN associates with the exonuclease ISG20 and is involved in 3′ trimming of pre-snRNAs (Espert et al. 2006).

The best characterized function of mammalian SMN is its role during assembly of the Sm core around the Sm site of the snRNAs, in which the SMN complex interacts with SmD1/D2-E/F/G and SmD3/B subcomplexes in an open ring formation. Interaction with the snRNAs leads to ring closure and dissociation of the SMN complex (Raker et al. 1996; Chari et al. 2008). Regarding protein–protein interactions, SMN binds through its internal Tudor domain with high affinity to symmetric dimethyl arginines (sDMAs) in the RG-rich C-termini of SmB/B′, D1 and D3. sDMA modifications on the Sm proteins take place in the cytoplasm and are catalyzed by the so-called methylosome; this 20S complex consists of the methyltransferase PRMT5, the WD40 protein MEP50, and the pICln protein (Friesen et al. 2001, 2002, for review, see Yong et al. 2004). Moreover, pICln also exists in a smaller, snRNA-free complex of 6S, in which it interacts with the Sm domains of a subset of Sm proteins (Friesen et al. 2001), specifically with the SmD1D2-EFG pentamer (Chari et al. 2008). In addition, the central Tudor domain of SMN resembles the classical Sm fold and may interact directly with the Sm core domain (Côté and Richard 2005; Selenko et al. 2001; Sprangers et al. 2003). In the

current model for the mammalian system, the Sm proteins are transferred from the 6S pICln complex to the SMN complex, which then allows snRNA binding, closure of the Sm ring, and RNP release (Chari et al. 2008).

To what extent is this classical model of SMN function valid in the trypanosome system? After the first Sm core variation in trypanosomatids was characterized (Wang et al. 2006), new questions arose regarding the specificity of biogenesis and assembly of Sm cores. By searching for specificity determining factors in *T. brucei*, tandem affinity purification of SmB followed by mass spectrometry revealed highly divergent orthologs of SMN and Gemin2. Since no other Gemins could be identified, the trypanosomatid SMN complex likely reflects a simpler ancestral core complex, similar to what had been demonstrated for *Drosophila melanogaster*. Here, such a minimal complex, composed only of SMN and Gemin2, discriminates against nontarget RNAs during Sm assembly (Kroiss et al. 2008).

In vitro Sm core reconstitutions of recombinant trypanosomal proteins revealed that Sm core assembly functions efficiently in the absence of SMN. However, if SMN is present, it adds a striking discriminatory activity: The canonical Sm core was efficiently assembled in vitro with SL RNA and U5 snRNA, but assembly was strongly reduced with U4 and U2 snRNAs, which associate with Sm core variants, or with a U5 snRNA carrying a mutated Sm site. This was consistent with findings that the SMN complex specifically bound the canonical SmD3/B subcomplex, directly contacting SmB, apparently a parasite-specific protein–protein interaction (Palfi et al. 2009). The above-mentioned interaction between SMN and the Sm proteins in mammals relies on the dimethylated arginines within the RG-rich C-termini of Sm proteins. Strikingly, trypanosomes lack both major interaction determinants of SMN, the SMN Tudor domain and the RG-rich domains in the Sm proteins (Palfi et al. 1991, 2009).

Another important difference relates to localization: Whereas the human SMN protein is predominantly cytoplasmic, its trypanosomatid counterpart is localized within the nucleus, suggesting that trypanosomal snRNP biogenesis, at least the Sm core assembly phase, proceeds in the nucleus. This is consistent with fluorescence in situ hybridization experiments that localized the SL RNA and U2 snRNA in the trypanosome nucleus (Biton et al. 2006; Tkacz et al. 2007; Jaé et al. 2011). Taken together, the trypanosome SMN complex represents a minimal Sm core assembly system, which mechanistically performs both chaperone- and specificity functions during Sm core assembly.

3.5 Non-snRNP Spliceosome Components and Their Putative Role in Splicing Regulation

The stepwise assembly pathway of the spliceosome on pre-mRNA is best characterized in the yeast and mammalian system and is initiated by the recognition of common pre-mRNA sequence elements: During the first step of spliceosome

assembly, the early complex (E complex) is formed by binding of the U1 snRNP to the 5′ splice site and recognition of the polypyrimidine tract and the 3′ splice site by the U2 auxiliary factors U2AF65 and U2AF35, respectively (reviewed by Wahl et al. 2009). Both splicing factors, U2AF65 and U2AF35 bind as a heterodimeric complex (Zamore and Green 1989). In addition, this early assembly phase of the spliceosome involves recognition of the branch point sequence by the branch point binding protein (BBP, also termed SF1) and cooperative interaction with U2AF65 facilitates the process of branch point recognition (Berglund et al. 1997, 1998).

Initial attempts to identify trypanosomal homologs of these early spliceosome assembly factors revealed only a U2AF35 candidate for *T. cruzi*. Surprisingly, residues predicted to be involved in the reciprocal U2AF35–U2AF65 interaction are not conserved (Vázquez et al. 2003). In a more recent study, the *T. cruzi*/*T. brucei* U2AF65 and SF1 proteins as well as the *T. brucei* U2AF35 protein were identified and characterized (Vázquez et al. 2009): RNAi-mediated silencing of TbU2AF35, TbU2AF65, and TbSF1 affected the first step of *trans* splicing. Moreover, as shown by yeast two-hybrid assays, fractionation of cell lysates and affinity selection, neither the *T. cruzi*, nor the *T. brucei* U2AF35 and U2AF65 proteins interact with each other, which stands in clear contrast to other eukaryotic systems. However, the interaction between U2AF65 and SF1 seems to be conserved in both species. Finally, sequence alignments revealed that the trypanosomal U2AF65 proteins strongly diverge from their human and yeast orthologs, whereas the Tb/TcSF1 proteins are more conserved, including the KH RNA-binding domain and the N-terminal Ser-Arg-Trp (SRW) motif, which are essential for U2AF65 interaction (Vázquez et al. 2009).

In addition to the necessity of splicing core elements like U2AF35–U2AF65, the selection of a precise splice site is determined by its relative strength and multiple factors, as shown for higher eukaryotes. In particular, this process is often regulated by intronic or exonic splicing silencers or enhancers (ISS, ESS, ISE, ESE), that are *cis*-acting elements within the pre-mRNA, and corresponding *trans*-acting protein factors that recognize the *cis* elements, thereby recruiting components of the splicing machinery to the splice sites or preventing their association (reviewed by Wang and Burge 2008). Most splicing regulators characterized so far belong either to the group of serine/arginine-rich (SR) proteins or to the group of heterogeneous nuclear ribonucleoproteins (hnRNPs). Remarkably, in lower eukaryotes such as yeast, SR proteins as well as exonic splicing enhancers appear to play only a very minor regulatory role. Moreover, in yeast the splicing-essential consensus sequences exhibit a relatively high degree of complementarity to U1 and U2 snRNAs, suggesting that degeneration of those consensus elements in higher eukaryotes correlates with their flexible recognition regulated by *trans* factors (reviewed by Izquierdo and Valcárcel 2006).

Are splicing processes in the trypanosomal system regulated? Research has focused so far on mRNA stability and translation, which are considered the major regulatory levels of gene expression in trypanosomes. Initial studies had identified the 3′ splice site and a polypyrimidine tract of variable length as *trans* splicing essential elements (Huang and van der Ploeg 1991; Matthews et al. 1994).

López-Estraño et al. (1998) suggested a model, in which *trans* splicing efficiency is modulated by exonic sequences. Moreover, a systematic study of those sequence requirements revealed that the presence of an AG dinucleotide in the 5′ UTR enhances *trans* splicing in constructs containing a suboptimal dinucleotide immediately upstream of the 3′ splice site (Siegel et al. 2005). Those findings suggested that in certain cases 5′ UTR sequences can contribute to efficient *trans* splicing and might function in an enhancer-like manner.

As mentioned earlier, such splicing enhancer elements are often bound by serine/arginine-rich proteins, which usually activate weak splice sites, as shown in the mammalian system (Schaal and Maniatis 1999). In general, all SR proteins exhibit a modular organization, containing an N-terminal RNA-binding domain and a C-terminal RS domain for protein–protein interactions (reviewed by Graveley 2006).

In trypanosomes, a protein termed RRM1 was discovered through a degenerate PCR approach, aiming at identifying genes that encode RNA recognition motifs (Manger and Boothroyd 1998). This protein is expressed in both procyclic and bloodstream forms and localizes in nuclear substructures, which may be functionally analogous to the speckles of higher eukaryotes. A second trypanosomal SR protein, termed TSR1, was identified thereafter by Ismaïli et al. (1999). TRS1 also exhibits a characteristic nuclear pattern in the bloodstream form and was shown to interact with the heterologous human U2AF complex and the SL RNA, leading to the assumption that TSR1 may be involved in *trans* splicing in *T. brucei*. Besides SR proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs) can also function as splicing regulators, usually involved in repression, and sometimes to activation (reviewed by Chen and Manley 2009). Characteristically, hnRNPs carry one or more RNA-recognition motifs and additional protein–protein interaction domains. For example, PTB, the polypyrimidine-tract binding protein also known as PTB1 or hnRNP I, represents a well-characterized splicing regulator of the hnRNP family. PTB can act as a repressor of numerous alternatively spliced pre-mRNAs (reviewed by Valcárcel and Gebauer 1997; Wagner and Garcia-Blanco 2001). Several modes for PTB-mediated repression, depending on the localization of the PTB binding sites, have been proposed; for example, PTB binding at the polypyrimidine tract can sterically interfere with other splicing factors such as U2AF65 and the U2 snRNP (Singh et al. 1995; Saulière et al. 2006).

In analogy to *cis* splicing, the 3′ splice sites of polycistronic pre-mRNAs in trypanosomes are also preceded by polypyrimidine tracts. However, in trypanosomatids, the polypyrimidine tracts are bifunctional determinants, affecting *trans* splicing as well as polyadenylation (LeBowitz et al. 1993; Matthews et al. 1994). Characteristically, trypanosomal polypyrimidine tracts often do not extend up to the 3′ splice site, and they appear to be associated with the branch point at variable distances from the 3′ splice site (Lücke et al. 2005).

An attempt by De Gaudenzi et al. (2005) to classify all known RRM-containing proteins of *T. brucei*, *T. cruzi*, and *L. major* revealed two RNA-binding proteins, DRBD3 and DRBD4, that resemble the mammalian PTB, which contains four RRMs, each of them capable of RNA binding (Oberstrass et al. 2005). A homology comparison of the individual RRMs in the trypanosomal DRBD3 and DRDB4

proteins versus the mammalian PTB suggested that the two trypanosomal proteins may jointly perform the function of the mammalian PTB (De Gaudenzi et al. 2005). Recently, the role of DRBD3 and DRBD4 in mRNA processing and mRNA stability was analyzed in more detail by RNAi knockdown and microarray analysis (Stern et al. 2009): Neither DRDB3 nor DRDB4 acted as general splicing factors for all pre-mRNAs, since no significant overlap was observed between the transcripts regulated by either protein. Instead, *trans* splicing of only a subset of genes seemed to depend on DRDB3 or DRDB4, respectively, and *cis* splicing was only affected by DRDB3 downregulation; in addition, both proteins affected the stability of different subsets of mRNAs. Taken together, these results suggested that the putative trypanosomal counterparts of PTB are multifunctional RNA-binding proteins.

Alternative splicing is fundamental for increasing mRNA complexity: A single pre-mRNA can be processed to yield multiple mRNAs, which occurs in at least 90% of the human protein-coding genes (Pan et al. 2008). Since *trans* splicing joins the noncoding SL mini-exon and the protein-coding sequence and since there are only very few *cis* spliced genes, the question arose whether alternative splicing processes play any role in trypanosomes. An initial study that used a reporter gene expressed from an rRNA locus demonstrated putative alternative splicing events in the different life stages of trypanosomes (Helm et al. 2008). However, based on a systematic genome-wide high-throughput RNA sequencing approach, Siegel et al. (2010) proposed that life cycle-dependent alternative splicing is not a widespread phenomenon in *T. brucei*. Most of the expressed genes use between one and three alternative 3′ splice sites. In contrast, another genome-wide RNA-Seq study provided evidence for alternative *trans* splicing in trypanosomes (Nilsson et al. 2010). In this study, four different consequences of alternative splicing were differentiated: first, the lack of a start codon, abolishing translation; second, variations in targeting signals, leading to differential localization of the corresponding gene products; third, heterogeneity in regulatory elements such as an upstream open-reading frame; fourth, the use of an alternative open-reading frame. In sum, over 2,500 life-cycle-dependent alternative splicing events were observed in this study, suggesting this provides an important regulatory mechanism during trypanosome development. For example, a recent study described that alternatively *trans*-spliced transcripts determine the dual localization of a trypanosomal protein in the cytosol versus the mitochondrium (Rettig et al. 2012).

References

Achsel T, Brahms H, Kastner B, Bachi A, Wilm M, Lührmann R (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. EMBO J 18:5789–5802

Avila ML, Bercovich N, Westergaard G, Levin MJ, Vázquez MP (2007) Mapping of the protein-binding interface between splicing factors SF3b155 and p14 of *Trypanosoma cruzi*. Biochem Biophys Res Commun 364:26–32

Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC (1992) Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J Biol Chem 267:9805–9815
Barth S, Shalem B, Hury A, Tkacz ID, Liang XH, Uliel S, Myslyuk I, Doniger T, Salmon-Divon M, Unger R, Michaeli S (2008) Elucidating the role of C/D snoRNA in rRNA processing and modification in *Trypanosoma brucei*. Eukaryot Cell 7:86–101
Bell M, Bindereif A (1999) Cloning and mutational analysis of the *Leptomonas seymouri* U5 snRNA gene: function of the Sm site in core RNP formation and nuclear localization. Nucleic Acids Res 27:3986–3994
Berglund JA, Chua K, Abovich N, Reed R, Rosbash M (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell 89:781–787
Berglund JA, Abovich N, Rosbash M (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev 12:858–867
Berriman M, Ghedin E, Hertz-Fowler C et al (2005) The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416–422
Biton M, Mandelboim M, Arvatz G, Michaeli S (2006) RNAi interference of XPO1 and Sm genes and their effect on the spliced leader RNA in *Trypanosoma brucei*. Mol Biochem Parasitol 150:132–143
Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72:291–336
Boelens WC, Jansen EJ, van Venrooij WJ, Stripecke R, Mattaj IW, Gunderson SI (1993) The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell 72:881–892
Boothroyd JC, Cross GA (1982) Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene 20:281–289
Bouveret E, Rigaut G, Shevchenko A, Wilm M, Séraphin B (2000) A Sm-like protein complex that participates in mRNA degradation. EMBO J 19:1661–1671
Brahms H, Raymackers J, Union A, de Keyser F, Meheus L, Lührmann R (2000) The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. J Biol Chem 275:17122–17129
Brahms H, Meheus L, de Brabandere V, Fischer U, Lührmann R (2001) Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B′ and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA 7:1531–1542
Branlant C, Krol A, Ebel JP, Lazar E, Haendler B, Jacob M (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J 1:1259–1265
Brow DA (2002) Allosteric cascade of spliceosome activation. Annu Rev Genet 36:333–360
Burge C, Tuschl T, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Campbell DA, Thornton DA, Boothroyd JC (1984) Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature 311:350–355
Campbell DA, Sturm NR, Yu MC (2000) Transcription of the kinetoplastid spliced leader RNA gene. Parasitol Today 16:78–82
Chari A, Golas MM, Klingenhäger M, Neuenkirchen N, Sander B, Englbrecht C, Sickmann A, Stark H, Fischer U (2008) An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. Cell 135:497–509
Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10:741–754
Côté J, Richard S (2005) Tudor domains bind symmetrical dimethylated arginines. J Biol Chem 280:28476–28483
Cross M, Günzl A, Palfi Z, Bindereif A (1991) Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. Mol Cell Biol 11:5516–5526

Cross M, Wieland B, Palfi Z, Günzl A, Röthlisberger U, Lahm HW, Bindereif A (1993) The *trans*-spliceosomal U2 snRNP protein 40K of *Trypanosoma brucei*: cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. EMBO J 12:1239–1248

De Gaudenzi J, Frasch AC, Clayton C (2005) RNA-binding domain proteins in kinetoplastids: a comparative analysis. Eukaryot Cell 4:2106–2114

De Lange T, Liu AY, Van der Ploeg LH, Borst P, Tromp MC, Van Boom JH (1983) Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell 34:891–900

De Lange T, Berkvens TM, Veerman HJ, Frasch AC, Barry JD, Borst P (1984) Comparison of the genes coding for the common 5′ terminal sequence of messenger RNAs in three trypanosome species. Nucleic Acids Res 12:4431–4443

Denker JA, Maroney PA, Yu YT, Kanost RA, Nilsen TW (1996) Multiple requirements for nematode spliced leader RNP function in *trans*-splicing. RNA 2:746–755

Djikeng A et al (2001) Characterization of a candidate *Trypanosoma brucei* U1 small nuclear RNA gene. Mol Biochem Parasitol 113:109–115

Du H, Rosbash M (2002) The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. Nature 419:86–90

Dungan JM, Watkins KP, Agabian N (1996) Evidence for the presence of a small U5-like RNA in active *trans*-spliceosomes of *Trypanosoma brucei*. EMBO J 15:4016–4029

El-Sayed NM et al (2003) The sequence and analysis of *Trypanosoma brucei* chromosome II. Nucleic Acids Res 31:4856–4863

El-Sayed NM et al (2005) Comparative genomics of trypanosomatid parasitic protozoa. Science 309:404–409

Espert L, Eldin P, Gongora C, Bayard B, Harper F, Chelbi-Alix MK, Bertrand E, Degols G, Mechti N (2006) The exonuclease ISG20 mainly localizes in the nucleolus and the Cajal (Coiled) bodies and is associated with nuclear SMN protein-containing complexes. J Cell Biochem 98:1320–1333

Fantoni A, Dare AO, Tschudi C (1994) RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. Mol Cell Biol 14:2021–2028

Friesen WJ, Paushkin S, Wyce A, Massenet S, Pesiridis GS, Van Duyne G, Rappsilber J, Mann M, Dreyfuss G (2001) The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol Cell Biol 21:8289–8300

Friesen WJ, Wyce A, Paushkin S, Abel L, Rappsilber J, Mann M, Dreyfuss G (2002) A novel WD repeat protein component of the methylosome binds Sm proteins. J Biol Chem 277:8243–8247

Gilinger G, Bellofatto V (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Res 29:1556–1564

Gottschalk A, Neubauer G, Banroques J, Mann M, Lührmann R, Fabrizio P (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. EMBO J 18:4535–4548

Grainger RJ, Beggs JD (2005) Prp8 protein: at the heart of the spliceosome. RNA 11:533–557

Graveley BR (2006) Sorting out the complexity of SR protein functions. RNA 6:1197–1211

Günzl A (2010) The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell 9:1159–1170

Günzl A, Cross M, Bindereif A (1992) Domain structure of U2 and U4/U6 small nuclear ribonucleoprotein particles from *Trypanosoma brucei*: identification of *trans*-spliceosomal specific RNA-protein interactions. Mol Cell Biol 12:468–479

Hall N et al (2003) The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. Nucleic Acids Res 31:4864–4873

Hartshorne T, Agabian N (1990) A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. Genes Dev 4:2121–2131

Helm JR, Wilson ME, Donelson JE (2008) Different trans RNA splicing events in bloodstream and procyclic *Trypanosoma brucei*. Mol Biochem Parasitol 159:134–137

Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Lührmann R (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. EMBO J 14:2076–2088

Holetz FB, Correa A, Avila AR, Nakamura CV, Krieger MA, Goldenberg S (2007) Evidence of P-body-like structures in *Trypanosoma cruzi*. Biochem Biophys Res Commun 356:1062–1067

Huang J, van der Ploeg LH (1991) Requirement of a polypyrimidine tract for *trans*-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3′ splice acceptor site. EMBO J 10:3877–3885

Hury A, Goldshmidt H, Tkacz ID, Michaeli S (2009) Trypanosome spliced-leader-associated RNA (SLA1) localization and implications for spliced-leader RNA biogenesis. Eukaryot Cell 8:56–68

Ingelfinger D, Arndt-Jovin DJ, Lührmann R, Achsel T (2002) The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. RNA 8:1489–1501

Ismaïli N, Pérez-Morga D, Walsh P, Mayeda A, Pays A, Tebabi P, Krainer AR, Pays E (1999) Characterization of a SR protein from *Trypanosoma brucei* with homology to RNA-binding *cis*-splicing proteins. Mol Biochem Parasitol 102:103–115

Ivens et al (2005) The genome of the kinetoplastid parasite, *Leishmania major*. Science 309:436–442

Izquierdo JM, Valcárcel J (2006) A simple principle to explain the evolution of pre-mRNA splicing. Genes Dev 20:1679–1684

Jaé N, Wang P, Gu T, Hühn M, Palfi Z, Urlaub H, Bindereif A (2010) Essential role of a trypanosome U4-specific Sm core protein in small nuclear ribonucleoprotein assembly and splicing. Eukaryot Cell 9:379–386

Jaé N, Preußer C, Krüger T, Tkacz ID, Engstler M, Michaeli S, Bindereif A (2011) snRNA-specific role of SMN in trypanosome snRNP biogenesis. RNA Biol 8:90–100

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468:664–668

Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, Nagai K (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 96:375–387

Kastner B, Bach M, Lührmann R (1990) Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. Proc Natl Acad Sci USA 87:1710–1714

Khusial P, Plaag R, Zieve GW (2005) LSm proteins form heptameric rings that bind to RNA via repeating motifs. Trends Biochem Sci 30:522–528

Konarska MM, Padgett RA, Sharp PA (1985) *Trans* splicing of mRNA precursors *in vitro*. Cell 42:165–171

Kooter JM, De Lange T, Borst P (1984) Discontinuous synthesis of mRNA in trypanosomes. EMBO J 3:2387–2392

Kroiss M, Schultz J, Wiesner J, Chari A, Sickmann A, Fischer U (2008) Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster*. Proc Natl Acad Sci USA 105:10045–10050

Laird PW, ten Asbroek AL, Borst P (1987) Controlled turnover and 3′ trimming of the trans splicing precursor of *Trypanosoma brucei*. Nucleic Acids Res 15:10087–10103

LeBowitz JH, Smith HQ, Rusche L, Beverley SM (1993) Coupling of poly(A) site selection and *trans*-splicing in *Leishmania*. Genes Dev 7:996–1007

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M et al (1995) Identification and characterization of a spinal muscular atrophy-determining gene. Cell 80:155–165

Leung AK, Nagai K, Li J (2011) Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. Nature 473:536–539

Lewis JD, Izaurralde E (1997) The role of the cap structure in RNA processing and nuclear export. Eur J Biochem 247:461–469

Liang XH, Liu Q, Liu L, Tschudi C, Michaeli S (2006) Analysis of spliceosomal complexes in *Trypanosoma brucei* and silencing of two splicing factors Prp31 and Prp43. Mol Biochem Parasitol 145:29–39

Liu Q, Liang XH, Uliel S, Belahcen M, Unger R, Michaeli S (2004) Identification and functional characterization of lsm proteins in *Trypanosoma brucei*. J Biol Chem 279:18210–18219

López-Estraño C, Tschudi C, Ullu E (1998) Exonic sequences in the 5′ untranslated region of alpha-tubulin mRNA modulate *trans* splicing in *Trypanosoma brucei*. Mol Cell Biol 18:4620–4628

Lücke S, Klöckner T, Palfi Z, Boshart M, 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

Lücke S, Jürchott K, Hung LH, Bindereif A (2005) mRNA splicing in *Trypanosoma brucei*: branch-point mapping reveals differences from the canonical U2 snRNA-mediated recognition. Mol Biochem Parasitol 142:248–251

Lund M, Kjems J (2002) Defining a 50 splice site by functional selection in the presence and absence of U1 snRNA 5′ end. RNA 8:166–179

Luz Ambrósio D, Lee JH, Panigrahi AK, Nguyen TN, Cicarelli RM, Günzl A (2009) Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors. Eukaryot Cell 8:990–1000

Maier AG, Webb H, Ding M, Bremser M, Carrington M, Clayton C (2001) The coatomer of *Trypanosoma brucei*. Mol Biochem Parasitol 115:55–61

Mair G, Shi HF, Li HJ et al (2000a) A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. RNA 6:163–169

Mair G, Ullu E, Tschudi C (2000b) Cotranscriptional cap 4 formation on the *Trypanosoma brucei* spliced leader RNA. J Biol Chem 275:28994–28999

Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, Lührmann R (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science 298:2205–2208

Mandelboim M, Barth S, Biton M, Liang XH, Michaeli S (2003) Silencing of Sm proteins in *Trypanosoma brucei* by RNA interference captured a novel cytoplasmic intermediate in spliced leader RNA biogenesis. J Biol Chem 278:51469–51478

Manful T, Cristodero M, Clayton C (2009) DRBD1 is the *Trypanosoma brucei* homologue of the spliceosome-associated protein 49. Mol Biochem Parasitol 166:186–189

Manger ID, Boothroyd JC (1998) Identification of a nuclear protein in *Trypanosoma brucei* with homology to RNA-binding proteins from *cis*-splicing systems. Mol Biochem Parasitol 97:1–11

Marchetti MA, Tschudi C, Kwon H, Wolin SL, Ullu E (2000) Import of proteins into the trypanosome nucleus and their distribution at karyokinesis. J Cell Sci 113:899–906

Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010:525241

Mattaj IW (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell 46:905–911

Matthews KR, Tschudi C, 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

Mayer MG, Floeter-Winter LM (2005) Pre-mRNA *trans*-splicing: from kinetoplastids to mammals, an easy language for life diversity. Mem Inst Oswaldo Cruz 100:501–513

Mayes AE, Verdone L, Legrain P, Beggs JD (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J 18:4321–4331

Meister G, Bühler D, Pillai R, Lottspeich F, Fischer U (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat Cell Biol 3:945–949

Michaeli S, Roberts TG, Watkins KP, Agabian N (1990) Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. J Biol Chem 265:10582–10588

Milhausen M, Nelson RG, Sather S, Selkirk M, Agabian N (1984) Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5′ sequences of trypanosomatid mRNAs? Cell 38:721–729

Moore MJ, Query CC, Sharp PA (1993) Splicing of precursors to mRNA by the spliceosome. In: Gesteland R, Atkins J (eds) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 303–357

Mottram J, Perry KL, Lizardi PM, Lührmann R, Agabian N, Nelson RG (1989) Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. *brucei*: identification of the U2, U4, and U6 RNA analogs. Mol Cell Biol 9:1212–1223

Mottram JC, Bell SD, Nelson RG, Barry JD (1991) tRNAs of *Trypanosoma brucei*. Unusual gene organization and mitochondrial importation. J Biol Chem 266:18313–18317

Mouaikel J, Narayanan U, Verheggen C, Matera AG, Bertrand E, Tazi J, Bordonne R (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. EMBO Rep 4:616–622

Murphy WJ, Watkins KP, Agabian N (1986) Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. Cell 47:517–525

Nakaar V, Dare AO, Hong D, Ullu E, Tschudi C (1994) Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. Mol Cell Biol 14:6736–6742

Nakaar V, Günzl A, Ullu E, Tschudi C (1997) Structure of the *Trypanosoma brucei* U6 snRNA gene promoter. Mol Biochem Parasitol 88:13–23

Nelissen RL, Will CL, van Venrooij WJ, Lührmann R (1994) The association of the U1-specific 70K and C proteins with U1 snRNPsis mediated in part by common U snRNP proteins. EMBO J 13:4113–4125

Neuenkirchen N, Chari A, Fischer U (2008) Deciphering the assembly pathway of Sm-class U snRNPs. FEBS Lett 582:1997–2003

Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. PLoS Pathog 6(8):e1001037

Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science 309:2054–2057

Palfi Z, Bindereif A (1992) Immunological characterization and intracellular localization of *trans*-spliceosomal small nuclear ribonucleoproteins in *Trypanosoma brucei*. J Biol Chem 267:20159–20163

Palfi Z, Günzl A, Cross M, Bindereif A (1991) Affinity purification of *Trypanosoma brucei* small nuclear ribonucleoproteins reveals common and specific protein components. Proc Natl Acad Sci USA 88:9097–9101

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

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

Palfi Z, Schimanski B, Günzl A, Lücke S, Bindereif A (2005) U1 small nuclear RNP from *Trypanosoma brucei*: a minimal U1 snRNA with unusual protein components. Nucleic Acids Res 33:2493–2503

Palfi Z, Jaé N, Preusser C, Kaminska KH, Bujnicki JM, Lee JH, Günzl A, Kambach C, Urlaub H, Bindereif A (2009) SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes. Genes Dev 23:1650–1664

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415

Parsons M, Nelson RG, Watkins KP, Agabian N (1984) Trypanosome mRNAs share a common 5′ spliced leader sequence. Cell 38:309–316

Pellizzoni L, Yong J, Dreyfuss G (2002) Essential role for the SMN complex in the specificity of snRNP assembly. Science 298:1775–1779

Plessel G, Fischer U, Lührmann R (1994) m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. Mol Cell Biol 14:4160–4172

Plessel G, Lührmann R, Kastner B (1997) Electron microscopy of assembly intermediates of the snRNP core: morphological similarities between the RNA-free (E.F.G) protein heteromer and the intact snRNP core. J Mol Biol 265:87–94

Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature 458:475–480

Preußer C, Palfi Z, Bindereif A (2009) Special Sm core complex functions in assembly of the U2 small nuclear ribonucleoprotein of *Trypanosoma brucei*. Eukaryot Cell 8:1228–1234

Query CC, Bentley RC, Keene JD (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell 57:89–101

Raker VA, Plessel G, Lührmann R (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. EMBO J 15:2256–2269

Rettig J, Wang Y, Schneider A, Ochsenreiter T (2012) Dual targeting of isoleucyl-tRNA synthetase in *Trypanosoma brucei* is mediated through alternative trans-splicing. Nucleic Acids Res 40(3):1299–1306

Rossi F, Forne T, Antoine E, Tazi J, Brunel C, Cathala G (1996) Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5′ splice site-U1 snRNP interaction. J Biol Chem 271:23985–23991

Ruan JP, Ullu E, Tschudi C (2007) Characterization of the *Trypanosoma brucei* cap hypermethylase Tgs1. Mol Biochem Parasitol 155:66–69

Saulière J, Sureau A, Expert-Bezançon A, Marie J (2006) The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. Mol Cell Biol 26:8755–8769

Schaal TD, Maniatis T (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol 19:261–273

Scherly D, Boelens W, van Venrooij WJ, Dathan NA, Hamm J, Mattaj IW (1989) Identificationof the RNA binding segment of human U1A protein and definition of its binding site on U1 snRNA. EMBO J 8:4163–4170

Schnare MN, Gray MW (1999) A candidate U1 small nuclear RNA for trypanosomatid protozoa. J Biol Chem 274:23691–23694

Selenko P, Sprangers R, Stier G, Bühler D, Fischer U, Sattler M (2001) SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol 8:27–31

Séraphin B (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. EMBO J 14:2089–2098

Siegel TN, Tan KS, Cross GA (2005) Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. Mol Cell Biol 21:9586–9594

Siegel TN, Hekstra DR, Wang X, Dewell S, Cross G (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res 38:4946–4957

Singh R, Valcárcel J, Green MR (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science 268:1173–1176

Solnick D (1985) Trans splicing of mRNA precursors. Cell 42:157–164

Sprangers R, Groves MR, Sinning I, Sattler M (2003) Highresolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. J Mol Biol 327:507–520

Stark H, Dube P, Lührmann R, Kastner B (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature 409:539–542
Stern MZ, Gupta SK, Salmon-Divon M, Haham T, Barda O, Levi S, Wachtel C, Nilsen TW, Michaeli S (2009) Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. RNA 15:648–665
Stevens SW, Abelson J (1999) Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci USA 96:7226–7231
Sutton RE, Boothroyd JC (1986) Evidence for trans splicing in trypanosomes. Cell 47:527–535
Takagi Y, Sindkar S, Ekonomidis D, Hall MP, Ho CK (2007) *Trypanosoma brucei* encodes a bifunctional capping enzyme essential for cap 4 formation on the spliced leader RNA. J Biol Chem 282:15995–16005
Tkacz ID, Lustig Y, Stern MZ, Biton M, Salmon-Divon M, Das A, Bellofatto V, Michaeli S (2007) Identification of novel snRNA-specific Sm proteins that bind selectively to U2 and U4 snRNAs in *Trypanosoma brucei*. RNA 13:30–43
Tkacz ID, Cohen S, Salmon-Divon M, Michaeli S (2008) Identification of the heptameric Lsm complex that binds U6 snRNA in *Trypanosoma brucei*. Mol Biochem Parasitol 160:22–31
Tkacz ID, Gupta SK, Volkov V, Romano M, Haham T, Tulinski P, Lebenthal I, Michaeli S (2010) Analysis of spliceosomal proteins in trypanosomatids reveals novel functions in mRNA processing. J Biol Chem 285:27982–27999
Tschudi C, Ullu E (1990) Destruction of U2, U4, or U6 small nuclear RNA blocks trans splicing in trypanosome cells. Cell 61:459–466
Tschudi C, Williams SP, Ullu E (1990) Conserved sequences in the U2 snRNA-encoding genes of Kinetoplastida do not include the putative branchpoint recognition region. Gene 91:71–77
Ullu E, Tschudi C (1991) Trans splicing in trypanosomes requires methylation of the 5′ end of the spliced leader RNA. Proc Natl Acad Sci USA 88:10074–10078
Urlaub H, Raker VA, Kostka S, Lührmann R (2001) Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. EMBO J 20:187–196
Valcárcel J, Gebauer F (1997) Post-transcriptional regulation: the dawn of PTB. Curr Biol 7: R705–R708
Van der Ploeg LH, Liu AY, Michels PA, De Lange T, Borst P, Majumder HK, Weber H, Veeneman GH, Van Boom J (1982) RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucleic Acids Res 10:3591–3604
Vázquez M, Atorrasagasti C, Bercovich N, Volcovich R, Levin MJ (2003) Unique features of the *Trypanosoma cruzi* U2AF35 splicing factor. Mol Biochem Parasitol 128:77–81
Vázquez MP, Mualem D, Bercovich N, Stern MZ, Nyambega B, Barda O, Nasiga D, Gupta SK, Michaeli S, Levin MJ (2009) Functional characterization and protein-protein interactions of trypanosome splicing factors U2AF35, U2AF65 and SF1. Mol Biochem Parasitol 164:137–146
Wagner EJ, Garcia-Blanco MA (2001) Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol 21:3281–3288
Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718
Wang Z, Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14:802–813
Wang P, Palfi Z, Preusser C, Lücke S, Lane WS, Kambach C, Bindereif A (2006) Sm core variation in spliceosomal small nuclear ribonucleoproteins from *Trypanosoma brucei*. EMBO J 25:4513–4523
Watkins KP, Agabian N (1991) In vivo UV cross-linking of U snRNAs that participate in trypanosome *trans*-splicing. Genes Dev 5:1859–1869
Watkins KP, Dungan JM, Agabian N (1994) Identification of a small RNA that interacts with the 5′ splice site of the *Trypanosoma brucei* spliced leader RNA *in vivo*. Cell T6:171–182
Westermann S, Weber K (2000) Cloning and recombinant expression of the La RNA-binding protein from *Trypanosoma brucei*. Biochim Biophys Acta 1492:483–487

Will CL, Lührmann R (2001) Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol 13:290–301

Will CL, Lührmann R (2006) Spliceosome structure and function. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA world, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Xu Y, Ben-Shlomo H, Michaeli S (1997) The U5 RNA of trypanosomes deviates from the canonical U5 RNA: the *Leptomonas collosoma* U5 RNA and its coding gene. Proc Natl Acad Sci USA 94:8473–8478

Yong J, Wan L, Dreyfuss G (2004) Why do cells need an assembly machine for RNA–protein complexes? Trends Cell Biol 14:226–232

Zamore PD, Green MR (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci USA 86:9243–9247

Zeiner GM, Foldynová S, Sturm NR, Lukes J, Campbell DA (2004) SmD1 is required for spliced leader RNA biogenesis. Eukaryot Cell 3:241–244

Chapter 4
mRNA Turnover in Trypanosomes

Christine Clayton

Abstract Most kinetoplastid mRNAs are synthesised by RNA polymerase II, then *trans* spliced, polyadenylated, and exported from the nucleus. The polycistronic nature of kinetoplastid transcription, however, means that—with a few exceptions—the amount of mRNA is determined by post-transcriptionally. Once in the cytoplasm, the amount of mRNA is determined by the rate of mRNA degradation and the amount of protein by the rates of translation initiation, elongation, and protein turnover. In many cases, regulation of mRNA turnover has been found to play a dominant role. This review discusses the mechanisms of mRNA degradation in kinetoplastids, and how that degradation is controlled.

4.1 The Role of mRNA Degradation in Determining mRNA Levels

4.1.1 Constitutive and Developmental Regulation of mRNA Abundance

High-throughput RNA Sequencing (RNA-Seq) analyses have shown that the most abundant polymerase II-transcribed mRNAs in trypanosomes, encoding alpha and beta tubulin, were present at about 500 copies per bloodstream-form trypanosome (Siegel et al. 2010). Another study, using quantitative Northern blotting, arrived at a figure of 72 molecules per cell (Haanstra et al. 2008). Since there are about 20 gene copies per haploid genome, the RNA-Seq data imply that there are nine molecules per gene. About three quarters of trypanosome mRNAs that arise from single-copy

C. Clayton (✉)
Zentrum für Molekulare Biologie ZMBH, Im Neuenheimer Feld 282, Postfach 106249, 69120 Heidelberg, Germany
e-mail: cclayton@zmbh.uni-heidelberg.de

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_4,

genes are present at between 1 and 4 copies per cell, while the least abundant mRNAs are present at less than 0.2 copies per cell; the number of mRNAs per gene therefore varies by at least 50-fold (Siegel et al. 2010).

In addition to constitutive control, the levels of some mRNAs vary during the parasite life cycles. Microarray analyses have demonstrated that between 2 and 15% of all mRNAs in kinetoplastid protists are subject to developmental regulation (Jensen et al. 2009; Kabani et al. 2009; Minning et al. 2009; Queiroz et al. 2009; Rochette et al. 2009). The differences in estimates arise from the number of life cycle stages examined and technical differences (different array platforms, hybridisation protocols, statistical approaches, and cut-offs). An analysis, using high-throughput RNA sequencing (RNA-Seq) in two stages of *Trypanosoma brucei*, indicated that 5.6% of mRNAs were at least twofold regulated (Siegel et al. 2010) and revealed the full extent of variations in mRNA copy numbers.

In African trypanosomes, a few highly expressed mRNAs are made by RNA polymerase I. This results in at least 500 times more copies per gene than are achieved by RNA polymerase II, presumably because of efficient transcription initiation and polymerase processivity (Haanstra et al. 2008; Siegel et al. 2010).

4.1.2 Responses to Specific Stimuli

In most organisms, nutrient deprivation or treatment with specific stimuli or inhibitors causes selective accumulation, or depletion, of mRNAs encoding relevant proteins. Kinetoplastids, in contrast, rarely show such responses. Most of the documented instances appear to involve subsets of developmentally regulated genes, affected by stimuli that are relevant to differentiation. For example, changes in the *T. brucei* transcriptome due to altered environmental glycerol (Vassella et al. 2000, 2004) or glucose (Haanstra et al. 2011) are likely to mirror responses to the environment in the Tsetse fly. In non-growing trypanosomes, such as the "stumpy" developmental form, the rates of both protein (Bass and Wang 1992) and RNA (Pays et al. 1993) synthesis are decreased; this is accompanied by a drastic loss of polysomes (Brecht and Parsons 1998) and reductions in various mRNAs required for growth (Queiroz et al. 2009).

Responses to other stimuli have so far been more difficult to find. For example, using a selective microarray that concentrated on genes involved in vesicular transport, Koumandou et al. (2008) found no evident specific changes in pol II-derived mRNAs after RNAi targeting clathrin or Rab5 mRNAs, or treatment with tunicamycin or DTT. Nevertheless, some specific effects have been documented. Heat shock caused strong destabilisation of most *T. brucei* mRNAs, though HSP70 and HSP83 mRNA were spared (Kramer et al. 2008). Analogous effects have been documented more extensively in *Leishmania* (Clayton and Shapira 2007). The most specific examples were seen in *Leishmania*. Starvation of procyclic *Leishmania major* for purines resulted in a threefold increase in the mRNA encoding the NT1 purine transporter; though expression of the protein was enhanced 200-fold

(Ortiz et al. 2010). This could indicate a strong increase in *NT1* mRNA translation efficiency, although degradation of the majority of the newly synthesised NT1 protein immediately after synthesis was not ruled out. Purine transporters NT1 and NT2 in *Leishmania donovani* show similar regulation (Carter et al. 2010).

4.1.3 Measuring RNA Half-Lives

If RNA levels were determined solely by gene copy number and half-life, the copies per gene should be directly proportional to the half-life (Haanstra et al. 2008). Most measurements of mRNA half-life in kinetoplastids rely on the use of the mRNA synthesis inhibitor actinomycin D, since other methods (pulse labelling with ^{3}H-adenine, tetracycline-inducible promoters) work too slowly relative to mRNA decay rates. For the most unstable mRNAs, measurements are compromised because the time to handle the cells—and possibly also penetration of inhibitors—exceeds the half-life. For the longest half-lives, the incubation times become a problem: Actinomycin D should not be used for more than about half a division time, because it stops growth and ultimately kills the cells. Also, oddly, for most of the more stable mRNAs that have been examined, there is an increase in the amount of mRNA (relative to total RNA or stable RNAs) for up to 2 h after addition of the inhibitor. The source of this increase is unresolved. As an alternative, or in addition, indirect inhibition of *trans* splicing, using sinefungin (which inhibits cap methylation) sometimes, but not always, prevents the initial rise in RNA (Colasante et al. 2007).

Inhibition of protein synthesis by cycloheximide or puromycin quite often stabilises kinetoplastid mRNAs. In one case, this has been shown to be a *trans* effect, caused, presumably, by loss of an unstable protein rather than through changes in ribosome loading (Webb et al. 2005). Such results are uninformative since the nature of the unstable protein(s) is unknown.

Most kinetoplastid mRNAs that have been studied in detail showed regulation of mRNA degradation. The most quantitative study was of two *T. brucei* mRNAs encoding developmentally regulated phosphoglycerate kinases. This showed that not only mRNA levels, but also the amounts of mature proteins, could be explained solely on the basis of mRNA degradation rates (Haanstra et al. 2008), which differed by more than an order of magnitude (Colasante et al. 2007); control by splicing was insignificant in comparison. Overall, using actinomycin D with sinefungin, half-lives ranging from 2 min to an hour have been observed in bloodstream-form trypanosomes (Colasante et al. 2007; Schwede et al. 2009). Half-lives for procyclic forms, and for *Leishmania*, are generally rather longer (for a detailed list, see Clayton and Shapira 2007). From the data so far, half-life might indeed be able to account for the variations in mRNA abundance seen. In proportion to the transcriptome, however, only a minute proportion of kinetoplastid mRNAs has been studied, and these are dominated by developmentally regulated transcripts. We therefore do not know if

mRNA degradation really is a major determinant of RNA abundance for constitutively expressed mRNAs.

To investigate this we undertook a global analysis of mRNA degradation rates in bloodstream-form trypanosomes. Half-lives varied from 5 min to at least 90 min, but most mRNAs have half-lives of less than 20 min (Manful et al. 2011). Although there was a strong correlation between mRNA half-life and abundance, degradation rates and copy number were not the only factor determining mRNA levels, and some mRNAs deviated strongly from expected patterns. These might be affected by differences in the efficiencies of splicing or mRNA export from the nucleus. Since no influence of gene position was observed, transcriptional effects can probably be ruled out in most cases.

4.2 The Enzymes that Degrade Trypanosome mRNA

4.2.1 Overview

In all eukaryotes so far studied, including trypanosomes and *Leishmania* (Clayton and Shapira 2007; Haile and Papadopoulou 2007; Ouellette and Papadopoulou 2009), the degradation of mRNAs follows a similar pattern (Houseley and Tollervey 2009). Initially, most mRNAs will be on polysomes, although highly unstable mRNAs might be degraded without any translation. Polysomal mRNAs are associated with several proteins, which need to be removed either before, or during, the degradation process. The poly(A) tail is bound by poly(A)-binding protein (PABP), while the cap binds to eIF4E. This binds to eIF4G, which interacts with eIF4A and PABP (Gebauer and Hentze 2004; Jackson et al. 2010). Translating mRNAs can therefore sometimes, although not always be circular (Brandt et al. 2010).

Degradation of most mRNAs is initiated by removal of the poly(A) tail; this can be initiated while translation is still ongoing (Hu et al. 2010). The next step is removal of the cap structure (decapping). At this point, translation initiation is no longer possible, and the RNA is susceptible to degradation by 5′–3′ exoribonucleases, and by the 3′–5′ exoribonucleolytic activity associated with the exosome complex (Houseley and Tollervey 2009). The degradation may take place in specialised cytoplasmic aggregates called "P-bodies." These individual steps are considered in more detail below.

4.2.2 Deadenylation

In mammalian cells and yeast, mRNAs are deadenylated by the CCR4/CAF1/NOT complex, the PAN2/PAN3 complex, or a poly(A) ribonuclease (PARN). All three are encoded in kinetoplastid genomes.

The major deadenylase of metazoans and yeast is the CCR4/CAF1/NOT complex. NOT1 is a large scaffold protein. It is associated directly with CAF1, which is the principal exoribonuclease of the complex in all organisms examined so far, apart from *S. cerevisiae*. The last relies on the activity of CCR4 protein, which is attached to CAF1; CCR4 is not so important in other organisms. Other subunits that are present in stoichiometric amounts are NOT2 and NOT3/5 (two similar proteins in *S. cerevisiae*). The ubiquitous helicase DHH1 is also associated, as are various other sub-stoichiometric components—yeast Caf130 or human CNOT10 (see, for example, Temme et al. 2010).

Kinetoplastids lack CCR4 but have most of the remaining components of the NOT complex (Schwede et al. 2009). RNAi targeting CAF1 inhibited trypanosome growth, delayed bulk deadenylation (as measured by examining all trypanosome poly(A) tails), and also strongly delayed deadenylation of several mRNAs. This indicated that, as in mammalian cells, CAF1 is a major deadenylase (Schwede et al. 2008). Highly unstable mRNAs were only partially stabilised by CAF1 RNAi (Schwede et al. 2009). More recent results suggest that additional components of the complex are also implicated in deadenylation and required for optimal parasite growth (V. Färber, unpublished results). A transcriptome-wide analysis showed that depletion of the CAF1/NOT complex almost completely inhibited mRNA degradation (V. Färber, unpublished results).

The trypanosome PAN2/PAN3 complex appears to be implicated mainly in the degradation of some unstable mRNAs (Schwede et al. 2009), but its depletion had only minor transcriptome-wide effects (V. Färber, unpublished results).

The *T. brucei* genome encodes at least two proteins related to PARN (Tb927.8.2850 and Tb927.10.8360). The genes are conserved in *Leishmania* species and trypanosomes and not developmentally regulated in *T. brucei*. These proteins do not seem to have essential roles in cultivatable forms of the parasite and depletion did not affect mRNA decay (Utter et al. 2011). Over-expression of one of them did result in changes in the levels of some transcripts (Utter et al. 2011), but interpreting such results is difficult, since over-expression can result in non-specific effects.

Deadenylation and polyadenylation are also implicated in quality control of nuclear RNAs; this is outside the scope of this review.

4.2.3 Decapping

In yeast and mammals, deadenylated mRNAs are subject to decapping by a protein called Dcp2, which has a MutT hydrolase domain. Dcp2 is associated with a less well-conserved protein called Dcp1. The complex is not active on short oligonucleotides. Various other proteins, including Edc3, Scd6, Pat1, and Dhh1, can enhance activity (Nissan et al. 2010). Trypanosomes have SCD6 and DHH1, but lack obvious homologues of Edc3 and Pat1. More mysteriously, the kinetoplastid genomes lack a homologue of Dcp2 (Schwede et al. 2009). Since 5′–3′ digestion is very important in

trypanosome mRNA degradation (see below), decapping of full-length mRNAs must nevertheless occur. One possible mechanism would be a specific endonucleolytic digestion of the spliced leader, which could generate a small stub (to be digested by the characterised decapping enzyme) and a naked 5′-end susceptible to exoribonuclease. There is, however, so far no evidence for such a process (Schwede et al. 2009).

Yeast and mammals have a second decapping enzyme, DcpS, which degrades very short capped RNA stubs produced by 3′–5′ exonucleases. An enzyme with this properties was recently described (Banerjee et al. 2009) and may be responsible for a decapping activity that was described for trypanosome extracts (Milone et al. 2002).

4.2.4 5′–3′ Digestion

In both yeast and mammalian cells, decapped, deadenylated mRNAs are degraded predominantly from the 5′-end by an exonuclease called Xrn1. Experiments on both trypanosomes and *Leishmania* have shown that 5′–3′ degradation has an important role in the degradation of unstable mRNAs. Moreover, the 5′–3′ degradation appears not to require prior deadenylation (Haile et al. 2008; Li et al. 2006). The kinetoplastid genomes encode four proteins with appropriate exoribonuclease domains, which were studied in detail in *T. brucei*. One, XRND, is located in the nucleus and is the homologue of an enzyme (Rat1 or Xrn2) that is involved in nuclear RNA quality control and transcription termination in yeast and mammals. Two of the proteins—XRNB and XRNC—have no known function: RNA interference does not affect trypanosome growth (Li et al. 2006), and bloodstream trypanosomes lacking XRNB grew normally (T. Manful, unpublished results). It is XRNA that is the functional homologue of Xrn1. RNAi targeting XRNA caused rapid lethality. It strongly inhibited the degradation of unstable, developmentally regulated *PGK* mRNAs (Li et al. 2006), causing accumulation of the mRNAs in an inappropriate life cycle stage. Recent RNA-Seq data have confirmed that XRNA is critical in the degradation of many of the more unstable mRNAs in trypanosomes, but not so important for very stable mRNAs (Manful et al. 2011).

4.2.5 The Exosome

The exosome is a 3′–5′ exoribonuclease complex that is responsible for 3′-trimming and quality control of many RNAs in the nucleus, and for 3′–5′ degradation of mRNAs in the cytoplasm. The complex consists of nine core subunits plus accessory proteins. The core is a hexamer of between two and six different proteins, each of which has an RNase PH-like domain. On top of this sit three more subunits with S1 RNA-binding domains. Exosomes are also found in Archaea and in these organisms, the RNase PH domains digest RNA as it is threaded through the hole

in the middle of the hexamer. In most eukaryotes studied to date, in contrast, the RNase PH subunits have inactivating mutations (Clayton and Estevez 2010). Instead of degrading RNAs itself, the exosome core may act as a sort of unwinding chaperone, binding RNA via the S1 subunits, unfolding it by threading through the core, and exposing the unfolded RNA to attached exonucleases (Bonneau et al. 2009). In the yeast cytoplasmic exosome, the main exoribonuclease activity is effected by the subunit Rrp44, which sits at the bottom of the structure (opposite the S1 subunits) (Schneider et al. 2009). Rrp44-like proteins are also important in human cytoplasmic exosome activity, although their association with the exosome is not as stable as in yeast (Staals et al. 2010; Tomecki et al. 2010). Meanwhile, in the nucleus, exoribonuclease activity is supplied by a different subunit, called Rrp6 (Bonneau et al. 2009).

The trypanosome exosome core contains nine different subunits, with the expected PH hexamer arrangement (Estévez et al. 2001, 2003). Trypanosomes have a homologue of Rrp44 but no interaction with the exosome has been detected. Unusually, however, the RRP6 subunit is a constitutive part of the exosome and even seems to be required for exosome stability (Haile et al. 2007). A structural analysis of the *Leishmania tarentolae* exosome revealed that RRP6 seemed to be bound on one side of the RNase PH hexamer (Cristodero et al. 2008). Oddly, even the purified RRP6-containing complex lacked exoribonuclease activity (Cristodero et al. 2008): it is not clear whether the isolated complex had been slightly degraded, or whether it requires specific buffer conditions, or substrates, or additional accessory proteins in order to function.

In *T. brucei*, RNAi targeting various exosome subunits (and, interestingly, also RNAi targeting RRP44) caused defects in ribosomal RNA processing (Estévez et al. 2001, 2003). Effects on mRNA degradation were, in contrast, very minor, being restricted to a slight delay in the degradation of very unstable mRNAs (Haile et al. 2003). This minor effect was recently confirmed in a transcriptome-wide analysis (V. Färber, unpublished results). This result—like all others from RNAi—suffers from the perennial disadvantage that the trypanosomes still contained some residual exosome activity, so the true contribution of the exosome to mRNA degradation may be greater than was revealed by the assay.

4.2.6 Endonucleases

As an alternative to deadenylation and decapping, mRNAs can be rendered susceptible to exoribonucleases by endonucleolytic cleavage. Besides RNAi, relatively few examples of this are known so far from mammalian cells, but recently, an example was found in *Leishmania*. In *Leishmania major*, the retroposon-like element SIDER2 is enriched in the 3′-UTRs of low-abundance mRNAs (Bringaud et al. 2007). In two different mRNAs, an endonucleolytic cut within the SIDER sequence was detected (Muller et al. 2010b): this cleavage effectively deadenylates the mRNA, leaving it open to attack from both ends. An mRNA containing SIDER2

is thus degraded in the 5′–3′ direction without any evidence for gradual poly(A) tail shortening (Muller et al. 2010a). It would be interesting to identify the endonuclease responsible and to find out to what extent this mechanism operates in other kinetoplastids.

4.2.7 P-Bodies and Stress Granules

P-bodies are thought to be localised centres of mRNA decay. Their contents vary depending on the cell type, but they typically contain mRNA, translation repressors, RNA-binding proteins that trigger mRNA decay, and some components of the mRNA degradation machinery—the decapping complex, deadenylases, Xrn1, Dhh1 (human Rck), and Scd6 (human Rap55), but not the exosome (Buchan and Parker 2009). P-body formation is enhanced by puromycin, and inhibited by cycloheximide, which suggests that P-bodies are sites for degradation of mRNAs that are not associated with ribosomes. Microscopically detectable P-bodies are, however, not essential for mRNA degradation (Eulalio et al. 2007), and recent results have shown that mRNA degradation can be initiated while the mRNA is on polysomes (Hu et al. 2010).

A second type of particle, the stress granule, can be induced by starvation, treatment with heavy metals or heat shock. These granules contain 40S ribosomal subunits and translation factors, in addition to some P-body components such as Xrn1, Dhh1, and Scd6, while lacking the decapping machinery and deadenylases. They are thought to be storage sites for mRNAs that may be associated with pre-initiation complexes and are temporarily sequestered from translation due to unfavourable conditions (Buchan and Parker 2009).

Prolonged (24 h) incubation of *T. cruzi* in buffer causes the development of granules containing DHH1 (Holetz et al. 2007) and poly(A) (Cassola et al. 2007). Similar granules were seen after a 3-h starvation of *T. brucei* procyclic forms (Cassola et al. 2007) (Table 4.1). Starvation granule formation was, as for metazoan and yeast P-bodies, inhibited by cycloheximide and promoted by puromycin treatment. In contrast, formation of the granules was not prevented by actinomycin D, which makes them more like stress granules. Using GFP fusion proteins or polyclonal antibodies, the authors found poly(A)-binding proteins, an eIF4E, XRNA, DHH1, but not two selected ribosomal proteins or eIF3D, in the granules. The absence of the small ribosomal protein would argue against the presence of translationally stalled mRNPs. Also present were various small proteins with a single RRM domain (Table 4.1). Similar granules were seen in parasites isolated from the vector. Immunoprecipitation with antibody to *T. cruzi* DHH1, followed by mass spectrometry, unfortunately yielded mainly extremely abundant proteins that are found as contaminants in many such studies (Holetz et al. 2010); in contrast, the CAF1/NOT complex components, which are known to be associated with DHH1 (Schwede et al. 2008), were not detected.

Table 4.1 P-bodies and stress granules in trypanosomes

Component	*T. cruzi* starvation	*T. brucei* starvation	*T. brucei* 41°C
poly(A)	SG	SG	low RNA
DHH1	SG*	nd	PB+HSG
SCD6	nd	nd	PB+HSG
PABP1	SG*	SG*	HSG
PABP2	SG*	–	HSG
XRNA	SG*	nd	XB
eIF2A	nd	nd	HSG
eIF3B	nd	nd	HSG
eIF3Db	SG+/–	nd	nd
eIF4E1	SG*	nd	HSG
eIF4E2	nd	nd	HSG
eIF4E3	nd	nd	HSG
eIF4E4	nd	nd	HSG
RPL3	–	nd	nd
RPS19	–	nd	nd
UBP1	SG*	SG*	nd
UBP2	SG*	nd	nd
RBP3	SG*+/–	nd	nd
RBP4	SG*+/	nd	nd
RBP5	SG*+/	nd	nd
RBP6	SG*	nd	nd
LA	–	nd	nd
CBP20	–	nd	nd

The locations of components shown in the left-hand column were determined by *in situ* hybridisation (poly(A), expression of GFP-tagged proteins, or using specific antisera. "–": not in stress granules; *SG** starvation granule with colocalisation with poly(A); *SG**+/− partial colocalisation; *nd* not done; *HSG* heat-shock granule; *PB* P-body; *XB* XRNA body. Starvation results are from Cassola et al. (2007) and heat-shock results from Kramer et al. (2008). The location of DHH1 in *T. cruzi* SG was also shown by Holetz et al. (2007), as was partial colocalisation with PABP1

A second study found not only DHH1 and XRNA, but also SCD6, in granules in procyclic *T. brucei*. Heat shock collapsed the polysomes, with decreases in most RNAs, but sparing *HSP70* and *HSP83* mRNAs. The number of DHH1- and SCD6-containing granules increased, and an unusual XRNA-containing structure, which lacked other P-body markers, was formed at the posterior of the cell. Various granules containing poly(A)-binding protein 1 (PABP1) also formed; some contained SCD6 and DHH1, while others did not (Table 4.1).

Procyclic *T. brucei* expressing a dominant-negative version of DHH1 selectively accumulated a small number of bloodstream-form-specific mRNAs, confirming the role of DHH1 in their degradation (Kramer et al. 2010b). Whether these effects are a consequence of P-body impairment, or defects in deadenylation or decapping, is not known—DHH1 has multiple functions. Similarly, DHH1 immunoprecipitation from *T. cruzi* selected for mRNAs that are normally unstable in epimastigotes (Holetz et al. 2010). Again, multiple explanations are possible.

4.3 Control of mRNA Degradation in Trypanosomes

4.3.1 Regulatory Sequences

The location of sequences that regulate mRNA abundances has been studied extensively using reporter genes. In nearly all cases, the sequence was found in the 3′-UTR, and, where examined, the 3′-UTR was found to regulate the mRNA half-life. The location makes sense: small ribosomal subunits, with associated tRNA and translation factors, will strip any RNA-bound proteins from the 5′-UTR, unless they are impeded by strong protein–protein interactions or RNA secondary structures. The 3′-UTR is, in contrast, free for binding of other proteins, which might have regulatory function.

Since the previous list of 3′ elements was published (Clayton and Shapira 2007), a variety of additional sequences has been described (see, for example, Abanades et al. 2009; Rodrigues et al. 2010). Nearly all of the regulatory sequences identified so far are much too long to be recognised by a single RNA-binding protein. If there are short (less than 40 nt) recognition elements within them, these must be highly context dependent since deletion mutations failed to delineate them. Most likely, recognition depends not only on primary sequence, but also on secondary structure.

A few short regulatory elements have nevertheless been described. U-rich elements in the 3′-UTR have been shown to be involved in the developmentally regulated degradation of various mRNAs: in *T. brucei*, for three mRNAs that are unstable in bloodstream forms, encoding EP procyclin, the B (cytosolic) isoform of phosphoglycerate kinase, and pyruvate phosphate dikinase (Quijada et al. 2002); and in mucin mRNAs in *T. cruzi* (D'Orso and Frasch 2001). These are of interest because they resemble the destabilising AU-rich elements found in the 3′-UTRs of many mammalian mRNAs. Two short elements that contain consensus motifs for recognition by Pumilio domain proteins control developmental regulation of a paraflagellar rod protein in *Leishmania* (Mishra et al. 2003) and cell cycle regulation in *T. brucei* (Archer et al. 2009), and an octamer sequence is required for cell cycle regulation of various mRNAs in *Crithidia fasciculata* (Avliyakulov et al. 2003).

The next step is the identification of effector proteins. The minimal criteria are the protein must bind to the recognition sequence in vivo and must fail to bind to non-functional mutated versions of the recognition sequence. The protein must also be required for regulation of the mRNA: in *T. brucei*, this can be shown by RNAi; in organisms lacking RNAi, it could be shown by knock-out or expression of dominant-negative versions. The ability to affect mRNA abundance when artificially tethered to a non-target mRNA would support the protein activity further. So far, unfortunately, no protein-target pair fulfils all of these criteria. Nevertheless, there are some indications of protein function. Most research so far has focussed on proteins with canonical RNA-binding domains: RRMs, Pumilio domains, and CCCH zinc fingers.

4.3.2 Proteins with RRM Domains

The RNA Recognition Motif (RRM) is a sequence of about 90 amino acids. The results of structural analyses suggest that proteins with only a single RRM domain are unlikely to show specific high-affinity RNA binding unless other domains (or associated proteins) contribute (Park et al. 2000); in contrast, two RRMs can recognise RNAs of about 11 residues (Wang and Tanaka Hall 2001) with nanomolar affinity (Park et al. 2000). RRM domains can also mediate protein–protein interactions (Kielkopf et al. 2004). Kinetoplastids have at least 70 proteins with RRM domains, only a few of which have known functions in RNA metabolism [e.g. poly(A)-binding protein, rRNA and mRNA processing factors; De Gaudenzi et al. 2006]. Many of the remainder have been targeted by RNAi and 17 were shown to be required for optimal growth (Wurst et al. 2009).

4.3.2.1 Proteins with a Single RRM Domain

UBP1 and UBP2 are related proteins which bind preferentially to U-rich sequences in vitro. They are found in starvation granules (Cassola et al. 2007) and shuttle between the nucleus and cytoplasm (Cassola and Frasch 2009); the RRM domain was required for both localisations, suggesting dependency on RNA binding. Over-expression of UBP1 in *T. cruzi* resulted in instability of a mucin mRNA to which UBP1 can bind. In *T. brucei*, RNAi and over-expression of UBP1 or UBP2 inhibited cell growth; over-expression increased the levels of mRNAs encoding the CFB1 cyclin-F-box protein family, while RNAi reduced the level of a related mRNA encoding CFB2 (Hartmann and Clayton 2008). The levels of mRNAs encoding a transmembrane protein family were also regulated in concert with UBP1 expression; in each case, the 3′-UTRs were required for the effects. However, there is no evidence that the effects were a direct consequence of UBP1/2 binding (Hartmann and Clayton 2008).

RNA pull-down assays using UBP1, UBP2, and RBP3 have given conflicting results. In *T. brucei*, there was no evidence for selective binding of UBP1 and UBP2 to particular mRNAs (Hartmann and Clayton 2008). UBP1 and UBP2 constitute 0.1% of the total cell protein, so are present in 30-fold molar excess over mRNA (Hartmann and Clayton 2008). The specificity of *T. cruzi* UBP1 and UBP2 was tested by a pull-down followed by cloning and sequencing of 40 clones. The authors were able to identify a putative consensus recognition motif in the RNA (Noe et al. 2008). Homodimerisation of UBP1 was documented in *T. cruzi*, which would allow sequence-specific binding (D'Orso and Frasch 2002), but neither homodimerisation nor heterodimerisation was found in *T. brucei*.

RBP3 is another small single-RRM protein related to the UBPs. Immunoprecipitation of *T cruzi* RBP3 pulled down several mRNAs encoding ribosomal proteins and a possible recognition sequence was deduced (Noe et al. 2008). In *T. brucei*, microarray hybridisation of RBP3-bound RNA revealed ten transcripts, not the

same as in *T. cruzi* but including *CFB1*; however, depletion of RBP3—which somewhat inhibited bloodstream trypanosome growth—had no effect on the transcriptome (Wurst et al. 2009).

The UBP1, UBP2, and RBP3 pull-downs all relied on rather insensitive, non-quantitative methodology; the immunoprecipitations gave no indication of what proportion of each target RNA was actually bound. It will be important to repeat these experiments using RNA-Seq and to compare the bound and unbound fractions in order to obtain quantitative information. Our results provide no evidence that UBP1, UBP2, or RBP3 is associated with polysomal RNA (Hartmann and Clayton 2008; C. Klein, unpublished). Their roles therefore remain enigmatic.

RBP10 is a bloodstream-form-specific cytoplasmic protein. RNAi targeting RBP10 in bloodstream forms causes up-regulation of procyclic-specific mRNAs and is lethal. Expression in procyclic forms similarly inhibits growth and turns on bloodstream-form-specific mRNAs, including that of native RBP10. The mechanism of action of RBP10 is not yet known. There is no evidence that RBP10 can bind to mRNAs, it shows no stable interactions with other proteins, and it is not associated with polysomes (M. Wurst, manuscript in preparation).

4.3.2.2 Proteins with Two RRM Domains

DRBD3 is a protein with two RRM domains that might be a homologue of polypyrimidine tract-binding protein, an Opisthokont splicing regulator. It is essential for *T. brucei* growth. Two different groups have shown that DRBD3 binds to, and stabilises, a subset of developmentally regulated mRNAs, including several encoding membrane proteins (Estévez 2008; Stern et al. 2009). Using reporter assays, it was demonstrated that for an mRNA encoding an amino acid transporter, DRBD3 acted via the 3′-UTR; the precise recognition sequence remains to be defined. A role of DRBD3 in splicing was also demonstrated (Stern et al. 2009).

4.3.2.3 Pumilio-Domain Proteins

In animal cells and yeast, some Pumilio (Puf) domain proteins negatively regulate mRNA translation and degradation, or affect mRNA localisation (Saint-Georges et al. 2008). In yeast, these are required for optimal growth under special conditions (Foat et al. 2005), but are not required for survival in rich medium. Other Puf-domain proteins are involved in rRNA processing or ribosome assembly (Thomson et al. 2007). Kinetoplastid genomes encode both types of Puf protein. Of the 10 kinetoplastid Puf proteins, PUF7 and PUF8 are in the nucleolus and implicated in rRNA processing (Droll et al. 2010).

T. brucei PUF1 (Luu et al. 2006) is equivalent to *T. cruzi* PUF6 (Dallagiovanna et al. 2005, 2008). PUF1/6 is located in the cytoplasm, in discrete spot that do not colocalise with DHH1 (Caro et al. 2006; Dallagiovanna et al. 2008) and not detected in polysomes (Dallagiovanna et al. 2008; Luu et al. 2006). In a *T. brucei*

puf1 knockout mutant, no defects in either growth or the transcriptome were detected. A pull-down of *T. cruzi* PUF6 preferentially enriched 6 mRNAs, and these were decreased when PUF6 was over-expressed (Dallagiovanna et al. 2008). However, the proportion of each mRNA that was bound to PUF6 was not measured, and a recognition site was not identified. Thus, PUF1/6 might be involved in destabilisation of a subset of target mRNAs, at least in *T. cruzi*, but there is no evidence that this function is essential.

T. brucei PUF9 is required for stabilisation of three RNAs in the S-phase of the cell cycle, and its own mRNA shows parallel regulation (Archer et al. 2009). The target RNAs were identified by a PUF9 pull-down, and all share a consensus Pumilio domain-binding motif, UUGUACC. In one case, this motif was also shown to be required for cell cycle regulation (Archer et al. 2009). So far, however, direct evidence for binding of PUF9 to the motif is missing, so it is not clear if the connection is direct or indirect.

As noted earlier, the negative regulatory element in the *Leishmania mexicana* PFR2 gene (encoding a component of the paraflagellar rod) contains a Puf consensus motif (AUGUAUAGUU) (Mishra et al. 2003). Whether this is recognised by a Pumilio protein is not known.

4.3.2.4 Proteins with CCCH Zinc Finger Domains

Proteins with $Cx_{4–15}Cx_{4–6}Cx_3H$ (CCCH) zinc finger domains are involved in various aspects of RNA metabolism. Usually, two domains are required for sequence-specific RNA binding. The individual sequenced kinetoplastid genomes encode 48–54 potential CCCH-domain proteins, 39 of which are present in all species, and few have known functions (Kramer et al. 2010a).

ZFP1, ZFP2, and ZFP3 have single CCCH domains and play a role in *T. brucei* differentiation. ZFP1 is up-regulated during differentiation and in the procyclic stage, and down-regulation of ZFP1 results in impaired kinetoplast repositioning during differentiation (Hendriks and Matthews 2005). Reduced *ZFP2* mRNA caused a delay in expression of the procyclic surface protein EP1 procyclin, while over-expression resulted in a morphological abnormality. Over-expression of ZFP3 resulted in enhanced differentiation (as assessed by the gain of EP1) and in morphological abnormalities in the procyclic form (Hendriks et al. 2001; Paterou et al. 2006). ZFP3 binds to EP procyclin mRNAs, and binding requires the presence of two previously characterised regulatory elements (Walrad et al. 2009); the evidence suggests that it regulates translation.

ZC3H18 is another single-RRM protein; it is not essential in bloodstream *T. brucei*, but depletion delays differentiation to the procyclic form. ZC3H18 is phosphorylated and can bind to 14-3-3 proteins (Benz et al. 2011). ZC3H20 has also recently been shown to stabilise a few developmentally regulated mRNAs (Ling et al. 2011).

In *Crithidia fasciculata*, an octamer sequence in the 5′-UTR is involved in cell cycle control. It can be bound by two different protein complexes. One, CSBPI, is a

dimer of CSPA (*T. brucei* ZC3H39) and CSPB (*T. brucei* ZC3H40), each of which contains a single CCCH domain. Although CSBPI cycles in concert with its target, and a knockout of CSPA caused loss of CSBP protein as well, the target mRNAs continue to cycle (Mittra et al. 2003). The second complex is a trimer consisting of poly(A)-binding protein and two other proteins whose phosphorylation was cell cycle dependent; it is not known if either of these is required for cell cycle regulation (Mittra and Ray 2004).

4.3.3 Nonsense-Mediated Decay

In metazoans, plants, and yeast, mRNAs with premature termination codons are degraded by a process known as "nonsense-mediated decay" (Stalder and Mühlemann 2008). The process depends on a protein called Upf1, which interacts with another protein, Upf2. Various possible mechanisms of premature codon recognition have been documented, including measurement of 3′-UTR length. Kinetoplastids have genes encoding both UPF1 and UPF2, and they interact (Delhi et al. 2011). In T. *brucei*, introduction of premature termination codons in either an endogenous gene or a reporter gene caused modest decreases in mRNA abundance, but so far, it was not possible to demonstrate that this depended on UPF1. The mechanism of nonsense-mediated decay, and the means by which the premature termination codon is recognised in trypanosomes, is unresolved, but it clearly did not depend on the length of the 3′-UTR (Delhi et al. 2011).

4.4 Outlook

Kinetoplastids have over 100 RNA-binding proteins whose functions are not known. Probably, most mRNAs are bound by several different proteins, some of which compete for binding to the same site. In order to understand control, it is necessary to understand the composition of the entire ribonucleoprotein particle. Alves et al. (2010) documented the composition of messenger ribonucleoprotein particles, washed with EDTA in order to remove the ribosomes. Although the washing removed some translation factors, many novel potential interacting proteins were identified, including some with RNA-binding motifs. We have also obtained polysomal proteomes (C. Klein, unpublished results) with additional RNA-binding proteins. Although these proteomes probably contain many contaminants, they will be an invaluable source of information about mRNA regulatory factors. Purification of specific mRNAs poses a much greater challenge, but potential methods exist, and their application will allow us to understand the complex interactions that determine the fates of individual mRNAs.

Acknowledgements Work in my laboratory is supported by the Land of Baden-Württemberg and the Deutsche Forschungsgemeinschaft.

References

Abanades DR, Ramirez L, Iborra S, Soteriadou K, Gonzalez VM, Bonay P, Alonso C, Soto M (2009) Key role of the 3′ untranslated region in the cell cycle regulated expression of the *Leishmania infantum* histone H2A genes: minor synergistic effect of the 5′ untranslated region. BMC Mol Biol 10:48

Alves L, Ávila A, Correa A, Holetz F, Mansur F, Manque P, de Menezes J, Buck G, Krieger M, Goldenberg S (2010) Proteomic analysis reveals the dynamic association of proteins with translated mRNAs in *Trypanosoma cruzi*. Gene 452:72–78

Archer SK, van Luu D, de Queiroz R, Brems S, Clayton CE (2009) *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog 5:e1000565

Avliyakulov N, Hines J, Ray D (2003) Sequence elements in both the intergenic space and the 3′ untranslated region of the *Crithidia fasciculata* KAP3 gene are required for cell cycle regulation of KAP3 mRNA. Eukaryot Cell 2:671–677

Banerjee H, Palenchar J, Lukaszewicz M, Bojarska E, Stepinski J, Jemielity J, Guranowski A, Ng S, Wah D, Darzynkiewicz E, Bellofatto V (2009) Identification of the HIT-45 protein from *Trypanosoma brucei* as an FHIT protein/dinucleoside triphosphatase: substrate specificity studies on the recombinant and endogenous proteins. RNA 15:1554–1564

Bass KE, Wang CC (1992) Transient inhibition of protein synthesis accompanies differentiation of *Trypanosoma brucei* from bloodstream from procyclic form. Mol Biochem Parasitol 56:129–140

Benz C, Mulindwa J, Ouna B, Clayton C (2011) The *Trypanosoma brucei* zinc finger protein ZC3H18 is involved in differentiation. Mol Biochem Parasitol 177:148–151

Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell 139:547–559

Brandt F, Carlson L, Hartl F, Baumeister W, Grünewald K (2010) The three-dimensional organization of polyribosomes in intact human cells. Mol Cell 39:560–569

Brecht M, Parsons M (1998) Changes in polysome profiles accompany trypanosome development. Mol Biochem Parasitol 97:189–198

Bringaud F, Müller M, Cerqueira G, Smith M, Rochette A, El-Sayed N, Papadopoulou B, Ghedin E (2007) Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. PLOS Pathog 3:e136

Buchan J, Parker R (2009) Eukaryotic stress granules: the ins and outs of translation. Mol Cell 36:932–941

Caro F, Bercovich N, Atorrasagasti C, Levin M, Vazquez M (2006) *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein family. Exp Parasitol 113:112–124

Carter N, Yates P, Gessford S, Galagan S, Landfear S, Ullman B (2010) Adaptive responses to purine starvation in *Leishmania donovani*. Mol Microbiol 78:92–107

Cassola A, Frasch A (2009) An RNA recognition motif mediates the nucleocytoplasmic transport of a trypanosome RNA-binding protein. J Biol Chem 284:35015–35028

Cassola A, De Gaudenzi J, Frasch A (2007) Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. Mol Microbiol 65:655–670

Clayton C, Estevez A (2010) The exosomes of trypanosomes and other protists. In: Jensen T (ed) The exosome. Landes Bioscience, Austin, TX

Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156:93–101

Colasante C, Robles A, Li C-H, Schwede A, Benz C, Voncken F, Guilbride DL, Clayton C (2007) Regulated expression of glycosomal phosphoglycerate kinase in *Trypanosoma brucei*. Mol Biochem Parasitol 151:193–204

Cristodero M, Böttcher B, Diepholz M, Scheffzeck K, Clayton C (2008) The exosome of *Leishmania tarentolae*: purification and structural analysis by electron microscopy. Mol Biochem Parasitol 159:24–29

D'Orso I, Frasch ACC (2001) Functionally different AU- and G-rich *cis* elements confer developmentally-regulated mRNA stability by interaction with specific RNA-binding proteins. J Biol Chem 276:15783–15793

D'Orso I, Frasch ACC (2002) TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and poly(A)-binding proteins forming a ribonucleoprotein complex. J Biol Chem 277:50520–50528

Dallagiovanna B, Perez L, Sotelo-Silveira J, Smircich P, Duhagon MA, Garat B (2005) *Trypanosoma cruzi*: molecular characterization of TcPUF6, a Pumilio protein. Exp Parasitol 109:260–264

Dallagiovanna B, Correa A, Probst C, Holetz F, Smircich P, de Aguiar A, Mansur F, Vieira da Silva C, Mortara R, Garat B, Buck G, Goldenberg S, Krieger M (2008) Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from *Trypanosoma cruzi*. J Biol Chem 283:8266–8273

De Gaudenzi JG, Frasch ACC, Clayton C (2006) RNA-binding domain proteins in kinetoplastids: a comparative analysis. Eukaryot Cell 4:2106–2114

Delhi P, Queiroz R, Inchaustegui D, Carrington M, Clayton C (2011) Is there a classical nonsense-mediated decay pathway in trypanosomes? PLoS One 6:e25112

Droll D, Archer S, Fenn K, Delhi P, Matthews K, Clayton C (2010) The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. FEBS Lett 84:1156–1162

Estévez A (2008) The RNA-binding protein *Tb*DRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. Nucleic Acids Res 36:4573–4586

Estévez A, Kempf T, Clayton CE (2001) The exosome of *Trypanosoma brucei*. EMBO J 20:3831–3839

Estévez AM, Lehner B, Sanderson CM, Ruppert T, Clayton C (2003) The roles of inter-subunit interactions in exosome stability. J Biol Chem 278:34943–34951

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E (2007) P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. Mol Cell Biol 27:3970–3981

Foat B, Houshmandi S, Olivas W, Bussemaker H (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. Proc Natl Acad Sci U S A 102:17675–17680

Gebauer F, Hentze M (2004) Molecular mechanisms of translation control. Nat Rev Mol Cell Biol 5:827–835

Haanstra J, Stewart M, Luu V-D, van Tuijl A, Westerhoff H, Clayton C, Bakker B (2008) Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form *Trypanosoma brucei*. J Biol Chem 283:2495–2507

Haanstra J, Kerkhoven E, van Tuijl A, Blits M, Wurst M, van Nuland R, Albert M-A, Michels P, Bouwman J, Clayton C, Westerhoff H, Bakker B (2011) A domino effect in drug action: from metabolic assault towards parasite differentiation. Mol Microbiol 79:94–108

Haile S, Papadopoulou B (2007) Developmental regulation of gene expression in trypanosomatid parasitic protozoa. Curr Opin Microbiol 10:569–577

Haile S, Estévez AM, Clayton C (2003) A role for the exosome in the initiation of degradation of unstable mRNAs. RNA 9:1491–1501

Haile S, Cristodero M, Clayton C, Estévez A (2007) The subcellular localisation of trypanosome RRP6 and its association with the exosome. Mol Biochem Parasitol 151:52–58

Haile S, Dupe A, Papadopoulou B (2008) Deadenylation-independent stage-specific mRNA degradation in *Leishmania*. Nucleic Acids Res 36:1634–1644

Hartmann C, Clayton C (2008) Regulation of a transmembrane protein gene family by the small RNA binding proteins TbUBP1 and TbUBP2. Mol Biochem Parasitol 157:112–115

Hendriks EF, Matthews KR (2005) Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. Mol Microbiol 57:706–716

Hendriks EF, Robinson DR, Hinkins M, Matthews KR (2001) A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. EMBO J 20:6700–6711

Holetz F, Correa A, Avila A, Nakamura C, Krieger M, Goldenberg S (2007) Evidence of P-body-like structures in *Trypanosoma cruzi*. Biochem Biophys Res Commun 356:1062–1067

Holetz F, Alves L, Probst C, Dallagiovanna B, Marchini F, Manque P, Buck G, Krieger M, Correa A, Goldenberg S (2010) Protein and mRNA content of TcDHH1-containing mRNPs in *Trypanosoma cruzi*. FEBS J 277:3415–3426

Houseley J, Tollervey D (2009) The many pathways of RNA degradation. Cell 136:763–776

Hu W, Petzold C, Coller J, Baker K (2010) Nonsense-mediated mRNA decapping occurs on polyribosomes in *Saccharomyces cerevisiae*. Nat Struct Mol Biol 17:244–247

Jackson R, Hellen C, Pestova T (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 10:113–127

Jensen B, Sivam D, Kifer C, Myler P, Parsons M (2009) Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. BMC Genomics 10:482

Kabani S, Fenn K, Ross A, Ivens A, Smith T, Ghazal P, Matthews K (2009) Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*. BMC Genomics 10:427

Kielkopf CL, Lucke S, Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. Genes Dev 18:1513–1526

Koumandou V, Natesan S, Sergeenko T, Field M (2008) The trypanosome transcriptome is remodelled during differentiation but displays limited responsiveness within life stages. BMC Genomics 9:e298

Kramer S, Queiroz R, Ellis L, Webb H, Hoheisel J, Clayton C, Carrington M (2008) Stress granules and the heat shock response in *Trypanosoma brucei*. J Cell Sci 121:3002–3014

Kramer S, Kimblin N, Carrington M (2010a) Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*. BMC Genomics 11:283

Kramer S, Queiroz R, Ellis L, Hoheisel J, Clayton C, Carrington M (2010b) The RNA helicase DHH1 is central to correct expression of many developmentally regulated mRNAs in trypanosomes. J Cell Sci 123:699–711

Li C-H, Irmer H, Gudjonsdottir-Planck D, Freese S, Salm H, Haile S, Estévez AM, Clayton CE (2006) Roles of a *Trypanosoma brucei* 5′→3′ exoribonuclease homologue in mRNA degradation. RNA 12:2171–2186

Ling A, Trotter J, Hendriks E (2011) A zinc finger protein, TbZC3H20, stabilises two developmentally regulated mRNAs in trypanosomes. J Biol Chem 286:20152–20162

Luu VD, Brems S, Hoheisel J, Burchmore R, Guilbride D, Clayton C (2006) Functional analysis of *Trypanosoma brucei* PUF1. Mol Biochem Parasitol 150:340–349

Manful T, Fadda A, Clayton C (2011) The role of the 5′-3′ exoribonuclease XRNA in transcriptome-wide mRNA degradation. RNA 17:2039–2047

Milone J, Wilusz J, Bellofatto V (2002) Identification of mRNA decapping activities and an ARE-regulated 3′ to 5′ exonuclease activity in trypanosome extracts. Nucleic Acids Res 30:4040–4050

Minning T, Weatherly D, Jr A, Orlando R, Tarleton R (2009) The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. BMC Genomics 10:370

Mishra KK, Holzer TR, Moore LL, LeBowitz JH (2003) A negative regulatory element controls mRNA abundance of the *Leishmania mexicana* Paraflagellar rod gene PFR2. Eukaryot Cell 2:1009–1017

Mittra B, Ray DS (2004) Presence of a poly(A) binding protein and two proteins with cell cycle-dependent phosphorylation in *Crithidia fasciculata* mRNA cycling sequence binding protein II. Eukaryot Cell 3:1185–1197

Mittra B, Sinha K, Hines J, Ray D (2003) Presence of multiple mRNA cycling sequence element-binding proteins in *Crithidia fasciculata*. J Biol Chem 278:26564–26571

Muller M, Padmanabhan PK, Papadopoulou B (2010a) Selective inactivation of SIDER2 retroposon-mediated mRNA decay contributes to stage- and species-specific gene expression in *Leishmania*. Mol Microbiol 77:471–491

Muller M, Padmanabhan PK, Rochette A, Mukherjee D, Smith M, Dumas C, Papadopoulou B (2010b) Rapid decay of unstable Leishmania mRNAs bearing a conserved retroposon signature 3′-UTR motif is initiated by a site-specific endonucleolytic cleavage without prior deadenylation. Nucleic Acids Res 38:5867–5883

Nissan T, Rajyaguru P, She M, Song H, Parker R (2010) Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms. Mol Cell 39:773–783

Noe G, De Gaudenzi J, Frasch A (2008) Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes. BMC Mol Biol 9:107

Ortiz D, Raquel Valdés R, Sanchez M, Hayenga J, Elya C, Detke S, Landfear S (2010) Purine restriction induces pronounced translational upregulation of the NT1 adenosine/pyrimidine nucleoside transporter in *Leishmania major*. Mol Microbiol 78:108–118

Ouellette M, Papadopoulou B (2009) Coordinated gene expression by post-transcriptional regulons in African trypanosomes. J Biol 8:100

Park S, Myszka DG, Yu M, Littler SJ, Laird-Offringa IA (2000) HuD RNA recognition motifs play distinct roles in the formation of a stable complex with AU-rich RNA. Mol Cell Biol 20:4765–4772

Paterou A, Walrad P, Craddy P, Fenn K, Matthews K (2006) Identification and stage-specific association with the translational apparatus of TbZFP3, a ccch protein that promotes trypanosome life cycle development. J Biol Chem 281:39002–39013

Pays E, Hanocq-Quertier J, Hanocq F, Van Assel S, Nolan D, Rolin S (1993) Abrupt RNA changes precede the first cell division during the differentiation of *Trypanosoma brucei* bloodstream forms into procyclic forms *in vitro*. Mol Biochem Parasitol 61:107–114

Queiroz R, Benz C, Fellenberg K, Hoheisel J, Clayton C (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics 10:495

Quijada L, Hartmann C, Guerra-Giraldez C, Drozdz M, Irmer H, Clayton CE (2002) Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* induces differentiation-related changes in the abundance of developmentally-regulated mRNAs. Nucleic Acids Res 30:1–11

Rochette A, Raymond F, Corbeil J, Ouellette M, Papadopoulou B (2009) Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of *Leishmania infantum*. Mol Biochem Parasitol 165:32–47

Rodrigues DC, Silva R, Rondinelli E, Urmenyi TP (2010) *Trypanosoma cruzi*: modulation of HSP70 mRNA stability by untranslated regions during heat shock. Exp Parasitol 126:245–253

Saint-Georges Y, Garcia M, Delaveau T, Jourdren L, Le Crom S, Lemoine S, Tanty V, Devaux F, Jacq C (2008) Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. PLoS One 3:e2293

Schneider C, Leung E, Brown J, Tollervey D (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res 37:1127–1140

Schwede A, Ellis L, Luther J, Carrington M, Stoecklin G, Clayton C (2008) A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. Nucleic Acids Res 36:3374–3388

Schwede A, Manful T, Jha B, Helbig C, Bercovich N, Stewart M, Clayton C (2009) The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. Nucleic Acids Res 37:5511–5528

Siegel T, Hekstra D, Wang X, Dewell S, Cross G (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res 38:4946–4957

Staals R, Bronkhorst A, Schilders G, Slomovic S, Schuster G, Heck A, Raijmakers R, Pruijn G (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome. EMBO J 29:2358–2367

Stalder L, Mühlemann O (2008) The meaning of nonsense. Trends Cell Biol 18:315–321

Stern M, Gupta S, Salmon-Divon M, Haham T, Barda O, Levi S, Wachtel C, Nilsen T, Michaeli S (2009) Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. RNA 15:648–665

Temme C, Zhang L, Kremmer E, Ihling C, Chartier A, Sinz A, Simonelig M, Wahle E (2010) Subunits of the Drosophila CCR4-NOT complex and their roles in mRNA deadenylation. RNA 16:1356–1370

Thomson E, Rappsilber J, Tollervey D (2007) Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. RNA 13:2165–2174

Tomecki R, Kristiansen M, Lykke-Andersen S, Chlebowski A, Larsen K, Szczesny R, Drazkowska K, Pastula A, Andersen J, Stepien P, Dziembowski A, Jensen T (2010) The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J 29:2342–2357

Utter C, Garcia S, Milone J, Bellofatto V (2011) Poly(A)-specific ribonuclease (PARN-1) function in stage-specific mRNA turnover in *Trypanosoma brucei*. Eukaryot Cell 10:1230–1240

Vassella E, Den Abbeele J, Butikofer P, Renggli CK, Furger A, Brun R, Roditi I (2000) A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes Dev 14:615–626

Vassella E, Probst M, Schneider A, Studer E, Renggli C, Roditi I (2004) Expression of a major surface protein of *Trypanosoma brucei* insect forms is controlled by the activity of mitochondrial enzymes. Mol Biol Cell 15:3986–3993

Walrad P, Paterou A, Acosta-Serrano A, Matthews K (2009) Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. PLoS Pathog 5:e1000317

Wang X, Tanaka Hall T (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat Struct Biol 8:141–145

Webb H, Burns R, Ellis L, Kimblin N, Carrington M (2005) Developmentally regulated instability of the GPI-PLC mRNA is dependent on a short-lived protein factor. Nucleic Acids Res 33:1503–1512

Wurst M, Robles A, Po J, Luu V, Brems S, Marentije M, Stoitsova S, Quijada L, Hoheisel J, Stewart M, Hartmann C, Clayton C (2009) An RNAi screen of the RRM-domain proteins of *Trypanosoma brucei*. Mol Biochem Parasitol 163:61–65

Chapter 5
tRNA Biogenesis and Processing

Jessica L. Spears, Mary Anne T. Rubio, Paul J. Sample, and Juan D. Alfonzo

Abstract tRNAs are essential in all domains of life; this becomes especially important in trypanosomatids, where for all purposes the same set of tRNAs are utilized for cytoplasmic and mitochondrial protein synthesis. What makes the system special is that although tRNA biogenesis starts in the nucleus, the resulting products will satisfy translational requirements in two very different compartments. The balance between intracellular tRNA transport and post-transcriptional modifications may modulate tRNA function in gene expression. This chapter will summarize what is currently known about various processes that a tRNA must undergo in a trypanosomatid cell to become fully functional. Whenever possible, we will highlight both commonalities and differences with other systems, while emphasizing open questions that may lead to new and surprising discoveries in this group of evolutionarily divergent organisms.

Keywords Editing • Mitochondria • Modification • Processing • tRNA transport • Trypanosomes

5.1 Introduction

Due to intricate membrane systems eukaryotic cells exhibit a great deal of intracellular compartmentalization, which provides an organizational hierarchy for various cellular and metabolic pathways. As a consequence, eukaryotic cells have evolved

J.L. Spears • M.A.T. Rubio • P.J. Sample
Department of Microbiology, The Ohio State Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

J.D. Alfonzo (✉)
Department of Microbiology, The Ohio State Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA
e-mail: alfonzo.1@osu.edu

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_5,

efficient transport systems that ensure that a given molecule reaches its final destination, which often differs from its original site of synthesis. This intracellular organization is highlighted especially in the maturation pathways of tRNAs in trypanosomatid parasites. In most eukaryotes, tRNAs are encoded in either of two DNA-containing compartments: the nucleus or the genome-containing organelles (chloroplast and mitochondria). However, in trypanosomatids, mitochondrial genomes do not appear to contain any tRNA genes. Therefore, the complete set of tRNAs used in both cytoplasmic and mitochondrial protein synthesis is encoded solely in the nuclear genome. In these organisms tRNAs are thus transcribed in the nucleus, exported to the cytoplasm, and later a subset of cytoplasmic tRNAs is actively imported into the mitochondria. However, before tRNAs can be rendered functional in any cellular compartment, they face many enzymatic reactions including end trimming, intron splicing, tRNA editing, and other modifications. Some of these processes, for example those involved in trimming of extraneous sequences at the tRNA ends, occur in the nucleus, usually preceding cytoplasmic export. Others, like editing and modification, may occur at any point in the tRNA maturation pathway and in any of the tRNA-containing compartments. The following sections will cover what is known about these maturation processes in trypanosomatids, focusing primarily on the genera *Trypanosoma* and *Leishmania* where most of the research has been performed. As tRNA transcription will be extensively covered elsewhere in this book, we will focus on steps immediately following transcription by polymerase III in the nucleus and steps that precede tRNA degradation in any compartment. Therefore, tRNA synthesis and stability will be covered only when needed to explain certain aspects of tRNA maturation. Special emphasis will be placed throughout this chapter on how transport dynamics may affect tRNA maturation, which in turn may have direct bearing on tRNA function.

5.2 tRNA Biogenesis and Processing

5.2.1 Nuclear Trimming of 5′ and 3′-Sequences to Generate a Full-Length tRNA

In trypanosomatids, like in most organisms, tRNAs are transcribed as precursor molecules, containing extra sequences at their 5′- and 3′-ends (5′-leader and 3′-trailer, respectively) that have to be removed as part of a tRNA's normal maturation process. The exact order of end-trimming events depends on the tRNA species, but most commonly in eukaryotic organisms, processing begins in the nucleus with 5′-leader removal followed by 3′-trailer cleavage (Fig. 5.1).

Maturation of the 3′-end of eukaryotic pre-tRNA involves two sequential events: removal of the 3′-extension followed by CCA addition. Cleavage of the 3′-trailer of pre-tRNAs is accomplished by the highly conserved tRNase Z, often found as two homologous proteins in eukaryotes. The smaller of the two enzymes, tRNase ZS,

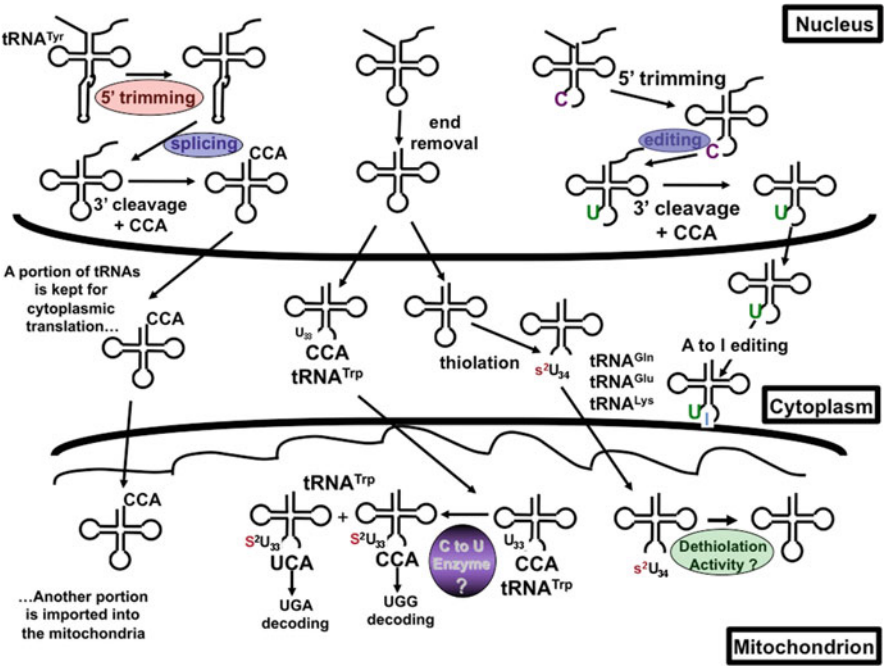


Fig. 5.1 tRNA maturation processes in trypanosomatids. Following transcription tRNAs undergo numerous processing events, those depicted here are the few pathways that have been studied in trypanosomatids. Shown on the *left* is the fate of a newly transcribed tRNA, specifically highlighting $tRNA^{Tyr}$, the only intron containing tRNA in trypanosomatids. The 5′-leader sequence is removed followed by intron splicing and 3′-trailer cleavage. CCA is added to the 3′-end of the tRNA, which is then exported into the cytoplasm where a portion of the tRNAs is kept for cytoplasmic translation and another portion is imported into the mitochondrion for mitochondrial translation. Equally important are tRNA editing and modification events that occur throughout the maturation process. Editing as exemplified by cytidine to uridine (C to U) and adenosine to inosine (A to I) can occur in all three compartments. *Highlighted* in the pathway in the *middle* is mitochondrial C to U editing that permits $tRNA^{Trp}$ to decode UGA as tryptophan. Highlighted in the maturation pathway on the *right* is the C to U editing of $tRNA^{Thr}$ that occurs in the nucleus before 3′-trailer cleavage and addition of the CCA end. This tRNA is further edited (A to I) after being exported to the cytoplasm. tRNA modifications such as thiolation (s^2U) have been observed in the cytoplasm (i.e., $tRNA^{Gln}$, $tRNA^{Glu}$, $tRNA^{Lys}$) as well as in the mitochondrion (i.e., $tRNA^{Trp}$) (*middle* pathway). Activities depicted as *colored* circles represent reactions known to occur in trypanosomatids; however, the enzymes responsible for those reactions remain unknown

appears in all domains of life (bacteria, eukarya, and archaea), while the larger, tRNase ZL, is only found in eukarya (formerly referred to as Elac1 and Elac2, respectively). It is not yet entirely clear why the two forms are present in some eukaryotes, but differences in substrate specificity (Takaku et al. 2004) and cellular localization of the two enzymes may explain it. tRNase ZL usually contains a predicted mitochondrial (Levinger et al. 2004) or chloroplastic (Vogel et al. 2005) targeting sequence, while tRNase ZS always localizes to the nucleus. In trypanosomatids, not much is known about the nature of the 3′-end maturation

enzyme(s). However, bioinformatic analysis in *T. brucei* reveals the presence of a potential homolog of the larger eukaryotic form of the enzyme, tRNase ZL (unpublished results). The other homolog is either absent or not easily detectable by database searches. Furthermore, attempts to identify a version of tRNA Z that contains a mitochondrial targeting sequence have been unsuccessful, suggesting that only tRNase ZL exists in trypanosomatids, but unexpectedly localizes to and functions only in the nucleus. However, one must admit this is just a computer-based prediction. If this protein is truly only nuclear, this should not be at all surprising due to growing evidence that mitochondrial tRNAs are imported from the cytoplasm as mature molecules (as described in the sections below). Thus tRNA import obviates the need for 3′-processing inside the mitochondria. It is possible that some version of tRNA Z may be involved in processing other RNAs in various organisms, including trypanosomatids, judging by the enzyme′s substrate specificity. It has been noted that the minimal substrate for in vitro cleavage by tRNase Z is an RNA molecule consisting of the T-stem loop, acceptor stem, and a minimum of an 11-nucleotide 3′-overhang (Mayer et al. 2000; Nashimoto et al. 1999; Schiffer et al. 2001), raising the possibility of alternative substrates.

Besides tRNA Z other cellular components, including the multisubstrate RNA-binding La protein, have been shown to play a less direct but important role in the removal of the 3′-trailer in eukaryotes (Yoo and Wolin 1997). In *T. brucei*, RNA polymerase III transcripts such as tRNAs end in a string of uridines that serves as the binding substrate for the La protein, which has been shown to prevent nuclease degradation of bound RNA in other systems (Wolin and Cedervall 2002). The *T. brucei* La protein homolog has been identified and recent attempts to characterize its function have produced interesting results. In yeast, downregulation of La is nonlethal; this is likely due to redundant mechanisms of 3′-trimming by exonucleases (Yoo and Wolin 1997). Conversely, RNAi of the *T. brucei* La protein arrested growth shortly after induction, indicating that it is essential for cell viability (Arhin et al. 2005; Foldynova-Trantirkova et al. 2005).

It has also been observed that the downregulation of La caused a 50% decrease in elongator $tRNA^{Met}$ and a fourfold increase in intron-containing $tRNA^{Tyr}$. Despite the various proposed interactions of La with different RNA species (for example other tRNAs), these were the only two to be noticeably affected. It should also be noted that immunoprecipitation experiments with La-specific antibodies and *T. brucei* extracts only pulled down the intron-containing $tRNA^{Tyr}$, pointing to its definite involvement in pre-$tRNA^{Tyr}$ processing. How it affects elongator $tRNA^{Met}$ and whether it interacts with other pre-tRNAs is still unknown (Arhin et al. 2005). It has been proposed that La plays a role in either recruiting the tRNA splicing machinery to the tRNA substrate and/or the recruitment of a tRNA Z to cleave the 3′-end trailer sequence. Still, the precise role the La protein plays in tRNA maturation remains a mystery, and subsequent experiments are required to elucidate its function in *T. brucei*. In a different study of *T. brucei* initiator-$tRNA^{Met}$, nuclear-localized polypeptides with similarity to an *S. cerevisiae* protein, involved in formation of SAM-dependent 1-methyladenosine at position 58 of the TΨC loop of tRNAs (Anderson et al. 1998), were shown to affect tRNA processing when

down-regulated. RNAi of the predicted adenosine methyltransferase, *Tb*MT40, produced an accumulation of pre-tRNAs with 5′-leaders but a trimmed 3′-end. This provides another example in trypanosomatids in which a 3′-to-5′ order of processing is observed. Curiously, downregulation of *Tb*MT40 and its partner protein *Tb*MT53, which is predicted to bind $tRNA^{iMet}$, produced an abundance of elongator $tRNA^{Met}$. The reason for this is unknown. The authors also show via immunoprecipitation that *Tb*MT40 and *Tb*MT53 are part of a 300 kDa multimeric complex with unknown components and function (Arhin et al. 2004).

Similarly, the mechanism for removal of the 5′-leader from tRNAs in trypanosomatids is not exactly clear. It was expected that, like in most eukaryotic tRNAs, 5′-removal is mediated by the highly conserved ribonuclease P (Xiao et al. 2002). RNase P was one of the first examples of a ribonucleoprotein using RNA as its catalytic moiety (Guerrier-Takada et al. 1983), cleaving its RNA substrate to produce 5′-monophosphate and 3′-hydroxyl end products. This metalloenzyme requires divalent ions such as Mg^{2+} for catalysis (Frank and Pace 1998) and correct folding of its RNA component (Baird et al. 2007; Hsieh and Fierke 2009; Kazantsev et al. 2009). It is believed that eukaryotic RNase P recognizes the pre-tRNA via the acceptor stem, the TΨC loop (Baird et al. 2007; Carrara et al. 1995; Levinger et al. 1995; Yuan and Altman 1995), and is aided by its interaction with pre-tRNA bound to the La protein (Esakova and Krasilnikov 2010; Yoo and Wolin 1997).

The exact nature of the enzyme(s) responsible for the removal of the 5′-leader sequence of trypanosomatid pre-tRNAs is not yet clear. However, *a T. brucei* mitochondrial RNase P-like activity has been partially characterized and shown to effectively cleave the 5′-leader of pre-tRNAs. Curiously, there is evidence suggesting that the RNA component of this mitochondrial RNase P is not needed for cleavage activity (Salavati et al. 2001). This observation is consistent with a recent report that the RNase P activity of human mitochondria is a protein-only enzyme (Holzmann et al. 2008). Interestingly, humans still encode an RNA-containing RNase P used for the processing of nucleus-encoded tRNAs. Since the presence of RNase P RNA has been difficult to discern by bioinformatic means, it is possible that in these organisms the requirement for the RNA component has been lost during evolution and that, in fact, even the nuclear RNase P is a protein-only enzyme. Paradoxically, it is not clear why an RNase P-like activity would be required in trypanosomatid mitochondria, since as mentioned above, tRNAs are imported as mature molecules. Likely, trypanosomatid mitochondrial RNase P has an alternative function, perhaps in the maturation of mitochondrial polycistronic transcripts, including mitochondrial rRNA.

Before the end-matured tRNA can participate in aminoacylation, a highly conserved three-nucleotide sequence, CCA, must be added posttranscriptionally by the 3′-specific CCA nucleotidyltransferase, the CCA-adding enzyme. In general, there are two classes of CCA-adding enzymes: the archaeal class I and the eukaryotic and bacterial class II (Xiong and Steitz 2006). Both facilitate the addition of CCA using CTP and ATP in a template-independent polymerization reaction. Currently, no published information is available for tRNA nucleotidyltransferases in *T. brucei*, but expectedly, a BLAST search of the *T. brucei* sequence database revealed the

presence of a putative class II CCA-adding enzyme and no matches for class I. However, this enzyme has not been genetically or biochemically characterized.

5.2.2 Intron Removal and Nuclear Export

In *T. brucei*, the only intron-containing tRNA is tyrosine tRNA ($tRNA^{Tyr}$), which has an 11 nucleotide intervening sequence. Intron-containing tRNAs can be found in all of the phylogenic kingdoms and usually consist of a short sequence of nucleotides that occur immediately 3′ of the anticodon loop in eukaryotes. The function of introns within tRNAs is largely unknown; however, in yeast $tRNA^{Tyr}$, the intron is required for conversion of the central uridine of the anticodon to pseudouridine (Johnson and Abelson 1983). The mechanism of tRNA intron splicing proceeds in two steps. The first involves the removal of the intron by a specific endonuclease that recognizes conserved elements of the anticodon stem. The two halves generated by the endonuclease are then joined by a ligase to complete the process (Greer et al. 1987; Fig. 5.1).

What is eye-catching about the *T. brucei* $tRNA^{Tyr}$ intron is its length; at 11 nucleotides it is one of the shortest tRNA introns known in nature. Assuming that tRNA splicing is as conserved in *T. brucei* as it is in other organisms, the *T. brucei* system may inherently provide insight into a very simplified minimal substrate for the *T. brucei* endonuclease. Although neither the endonuclease nor the ligase have been identified in trypanosomatids, earlier work showed that unlike the yeast endonuclease, site-specific mutations in the anticodon loop of *T. brucei* $tRNA^{Tyr}$ disrupted splicing, implying the strict necessity for a specific structure for cleavage. This is reminiscent of tRNA splicing in plants (Stange et al. 1988). Notably, $tRNA^{Tyr}$ mutants deficient in tRNA splicing were observed to accumulate with 3′-extensions, while 5′-extensions were not detected with probes specific for the 5′-leader sequence, suggesting that the 5′-mature and unspliced intermediate serves as the substrate for 3′-end processing. This order of events is different from that observed in other organisms where 5′- and 3′-maturation occur before splicing (De Robertis and Olson 1979; Melton et al. 1980).

Once the pre-tRNA ends have been cleaved off and in the case of $tRNA^{Tyr}$, once the intron has been removed, the mature tRNA must be exported from the nucleus to the cytoplasm. Again, little is known about this process in *T. brucei*, but in other eukaryotic systems a Ran-GTP dependent protein, exportin T, carries tRNA across the nuclear membrane and releases it after Ran-GTP hydrolysis to form Ran-GDP (Hopper et al. 2010). It has been proposed that exportin T contributes to tRNA proofreading before export by assessing the condition of the 5′- and 3′-ends as well as the structure provided by the acceptor stem and T-stem loop (Arts et al. 1998; Cook et al. 2009; Lipowsky et al. 1999). The maturation status of tRNAs destined for mitochondrial import is currently a point of contention, so it will be interesting to see how exportin T discriminates substrates for different localization or if other factors are also involved in tRNA export, as has been seen in fungi (Hopper et al. 2010).

In summary, although the various activities that mature a tRNA via end trimming and CCA addition are predictably present in trypanosomatids, little is known about the biochemical nature and the mechanisms of any of the enzymes or processing events.

5.2.3 tRNA Editing and Modifications

Maturation of tRNAs involves more than the typical splicing of introns and the removal of the 5′-leader and 3′-trailer. Posttranscriptional editing and modification of many nucleotides throughout the tRNA sequence represents in all domains of life another necessary step before a tRNA becomes fully functional. Over 100 different modified nucleotides have been characterized with a median of 8 modifications per tRNA (Phizicky and Alfonzo 2009). While individually many of the modifications are dispensable, with functions ranging from tRNA stability, to tRNA structure and folding to translational fidelity, editing and modification events can be integral parts of normal cell function, some even being essential for viability (Phizicky and Alfonzo 2009). In a few cases, trypanosomatid tRNAs provide the only example for a particular modification or editing event. There are, of course, the modifications that are nearly universally conserved and after which tRNA regions are named, including dihydrouridine (D) in the D arm and pseudouridine (Ψ) in the TΨC arm, but perhaps of greater interest are those modifications that may be unique to trypanosomatid physiology. Alternatively, equally interesting are those modifications that despite being in common with other systems, in trypanosomatids, still provide nuances in their mechanisms of synthesis. These may promise to become potential chemotherapeutic targets. In general, little is known on tRNA editing and modification in trypanosomatids. The following sections will explore various themes that are slowly developing in the study of modifications in these organisms, but unfortunately the discussion will be limited to the few examples of modifications that have been studied. These include cytidine-to-uridine (C to U) and adenosine-to-inosine editing (A to I), inside and outside of organelles (Alfonzo et al. 1999; Crain et al. 2002; Rubio et al. 2006, 2007), and tRNA thiolation (Bruske et al. 2009; Crain et al. 2002; Wohlgamuth-Benedum et al. 2009).

5.2.3.1 tRNA Editing

As originally stated, “RNA editing” is broadly defined as a programmed posttranscriptional alteration of sequence information in mRNA beyond what is encoded in the DNA genome (Benne et al. 1986; Gray 2003). This definition was originally used to describe only canonical nucleotide changes, but in the following subsections we will use an even more generalized definition of editing that includes the replacement of canonical nucleotides for noncanonical ones (e.g., adenosine for inosine) that have a direct effect on either tRNA structure or function.

C to U Editing

The first studies of tRNA editing in trypanosomatids came from the fact that in these organisms, all the tRNAs used in mitochondrial translation are nucleus encoded and subsequently imported into the mitochondrion. A problem is encountered when considering the translation of tryptophan codons because, like with many eukaryotic organisms, in the trypanosomatid mitochondrial genome the canonical UGG tryptophan codon is often replaced by UGA, a stop codon in cytoplasmic translation (Alfonzo et al. 1999). This led to the question of how organisms with a single $tRNA^{Trp}$ with anticodon CCA could decode UGG as tryptophan and UGA as stop in the cytoplasm, while decoding UGA as tryptophan once imported into the mitochondrion. As first shown in *Leishmania tarentolae* (Alfonzo et al. 1999) more than a decade ago and later corroborated in *Trypanosoma brucei* (Charriere et al. 2006), these organisms have solved this decoding conundrum in a simple yet rather elegant way. The tRNA undergoes cytidine (C) to uridine (U) editing at the first position of the anticodon, thus changing the CCA anticodon to UCA which can now decode UGA by canonical base pairing and even UGG by wobbling (Alfonzo et al. 1999). Clearly, mitochondrial compartmentalization of such an activity avoids the potential generation of a high-copy suppressor tRNA (with anticodon UCA) in the cytoplasm (Fig. 5.1).

While the reason for this type of editing is clear, its mechanism remains uncertain. The simplest activity accounting for the observed C to U editing would involve a tRNA-specific cytidine deaminase. A precedent for a polynucleotide-specific cytidine deaminase already exists in the editing of mRNA in mammals (Navaratnam et al. 1993). The possibility of such an enzyme in trypanosomatid mitochondria has been most recently reinforced with the discovery of an analogous activity in archaea, where the first *c*ytidine *d*eaminase *a*cting on *t*RNAs (CDAT8) has been described (Randau et al. 2009). This enzyme, a member of the larger cytidine deaminase family, catalyzes C to U via a conserved hydrolytic deamination reaction in a zinc-dependent manner. Unfortunately, no protein similar to CDAT8 in *Trypanosoma* or *Leishmania* could be identified by bioinformatic analysis. Therefore, the enzyme responsible for this mitochondrial editing activity remains elusive. However, an in vivo approach still allowed defining certain features of the tRNA substrate that are important for $tRNA^{Trp}$ editing. We took advantage of the fact that tRNA variants expressed from a plasmid in *L. tarentolae*, when transcribed could still be imported into the mitochondria. We used this approach to establish that a single base pair reversal, at the last position of the $tRNA^{Trp}$ anticodon stem, abolished C to U editing in vivo (Crain et al. 2002). Beyond this position, it is not known what other determinants of editing exist in the natural $tRNA^{Trp}$ substrate in *Leishmania,* but in *T. brucei* (as discussed below) modifications may serve as an antideterminant (Wohlgamuth-Benedum et al. 2009). Clearly, C to U editing has direct bearing on translation and because an estimated 88% of the conserved mitochondrial tryptophan codons are UGA, it has been concluded that this editing is essential for mitochondrial translation and cell survival.

The only other example of C to U editing in trypanosomatids occurs outside the mitochondrion and is, in fact, the first example of C to U editing outside organelles in eukarya (Gaston et al. 2007; Rubio et al. 2006). Occurring just 5′ of the tRNA anticodon (position 32) in all three isoacceptors of $tRNA^{Thr}$, this editing event takes place while the tRNA is still in the nucleus and before the removal of the 5′-leader from the tRNA (Fig. 5.1) (Gaston et al. 2007). Unlike the mitochondrial editing of $tRNA^{Trp}$, C_{32} to U_{32} editing has no direct bearing on decoding; however, in vitro, C_{32} to U_{32} stimulates, although is not required for, the essential adenosine-to-inosine formation at the wobble position of $tRNA_{AGU}^{Thr}$. As A to I has a direct effect on translation and is essential for cell viability, it may stand to reason that C_{32} to U_{32} indirectly affects translational efficiency of some, if not all, threonine codons.

A to I Editing

By far, the most common form of editing occurring in eukaryotic tRNAs is the adenosine to inosine substitution at the first position of the anticodon. Mechanistically, this type of editing proceeds via a conserved hydrolytic deamination reaction, much the same as that described for the C to U activity of archaeal CDAT8 (Randau et al. 2009). Unlike the case of C to U editing of $tRNA^{Trp}$, the enzyme responsible for A to I editing has been identified in *T. brucei* in our laboratory (*Tb*ADAT2/3) (Rubio et al. 2007). This enzyme functions as a heterodimer, comprised of two subunits, ADAT2 and ADAT3, which harbor the Zn-coordinating motifs, H(C)xE and PCxxC, characteristic of all members of the cytidine deaminase superfamily. In these enzymes, the conserved histidine and two cysteines coordinate a zinc ion, while the fourth ligand is an activated water molecule. A conserved glutamate in ADAT2 then acts as a proton shuttle between the activated water and the exocyclic nitrogen at C6 of the purine ring. This set of ligands and cofactors thus act in concert to hydrolyze the amino group at C6 of the adenosine ring releasing ammonia as the leaving group. Although inosine deamination has been studied for several years in various systems, the enzyme from *T. brucei* still produced a few interesting surprises. For instance, the core sequences of these deaminases, including the residues involved in catalysis, resemble those found in cytidine deaminases despite the fact that the enzyme catalyzes an adenosine deamination reaction (Gerber and Keller 1999). This observation in itself may be of no consequence, but given our finding that tRNAs may undergo both A to I and C to U editing (as discussed briefly in the preceding section), we decided to explore the possibility that this enzyme could perform both reactions. To our surprise, downregulation of TbADAT2 (the presumed catalytic subunit) in *T. brucei* led to a reduction in both A to I and C to U editing of $tRNA^{Thr}$, suggesting that the enzyme is involved in both reactions in vivo. However, in vitro we found that TbADAT2/3 could robustly catalyze A to I, but not C to U, editing in tRNA. This raises the question of what factor may be missing in the in vitro reaction with recombinant proteins. We could, however, demonstrate that TbADAT2/3 efficiently catalyzes C to U deamination of DNA (Rubio et al. 2007). Although the biological significance of the DNA reaction

is not clear, the observed activity demonstrated that the enzyme had the inherent capacity to catalyze both reactions as previously hypothesized. The question still remains, why it cannot perform the reaction on tRNA? One explanation may be that there are posttranscriptional modifications in the natural substrate that may be important specificity determinants and are missing in our in vitro substrate. Alternatively, the enzyme itself may be missing some additional cofactor found in vivo. Here it is worth noting that the A to I reaction at the wobble position occurs after tRNA export from the nucleus, but interestingly, both subunits (*Tb*ADAT2 and 3) localize to both the nucleus and the cytoplasm (Gaston and Alfonzo, unpublished results). This raises the possibility that within a cellular compartment, the substrate specificity can change by virtue of other proteins associating with ADAT2 within a given compartment. In this realm, compartmentalization will determine enzyme specificity in a manner reminiscent of the previously described C to U editing of $tRNA^{Trp}$ in mitochondria of trypanosomatids. Currently, the answers to many of these questions are still far from clear.

5.2.4 Modifications

Beyond C to U and A to I editing, trypanosomatid tRNAs also undergo a number of posttranscriptional modifications. Initially, a number of modifications were identified by a combination of liquid chromatography and mass spectrometry analysis of cytoplasmic and mitochondrial $tRNA^{Trp}$ (Crain et al. 2002). This of course was prompted by the discovery of the C to U editing event described above. From these studies, 12 different modifications were identified in *Leishmania* $tRNA^{Trp}$ including dihydrouridine (D), pseudouridine (Ψ), 2-thiouridine (s^2U), N^6-isopentenyladenosine, and a number of sugar and base methylations (i.e., 2′-O-methylcytidine, 2′-O-methyluridine, 2′-O-methylguanosine, 5-methyluridine, 7-methylcytidine, and N^2-methylguanosine). These are by no mean unusual and occur in tRNAs from other organisms (Juhling et al. 2009; Sprinzl and Vassilenko 2005). What generated interest was the finding that many modifications were added to the tRNA following mitochondrial import.

Most surprising was the discovery that this tRNA undergoes thiolation at position 33 of the anticodon (adjacent to the edited nucleotide). Although previous work had raised the possibility of this tRNA position being modified, this finding, however, constituted the first demonstration of modification at this position in any organism. U_{33} plays an important function in shaping the anticodon loop structure allowing a tRNA to splay the anticodon nucleotides, priming them for translation. This unusual thiolation of $tRNA^{Trp}$ eventually led to further exploration of the thiolation pathways of trypanosomatids. In *T. brucei*, like in most eukaryotes, there are two places where tRNAs can be thiolated: the cytoplasm and the mitochondria (Fig. 5.1). Cytoplasmic thiolation seems to require the same components as in yeast, but the specific contributions of factors like Urm1, Uba, Ncs 1, and Ncs 6 have not been formally tested (Leidel et al. 2009). These are identifiable by genomic

searches and are expected to provide similar functions as in the yeast system. Trypanosomatids and other eukaryotes differ in the nature of the tRNAs used for mitochondrial thiolation. In the *T. brucei* cytoplasm, $tRNA_{UUG}^{Gln}$, $tRNA_{UUC}^{Glu}$, and $tRNA_{UUU}^{Lys}$ are the only known targets for thiolation, but because of tRNA import, these tRNAs enter the mitochondria already containing the thiol group added by the cytoplasmic thiolation system. So far, the only tRNA known to undergo mitochondrial thiolation in these organisms is $tRNA^{Trp}$ (Alfonzo et al. 1999; Charriere et al. 2006), which as described above also undergoes C to U editing. Surprisingly, in our studies of thiolation, we found that RNAi downregulation of any of the mitochondrial thiolation factors (including Nfs, the key desulfurase essential for thiolation) led to upregulation of tRNA editing to almost 100% (Wohlgamuth-Benedum et al. 2009). This observation implies that s^2U_{33} acts as a negative determinant for tRNA editing and helps maintain the levels of the two isoacceptors as required for UGG and UGA decoding. Notably, $tRNA^{Trp}$ is not thiolated in the cytoplasm during transit. This raised the question of how this tRNA avoids cytoplasmic thiolation. We showed that editing is not required for thiolation at U_{33} in *L. tarentolae*, a close relative of *T. brucei* (Crain et al. 2002). Therefore, the only viable explanation is that the cytoplasmic and mitochondrial tRNA thiolation systems differ in their substrate recognition and that, in fact, there are features common to $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Lys}$ required for thiolation that are not present in $tRNA^{Trp}$. Recently, it was shown that following import, $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Lys}$ become dethiolated by an unidentified activity, raising the possibility that the mitochondrial thiolation may play a "repair" role for this tRNA set; again, this has not been formally tested (Bruske et al. 2009).

An additionally surprising discovery with the *T. brucei* system involves the fate of the cytoplasmic tRNAs in the absence of thiolation. We showed that $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Lys}$ become unstable and are quickly degraded if thiolation is impaired (Wohlgamuth-Benedum et al. 2009). This instability was specific only to the thiolated tRNA set and differs in this respect from a more general rapid tRNA degradation pathway described (Alexandrov et al. 2006; Engelke and Hopper 2006). The nature of the enzymes or factors mediating this degradation is currently unknown. Overall, the thiolation system of *T. brucei* shows not only how intracellular compartmentalization affects tRNA modification, but it even exemplifies how location may affect modification enzyme substrate specificity. A curious corollary of the thiolation story is the remarkable finding that the same desulfurase required for iron–sulfur (FeS) assembly is also required for tRNA thiolation in both the cytoplasm and the mitochondria. In the mitochondria, subunits of respiratory complex III require an FeS cluster, therefore downregulation of Nfs could lead to downregulation of respiratory rates. We suggest a model by which the divergence of the two pathways (FeS assembly and tRNA editing/thiolation) from a common key enzyme may be exploited by these cells to carefully match respiratory rates to mitochondrial translation perhaps by offsetting the 50/50 ratio of edited and unedited $tRNA^{Trp}$. It is also worth mentioning that a number of cytoplasmic modification enzymes require FeS clusters for activity, thus this hypothesis may even include cytoplasmic modification systems in connection with FeS-cluster

assembly for global metabolic regulation (Lill and Muhlenhoff 2006). These proposals are of course largely speculative, but their exploration may reveal important global aspects in the coordination of these facets of cellular metabolism.

5.3 tRNA Import into the Mitochondria

The mitochondrion is the powerhouse of the cell in that it provides the bulk of the energy required for various cellular transactions in the form of ATP. The mitochondrial genome of kinetoplastids encodes only a small subset of genes required for the assembly of complete respiratory complexes and generation of ATP. Glaringly missing from the mitochondrial genomes of trypanosomatids are any recognizable tRNA genes. This is not unique to the trypanosomatids and indeed most, if not all, mitochondria-containing organisms have lost tRNA genes from their mitochondrial genomes during evolution. Historically, the case of missing tRNA genes in mitochondria dates back to the 1960s when Suyama and coworkers while working with the single-cell protist *Tetrahymena pyriformis* first introduced the concept of mitochondrial import of tRNAs from the cytosol (Suyama 1967). This was needed in order to supplement the protein synthesis machinery with the missing tRNAs, which in conjunction with those tRNAs still encoded in the organelle, could account for the decoding of all the codons used in mitochondrial translation (Suyama 1967).

To date, multiple occurrences of tRNA import have been described in diverse organisms: the ciliate *Tetrahymena* (Rusconi and Cech 1996; Suyama 1967), kinetoplastid flagellates *Trypanosoma* and *Leishmania* (Hancock and Hajduk 1990; Mottram et al. 1991; Simpson et al. 1989), yeast (Martin et al. 1979; Rinehart et al. 2005), various marsupials (Dorner et al. 2001), the apicomplexan *Toxoplasma* (Esseiva et al. 2004), plants (Marechal-Drouard et al. 1988), and, most recently, mammals (Rubio et al. 2008). The number of imported tRNAs varies greatly from as few as one tRNA in mammals to a full set of tRNAs in kinetoplastids and apicomplexans (Salinas et al. 2008). The kinetoplastids *Leishmania* and *Trypanosoma* represent the most extreme case of missing tRNA genes, in that all cytoplasmic tRNAs, with the exception of initiator $tRNA^{Met}$, have to be imported from the cytoplasm. In this section our discussion will focus on the studies of tRNA import into the mitochondria of *Leishmania* and *Trypanosoma*, two systems that cover the lion's share of what is known about tRNA import in this group of evolutionary diverse organisms.

The disappearance of a varying number of tRNA genes from mitochondrial genomes among different organisms has perhaps driven the independent evolution of systems that permit the import of tRNAs from the cytoplasm (Lithgow and Schneider 2010; Salinas et al. 2008); still, the factors that control tRNA import, specifically the nature of the import machinery, remain to be fully understood. In the absence of clear knowledge of the import machinery, here we will refrain from the relaxed use of the term "mechanism," and instead we will divide import

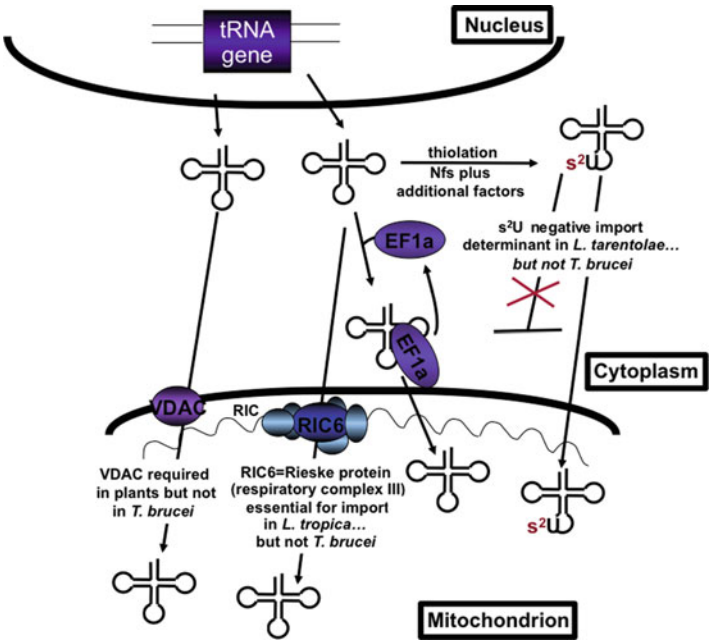


Fig. 5.2 Pathway for tRNA import into mitochondria independent of the protein import pathway. The pathways shown highlight important features of tRNA import into the mitochondria that occurs independently of the protein import pathway. Specific factors and complexes affect import differently in different organisms. For example, thiolation (s^2U) is a negative import determinant for tRNA import in *L. tarentolae*, but is not a determinant in *T. brucei*. Likewise, RIC6 (Rieske protein) of the RNA import complex (RIC) is necessary for import in *L. tropica*, but not in *T. brucei*. Elongation factor EF1a is also involved in import in *T. brucei*. Other factors, like VDAC, play a role in other organisms, but not in *T. brucei* (as shown)

systems into two broadly defined types (a) those utilizing the known protein import pathway, and (b) those in which tRNA import is shown to occur independent of the protein import machinery (Alfonzo and Soll 2009). For clarity, type I import, is limited to the single known example from yeast where direct participation of the canonical and well-characterized protein import machinery itself helps drive the tRNA across the mitochondrial double membrane. Since this type of import has not been described in trypanosomatids, for all practical purposes in these organisms tRNA import occurs independently of the protein import machinery (Fig. 5.2). Clearly tRNA import components themselves have to be transported and inserted in the membranes; careful teasing of the secondary effects of protein import from that of the downstream import of the tRNA remains a delicate nuance in the identification of proteins involved in the transport of tRNAs.

Common to all organisms that import tRNAs into the mitochondrion, including trypanosomatids, is the need for tRNAs to traverse both the outer and inner mitochondrial membranes to reach their final destination in the mitochondrial matrix, where translation takes place. There are various hypotheses that explain

how a given tRNA is recognized in the cytosol and delivered to the mitochondrion, but currently no single proposal for tRNA delivery explains all the data published so far. While the tRNA import process can be efficiently reproduced in vitro in the absence of cytoplasmic factors, the contribution of cytoplasmic factor(s) to the process also has to be addressed. Given that tRNAs in the cell are integrally associated with components of the translation machinery, they need to be recycled during protein synthesis. The missing link in the series of events is how the tRNA frees itself from the translation machinery and the mode by which the tRNA reaches the mitochondrion. One hypothesis proposes that a fraction of the nuclear-encoded tRNAs escapes the cytoplasmic translation machinery by interacting with protein factors that will consequently direct it to the mitochondrion. In *Trypanosoma brucei*, the best example is provided by the cytoplasmic translation elongation factor 1α (eEF1α), which plays a role as a specificity determinant for a small subset of imported tRNAs (Bouzaidi-Tiali et al. 2007), but by extrapolation may play a similar role with all cytoplasmic tRNAs prior to transport (Fig. 5.2). In vivo experiments suggested that in *T. brucei*, $tRNA^{Met}$, $tRNA^{Ile}$, and $tRNA^{Lys}$ contain a major localization determinant within the TΨC-stem at nucleotides 51 and 63, which form a base pair in the canonical tRNA structure. This finding was a logical derivation of an earlier observation that in the initiator $tRNA^{Met}$, one of two tRNAs that is not imported into the *T. brucei* mitochondrion, this base pair is identical to the main antideterminant preventing interaction with elongation factor (eEF1α). This is different from all the imported tRNAs, which contain the base pair needed for eEF1α binding, implicating this factor in tRNA delivery to the mitochondrial surface. Indeed, knockdown of cytoplasmic eEF1α, but not of translation initiation factor 2, was found to inhibit mitochondrial import of newly synthesized tRNAs, long before cytoplasmic translation or cell growth is affected. Additionally, $tRNA^{Sec}$, the only other cytosol-specific tRNA in *T. brucei*, which has its own elongation factor and does not bind eEF1α, could be redirected to the mitochondrion by simple introduction in its sequence the same base pair required for eEF1α binding. Therefore, for tRNAs that are imported, binding to eEF1α provides an additional level of import specificity beyond what is already provided by the import machinery itself (Lye et al. 1993; Shi et al. 1994; Bouzaidi-Tiali et al. 2007).

In terms of actual mitochondrial import factors intrinsically associated with the membranes, a system has been described in *Leishmania tropica*. This system is independent of the mitochondrial protein import pathway and involves the so-called *R*NA *i*mport *c*omplex (RIC) (Mukherjee et al. 2007). The RIC was derived from a detergent-solubilized extract of *L. tropica* inner mitochondrial membranes that was subjected to chromatography. This affinity column harbored an immobilized RNA oligonucleotide corresponding to what has been identified previously as a legitimate import signal by the same authors. This affinity substrate corresponded to the D arm of $tRNA^{Tyr}$ and when used for affinity chromatography yielded a massive multiprotein aggregate (Bhattacharyya et al. 2000). Sequences resulting from mass spectrometry analysis of the resulting complex produced a total of 122 nuclear-encoded ORFs. These were then used to mine the *Leishmania* genome sequence database, and the role of specific proteins in tRNA import was further analyzed by

Western blotting and gene knockdown using RNAi (Mukherjee et al. 2007). The RIC is comprised of an 11-protein core complex assembled at the mitochondrial inner membrane with a stoichiometry adding up to a total mass of ~580 kDa. The import complex requires ATP and membrane potential to import tRNAs. Within the complex, there are three mitochondrial- and eight nuclear-encoded subunits. Analyses by knockdown and in vitro reconstitution experiments indicated that six of the eight nuclear-encoded subunits, RIC 1, 4A, 6, 8A, 8B, and 9, are essential for import. The RIC has been affinity-purified and resolved from other mitochondrial complexes by native gel electrophoresis (Goswami et al. 2006). Functional complexes could be reconstituted with recombinant subunits expressed in *Escherichia coli*. Several essential RIC subunits are identical to specific subunits of respiratory complexes. The two nonessential subunits were identified as RIC3, an M16 metalloproteinases, and RIC5, a trypanosomatid-specific protein. It is proposed that RIC1 and RIC8A are the two receptors involved in initial tRNA binding. Then, trimeric RIC6 and RIC9 form the translocation pore, while RIC4A and RIC8B anchor the complex to the membrane. Membrane-embedded mitochondrial-encoded subunits 2 (dimeric), 4B (substoichiometric), and 7 interact with RIC4A (Fig. 5.2). The dispensable subunits RIC3 and RIC5 are assembled peripherally (Goswami et al. 2006).

Despite many reports, the relevance of the *Leishmania* import complex is still controversial. For instance, one of its essential components is the Rieske protein, but downregulation of its expression by RNA interference in *T. brucei* had no effect on tRNA import (Paris et al. 2009), despite the predicted effects on membrane potential and mitochondrial function traditionally ascribed to Rieske function. Thus in the kinetoplastid system, the true nature of the tRNA import machinery is not yet clear. In plants, inhibition of the VDAC (voltage-dependent anion channel) by the addition of reuthenium red impaired tRNA import into mitochondria (Salinas et al. 2006), suggesting that the VDAC plays a critical role for tRNA transport across the outer membrane. However, the same is not true in *T. brucei*, where a knockout of the gene encoding the VDAC protein had no effect on tRNA transport (Pusnik et al. 2009) (Fig. 5.2). In conclusion, to date no single protein has been identified and characterize that directly contributes to tRNA translocation across the mitochondrial membranes in trypanosomatids. Thus far, what all protein-import independent mechanisms, including those of trypanosomatids, have in common is the requirement for ATP (although the exact role of ATP is still not clear) and lack of requirement for membrane potential (Alfonzo and Soll 2009). However, since the actual transporters have not been identified, it is difficult to conclude more on the actual transport mechanism.

5.3.1 Analyses of tRNA Molecules as Import Substrates

Earlier observations suggested that in *Trypanosoma brucei*, the 5′-leader-containing precursor tRNA is the substrate for in vivo import. The leader sequence

was, in fact, supposed to contain sequence information needed for import. Specifically, the 5′-flanking sequence of the precursor $tRNA^{Leu}$ was shown to be important for its in vivo localization in *T. brucei* mitochondria (Sherrer et al. 2003). The proposed model was analogous to the protein import mechanism, where the preprotein N terminus serves as a "zip code" for organellar targeting. However, conflicting experiments showed that at least in the case of $tRNA^{Leu}$ (CAA) isoacceptor, import occurs regardless of the sequence context of the imported gene, implying that the tRNA presequence could not be an import determinant for every tRNA in *T. brucei* (Tan et al. 2002).

To address the contribution of sequence determinants within the mature tRNA molecule, the in vivo import of tRNAs into the mitochondrion *of L. tarentolae* was studied using two tRNAs that differentially localize within the cell (Lima and Simpson 1996). In *L. tarentolae*, $tRNA^{Ile}$(UAU) is mostly localized within the mitochondrion while $tRNA^{Gln}$(CUG) is primarily in the cytosol (Lye et al. 1993; Shi et al. 1994). To permit discrimination of the exogenously transfected tRNA from that of the endogenous tRNAs in these in vivo import experiments, plasmids encoding sequence-tagged variants at the D loop of either $tRNA^{Ile}$ or $tRNA^{Gln}$ were transfected into *L. tarentolae* cells. The import of RNAs, assessed by primer extension analysis, revealed that the plasmid-encoded genes were expressed and that the tagged tRNAs showed a similar intracellular localization as the endogenous tRNAs. The in vivo import experiments in *L. tarentolae* further demonstrated that the exchange or deletion of the 5′-flanking genomic sequences had no effect on the expression or mitochondrial localization of the tagged $tRNA^{Ile}$ or on the expression or cytoplasmic localization of the tagged $tRNA^{Gln}$, suggesting that the signals for importation are localized within the tRNA itself. Swapping the D stem and loop from the mainly cytoplasmic $tRNA^{Gln}$ with that from the $tRNA^{Ile}$ produced an increased mitochondrial localization of the plasmid-expressed mutated $tRNA^{Gln}$. Given that the D loop exchange between the two differentially localized tRNAs did not eliminate the mitochondrial localization of the plasmid-expressed tagged $tRNA^{Ile}$, the role of tertiary tRNA structure or additional sequence elements were proposed to contribute an essential role in signaling the mitochondrion to import the tRNA (Lima and Simpson 1996).

Similar experiments were performed in vitro with the same tRNAs used for the in vivo studies, namely $tRNA^{Ile}$ and $tRNA^{Gln}$. Worth highlighting with these studies in the *L. tarentolae* in vitro import system is the observation that the system is saturable. This strongly implies that the in vitro assay exhibits dynamics akin to receptor-mediated systems. Moreover, the amount of tRNA protected from nuclease digestion in the presence of isolated mitochondria reaches a plateau at different concentrations, 25 μM for $tRNA^{Ile}$ and 3.8 μM for $tRNA^{Gln}$ (Rubio et al. 2000). This difference seen in the in vitro saturation levels of imported tRNA is consistent with the levels of in vivo localization of these tRNAs, where $tRNA^{Ile}$ is primarily mitochondrial and $tRNA^{Gln}$ is primarily cytoplasmic (Kapushoc et al. 2002; Lye et al. 1993). In vitro, swapping the D-loops of full-length $tRNA^{Ile}$ and $tRNA^{Gln}$ led to the reversal of their import efficiency. This was also consistent with the previous in vivo import studies, which strongly supports the fact that at least for these two

tRNAs the in vitro import system was a valid mimic of the in vivo situation. This observation not only argues for the specificity of this process, but also implicates the importance of structural interaction between the D-arm and the TΨC-arm in the tRNA in providing discrimination for mitochondrial import. Again, in the absence of any additional cytoplasmic factors, the mitochondrion can itself mediate imported levels of tRNA reminiscent of the in vivo situation. This does not, of course, discount the possibility that cytoplasmic factors play a role in vivo.

In *Leishmania tropica,* the import of tRNAs into mitochondria has been proposed to involve two signature consensus sequences dividing tRNAs into either class I tRNAs or class II tRNAs. Type I tRNAs are proposed to contain a conserved sequence motif in the D arm needed for efficient import across the inner mitochondrial membrane, and positively stimulates import of type II molecules into the mitochondrial matrix. In contrast, type II tRNAs exhibit a conserved sequence motif within the variable region and the TΨC domain and are poorly imported into the mitochondrion. Additionally, type II tRNAs inhibit import of type I RNAs into the mitochondrial matrix. It is further speculated that a limited number of receptors regulate the rate of import of individual tRNAs and that regulatory interactions exist between the two types of tRNA molecules (Mahapatra et al. 1998). Furthermore, Mahapatra and Adhya suggest that in vitro, an import protein receptor at the mitochondrial surface specifically recognizes a sequence motif only at the D-arm for $tRNA^{Tyr}$ (Mahapatra et al. 1994, 1998). Specifically, the D-arm sequence motif AUGGCAGAG is proposed to interact with *L. tropica* RIC. This D-loop has been proposed as the receptor recognition motif and a key import determinant. The import probe was quite unique, in that it consisted of antisense RNA transcripts between −53 and +25 nucleotides of the *Leishmania* β-tubulin gene. Experimental evidence showed that mitochondrial uptake in vitro of this transcript could be competitively displaced by *Leishmania* tRNA. Accordingly, they suggested that the nucleotide sequence AUGGCAGAG or its motif within the antisense RNA might account for this competitive displacement, since this sequence motif exhibited homology to the sequence AU(U)GGC/UA within the D-loop region of two imported tRNA species, $tRNA^{Thr}$ and $tRNA^{Tyr}$ (Bhattacharyya et al. 2000).

In order to further test the proposal that an import signal resides within the D loop, a different study explored the possible presence of the proposed sequence determinant in all tRNA sequences available at the time. These sequences were extensively surveyed and examined for the existence of this D-loop motif and whether the motif could determine the import fate of a given tRNA (Suyama et al. 1998). No obvious consensus sequence at the D-loop was obtained from the analysis that could correlate tRNA localization within the cell to the presence of the D-loop motif, calling into question its in vivo importance. Therefore, the nucleotide sequence at the D-loop alone does not explain the observation of variable import phenotypes between tRNAs and in these arguments one must include the contribution of tertiary contacts between different regions of the highly conserved L-shaped tRNA structure (Suyama et al. 1998).

In *L. tarentolae*, further studies on the import of short RNAs, containing either the class I or II D-loop consensus demonstrated that both were imported in vitro into *L. tarentolae* mitochondria (Rubio et al. 2000). Strikingly, the in vitro import of the short RNA substrates led to the loss of discrimination for mitochondrial localization, which is in contrast to the import behavior observed with the full-length tRNA. This indicated that even though shorter RNA molecules may be imported in vitro, the ability of the mitochondria to discriminate between the substrates to be imported becomes compromised. To ascertain the limits of import specificity, smaller RNAs including five different 16- to 17-nucleotide mini-helix RNAs and one unstructured 17-nucleotide substrate (nonhelix forming) were also tested; all could be efficiently imported in vitro. However, unstructured RNAs of greater sizes (19, 24, and 33 nucleotides) failed to support import. The fact that in vitro import loses its specificity when the RNA substrate becomes less than ~17 nucleotides in length suggests that although a feature contributed by the D-loop may be important, it is within the context of a full-length tRNA that import discrimination is achieved.

One common feature between the in vitro tRNA import systems in *L. tarentolae* and *L. tropica* mitochondria is the observation of in vitro import in the absence of added cytoplasmic factors. These two systems, however, still yield conflicting data; as discussed above, a membrane potential is not required for import in *L. tarentolae*, but it is absolutely required in *L. tropica*. However, in both systems there is a need for ATP hydrolysis. Significantly, the lack of requirement for membrane potential suggests that tRNAs are actively imported via a route other than the protein import pathway, which has a strict requirement for a membrane potential. Coincidentally, the lack of requirement for membrane potential in *Leishmania* is similar to that of the import of $tRNA^{Gln}$ in yeast and human mitochondria (Rubio et al. 2008). These systems radically differ from that of yeast $tRNA^{Lys}$, which has a strict dependence on membrane potential due to the involvement of the protein import pathway. Despite these differences, it is clear from experiments where mitochondria were pretreated with proteinase K prior to import, that the machinery that imports tRNAs is proteinaceous in nature.

Aside from studies focused on determining positive elements contributing to tRNA import, possible negative import determinants have also been described. Given that posttranscriptional modifications are known to affect tRNA structure, the possibility that nucleotide modifications affect the tRNA in a positive or negative manner is likely. One report suggested that tRNA thiolation of uridines to form 2-thiouridine (s^2U) is a negative determinant for import in *L. tarentolae* (Fig. 5.2; Kaneko et al. 2003; Paris et al. 2009). Using RNA interference (RNAi), we knocked down a key component of the tRNA thiolation pathway in *T. brucei*, Nfs. We showed that although the inhibition of this protein led to a concomitant decrease of $tRNA^{Glu}$ thiolation, it had no effect on the distribution of this tRNA species in vivo or in vitro, suggesting that s^2U is not a negative determinant for tRNA import in *T. brucei* (Paris et al. 2009).

5.4 Concluding Remarks

If history has taught us something it is that with trypanosomatids the only thing that can be expected is often the unexpected. This highlights a more complex truth that reflects the evolutionary position of these medically important single-cell protists, highly divergent from the more commonly studied eukaryotic organisms. In this chapter, we tried our best to highlight recent developments in the field of tRNA biogenesis in trypanosomatids. What should become evident to the reader is that in fact, little is known about most of the events leading to formation of mature tRNAs in these cells, despite much progress made in other systems. One should, however, never discard the importance of digging further into biological processes in organisms that have been a constant source of awesome surprises.

References

Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, Phizicky EM (2006) Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell 21:87–96

Alfonzo JD, Soll D (2009) Mitochondrial tRNA import–the challenge to understand has just begun. Biol Chem 390:717–722

Alfonzo JD, Blanc V, Estevez AM, Rubio MA, Simpson L (1999) C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. EMBO J 18:7056–7062

Anderson J, Phan L, Cuesta R, Carlson BA, Pak M, Asano K, Bjork GR, Tamame M, Hinnebusch AG (1998) The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. Genes Dev 12:3650–3662

Arhin GK, Shen S, Irmer H, Ullu E, Tschudi C (2004) Role of a 300-kilodalton nuclear complex in the maturation of *Trypanosoma brucei* initiator methionyl-tRNA. Eukaryot Cell 3:893–899

Arhin GK, Shen S, Perez IF, Tschudi C, Ullu E (2005) Downregulation of the essential *Trypanosoma brucei* La protein affects accumulation of elongator methionyl-tRNA. Mol Biochem Parasitol 144:104–108

Arts GJ, Kuersten S, Romby P, Ehresmann B, Mattaj IW (1998) The role of exportin-t in selective nuclear export of mature tRNAs. EMBO J 17:7430–7441

Baird NJ, Fang XW, Srividya N, Pan T, Sosnick TR (2007) Folding of a universal ribozyme: the ribonuclease P RNA. Q Rev Biophys 40:113–161

Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46:819–826

Bhattacharyya SN, Mukherjee S, Adhya S (2000) Mutations in a tRNA import signal define distinct receptors at the two membranes of Leishmania mitochondria. Mol Cell Biol 20:7410–7417

Bouzaidi-Tiali N, Aeby E, Charriere F, Pusnik M, Schneider A (2007) Elongation factor 1a mediates the specificity of mitochondrial tRNA import in T. brucei. EMBO J 26:4302–4312

Bruske EI, Sendfeld F, Schneider A (2009) Thiolated tRNAs of *Trypanosoma brucei* are imported into mitochondria and dethiolated after import. J Biol Chem 284:36491–36499

Carrara G, Calandra P, Fruscoloni P, Tocchini-Valentini GP (1995) Two helices plus a linker: a small model substrate for eukaryotic RNase P. Proc Natl Acad Sci U S A 92:2627–2631

Charriere F, Helgadottir S, Horn EK, Soll D, Schneider A (2006) Dual targeting of a single tRNA (Trp) requires two different tryptophanyl-tRNA synthetases in *Trypanosoma brucei*. Proc Natl Acad Sci U S A 103:6847–6852

Cook AG, Fukuhara N, Jinek M, Conti E (2009) Structures of the tRNA export factor in the nuclear and cytosolic states. Nature 461:60–65

Crain PF, Alfonzo JD, Rozenski J, Kapushoc ST, McCloskey JA, Simpson L (2002) Modification of the universally unmodified uridine-33 in a mitochondria-imported edited tRNA and the role of the anticodon arm structure on editing efficiency. RNA 8:752–761

De Robertis EM, Olson MV (1979) Transcription and processing of cloned yeast tyrosine tRNA genes microinjected into frog oocytes. Nature 278:137–143

Dorner M, Altmann M, Paabo S, Morl M (2001) Evidence for import of a lysyl-tRNA into marsupial mitochondria. Mol Biol Cell 12:2688–2698

Engelke DR, Hopper AK (2006) Modified view of tRNA: stability amid sequence diversity. Mol Cell 21:144–145

Esakova O, Krasilnikov AS (2010) Of proteins and RNA: the RNase P/MRP family. RNA 16:1725–1747

Esseiva AC, Naguleswaran A, Hemphill A, Schneider A (2004) Mitochondrial tRNA import in *Toxoplasma gondii*. J Biol Chem 279:42363–42368

Foldynova-Trantirkova S, Paris Z, Sturm NR, Campbell DA, Lukes J (2005) The *Trypanosoma brucei* La protein is a candidate poly(U) shield that impacts spliced leader RNA maturation and tRNA intron removal. Int J Parasitol 35:359–366

Frank DN, Pace NR (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. Annu Rev Biochem 67:153–180

Gaston KW, Rubio MA, Spears JL, Pastar I, Papavasiliou FN, Alfonzo JD (2007) C to U editing at position 32 of the anticodon loop precedes tRNA 5′ leader removal in trypanosomatids. Nucleic Acids Res 35:6740–6749

Gerber AP, Keller W (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science 286:1146–1149

Goswami S, Dhar G, Mukherjee S, Mahata B, Chatterjee S, Home P, Adhya S (2006) A bifunctional tRNA import receptor from Leishmania mitochondria. Proc Natl Acad Sci U S A 103:8354–8359

Gray MW (2003) Diversity and evolution of mitochondrial RNA editing systems. IUBMB Life 55:227–233

Greer CL, Soll D, Willis I (1987) Substrate recognition and identification of splice sites by the tRNA-splicing endonuclease and ligase from *Saccharomyces cerevisiae*. Mol Cell Biol 7:76–84

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35:849–857

Hancock K, Hajduk SL (1990) The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. J Biol Chem 265:19208–19215

Holzmann J, Frank P, Loffler E, Bennett KL, Gerner C, Rossmanith W (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell 135:462–474

Hopper AK, Pai DA, Engelke DR (2010) Cellular dynamics of tRNAs and their genes. FEBS Lett 584:310–317

Hsieh J, Fierke CA (2009) Conformational change in the *Bacillus subtilis* RNase P holoenzyme–pre-tRNA complex enhances substrate affinity and limits cleavage rate. RNA 15:1565–1577

Johnson PF, Abelson J (1983) The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. Nature 302:681–687

Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res 37:159–162

Kaneko T, Suzuki T, Kapushoc ST, Rubio MA, Ghazvini J, Watanabe K, Simpson L (2003) Wobble modification differences and subcellular localization of tRNAs in *Leishmania tarentolae*: implication for tRNA sorting mechanism. EMBO J 22:657–667

Kapushoc ST, Alfonzo JD, Simpson L (2002) Differential localization of nuclear-encoded tRNAs between the cytosol and mitochondrion in *Leishmania tarentolae*. RNA 8:57–68

Kazantsev AV, Krivenko AA, Pace NR (2009) Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. RNA 15:266–276
Leidel S, Pedrioli PG, Bucher T, Brost R, Costanzo M, Schmidt A, Aebersold R, Boone C, Hofmann K, Peter M (2009) Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. Nature 458:228–232
Levinger L, Greene V, Birk A, Bourne R, Kolla S, Whyte S (1995) RNase P and 3′-tRNase processing matrices in the analysis of Drosophila transfer RNA D/T loop tertiary contacts. Nucleic Acids Symp Ser 33:82–84
Levinger L, Morl M, Florentz C (2004) Mitochondrial tRNA 3′ end metabolism and human disease. Nucleic Acids Res 32:5430–5441
Lill R, Muhlenhoff U (2006) Iron-sulfur protein biogenesis in eukaryotes: components and mechanisms. Annu Rev Cell Dev Biol 22:457–486
Lima BD, Simpson L (1996) Sequence-dependent in vivo importation of tRNAs into the mitochondrion of *Leishmania tarentolae*. RNA 2:429–440
Lipowsky G, Bischoff FR, Izaurralde E, Kutay U, Schafer S, Gross HJ, Beier H, Gorlich D (1999) Coordination of tRNA nuclear export with processing of tRNA. RNA 5:539–549
Lithgow T, Schneider A (2010) Evolution of macromolecular import pathways in mitochondria, hydrogenosomes and mitosomes. Philos Trans R Soc Lond B Biol Sci 365:799–817
Lye LF, Chen DH, Suyama Y (1993) Selective import of nuclear-encoded tRNAs into mitochondria of the protozoan *Leishmania tarentolae*. Mol Biochem Parasitol 58:233–245
Mahapatra S, Ghosh T, Adhya S (1994) Import of small RNAs into Leishmania mitochondria in vitro. Nucleic Acids Res 22:3381–3386
Mahapatra S, Ghosh S, Bera SK, Ghosh T, Das A, Adhya S (1998) The D arm of tRNATyr is necessary and sufficient for import into Leishmania mitochondria in vitro. Nucleic Acids Res 26:2037–2041
Marechal-Drouard L, Weil JH, Guillemaut P (1988) Import of several tRNAs from the cytoplasm into the mitochondria in bean *Phaseolus vulgaris*. Nucleic Acids Res 16:4777–4788
Martin RP, Schneller JM, Stahl AJ, Dirheimer G (1979) Import of nuclear deoxyribonucleic acid coded lysine-accepting transfer ribonucleic acid (anticodon C–U–U) into yeast mitochondria. Biochemistry 18:4600–4605
Mayer M, Schiffer S, Marchfelder A (2000) tRNA 3′ processing in plants: nuclear and mitochondrial activities differ. Biochemistry 39:2096–2105
Melton DA, De Robertis EM, Cortese R (1980) Order and intracellular location of the events involved in the maturation of a spliced tRNA. Nature 284:143–148
Mottram JC, Bell SD, Nelson RG, Barry JD (1991) tRNAs of *Trypanosoma brucei*. Unusual gene organization and mitochondrial importation. J Biol Chem 266:18313–18317
Mukherjee S, Basu S, Home P, Dhar G, Adhya S (2007) Necessary and sufficient factors for the import of transfer RNA into the kinetoplast mitochondrion. EMBO Rep 8:589–595
Nashimoto M, Tamura M, Kaspar RL (1999) Selection of cleavage site by mammalian tRNA 3′ processing endoribonuclease. J Mol Biol 287:727–740
Navaratnam N, Morrison JR, Bhattacharya S, Patel D, Funahashi T, Giannoni F, Teng BB, Davidson NO, Scott J (1993) The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem 268:20709–20712
Paris Z, Rubio MA, Lukes J, Alfonzo JD (2009) Mitochondrial tRNA import in *Trypanosoma brucei* is independent of thiolation and the Rieske protein. RNA 15:1398–1406
Phizicky EM, Alfonzo JD (2009) Do all modifications benefit all tRNAs? FEBS Lett 584:265–271
Pusnik M, Charriere F, Maser P, Waller RF, Dagley MJ, Lithgow T, Schneider A (2009) The single mitochondrial porin of *Trypanosoma brucei* is the main metabolite transporter in the outer mitochondrial membrane. Mol Biol Evol 26:671–680
Randau L, Stanley BJ, Kohlway A, Mechta S, Xiong Y, Soll D (2009) A cytidine deaminase edits C to U in transfer RNAs in Archaea. Science 324:657–659
Rinehart J, Krett B, Rubio MA, Alfonzo JD, Soll D (2005) *Saccharomyces cerevisiae* imports the cytosolic pathway for Gln-tRNA synthesis into the mitochondrion. Genes Dev 19:583–592

Rubio MA, Liu X, Yuzawa H, Alfonzo JD, Simpson L (2000) Selective importation of RNA into isolated mitochondria from *Leishmania tarentolae*. RNA 6:988–1003

Rubio MA, Ragone FL, Gaston KW, Ibba M, Alfonzo JD (2006) C to U editing stimulates A to I editing in the anticodon loop of a cytoplasmic threonyl tRNA in *Trypanosoma brucei*. J Biol Chem 281:115–120

Rubio MA, Pastar I, Gaston KW, Ragone FL, Janzen CJ, Cross GA, Papavasiliou FN, Alfonzo JD (2007) An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc Natl Acad Sci U S A 104:7821–7826

Rubio MA, Rinehart JJ, Krett B, Duvezin-Caubet S, Reichert AS, Soll D, Alfonzo JD (2008) Mammalian mitochondria have the innate ability to import tRNAs by a mechanism distinct from protein import. Proc Natl Acad Sci U S A 105:9186–9191

Rusconi CP, Cech TR (1996) The anticodon is the signal sequence for mitochondrial import of glutamine tRNA in Tetrahymena. Genes Dev 10:2870–2880

Salavati R, Panigrahi AK, Stuart KD (2001) Mitochondrial ribonuclease P activity of *Trypanosoma brucei*. Mol Biochem Parasitol 115:109–117

Salinas T, Duchene AM, Delage L, Nilsson S, Glaser E, Zaepfel M, Marechal-Drouard L (2006) The voltage-dependent anion channel, a major component of the tRNA import machinery in plant mitochondria. Proc Natl Acad Sci U S A 103:18362–18367

Salinas T, Duchene AM, Marechal-Drouard L (2008) Recent advances in tRNA mitochondrial import. Trends Biochem Sci 33:320–329

Schiffer S, Helm M, Theobald-Dietrich A, Giege R, Marchfelder A (2001) The plant tRNA 3′ processing enzyme has a broad substrate spectrum. Biochemistry 40:8264–8272

Sherrer RL, Yermovsky-Kammerer AE, Hajduk SL (2003) A sequence motif within trypanosome precursor tRNAs influences abundance and mitochondrial localization. Mol Cell Biol 23:9061–9072

Shi X, Chen DH, Suyama Y (1994) A nuclear tRNA gene cluster in the protozoan *Leishmania tarentolae* and differential distribution of nuclear-encoded tRNAs between the cytosol and mitochondria. Mol Biochem Parasitol 65:23–37

Simpson AM, Suyama Y, Dewes H, Campbell DA, Simpson L (1989) Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also contain small minicircle and maxicircle transcripts of unknown function. Nucleic Acids Res 17:5427–5445

Sprinzl M, Vassilenko KS (2005) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 33(Database Issue):D139–D140

Stange N, Gross HJ, Beier H (1988) Wheat germ splicing endonuclease is highly specific for plant pre-tRNAs. EMBO J 7:3823–3828

Suyama Y (1967) The origins of mitochondrial ribonucleic acids in *Tetrahymena pyriformis*. Biochemistry 6:2829–2839

Suyama Y, Wong S, Campbell DA (1998) Regulated tRNA import in Leishmania mitochondria. Biochim Biophys Acta 1396:138–142

Takaku H, Minagawa A, Takagi M, Nashimoto M (2004) A novel 4-base-recognizing RNA cutter that can remove the single 3′ terminal nucleotides from RNA molecules. Nucleic Acids Res 32:91

Tan TH, Pach R, Crausaz A, Ivens A, Schneider A (2002) tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. Mol Cell Biol 22:3707–3717

Vogel A, Schilling O, Spath B, Marchfelder A (2005) The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. Biol Chem 386:1253–1264

Wohlgamuth-Benedum JM, Rubio MA, Paris Z, Long S, Poliak P, Lukes J, Alfonzo JD (2009) Thiolation controls cytoplasmic tRNA stability and acts as a negative determinant for tRNA editing in mitochondria. J Biol Chem 284:23947–23953

Wolin SL, Cedervall T (2002) The La protein. Annu Rev Biochem 71:375–403

Xiao S, Scott F, Fierke CA, Engelke DR (2002) Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. Annu Rev Biochem 71:165–189

Xiong Y, Steitz TA (2006) A story with a good ending: tRNA 3′-end maturation by CCA-adding enzymes. Curr Opin Struct Biol 16:12–17
Yoo CJ, Wolin SL (1997) The yeast La protein is required for the 3′ endonucleolytic cleavage that matures tRNA precursors. Cell 89:393–402
Yuan Y, Altman S (1995) Substrate recognition by human RNase P: identification of small, model substrates for the enzyme. EMBO J 14:159–168

Chapter 6
rRNA Biogenesis in Trypanosomes

Shulamit Michaeli

Abstract rRNA processing is a complex and essential process that starts in the nucleolus, continues in the nucleoplasm, and is completed in the cytoplasm. The process involves the concerted action of the small nucleolar RNAs that direct cleavage at distinct sites in the intergenic spaces and modify the rRNA by 2′-*O*-methylation and pseudouridylation. The process also requires protein factors such as endonucleases, ATPases, GTPases, and helicases that mediate the precise cleavage of the pre-rRNA in the intergenic regions and ensure the correct assembly of the two ribosomal subunits. In this chapter, I describe unique properties of this process in trypanosomes. As opposed to most eukaryotes, the large subunit of the rRNA is cleaved into two large subunits and additional small rRNA fragments. The unique repertoire of small nucleolar RNAs (snoRNAs) is described, emphasizing the lack of conventional snoRNAs involved in rRNA processing and the presence of trypanosome-specific snoRNAs. Finally, the protein factors involved in this process are described, focusing on factors whose function was elucidated in trypanosomes, and on a bioinformatics survey to detect protein factors involved in this process in other eukaryotes.

Keywords Large subunit rRNA • Processosome • Small nucleolar RNAs • Small rRNA • Small subunit rRNA • Trypanosomes

S. Michaeli (✉)
The Mina and Everard Goodman Faculty of Life Sciences, and Advanced Materials and Nanotechnology Institute, Bar-Ilan University, Ramat-Gan 52900, Israel
e-mail: michaes@mail.biu.ac.il

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_6,

6.1 Trypanosome rRNA Processing is Unique: Cleavage of a Large Subunit to Small rRNA Fragments

rRNA processing is a complex process that involves several small nucleolar RNAs (snoRNAs) and hundreds of factors including enzymes such as endo- and exonucleases, and a variety of factors that control this complex process, such as AAA-type ATPases, helicases, and GTPases. The processing of rRNA starts in the nucleolus, continues in the nucleoplasm, and is completed in the cytoplasm (Tschochner and Hurt 2003). The nucleolus is compartmentalized into three sections: the fibrillar center (FC), dense fibrillar component (DFC), and granular component (GC). Pre-RNA is first localized to the interspaces between the FC and DFC and then moves to the DFC where rRNA processing begins, and continues in GC. SnoRNP proteins such as fibrillarin are localized in DFC, but factors involved in late stages of processing are confined to the GC (Puvion-Dutilleul et al. 1991).

In eukaryotes, such as yeast and mammals the pre-rRNA is transcribed as a long precursor that is composed of external and internal transcribed spacers (ETS and ITS, respectively). The precursor has the following structure: 5′-ETS/SSU/ITS1/5.8S/ITS2/LSU/3′-ETS. The SSU encodes for the 18S small subunit and the LSU for the 28S rRNA large subunit rRNA. rRNA processing starts with cleavages in the 5′ ETS at position A0, followed by consecutive cleavages at sites A1 and A2 (ITS1), which removes the 5′ external transcribed spacer (5′ ETS) and separates the 18S precursor (20S) from the 27S precursor, which is composed of 25S and 5.8S precursors. In yeast, in 85% of the cases, the 27SA2 pre-rRNA is cleaved at A3 located in the ITS1 by MRP (Fig. 6.1a). In 15% of the cases, the 27SA2 precursor is cleaved at position B1. The last step of LSU processing involves cleavage at positions C1 and C2, located at the ITS2, to generate the 5.8S and 25S rRNA (Eichler and Craig 1994; Venema and Tollervey 1999). The pre-rRNA SSU is transported to the cytoplasm where it is further cleaved at site D, yielding the mature 18S rRNA molecule (Fig. 6.1a).

In trypanosomes, the LSU is uniquely cleaved, and these cleavages yield two large fragments of 1,900 nt (LSUα) and 1,600 nt (LSUβ), as well as four discrete small ribosomal RNAs (srRNA) of about 70–220 nt (Campbell et al. 1987; Cordingley and Turner 1980; Schnare et al. 1983; White et al. 1986). A similar multiple fragmentation of rRNA also occurs in *Euglena gracilis*, in which the rRNA is fragmented into 16 discrete RNA components (Schnare and Gray 1990).

In *T. brucei*, three predominant rRNA processing intermediates were identified: ETS-18S-ITS1 (3.7 Kb), 5.8S-ITS2-28S (5.9 Kb), and a 5.1 Kb species containing pre-28S (Fig. 6.1b). The first cleavage of the 9.7 Kb pre-rRNA takes place at B1, which is located upstream of 5.8S; then, cleavages take place at sites A', A0, A1, and A2 in the ETS (Hartshorne and Toyofuku 1999; Hartshorne et al. 2001). The pre-SSU (3.7 Kb) is further processed to yield the mature 2.2 Kb SSU. Although the LSU cleavages that form the two large subunits and the small RNA species (srRNAs) were reported almost 2 decades ago, relatively little is known about the machinery that mediates these cleavages or about their biological role. However,

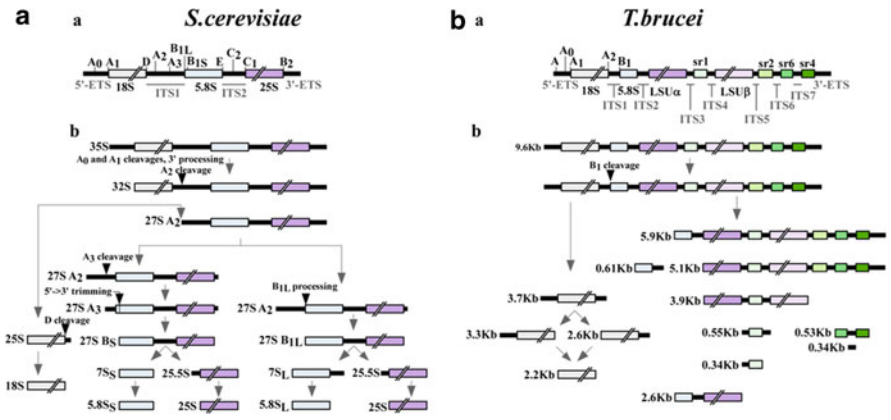


Fig. 6.1 Schematic representation of the rRNA processing steps in yeast and trypanosomes. (**a**) The rRNA processing scheme, derived from data presented by Venema and Tollervey (1999). (**b**) The *T. brucei* schemes are derived from data presented by Barth et al. (2008), Hartshorne (1998); Hartshorne and Toyofuku (1999), and Hartshorne et al. (2001). In both panels, the pre-rRNA of the full-sized rRNA precursor is presented on *top*. The location of the cleavage sites with the ITS and ETS are indicated. Below, the different intermediates and products generated during processing are shown. The sizes of the precursors are given in Kb, or as the S value

since these cleavages are trypanosome-specific, they must be mediated by parasite-specific functions. Only in recent years have we begun to understand these complex processes.

6.2 The Conserved snoRNAs Involved in rRNA Processing in Other Eukaryotes and Their Presence in Trypanosomes

Small nucleolar RNAs (snoRNAs) are known to participate in the pre-rRNA cleavages. The majority of these snoRNAs belong to C/D and H/ACA RNA classes (see below). Five major snoRNAs were reported to function in the processing of LSU: U3, and U14, which are present in both higher and lower eukaryotes, U22, U17 (snR30 in yeast), and snR10, which is only present in yeast (Venema and Tollervey 1999).

U3 snoRNP is believed to be the first to bind to nascent pre-rRNA transcript, near the 5′ end, and it is thought to be critical in organizing the active processing complex (Dragon et al. 2002). This snoRNP is more abundant and larger than most of the snoRNAs, has an atypical structure, and it contains more proteins than all other snoRNPs. In yeast, U3 was shown to be essential for cleavage at the A0 site within the 5′-ETS, followed by A1 at the 5′ETS/18S rRNA boundary, and then A2, within ITS1 (Venema and Tollervey 1999). A0 cleavage occurs near the 3′ end of the 5′-ETS, only 90 nt upstream of the A1 cleavage that forms the mature end of SSU.

The vertebrate primary pre-rRNA cleavage occurs near the 5′ end of the 5′ETS (Kass and Sollner-Webb 1990). In mammals, U3 is required for cleavage at site A0 (5′ end of 18S), site A1 (3′ end of 18S), and site A2 (5′ end of the 5.8S RNA) (Kass and Sollner-Webb 1990). It was proposed that a single-stranded 5′ hinge and 3′ hinge, which separate domains I and II of U3 snoRNA, base pair with ETS, thus docking U3 snRNPs on the pre-rRNA substrate (Savino and Gerbi 1990). Indeed, direct contacts between the 5′ domain of U3, including conserved box A and the variable hinge region, and ETS sequences were demonstrated, using psoralen crosslinking in yeast and mammalian cells (Beltrame and Tollervey 1992; Tyc and Steitz 1992). The interaction of U3 with the ETS sequence was shown to be essential for cleavages at the A0, A1, and A2 sites (Mereau et al. 1997).

Trypanosomes possess a U3 homologue that contains the conserved boxes typical of this RNA in other eukaryotes including the A box (Hartshorne and Agabian 1993). This gene is transcribed by RNA polymerase III from an extragenic tRNA promoter (Nakaar et al. 1994). The primary trypanosomatid rRNA transcript is first cleaved at the ITS1/5.8S boundary, and then the 5′ ETS sequence is removed by consecutive cleavages at A', A0, and A1 at the 5′ ETS/SSU junction (Fig. 6.1b). Using psoralen crosslinking, it was demonstrated that the trypanosome U3 interacts with ETS sequences (Hartshorne 1998). One of the interaction sites is located downstream of A' (s1a and s1b) and one is near A0 (s3) (depicted in Fig. 6.1b). Uniquely, in *T. brucei*, two U3-crosslinkable 5′ ETS sites were shown to be essential for SSU rRNA production, using mutations introduced to a tagged rRNA cluster (Hartshorne et al. 2001). Site 1b is novel and located 3′ to the A' site, and site 3 lies upstream of A0, in a position analogous to the yeast U3-binding site. Although the site near A' or site 3 (A0) have no effect on A' cleavage, each of these sites is required for A0 and A1 cleavage, and SSU production (Hartshorne and Toyofuku 1999; Hartshorne et al. 2001). Evidence suggests that A'-cleavage-site-associated factors stabilize U3-binding to the adjacent site1b element, which may work together with the downstream U3 site3 (A0) to promote pre-SSU processing. These data suggest that processing of 5′ ETS is a three-step process, versus the two-step process noted in yeast. Inspection of processing intermediates indicates that A' cleavage precedes that at A0, followed by A1 cleavage (Hartshorne and Toyofuku 1999; Hartshorne et al. 2001). This might suggest that the U3-associated complex assembled at A' translocates to the A0 region, or alternatively, that individual complexes assembled at each site communicate with one another to coordinate the ordered pre-SSU rRNA cleavages (Hartshorne and Toyofuku 1999; Hartshorne et al. 2001).

snR30—also known as U17—belongs to the H/ACA family and is essential for SSU processing. This snoRNA is essential for life in yeast (Bally et al. 1988; Tollervey 1987). Depletion of snR30 in yeast abolished cleavages at A0, A1, and A2, leading to reduction in the mature 18S rRNA (Morrissey and Tollervey 1993). snR30 possesses distinct structural motifs such as m1 and m2 sequences located in the 3′ hairpin, which are essential for its function in 18S processing (Atzorn et al. 2004). snR30, like many other snoRNAs, associates tightly with its rRNA target; the helicase Rok1 is needed to detach the RNA from its target (Bohnsack et al. 2008).

The trypanosome snR30 (Barth et al. 2005), like its homologues, is composed of three stem-loop structures. The H box located between the 5′ hairpin and the internal loop deviates from the canonical sequence (ANANNA). This is the first and maybe the only H/ACA RNA in trypanosomes that is composed of three hairpin loops. In addition, as opposed to all trypanosome H/ACA RNAs that contain an AGA box at their 3′ end (Liang et al. 2005), snR30 contains an ACA box. The RNA is much shorter than its yeast homologue (270 nt compared to 600 nt in yeast). As expected from a bona fide H/ACA RNA, the snR30 level is reduced in CBF5-silenced cells and the RNA binds to the H/ACA core protein, NHP2 (Barth et al. 2005). However, functional data are still lacking demonstrating the role of the trypanosome homologue in rRNA processing at sites AO, A1, and A2 in analogy to the role of this snoRNA in other eukaryotes.

Two other conserved snoRNAs involved in SSU processing are U14 and U22. U14 contains two distinct domains, A and B. Domain A but not domain B is essential for rRNA processing at the A1 and A2 sites (Jarmolowski et al. 1990; Li et al. 1990; Liang et al. 1997; Liang and Fournier 1995). Another characteristic domain of this RNA is the Y domain, situated between the A and B boxes and interacting with pre-rRNA by base-pairing (Morrissey and Tollervey 1997). The U14 snoRNA harbors complementarity to both pre-rRNA and 18S rRNA for guiding methylation, and thus has a dual role in ribosome synthesis. Bioinformatic searches failed to reveal a trypanosome homologue of U14 carrying these very distinct and characteristic domains. However, TB11Cs2C1 (see below) might be a functional homologue of U14, based on its interaction with SSU and pre-rRNA sequences upstream of A0 (Gupta et al. 2010). U22 is another snoRNA essential for SSU processing. Depletion of U22 in *Xenopus* leads to accumulation of SSU defective at both its 5′ and 3′ ends (Tycowski et al. 1994). A U22 homologue was also not detected in the trypanosomatid genomes.

Two major snoRNAs were shown to participate in LSU processing, MRP and U8. MRP directs cleavage at A3 situated in ITS1. Nine proteins are bound to this RNA, and eight are shared with RNaseP, which is involved in pre-tRNA cleavage (Venema and Tollervey 1999). Although cleavage to release 5.8S at the B1 site can circumvent the cleavage by MRP, MRP is essential in yeast (Venema and Tollervey 1999). A trypanosome homologue to MRP was identified (Barth et al. 2008). In trypanosomes, the cleavage at the B1 site was not affected in cells depleted of either C/D and H/ACA, suggesting that MRP might be involved in this cleavage (Barth et al. 2005; Barth et al. 2008). The MRP homologue (521 nt) is divided into two major domains. MRP carries the canonical conserved region (CR); its CR-IV domain contains the three conserved nts found in all MRP RNAs, but its CR-V domain is less conserved. The stem-loops of the second domain are related to equivalent domains in other MRPs, but the trypanosome RNA carries an additional domain (P10) and lacks the subdomain eP19. No functional data exist to demonstrate the interaction of MRP with ITS1 or prove that this RNA cleaves at the ITS1 site (Barth et al. 2008).

The other snoRNA that functions in LSU processing is U8. U8 is involved in cleavage at ITS2 and has the potential to interact both with ITS2 and 5.8S rRNA.

These two sites simultaneously bind to the snoRNA (Morrissey and Tollervey 1997). In *Xenopus*, antisense RNA directed to U8 inactivated the function of the molecule, resulting in defects in cleavage between LSU and 5.8S (Peculis 1997; Peculis and Steitz 1994). No homologue of U8 was found in trypanosomes. However, we have recently identified a snoRNA (TB10Cs4C4) that mediates cleavage at ITS2 (see below). This RNA, although lacking homology to U8, might serve a similar function.

In yeast, snR10 is involved in 25S processing, and its disruption delays cleavages at A1 and A2 as well as pseudouridylation at position 2919 in 25S rRNA (Venema and Tollervey 1999). So far, no homologue of this RNA was found in trypanosomes.

The data that accumulated thus far suggest that trypanosomes possess a few of the highly conserved snoRNAs such as U3, MRP, and snR30, but lack snoRNAs that were shown to function in rRNA processing in other eukaryotes, such as U14, U8, and snR10, suggesting that trypanosomes may possess species-specific functional homologues of these RNAs.

6.3 Trypanosome-Specific snoRNAs Involved in rRNA Processing

The failure to detect the most conserved snoRNAs except U3, snR30, and MRP, and the unique processing events involved in LSU processing suggest that trypanosomes may possess trypanosome-specific snoRNAs (Fig. 6.2). To evaluate the role of snoRNAs in this process, the defects in rRNA processing were examined under silencing of NOP1 and NOP58, the C/D core proteins (Barth et al. 2008; Gupta et al. 2010), and the H/ACA core protein, CBF5 (Barth et al. 2005). The results indicate that under NOP58 depletion, accumulation of 5′ ETS and all the ITSs were detected, suggesting that C/D snoRNAs operate in the cleavages of SSU, the two large LSU species, and the small rRNA fragments 1, 2, 4, and 6. Inspection of the rRNA precursors that accumulate during NOP1 silencing demonstrated massive accumulation of the unprocessed precursor (9.6 Kb), a precursor comprising 5.8S and LSU (5.9 Kb), a precursor carrying only LSU and missing 5.8S (5.1 Kb), as well as a 3.9 Kb precursor that separates the LSUα from sRNA-2, -4, and -6 by cleavage at ITS5. In addition, precursors were detected covering the srRNA domains; these include one comprised of sRNA 6 and 4 (0.53 Kb) and only ITS7 (0.34 Kb), as well as srRNA-1 with its flanking sequences (0.55 Kb), or srRNA-1 with only the ITS3 (0.34 Kb) (see Fig. 6.1b). Depletion of CBF5 affected the processing of srRNA-2, -4, and -6. In addition, it affected cleavages at positions A0, A1, and A2, most probably due to a decreased level of snR30, which directs the cleavage at these positions. However, our data strongly suggest that no H/ACA snoRNA exists to mediate cleavages to liberate srRNA-1 (in ITS3 and ITS4) but should possess snoRNA (s) that mediate clevage(s) at ITS5, which is essential for

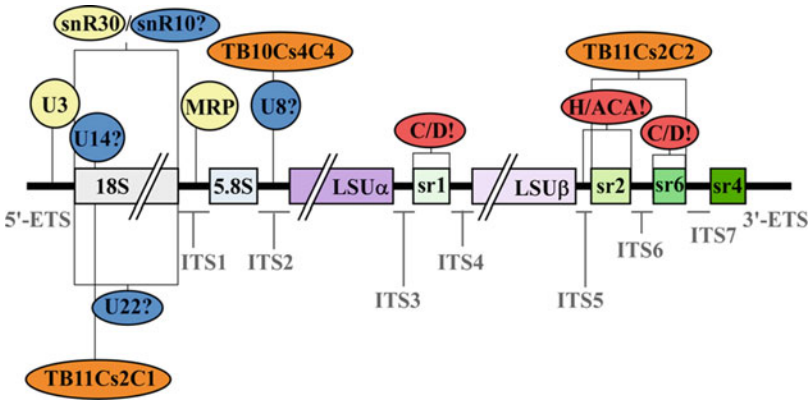


Fig. 6.2 The unique repertoire of *T. brucei* snoRNA involved in rRNA processing. The location of cleavages by the snoRNAs that were identified in trypanosomes, such as U3 (Hartshorne 1998; Hartshorne and Toyofuku 1999; Hartshorne et al. 2001), snRN30 (Barth et al. 2005), TB11Cs2C1 and TB11Cs2C2 (Gupta et al. 2010), and TB10Cs4C4, TB6Cs1C3, and TB9Cs2C1 (Michaeli et al. 2012). The prediction of the snoRNAs involved in rRNA processing is based on data presented by Barth et al. (2005, 2008). The positions where C/D or H/ACA are expected to function are marked with exclamation points; the snoRNAs for which no homologues were identified in trypanosomes are marked with a *question mark*

releasing srRNA-2, -6, and -4 (Barth et al. 2005). In contrast, C/D snoRNA most probably operates to separate the two LSU subunits, and thus affects cleavages at ITS3 and ITS4 (Fig. 6.2).

The presence of trypanosome snoRNAs that specifically direct cleavages at the trypanosome-specific sites was recently supported by the description of two novel snoRNAs (TB11Cs2C1 and C2). These snoRNAs are among the most abundant snoRNAs in the cell and are located in the same gene cluster as the SLA1 and snR30 (Gupta et al. 2010; Roberts et al. 1996; Roberts et al. 1998). In vivo crosslinking and bioinformatic predictions suggest that TB11Cs2C1 (76 nt) interacts both with the SSU (at the 3′ end, U1900) and at a position upstream of A0 in the ETS. Indeed, two different crosslinked sites were mapped on the snoRNA, one corresponding to interaction with SSU (position U22) and one with ETS (position U32 and U35). Although there is no structural resemblance between TB11Cs2C1 and U14, TB11Cs2C1 may nevertheless function as U14 ortholog because U14 interacts with 18S rRNA, and this RNA is essential for SSU processing (Li et al. 1990). The highly abundant snoRNA present in this cluster is TB11Cs2C2, whose function was analyzed by snoRNAi. snoRNAi is induced by expressing dsRNA leading to knockdown of the RNA (see below). Depletion of TB11Cs2C2 resulted in specific accumulation of ITS5 and ITS7, but no accumulation of ITS6 was observed. No major changes were observed in the level of the other ITSs, suggesting that TB11Cs2C2 functions in specific processing events required to release srRNA-2 and -6. Psoralen-UV crosslinking supported the interaction between the snoRNA and srRNA-2 and -6, suggesting that this snoRNA functions in cleavages that release the trypanosome-specific srRNAs (Gupta et al. 2010).

RNA-seq performed recently on small RNP particles of *T. brucei* identified highly not only abundant snoRNAs including the SLA1, snR30, TB11CsC2C1, and C2 species, but also additional snoRNAs imbedded within the known snoRNA clusters (Michaeli et al. 2012). snoRNAi of one of these abundant RNAs (TB10Cs4C4) demonstrated that this RNA functions in cleavage at ITS2, like U8. Cumulative data therefore support the notion that trypanosomes possess unique snoRNAs to replace the highly conserved snoRNAs that mediate the same cleavages in other eukaryotes. Thus, U14 might be replaced by TB11Cs2C1 and U8 by TB10Cs2C1. Trypanosomes also possess trypanosome-specific snoRNAs that carry out the trypanosome-specific cleavages. One such example is TB11Cs2C2, which is involved in the cleavages of srRNA-2 and -6. However, other snoRNAs must exist to perform cleavages to release sRNA-1, and to cleave downstream of LSUβ, and upstream and downstream of srRNA-4. RNA-seq data identified 12 abundant snoRNAs including SLA1, snR30, TB11Cs2C1, TB11Cs2C1, and TB11Cs2C2. Interestingly, only two of these 12 abundant snoRNAs represent H/ACA RNAs. Indeed, most of yeast snoRNAs involved in rRNA processing also belong to the C/D family. Studies are in progress to decipher the function of these seven additional abundant snoRNAs, using snoRNAi and psoralen-UV-crosslinking mapping. We performed mapping of snoRNA interactions using the recently described "RNA walk" method (Lustig et al. 2010). This method was used to map the interaction of TB11Cs2C1/C2 with its rRNA targets (Gupta et al. 2010) and is based on in vivo crosslinking with AMT-psoralen, followed by affinity selection of the snoRNA–rRNA duplex and mapping of the crosslink sites by primer extension. The same methodology in conjunction with snoRNAi was recently used to decipher the function of three additional snoRNAs: TB10Cs4C4, TB6Cs1C3, and TB9Cs2C1 (Michaeli et al. 2012). Given the bioinformatic prediction and the rRNA defects observed under silencing, TB10Cs4C4 is most probably involved in cleavage at ITS2, thus separating 5.8S from LSU rRNA. TB10Cs4C4 might therefore represent a functional homologue of U8 that cleaves within ITS2. TB6Cs1C3 was shown to direct cleavage at ITS6, separating srRNA-2 and srRNA-6. TB9Cs2C1 was shown to be involved in cleavage at the boundaries of LSUα and LSUβ as well as ITS6 [depicted in Fig. 6.2; Michaeli et al. (2012)]. The exact role of the additional abundant snoRNA described in Michaeli et al. (2012) is currently under investigation.

6.4 snoRNAs in rRNA Modification: C/D snoRNAs

The two major modifications on rRNA are 2′-*O*-methylation and pseudouridylation (Decatur and Fournier 2003). In eukaryotes and *Archaea,* the modifications are guided by two types of small RNAs that dictate the exact sites of modification by the formation of a specific duplex with their large RNA partners. The C/D snoRNAs that guide 2′-*O*-methylation are named after short sequence motifs, the C box (RUGAUGA; R designates any purine) and the D box (CUGA). These boxes are short sequences near the 5′and 3′ end of the snoRNA, that are essential for processing, localization, and function of the snoRNA (Watkins et al. 2002). It

was found that a region of perfect complementarity (10–21 nt) between the target RNA and the snoRNA lies upstream of the D box. The methylated nucleotide is always located 5 nt upstream from the D box. This is known as the +5 rule (Kiss-Laszlo et al. 1996). The snoRNA usually contains two C and D boxes, and hence has the potential to guide the modifications on two sites. C/D snoRNAs are present in stable complexes formed together with four core proteins: fibrillarin/NOP1, NOP56, NOP58, and SNU13. These proteins were identified in *T. brucei* and are highly homologous to their counterparts in other eukaryotes (Barth et al. 2008; Dunbar et al. 2000b). The association of fibrillarin with C/D guide RNAs depends on a bridge formed by the other two core proteins, SNU13 and NOP56/58. SNU13 nucleates the assembly of C/D RNP. SNU13 binds the K-turn motif in the C/D guide RNAs formed by interaction of the C and D boxes. This interaction forms the binding site for NOP56/58 protein. NOP56/58 in turn recruits the catalytic protein, fibrillarin, to complete the assembly of the functional complex (Rashid et al. 2003). The trypanosome C/D snoRNAs have the potential to form the K turns (A–G and U–U pairing), which is important for SNU13 binding (Liang et al. 2007a).

Early studies suggested the existence of ~100 2′-*O*-methylation sites in trypanosomatids (Gray 1979). The first trypanosome C/D snoRNA was discovered in *Leptomonas* (Levitan et al. 1998). Later, snoRNAs and reiterated gene clusters encoding C/D snoRNA species were identified in *T. brucei* (Dunbar et al. 2000a). The first whole-genome search conducted in *T. brucei* identified 57 C/D genes (Liang et al. 2005), and 63 C/D genes were identified in *Leishmania major* (Liang et al. 2007a), with snoRNAs ranging in size between 67 and 118 nt. The only unique property of trypanosome C/D snoRNAs is that many of them are double-guiders and can potentially guide adjacent sites on rRNA. However, mapping the 2′-*O*-methylation sites (Nms) on rRNA by primer extension identified at least 131 Nms in *T. brucei* (Barth et al. 2008), suggesting that at least 65 C/D snoRNAs must exist in the *T. brucei* genome. A study from *Euglena gracilis*, which is closely related to trypanosomes, suggested the existence of 200 Nms on its rRNA, indicating the existence of even more C/D snoRNA species in this organism (Russell et al. 2006). So far, 79 C/D snoRNAs—with the potential to guide ~150 Nms—were identified in *T. brucei* (Michaeli et al. 2012). However, not all the newly described snoRNAs have known rRNA targets, and additional snoRNAs must therefore be present in the genome to guide the previously described modifications. About 40 Nms remain, for which we failed to find the corresponding snoRNAs, suggesting that additional snoRNAs are likely to exist. Interestingly, ~40% of the Nms are species-specific modifications that do not have counterparts in yeast, humans, or plants, and 40% of species-specific predicted modifications are located in unique positions outside the highly conserved richly modified domains in other eukaryotes (Liang et al. 2005). The location of 84 modifications potentially guided by the snoRNAs was compared to these modifications in yeast, humans, and plants. Among these modifications, 44 predicted sites are modified in at least one additional organism, and of these modifications, 23 were found in all organisms. These results suggest that many of the modifications on the trypanosome rRNA are species-specific. These modifications are clustered together in the most conserved structural domains. In

several cases, such as TB5Cs1C1, TB3Cs2C1, and TB8Cs2C1, the same snoRNA can direct methylation on adjacent sites (Liang et al. 2005). The increased number of modifications in the conserved functional domains may stabilize the ribosome and help it to function even under adverse conditions. Indeed, modifications are clustered in domains where specific translation events take place. Recent data from yeast suggest that blocking one or two modifications in the same rRNA domain does not significantly affect the cells, but loss of three to five modifications, especially in domains such as helix69, impairs growth and results in decreased amino acid incorporation, increased stop codon readthrough, and reduced rRNA level due to faster turnover of rRNA (Liang et al. 2007b).

One characteristic feature of trypanosome rRNA modifications and their guide RNAs is that, in contrast to the yeast and human genomes, the number of predicted Nms exceeds the number of pseudouridines. Hypermethylation exists in plants and thermophiles, which helps sustain ribosome function at high temperatures (Dennis et al. 2001). Similarly, the hypermethylation in the parasite may enable it to cope with the temperature shifts during the cycling between the procyclic stage (26°C) and the bloodstream form of the parasite (37°C). Indeed, we demonstrated higher expression of C/D snoRNAs in the bloodstream form, possibly leading to elevated methylation during this life-cycle stage, and thereby supporting the notion that hypermethylation may be an important determinant of ribosome function at elevated temperature (Barth et al. 2008). Interestingly, although modifications and the corresponding snoRNAs that guide these were elevated in bloodstream from, no modifications were detected that are exclusively stage-specific (Barth et al. 2008). RNA-seq analysis of snoRNAs in the two life-cycle stages is in progress to identify snoRNAs that are differentially regulated. Note, however, that despite the correlation discussed above between Nm modifications and ribosome function at high temperature, in several organisms not undergoing these temperature shifts the number of C/D snoRNAs also exceeds that of H/ACA snoRNAs (Dieci et al. 2009).

6.5 H/ACA snoRNAs

The snoRNA species that guide pseudouridylation consist in most eukaryotes of two hairpin domains connected by a single-stranded hinge, the H (AnAnnA) domain, and a tail region, the ACA box. Two short rRNA recognition motifs of the soRNA base-pair with rRNA sequences flanking the uridine to be converted to pseudouridine (Ganot et al. 1997; Tollervey and Kiss 1997). The pseudouridine is always located 14–16 nt upstream of the H box or ACA box of the snoRNA.

Whereas the trypanosome C/D snoRNA fits the consensus, the trypanosome H/ACA possesses unique features. All the snoRNAs (except snR30) are single-hairpin RNAs, and in addition, these guide RNAs carry an AGA rather than an ACA box (Liang et al. 2001; Liang et al. 2007b; Liang et al. 2005; Liang et al. 2002; Uliel et al. 2004). The H/ACA are smaller in size compared to C/D, ranging from 57 to 78 nt. After the discovery of the single-hairpin H/ACA in trypanosomes, such

molecules were also found in *Archaea* (Rozhdestvensky et al. 2003; Tang et al. 2002) and in *Euglena* (Russell et al. 2004). The single-hairpin may represent the primordial H/ACA structure, since it is common to *Archeae* and the lower eukaryotes. It is likely that the single-hairpin fused only later in evolution. During this process, the ANA box was converted to an H box. The finding that each hairpin of H/ACA binds autonomously to the four core proteins supports the fusion hypothesis. Interestingly, however, *Giardia*, which is a very ancient eukaryote, possesses double-hairpin RNAs (Yang et al. 2005).

All trypanosome H/ACA RNA except snR30 carry the AGA box. Mutations introduced into the AGA sequence showed the absolute requirement for this box for the stability of the RNA (Myslyuk et al. 2008). To find H/ACA snoRNAs in the trypanosome genome, an algorithm that was termed PsiScan was developed to perform a genome-wide search for H/ACA snoRNAs in trypanosome genomes. The program selected sequences conserved among three species (*T. brucei*, *T. cruzi*, *L. major*). In the next step, a minimal-energy calculation was used to select molecules predicted to fold into the lower-energy structure of a single stem-loop structure, and in the last step, a support vector machine learning algorithm was used for prediction. Known H/ACA molecules were used as positive examples, and negative examples were molecules identified by the computational analysis, but that were shown experimentally not to be H/ACA RNAs. This program increased the repertoire by identifying 14 additional H/ACA RNAs (Myslyuk et al. 2008). Recent RNA-seq revealed a total of 63 H/ACA molecules in *T. brucei*. Many of these snoRNAs were not detected in the genome-wide analysis (Liang et al. 2005), because the initial screen for H/ACA was based on the genomic co-localization of this RNA in clusters with C/D snoRNAs (Liang et al. 2005). The recent analysis identified clusters that contain only H/ACA molecules, as well as "solitary" H/ACA molecules (Michaeli et al. 2012). Since no pseudouridine sites have been mapped so far in trypanosomes, the extent of this modification will be determined by describing the complete repertoire of these guide RNAs.

Four proteins bind to the H/ACA RNA. The centromere binding protein factor-5 (CBF5 in yeast known as dyskerin in humans) is a pseudouridine synthase with a catalytic and a PUA (pseudouridine and archaeosine transglycosylase) domain. Both ACA and the lower stem of the guide RNA are bound by a PAU domain, anchoring the antisense elements near the catalytic site (Rashid et al. 2006). Silencing of *T. brucei* CBF5 resulted in marked reduction in the level of H/ACA RNA (Barth et al. 2005). Other proteins that bind to the particles are GAR1, NOP10, and NHP2 (or L7Ae in *Archaea*). Current evidence from the *Archaea* complex indicates that GAR1 is involved in binding and/or release of the target RNA. NOP10 is positioned along the upper stem of the guide RNA between CBF5 and NHP2, and appears to interact with both NHP2 and the guide RNA. NHP2 binds the kink-turn (k-turn) motif located in the upper stem-apical loop of the guide H/ACA and induces a major bend (or kink) in the RNA (Li and Ye 2006; Rashid et al. 2006). Although we could not find conserved k-turn motifs in the trypanosome snoRNA mutations introduced into the *L. collosoma*, H/ACA

indicated that the size and the structure of the apical loop cannot tolerate changes (Myslyuk et al. 2008).

6.6 Genomic Organization and Processing of snoRNAs

In eukaryotes from yeast to vertebrate, all modification-guiding snoRNAs are transcribed by RNA polymerase II, whereas snoRNA involved in rRNA processing can also be transcribed by RNA polymerase II (Matera et al. 2007). Vertebrates, plants, and yeast all contain independently transcribed snoRNA genes flanked by promoter, enhancer, and termination sequences (Brown et al. 2003; Matera et al. 2007). However, snoRNAs have also been identified in introns, and these are transcribed from the host gene promoter. In vertebrates and yeast, having only a single snoRNA in any intron, the processing is splicing-dependent and is mediated by exonucleases after debranching of the lariat (Filipowicz and Pogacic 2002). In yeast, polycistronic snoRNAs are processed by the RNase III homologue, Rnt1p, which cuts duplex RNA carrying an AGNN tetraloop (Chanfreau et al. 1999). In plants, many snoRNAs are transcribed as polycistronic units that are processed by the RNase III enzyme and by 5′ and 3′ exonucleases (Chekanova et al. 2002).

All the trypanosome snoRNAs studied so far are organized in clusters and are transcribed by polymerase II (Xu et al. 2001), except for U3, which is transcribed by polymerase III (Nakaar et al. 1994). The genome organization of trypanosome snoRNAs resembles that in plants (Brown et al. 2003), because the genes are clustered and each cluster carries a mixture of both C/D and H/ACA RNAs (Liang et al. 2007a; Liang et al. 2005). Initially, 21 clusters were identified in *T. brucei* (Liang et al. 2005). The recent RNA-seq data expanded the repertoire and 50 snoRNA clusters were revealed. Although most clusters carry both C/D and H/ACA, clusters carrying only C/D or only H/ACA snoRNAs were identified as well, and not all the clusters were found to be repeated (Liang et al. 2005; Michaeli et al. 2012). In *Leishmania*, 23 clusters were revealed that code for 62 C/D snoRNAs that can potentially guide 79 methylations, and 37 H/ACA RNAs were identified that can direct 30 pseudouridylations on rRNA (Liang et al. 2007a).

In *Leptomonas collosoma*, the signals for snoRNA processing are located in the intergenic region, which has the potential to form long stems. In expression studies, it was demonstrated that flanking sequences of at least 10 nt are essential for efficient processing (Xu et al. 2001). Studies from mammals and yeast suggest that base-pairing between the 5′ and 3′ terminal stem of the C/D snoRNA-coding sequence is required for processing and accumulation of these RNAs (Huang et al. 1992). However, many intron-encoded box C/D snoRNAs expressed in mammalian cells or polycistronic snoRNAs in yeast lack the canonical 5′–3′-terminal helix (Darzacq and Kiss 2000). Studies on these snoRNAs indicated that their processing is supported by external intronic stem structures that are fully or partially degraded during exonucleolytic cleavage. The trypanosome C/D snoRNAs resemble such

snoRNAs, lacking the 5′–3′ stem, but instead possessing extragenic flanking helices. These long stems were observed mainly in *L. collosoma* (Xu et al. 2001) and in *L. major* (Liang et al. 2007a). Interestingly, although the external stems are thermodynamically favorable, no significant correlation can be found between the abundance of the snoRNA and the stability of the external duplex. Other factors such as the conservation of the C and D boxes and the presence of a canonical K-turn play a major role in snoRNA processing and stability (Liang et al. 2007a). A machine learning approach was used to predict the parameters that affect snoRNA abundance. It was found that in C/D snoRNAs several factors such as conservation of the C box, the number of modifications guided by the snoRNA, the location of the modification, and whether the modification is conserved over evolution correlate with the abundance of the RNA. This machine learning approach correctly predicted the abundance of snoRNAs in 65% of the cases (Myslyuk et al. 2008).

The snoRNAs are most probably released from the precursors by an endonuclease similar to the yeast RNase III enzyme, RNT1 (Chanfreau et al. 1999). Two RNase III-like enzymes, Tb11.46.0012 and Tb.10.01.0150, present in the *T. brucei* genome as well as the two Dicers, DCL1 and DCL2 (Shi et al. 2006), were examined for their role in snoRNA processing by RNAi, but the depletion of these factors had no effect on snoRNA levels (Hury and Michaeli, unpublished results). Relatively little is known regarding nuclear functions that affect snoRNA processing. The nuclear exosome RRP6 as well as the MTR4 helicase were shown to be required for snoRNA biogenesis (Gupta and Michaeli, unpublished data).

Almost all snoRNA clusters are repeated several times in the genome. Like many protein-coding genes, the repeated nature of snoRNA clusters represents a mechanism of coping with the absence of polymerase II promoters. The high expression of these RNAs is therefore mediated by gene multiplicity. The RNA-seq data revealed 14 “solitary” snoRNAs among the 50 snoRNA loci. Some of the “solitary” snoRNAs were found near abundant mRNAs, suggesting that another strategy to highly express some snoRNAs is based on their location near a highly expressed transcript. The best example is a snoRNA located in the intergenic region of α and β tubulin genes. Interestingly, snoRNAs were also found in the 3′ UTR of genes (Kolev et al. 2010).

Most recently, snoRNA clusters were shown to undergo *trans*-splicing and polyadenylation (Kolev et al. 2010). We have recently obtained evidence that the most abundant snoRNAs, such as the snoRNA present in the SLA1 cluster (snR30, TB11Cs2C1, and C2), are polyadenylated in a complex manner by two different polymerases. The level of such snoRNAs emerges from the interplay between the two processes that act to generate the snoRNP, namely pre-snoRNA trimming by exonucleolytic activities and snoRNP assembly. The complex polyadenylation delays the recruitment of the exosome that trims the snoRNA from the 3′ end, thus enabling efficient assembly before the exosome can degrade the RNA. This mechanism can explain how the abundance of snoRNAs transcribed from the same

locus can vary by 10- to even a 100-fold. These highly abundant snoRNAs were shown to function in rRNA processing (Michaeli et al. 2012).

Recent data emerging from studies in yeast and mammals suggest that assembly and trafficking of C/D and H/ACA RNPs are highly regulated processes. The complexes are assembled on pre-RNPs on nascent guide RNA transcripts, which are transported in mammalian cells to Cajal bodies, where maturation to active RNP complexes occurs. Cajal bodies are complex intranuclear structures enriched in factors involved in RNA modification and in the assembly of RNA–protein complexes (Nizami et al. 2010). The bodies contain proteins such as coilin and the small Cajal body (sca) RNAs that guide modifications on snRNAs (Matera et al. 2007). From the Cajal bodies, snoRNPs are forwarded to the sites where they function. H/ACA molecules associate with NAF1 at the site of transcription, but this protein is exchanged by GAR1 within the Cajal bodies (Dez et al. 2002). NAF1 exists in trypanosomes but its role in snoRNA biogenesis has not yet been explored. So far, no evidence exists for the presence of Cajal bodies in trypanosomes. However, SLA1 was found in a distinct nuclear domain that functionally resembles Cajal bodies (Hury et al. 2009). Indeed, SLA1 was shown to bind a protein, MTAP (Zamudio et al. 2009), whose human homologue binds scaRNAs and telomerase RNA found in Cajal bodies (Tycowski et al. 2009). These data suggest that trypanosomes carry a nuclear body analogous to the Cajal body. As in mammals, this site could be the processing site of snoRNA including SLA1. The mature snoRNAs either migrate to the nucleolus or in the case of SLA1, the RNP migrates to the site of SL RNA biogenesis known also as the SL RNP factory (Hury et al. 2009).

6.7 Conservation of snoRNAs Among Trypanosomatid Species

The most extensive comparisons were performed between *T. brucei* and *L. major* snoRNAs (Liang et al. 2007a). The pattern of Nm modification is highly conserved between *L. major* and *T. brucei*. The *L. major* clusters are highly repeated compared to those of *T. brucei* (Liang et al. 2007a). As discussed above and noted for *T. brucei*, snoRNAs that exist in the same gene cluster may be differentially expressed (Liang et al. 2007a). Sequence conservation of H/ACA among *Leishmania* species and other trypanosomatids allowed structural features to be identified that are conserved in these RNAs. For instance, it was found that stem I of the H/ACA RNA is usually perfect and ranges from 4 to 8 nt in length, and many compensatory changes support the integrity of the stem. The pseudouridylation pocket varies in size from 12 to 17 nts. Stem II also varies in size, but a perfect stem of 4–7 nt exists immediately adjacent to the pseudouridylation pocket. The conserved length of the stems is supported by the presence of compensatory changes (Liang et al. 2007a; Uliel et al. 2004). A consensus structure for the trypanosomatid H/ACA RNA was suggested (Myslyuk et al. 2008).

6.8 snoRNAi

Silencing of snoRNA genes was reported for three trypanosomatid species: *L. collosom*a, *L. major*, and *T. brucei*. Silencing was achieved in *L. collosoma* and *L. major* by expressing antisense RNA from an episomal vector, and in *T. brucei* by expressing dsRNA from a vector carrying two T7 opposing promoters (Liang et al. 2003). Silencing in *L. collosoma* of a C/D guide rRNA (snoRNA-2) reduced the level of the RNA by 90% and eliminated the Nm modifications guided by this RNA. Silencing of H/ACA was less efficient compared to C/D snoRNAs. In *L. major*, expression of dsRNA to sno270 (snR30) resulted in partial silencing. Recent data also demonstrated efficient silencing of TB11Cs2C2 (~90%) which resulted in defects in rRNA processing, suggesting the role of this RNA in srRNA processing (Gupta et al. 2010). The mechanism of snoRNAi was investigated in *T. brucei* and it was found that the process relies on the RNAi machinery, requiring TbDCL2, which is the nuclear DICER, but not the cytoplasmic DICER, TbDCL1, which mediates the silencing of mRNAs (Gupta et al. 2010). In addition, Argonaute 1 (AGO1) was shown to be required for this process. Indeed, cells in which AGO1 expression was knocked down failed to perform snoRNAi (Gupta et al. 2010). The mechanism of snoRNA silencing in *Leishmania* must be different, since *L. major* lacks the RNAi machinery. However, RNAs of a size appropriate for siRNA are produced in cells expressing the antisense RNAs and these can induce degradation of the RNA by an as yet unknown mechanism. Interestingly, silencing was observed for the mature snoRNA, but the level of pre-snoRNA was unchanged (Liang et al. 2003), suggesting that the degradation must take place after snoRNA processing. snoRNAi was recently used to decipher the role of the abundant *T. brucei* snoRNAs in rRNA processing (Gupta et al. 2010; Michaeli et al. 2012).

6.9 Enzymes and Protein Factors Involved in rRNA Processing

The yeast system contributed most to the identification of protein factors involved in rRNA processing. Using classical yeast genetics, many *trans*-acting factors were identified in the last 3 decades. However, this methodology has limitations, and this approach only provided a glimpse of the complexity of the process (Fromont-Racine et al. 2003). A real breakthrough in describing the machinery that is involved in rRNA processing and assembly resulted from the application of the tandem-affinity purification method (TAP) that allowed isolation of native particles and their analysis by advanced mass-spectrometry. This approach led to the identification of many novel proteins and to the definition of the SSU processome (Dragon et al. 2002; Gavin et al. 2002).

Unexpectedly, the purified SSU processome contains very few ribosomal proteins and factors required for the synthesis of the 60S ribosomal subunit, suggesting that the purification enriched a stable preribosomal complex that was

separated from the LSU by cleavage at the A3 site. Sixty proteins were detected in this processome including the U3 snRNP. Novel proteins associated with this complex were termed Utp1-17, and the purification also detected enzymes such as helicases and ATPases. The helicases and ATPases are required for structural changes that take place within the pre-rRNA and in the complex protein–RNA interactions that take place during the different processing events (Dragon et al. 2002; Gavin et al. 2002).

To examine if trypanosomes like yeast carry similar processome factors, a bioinformatic search was carried out, searching for protein factors that comprise the processome in yeast. Our analysis identified the following proteins: UTP 3, 6, 7, 10, 11, 13, 14, 15, 18, 19, and 20 (Table 6.1). No homologues could be identified for UTP 4, 5, 8, and 9, either because these are missing or are poorly conserved. PRP9 that was shown to associate with U3 snRNP and whose depletion affects the biogenesis of both the small and large subunit (Dragon et al. 2002; Venema et al. 2000) was identified in trypanosomes (Table 6.1). A trypanosome homologue also exists for PRP5, a factor whose depletion specifically affects cleavage at A2 (Grandi et al. 2002; Nissan et al. 2002; Venema and Tollervey 1996). The 43S particle containing the SSU precursor (20S) is exported to the cytoplasm where it is cleaved to the 18S RNA. Two kinases were shown to be important for this process, RIO1 and RIO2. Both these proteins are present in trypanosomes (Table 6.1). These yeast kinases are known to shuttle between the nucleus and the cytoplasm (Vanrobays et al. 2001). Not all the factors that were shown to be essential for 40S processing were found in trypanosomes, including HCR1, which is believed not only to influence the cleavage reaction to produce the mature SSU but also plays a role in translation initiation (Asano et al. 2001).

As described above, maturation of the pre-40S ribosomal subunit involves a single cleavage at the 3′ end of the 18 rRNA (cleavage D), which requires about ten different nonribosomal factors, including ENP1, TSR1, LTV1, DIM1, NOB1, and RIO2 (Fromont-Racine et al. 2003) (Table 6.1). Several homologous factors exist in trypanosomes including ENP1, TSR1, DIM1, NOB1, and RIO2. Of interest is the presence of NOB1 which is the endonuclease that performs the cleavage at site D (Fig. 6.1a), suggesting that this processing step is highly conserved in evolution (Fatica et al. 2003).

The processing of the 60S subunit involves cleavages of the 3′ end of ITS1 and excision of ITS2. Maturation of LSU is even more complex in trypanosomes and is thought to involve trypanosome-specific factors, as stated above. In yeast, over 60 proteins are associated with different steps of 60S maturation. The role of the different preribosomal factors in these complexes is currently unknown. A yeast factor that seems to be associated with early 60S particles is SSF1 (Wehner et al. 2002). Complexes associated with this factor contain the pre-LSU fragments. Another important factor that is found in SSF1-containing complexes is RPF1 (Wehner et al. 2002). SSF1 and RPF1 are both present in trypanosomes (Table 6.1). Later in this process, factors such as NOG2 and NUG1 join the complex (Saveanu et al. 2001; Saveanu et al. 2003). A trypanosome NOG2 homologue was identified, but not one for NUG1. As soon as cleavage takes place at ITS2, and the pre-5.8S is

Table 6.1 Trypanosome homologues to protein factors involved in ribosome biogenesis

Gene	Yeast annotation	*T. brucei* gene	Annotation in *T. brucei* genome
SnoRNPs			
NOP1	YSL014W	Tb10.6k15.3160, Tb10.61.2040, Tb10.61.1920	Fibrillarin
SNU13	YELO26W	Tb09.160.3670	Ribosomal protein S6
NOP56	YLR197W	Tb927.8.3750	Nucleolar protein
NOP58 NOP5	YOR310C		
	Tb09.160.3820	Hypothetical protein	
NHP2	YDL208W	Tb927.4.750	50S ribosomal protein L7Ae
NOP10	YHR072W	Tb10.70.2465	Nucleolar RNA-binding protein
GAR1	YHR089C	Tb927.4.470	snoRNP protein gar1
SHQ1	YIL104C	Tb11.01.8060	Hypothetical protein
CBF5	YLR175W	Tb10.100.0060	Centromere/microtubule binding protein
NAF1	YNL124W	Tb927.2.3160	Hypothetical protein
90S particles involved in SSU processing			
UTP20	YBL004W	Tb10.70.3570	Hypothetical protein
KRR1	YCL059C	Tb927.6.4350	Ribosomal RNA assembly protein
PWP2 UTP1	YCR057C	Tb10.389. 0160	Periodic tryptophan protein 2
NOP14 UTP2	YDL148C	Tb927.2.2950	Hypothetical protein
UTP3 SAS10	YDL153C	Tb11.01.0490	Hypothetical protein
FAL1	YDR021W	TB11.12.0011	ATP-dependent DEAD/H RNA helicase
UTP6	YDR449C	Tb09.244.2650	Hypothetical protein
UTP7 KRE31	YER082C	Tb927.8.2600	Hypothetical protein
ROK1	YGL171W	Tb11.01.5370	ATP-dependent DEAD/H RNA helicase
UTP22	YGR090W	Tb10.389.0280	Hypothetical protein
ENP2	YGR145W	Tb927.7.700	Hypothetical protein
GNO1	YGR280C	Tb927.7.5640	Hypothetical protein
IMP3	YHR148W	Tb11.01.5500	U3 small nucleolar ribonucleoprotein IMP3
UTP18	YJL069C	Tb10.6k15.0360	Hypothetical protein
UTP10	YJL109C	Tb09.160.1560	Hypothetical protein
MPP10	YJR002W	Tb927.3.3590	U3 small nucleolar ribonucleoprotein protein MPP10
UTP11	YKL099C	Tb10.61.1030	U3 snoRNA-associated protein
SOF1	YLL011W	Tb09.211.2550	Hypothetical protein
DIP2 UTP12	YLR129W	Tb927.7.4220	Hypothetical protein
EMG1 NEP1	YLR186W	Tb927.8.5040	Hypothetical protein
UTP13	YLR222C	Tb11.03.0680	Hypothetical protein
UTP21	YLR409C	Tb11.01.2250	Hypothetical protein
UTP14	YML093W	Tb10.61.2440	Hypothetical protein
UTP15	YMR093W	Tb927.8.1980	Hypothetical protein
DHR1 ECM16	YMR128W	Tb927.6.4600	pre-mRNA splicing factor ATP-dependent RNA helicase

(continued)

Table 6.1 (continued)

Gene	Yeast annotation	*T. brucei* gene	Annotation in *T. brucei* genome
RRP5	YMR229C	Tb927.1.1370	rRNA biogenesis protein
IMP4	YNL075W	Tb927.6.2780	U3 small nuclear ribonucleoprotein (snRNP)
KRE33	YNL132W	Tb927.5.2530	Hypothetical protein
KRI1	YNL308C	Tb927.5.840	Hypothetical protein
RCL1	YOL010W	Tb09.211.4870	RNA 3′-terminal phosphate cyclase-like protein
BMS1	YPL217C	Tb11.01.0820	Hypothetical protein
MRD1	YPR112C	Tb927.8.4170	RNA-binding protein
RRP9	YPR137W	Tb10.70.4960	Hypothetical protein
UTP19 NOC4	YPR144C	Tb927.4.3670	Hypothetical protein
Pre-40S particles			
DRS2	YAL026C	Tb927.6.3550	Phospholipid-translocating P-type ATPase (flippase)
PTC3	YBL056W	Tb10.70.2270	Protein phosphatase 2C-like
RIO2	YNL207W	Tb927.6.2840	Hypothetical protein
RIO1 RRP10	YOR119C	Tb927.3.5400	Hypothetical protein
TSR1	YDL060W	Tb927.8.1410	Hypothetical protein
NOB1	YOR056C	Tb11.01.2630	Hypothetical protein
Pre-40S	YOR145C	Tb09.211.2960	Hypothetical protein
DIM1	YPL266W	Tb927.6.1610	Ribosomal RNA adenine dimethylase family protein, conserved
ENP1 MEG1	YBR247C	Tb927.6.1900	Bystin
Pre-60S and 90S particles			
PWP1	YLR196W	Tb927.8.6770	Hypothetical protein
HAS1	YMR290C	Tb10.70.0570	ATP-dependent DEAD/H RNA helicase
RRP12	YPL012W	Tb11.02.0110	Hypothetical protein
Pre-60S particles			
ERP2	YAL007C	Tb11.01.6880	Cytosolic coat protein
MAK16	YAL025C	Tb927.7.3380	Hypothetical protein
SPB1	YCL054W	Tb927.4.4170	rRNA methyltransferase
DBP10	YDL031W	Tb927.4.2630	ATP-dependent DEAD/H RNA helicase
MAK21 NOC1	YDR060W	Tb11.18.0014	Hypothetical protein
RRP8	YDR083W	Tb927.1.1120	Hypothetical protein
RRP1	YDR087C	Tb927.8.5490	Hypothetical protein
PUF6	YDR496C	Tb927.3.2470	Pumilio/PUF RNA binding protein 8
NOP16	YER002W	Tb927.8.4290	Hypothetical protein
NSA2	YER126C	Tb10.6k15.2890	40S ribosomal protein S8
NSA1	YGL111W	Tb11.01.3850	Hypothetical protein
KEM1 XRN1	YGL173C	Tb927.7.4900	5′–3′ exonuclease XRNA
NOP7 YPH1	YGR103W	Tb09.211.0180	Hypothetical protein
SDA1	YGR245C	Tb927.5.2660	Hypothetical protein

(continued)

Table 6.1 (continued)

Gene	Yeast annotation	*T. brucei* gene	Annotation in *T. brucei* genome
REX1 RNH70	YGR276C	Tb927.8.3710	Hypothetical protein
SSF1	YHR066W	Tb927.3.2830	Peter pan protein, putative
RPF1	YHR088W	Tb11.03.0050	RNA processing factor 1
MTR4 DOB1	YJL050W	Tb10.6k15.3220	ATP-dependent DEAD/H RNA helicase
MRT4 DOB1	YKL009W	Tb10.6k15.0270	60S acidic ribosomal protein
EBP2	YKL172W	Tb09.244.2790	rRNA processing protein
DBP7	YKR024C	Tb927.3.620	ATP-dependent DEAD/H RNA helicase
RPF2	YKR081C	Tb927.7.270	Ribosome biogenesis protein
NOC3	YLR002C	Tb10.70.0510	Hypothetical protein
RLP24	YLR009W	Tb11.02.4220	60S ribosomal protein L24
MDN1 REA1	YLR106C	Tb927.1.880	Hypothetical protein
DBP9	YLR276C	Tb11.02.2620	ATP-dependent DEAD/H RNA helicase
ERB1	YMR049C	Tb11.02.4620	Hypothetical protein
RRB1	YMR131C	Tb11.02.0570	Hypothetical protein
RLP7 RIX9	YNL002C	Tb10.61.3190	60S ribosomal protein L7
NOP2 YNA1	YNL061W	Tb927.4.3840	Nucleolar protein
NOP15	YNL110C	Tb11.01.8310	Hypothetical protein
NOG2 NUG2	YNR053C	Tb927.7.7450	GTP-binding protein
NOP12	YOL041C	Tb09.211.4120	Hypothetical protein
RAT1 XRN2	YOR048C	Tb10.70.0610	5′–3′ exoribonuclease XRND
NOC2 RIX3	YOR206W	Tb10.389.1290	Hypothetical protein
YTM1	YOR272W	Tb927.4.3850	Hypothetical protein
RRS1	YOR294W	Tb927.6.2050	Ribosome biogenesis regulatory protein (RRS1)
NOP4 NOP77	YPL043W	Tb11.01.7680	Hypothetical protein
NOG1	YPL093W	Tb11.02.0620	Nucleolar GTP-binding protein 1
RSA1	YPL193W	Tb11.01.0090	Hypothetical protein
NIP7	YPL211W	Tb10.389.1790	Hypothetical protein
TIF6 CDC95	YPR016C	Tb10.70.1770	Eukaryotic translation initiation factor 6
Pre-60S late particles			
60s WD ND	YCR072C	Tb927.8.5990	Hypothetical protein
LSG1 KRE35	YGL099W	Tb927.5.4310	GTP-binding protein
NMD3 SRC5	YHR170W	Tb927.7.970	Hypothetical protein
EFL1 RIA1	YNL163C	Tb927.3.2170	Translation elongation factor EF-2

T. brucei homologues of all known or predicted factors involved in yeast ribosome maturation are listed, based on Table 6.1 (Fromont-Racine et al. 2003). One hundred *T. brucei* potential protein homologues (out of 153) were identified by performing BLASTP against *T. brucei* TREU927 (taxid:185431) NR database [4.2010]. Potential *T. brucei* homologues with the highest query coverage and lowest e-value were BLASTP against *Saccharomyces cerevisiae* (taxid:4932) NR database [4.2010], confirming the reliability of the homologues (Reciprocal BLAST). Factors are grouped according to their putative function: snoRNP proteins, factors of the 90S processome involved in SSU production, and factors necessary for the assembly of 40 and 60S ribosomes

separated from pre-LSU, other proteins such as ARX1 join. ARX1 proteins are present both in the cytoplasm and in the nucleus and are associated with export factors such as NMD3 and MTR2 (Nissan et al. 2002). ARX1 could not be easily detected in the *T. brucei* genome, but homologues to its interacting protein, NMD3, were identified. Maturation of the pre-60S particle is completed in the cytoplasm. One factor involved in this stage is the EF1 GTPase, which is genetically and functionally linked to TIF6 (Sanvito et al. 1999). TIF6 leaves the nucleus as a component of the pre-60S particle and probably acts as an RNP chaperone during both early pre-RNA processing steps and at a late cytoplasmic maturation step. EFL1 is also present in trypanosomes (Table 6.1). GTP hydrolysis by EF1 may serve as a checkpoint to verify that 60S subunit is properly assembled before it joins to the 40S to initiate translation, suggesting that this type of checkpoint also operates in trypanosomes. Other factors, such as NMD3 and LSG1, bind to recycling mature-free 60S subunits in addition to their binding to nascent subunits (Nissan et al. 2002). An LSG1 homologue exists in trypanosomes, and it will be interesting to determine whether it plays a role in transport of the pre-60S and recruitment of the 60S complexes to the translating ribosomes.

Protein–RNA interactions are essential for the association of several factors with preribosomal rRNA. Interestingly, several of the preribosomal factors resemble ribosomal proteins and in fact the factors IMP3, MRT4, RLP7, and RLP24 are homologous to ribosomal proteins, RPS9, RPP0, RPL7, and RPL24, respectively. Homologues of the preribosomal factors are present in trypanosomes (Table 6.1). Another protein, which contains four RRM domains and binds directly to rRNA, is NOP4 (Sun and Woolford 1997); this protein also exists in trypanosomes and might also function in 60S subunit maturation.

Of special interest are three factors which were extensively studied in trypanosomes: NOG1 (Jensen et al. 2003), NOP44/66 (Jensen et al. 2005), and the proteins p34/p37 (Prohaska and Williams 2009). NOG1, as in other eukaryotes, was shown to be involved in processing of LSU; its depletion leads to accumulation of typical rRNA precursors and specific inhibition in cleaving ITS2 (Jensen et al. 2003). Of special interest are the two phosphoproteins NOP44/46; these proteins are trypanosome-specific and their depletion by RNAi resulted in a significant decrease in srRNA-1 processing. In addition, their absence also resulted in aberrant processing intermediates emerging from cleavages at LSUα and LSUβ. These proteins therefore appear to conduct a trypanosome-specific rRNA processing event (Jensen et al. 2003, Jensen et al. 2005). In addition, PUF proteins, especially PUF7, were shown recently to be involved in processing of the large pre-RNA precursor as well as the pre-SSU (Droll et al. 2010).

Of interest are the trypanosome-specific proteins P34/37, which are involved in nuclear export of the 60S ribosomal subunit (Prohaska and Williams 2009). The bioinformatic analysis presented here awaits verification by purification of the relevant complexes and identifying rRNA defects that may result from depletion of these factors by RNAi.

6.10 Conclusions

rRNA processing in trypanosomes is unique, and involves fragmentation of the LSU to two large fragments and four small rRNA species. Nevertheless, the biological role of this unique cleavage process is currently unknown. We have begun to dissect the unique machinery that mediates the cleavages generating these species. Currently, 142 snoRNAs were found in trypanosomes and this number is higher than the repertoire of 100 snoRNAs described in yeast. RNA-seq of snoRNA complexes revealed highly abundant snoRNAs, of which a few were already shown to function in trypanosome-specific cleavages (Gupta et al. 2010; Michaeli et al. 2012). Although trypanosomes possess a few of the highly conserved snoRNAs present in other eukaryotes such as U3, snR30, and MRP RNAs, on the other hand, they also lack snoRNAs, such as U8 and U14, that are known from other eukaryotes. However, trypanosomes contain trypanosome-specific, functional homologues to these RNA species. Future studies should unravel the function of these snoRNA using a variety of tools, including snoRNAi knockdown and "RNA walk" (Liang et al. 2003; Lustig et al. 2010). Our knowledge of the protein factors that participate in rRNA processing and maturation is currently very limited. Here, I presented a survey of such factors in the *T. brucei* genome and identified many of these conserved factors. However, future studies should be focused on purification and mass-spectrometric analysis of the processome complex. Such studies are expected to identify both conserved factors and trypanosome-specific factors.

Acknowledgements Our work on rRNA processing was supported by a grant from the Israel-US Binational Science Foundation (BSF). S.M. holds the David and Inez Myers Chair in RNA silencing of diseases. I wish to thank Fr. Hiba Asher-Waldman for the bioinformatics analysis of the protein factors involved in rRNA processing and maturation.

References

Asano K, Phan L, Valasek L, Schoenfeld LW, Shalev A, Clayton J, Nielsen K, Donahue TF, Hinnebusch AG (2001) A multifactor complex of eIF1, eIF2, eIF3, eIF5, and tRNA(i)Met promotes initiation complex assembly and couples GTP hydrolysis to AUG recognition. Cold Spring Harb Symp Quant Biol 66:403–415

Atzorn V, Fragapane P, Kiss T (2004) U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. Mol Cell Biol 24:1769–1778

Bally M, Hughes J, Cesareni G (1988) SnR30: a new, essential small nuclear RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res 16:5291–5303

Barth S, Hury A, Liang XH, Michaeli S (2005) Elucidating the role of H/ACA-like RNAs in trans-splicing and rRNA processing via RNA interference silencing of the *Trypanosoma brucei* CBF5 pseudouridine synthase. J Biol Chem 280:34558–34568

Barth S, Shalem B, Hury A, Tkacz ID, Liang XH, Uliel S, Myslyuk I, Doniger T, Salmon-Divon M, Unger R, Michaeli S (2008) Elucidating the role of C/D snoRNA in rRNA processing and modification in *Trypanosoma brucei*. Eukaryot Cell 7:86–101

Beltrame M, Tollervey D (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. EMBO J 11:1531–1542

Bohnsack MT, Kos M, Tollervey D (2008) Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase. EMBO Rep 9:1230–1236

Brown JW, Echeverria M, Qu LH (2003) Plant snoRNAs: functional evolution and new modes of gene expression. Trends Plant Sci 8:42–49

Campbell DA, Kubo K, Clark CG, Boothroyd JC (1987) Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. J Mol Biol 196:113–124

Chanfreau G, Gouyette C, Schwer B, Jacquier A (1999) Trans-complementation of the second step of pre-mRNA splicing by exogenous 5′ exons. RNA 5:876–882

Chekanova JA, Dutko JA, Mian IS, Belostotsky DA (2002) *Arabidopsis thaliana* exosome subunit AtRrp4p is a hydrolytic 3′→5′ exonuclease containing S1 and KH RNA-binding domains. Nucleic Acids Res 30:695–700

Cordingley JS, Turner MJ (1980) 6.5S RNA; preliminary characterisation of unusual small RNAs in *Trypanosoma brucei*. Mol Biochem Parasitol 1:91–96

Darzacq X, Kiss T (2000) Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5′,3′-terminal stem structure. Mol Cell Biol 20:4522–4531

Decatur WA, Fournier MJ (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. J Biol Chem 278:695–698

Dennis PP, Omer A, Lowe T (2001) A guided tour: small RNA function in Archaea. Mol Microbiol 40:509–519

Dez C, Noaillac-Depeyre J, Caizergues-Ferrer M, Henry Y (2002) Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. Mol Cell Biol 22:7053–7065

Dieci G, Preti M, Montanini B (2009) Eukaryotic snoRNAs: a paradigm for gene expression flexibility. Genomics 94:83–88

Dragon F, Gallagher JE, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlage RE, Shabanowitz J, Osheim Y, Beyer AL, Hunt DF, Baserga SJ (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature 417:967–970

Droll D, Archer S, Fenn K, Delhi P, Matthews K, Clayton C (2010) The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. FEBS Lett 584:1156–1162

Dunbar DA, Chen AA, Wormsley S, Baserga SJ (2000a) The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. Nucleic Acids Res 28:2855–2861

Dunbar DA, Wormsley S, Lowe TM, Baserga SJ (2000b) Fibrillarin-associated box C/D small nucleolar RNAs in *Trypanosoma brucei*. Sequence conservation and implications for 2′-O-ribose methylation of rRNA. J Biol Chem 275:14767–14776

Eichler DC, Craig N (1994) Processing of eukaryotic ribosomal RNA. Prog Nucleic Acid Res Mol Biol 49:197–239

Fatica A, Oeffinger M, Dlakic M, Tollervey D (2003) Nob1p is required for cleavage of the 3′ end of 18S rRNA. Mol Cell Biol 23:1798–1807

Filipowicz W, Pogacic V (2002) Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol 14:319–327

Fromont-Racine M, Senger B, Saveanu C, Fasiolo F (2003) Ribosome assembly in eukaryotes. Gene 313:17–42

Ganot P, Bortolin ML, Kiss T (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell 89:799–809

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G,

Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415:141–147

Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, Marzioch M, Schafer T, Kuster B, Tschochner H, Tollervey D, Gavin AC, Hurt E (2002) 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol Cell 10:105–115

Gray MW (1979) The ribosomal RNA of the trypanosomatid protozoan *Crithidia fasciculata*: physical characteristics and methylated sequences. Can J Biochem 57:914–926

Gupta SK, Hury A, Ziporen Y, Shi H, Ullu E, Michaeli S (2010) Small nucleolar RNA interference in *Trypanosoma brucei*: mechanism and utilization for elucidating the function of snoRNAs. Nucleic Acids Res 38:7236–7247

Hartshorne T (1998) Distinct regions of U3 snoRNA interact at two sites within the 5′ external transcribed spacer of pre-rRNAs in *Trypanosoma brucei* cells. Nucleic Acids Res 26:2541–2553

Hartshorne T, Agabian N (1993) RNA B is the major nucleolar trimethylguanosine-capped small nuclear RNA associated with fibrillarin and pre-rRNAs in *Trypanosoma brucei*. Mol Cell Biol 13:144–154

Hartshorne T, Toyofuku W (1999) Two 5′-ETS regions implicated in interactions with U3 snoRNA are required for small subunit rRNA maturation in *Trypanosoma brucei*. Nucleic Acids Res 27:3300–3309

Hartshorne T, Toyofuku W, Hollenbaugh J (2001) *Trypanosoma brucei* 5′ETS A'-cleavage is directed by 3′-adjacent sequences, but not two U3 snoRNA-binding elements, which are all required for subsequent pre-small subunit rRNA processing events. J Mol Biol 313:733–749

Huang GM, Jarmolowski A, Struck JC, Fournier MJ (1992) Accumulation of U14 small nuclear RNA in *Saccharomyces cerevisiae* requires box C, box D, and a 5′, 3′ terminal stem. Mol Cell Biol 12:4456–4463

Hury A, Goldshmidt H, Tkacz ID, Michaeli S (2009) Trypanosome spliced-leader-associated RNA (SLA1) localization and implications for spliced-leader RNA biogenesis. Eukaryot Cell 8:56–68

Jarmolowski A, Zagorski J, Li HV, Fournier MJ (1990) Identification of essential elements in U14 RNA of *Saccharomyces cerevisiae*. EMBO J 9:4503–4509

Jensen BC, Wang Q, Kifer CT, Parsons M (2003) The NOG1 GTP-binding protein is required for biogenesis of the 60 S ribosomal subunit. J Biol Chem 278:32204–32211

Jensen BC, Brekken DL, Randall AC, Kifer CT, Parsons M (2005) Species specificity in ribosome biogenesis: a nonconserved phosphoprotein is required for formation of the large ribosomal subunit in *Trypanosoma brucei*. Eukaryot Cell 4:30–35

Kass S, Sollner-Webb B (1990) The first pre-rRNA-processing event occurs in a large complex: analysis by gel retardation, sedimentation, and UV cross-linking. Mol Cell Biol 10:4920–4931

Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell 85:1077–1088

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 6: e1001090

Levitan A, Xu YX, Ben Dov C, Ben Shlomo H, Zhang YF, Michaeli S (1998) Characterization of a novel trypanosomatid small nucleolar RNA. Nucleic Acids Res 26:1775–1783

Li L, Ye K (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. Nature 443:302–307

Li HD, Zagorski J, Fournier MJ (1990) Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. Mol Cell Biol 10:1145–1152

Liang WQ, Fournier MJ (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. Genes Dev 9:2433–2443

Liang WQ, Clark JA, Fournier MJ (1997) The rRNA-processing function of the yeast U14 small nucleolar RNA can be rescued by a conserved RNA helicase-like protein. Mol Cell Biol 17:4124–4132

Liang XH, Liu L, Michaeli S (2001) Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. J Biol Chem 276:40313–40318

Liang XH, Xu YX, Michaeli S (2002) The spliced leader-associated RNA is a trypanosome-specific sn(o) RNA that has the potential to guide pseudouridine formation on the SL RNA. RNA 8:237–246

Liang XH, Liu Q, Michaeli S (2003) Small nucleolar RNA interference induced by antisense or double-stranded RNA in trypanosomatids. Proc Natl Acad Sci USA 100:7521–7526

Liang XH, Uliel S, Hury A, Barth S, Doniger T, Unger R, Michaeli S (2005) A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. RNA 11:619–645

Liang XH, Hury A, Hoze E, Uliel S, Myslyuk I, Apatoff A, Unger R, Michaeli S (2007a) Genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in Leishmania major indicates conservation among trypanosomatids in the repertoire and in their rRNA targets. Eukaryot Cell 6:361–377

Liang XH, Liu Q, Fournier MJ (2007b) rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. Mol Cell 28:965–977

Lustig Y, Wachtel C, Safro M, Liu L, Michaeli S (2010) 'RNA walk' a novel approach to study RNA-RNA interactions between a small RNA and its target. Nucleic Acids Res 38:e5

Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol 8:209–220

Mereau A, Fournier R, Gregoire A, Mougin A, Fabrizio P, Lührmann R, Branlant C (1997) An in vivo and in vitro structure-function analysis of the *Saccharomyces cerevisiae* U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. J Mol Biol 273:552–571

Michaeli S, Doniger T, Gupta SK, Wurtzel O, Romano M, Visnovezky D, Sorek R, Unger R, Ullu E (2012) RNA-seq analysis of small RNPs in *Trypanosoma brucei* reveals a rich repertoire of non-coding RNAs. Nucleic Acids Res 40(3):1282–1298

Morrissey JP, Tollervey D (1993) Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol Cell Biol 13:2469–2477

Morrissey JP, Tollervey D (1997) U14 small nucleolar RNA makes multiple contacts with the pre-ribosomal RNA. Chromosoma 105:515–522

Myslyuk I, Doniger T, Horesh Y, Hury A, Hoffer R, Ziporen Y, Michaeli S, Unger R (2008) Psiscan: a computational approach to identify H/ACA-like and AGA-like non-coding RNA in trypanosomatid genomes. BMC Bioinformatics 9:471

Nakaar V, Dare AO, Hong D, Ullu E, Tschudi C (1994) Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. Mol Cell Biol 14:6736–6742

Nissan TA, Bassler J, Petfalski E, Tollervey D, Hurt E (2002) 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. EMBO J 21:5539–5547

Nizami Z, Deryusheva S, Gall JG (2010) The Cajal body and histone locus body. Cold Spring Harb Perspect Biol 2:a000653

Peculis BA (1997) The sequence of the 5′ end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. Mol Cell Biol 17:3702–3713

Peculis BA, Steitz JA (1994) Sequence and structural elements critical for U8 snRNP function in Xenopus oocytes are evolutionarily conserved. Genes Dev 8:2241–2255

Prohaska K, Williams N (2009) Assembly of the *Trypanosoma brucei* 60S ribosomal subunit nuclear export complex requires trypanosome-specific proteins P34 and P37. Eukaryot Cell 8:77–87

Puvion-Dutilleul F, Bachellerie JP, Puvion E (1991) Nucleolar organization of HeLa cells as studied by in situ hybridization. Chromosoma 100:395–409

Rashid R, Aittaleb M, Chen Q, Spiegel K, Demeler B, Li H (2003) Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. J Mol Biol 333: 295–306

Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. Mol Cell 21:249–260

Roberts TG, Dungan JM, Watkins KP, 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

Roberts TG, Sturm NR, Yee BK, Yu MC, Hartshorne T, Agabian N, Campbell DA (1998) Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans. Mol Cell Biol 18:4409–4417

Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Hüttenhofer A (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. Nucleic Acids Res 31:869–877

Russell AG, Schnare MN, Gray MW (2004) Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. RNA 10:1034–1046

Russell AG, Schnare MN, Gray MW (2006) A large collection of compact box C/D snoRNAs and their isoforms in *Euglena gracilis*: structural, functional and evolutionary insights. J Mol Biol 357:1548–1565

Sanvito F, Piatti S, Villa A, Bossi M, Lucchini G, Marchisio PC, Biffo S (1999) The beta4 integrin interactor p27(BBP/eIF6) is an essential nuclear matrix protein involved in 60S ribosomal subunit assembly. J Cell Biol 144:823–837

Saveanu C, Bienvenu D, Namane A, Gleizes PE, Gas N, Jacquier A, Fromont-Racine M (2001) Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. EMBO J 20:6475–6484

Saveanu C, Namane A, Gleizes PE, Lebreton A, Rousselle JC, Noaillac-Depeyre J, Gas N, Jacquier A, Fromont-Racine M (2003) Sequential protein association with nascent 60S ribosomal particles. Mol Cell Biol 23:4449–4460

Savino R, Gerbi SA (1990) In vivo disruption of Xenopus U3 snRNA affects ribosomal RNA processing. EMBO J 9:2299–2308

Schnare MN, Gray MW (1990) Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. J Mol Biol 215:73–83

Schnare MN, Spencer DF, Gray MW (1983) Primary structures of four novel small ribosomal RNAs from *Crithidia fasciculata*. Can J Biochem Cell Biol 61:38–45

Shi H, Tschudi C, Ullu E (2006) An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. RNA 12:2063–2072

Sun C, Woolford JL Jr (1997) The yeast nucleolar protein Nop4p contains four RNA recognition motifs necessary for ribosome biogenesis. J Biol Chem 272:25345–25352

Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Hüttenhofer A (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. Proc Natl Acad Sci USA 99:7536–7541

Tollervey D (1987) A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. EMBO J 6:4169–4175

Tollervey D, Kiss T (1997) Function and synthesis of small nucleolar RNAs. Curr Opin Cell Biol 9:337–342

Tschochner H, Hurt E (2003) Pre-ribosomes on the road from the nucleolus to the cytoplasm. Trends Cell Biol 13:255–263

Tyc K, Steitz JA (1992) A new interaction between the mouse 5′ external transcribed spacer of pre-rRNA and U3 snRNA detected by psoralen crosslinking. Nucleic Acids Res 20:5375–5382

Tycowski KT, Shu MD, Steitz JA (1994) Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. Science 266:1558–1561

Tycowski KT, Shu MD, Kukoyi A, Steitz JA (2009) A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. Mol Cell 34:47–57

Uliel S, Liang XH, Unger R, Michaeli S (2004) Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. Int J Parasitol 34:445–454

Vanrobays E, Gleizes PE, Bousquet-Antonelli C, Noaillac-Depeyre J, Caizergues-Ferrer M, Gelugne JP (2001) Processing of 20S pre-rRNA to 18S ribosomal RNA in yeast requires Rrp10p, an essential non-ribosomal cytoplasmic protein. EMBO J 20:4204–4213

Venema J, Tollervey D (1996) RRP5 is required for formation of both 18S and 5.8S rRNA in yeast. EMBO J 15:5701–5714

Venema J, Tollervey D (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. Annu Rev Genet 33:261–311

Venema J, Vos HR, Faber AW, van Venrooij WJ, Raue HA (2000) Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association. RNA 6:1660–1671

Watkins NJ, Dickmanns A, Lührmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. Mol Cell Biol 22:8342–8352

Wehner KA, Gallagher JE, Baserga SJ (2002) Components of an interdependent unit within the SSU processome regulate and mediate its activity. Mol Cell Biol 22:7258–7267

White TC, Rudenko G, Borst P (1986) Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Res 14:9471–9489

Xu Y, Liu L, Lopez-Estrano C, Michaeli S (2001) Expression studies on clustered trypanosomatid box C/D small nucleolar RNAs. J Biol Chem 276:14289–14298

Yang CY, Zhou H, Luo J, Qu LH (2005) Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*. Biochem Biophys Res Commun 328:1224–1231

Zamudio JR, Mittra B, Chattopadhyay A, Wohlschlegel JA, Sturm NR, Campbell DA (2009) *Trypanosoma brucei* spliced leader RNA maturation by the cap 1 2′-O-ribose methyltransferase and SLA1 H/ACA snoRNA pseudouridine synthase complex. Mol Cell Biol 29:1202–1211

Chapter 7
RNA Editing in African Trypanosomes: A U-ser's G-U-ide

H. Ulrich Göringer

Abstract Of the chemical repertoire of RNA processing pathways, one reaction cycle stands out due to its idiosyncratic features. The process has evolved within the mitochondria of only one clad of organisms, the kinetoplastid protozoa (*Excavata*, *Euglenozoa*), which include the African trypanosome, the causative agent of sleeping sickness. The process can be categorized as an RNA editing-type reaction. It is characterized by the insertion and/or deletion of uridine nucleotides into otherwise nontranslatable mRNAs. Kinetoplastid RNA editing relies on an exclusive class of *trans*-acting, small noncoding RNA molecules known as guide RNAs. Furthermore, a unique molecular machinery, the editosome, catalyzes the process. Editosomes represent high-molecular-mass multienzyme assemblies that provide a catalytic surface for the individual steps of the editing reaction cycle. Here I review the current mechanistic model of the editing reaction and I summarize the molecular components and accessory factors of the editosome machinery. Special emphasis is put on the recent discovery of the molecular structure of the editing complex, which allows a first correlation of structural features with functional characteristics.

Keywords African trypanosomes • Editosome • Guide RNA • Kinetoplastids • Mitochondria • Pre-mRNA • RNA editing

7.1 Introduction

The RNA editing reaction in kinetoplastid protozoa such as African trypanosomes is a unique and probably the most bizarre posttranscriptional modification reaction in any living system (Benne et al. 1986). The reaction takes place within the single mitochondrion of the protozoan organism, which in itself is characterized by

H.U. Göringer
Genetics, Darmstadt University of Technology, Schnittspahnstr. 10, 64287 Darmstadt, Germany
e-mail: goringer@hrzpub.tu-darmstadt.de

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_7,

several exceptional features. Unlike other eukarya the mitochondrial genome is organized in a very large "disc-shaped" molecular assembly known as the kinetoplast (k). It is located within the mitochondrial matrix and consists of two classes of double-stranded circular DNA molecules, so-called "maxicircles" and "minicircles." African trypanosomes contain about 25–50 presumably identical 22 kbp maxicircles and roughly 10^4 minicircles with an average size of 1 kbp. Both kDNA elements are intertwined by concatenation thereby forming a massive DNA network, which exhibits a discrete cell cycle DNA synthesis stage (reviewed in McKean 2003) and accounts for up to 20% of the total cellular DNA content. Similar to other eukaryotes, DNA maxicircles encode ribosomal (r)RNA genes and subunits of the oxidative phosphorylation system. However, unlike in other organisms more than 50% of these genes are "incompletely" encoded: Substantial sequence information is missing—in some cases more than 50%—in addition to the presence of frameshifts and nonexisting translational start and stop codons. As a consequence, precursor (pre) mRNAs of these genes require RNA editing to be converted to functional mRNAs for protein synthesis.

Biochemically, the reaction is characterized by the insertion and/or deletion of exclusively U-nucleotides. In African trypanosomes, more than 3,000 U-residues are inserted and about 300 U's deleted at hundreds of editing sites in 12 different pre-mRNAs (for a recent review see Hajduk and Ochsenreiter 2010). The process is site-specific and relies on small, noncoding RNAs, known as guide RNAs (gRNAs) (Blum et al. 1990; Blum and Simpson 1990; Schmid et al. 1995; Hermann et al. 1997). Guide RNAs are DNA minicircle transcripts and act as "quasi templates" in the process. They initiate the editing reaction by the formation of a hybrid gRNA/pre-mRNA molecule, which adopts a three-way helix junction topology (Leung and Koslowsky 2001; Yu and Koslowsky 2006; Reifur and Koslowsky 2008). The structure includes a short "anchoring" duplex that borders the sequence to be edited. Unpaired gRNA nucleotides next to the anchor region specify U-insertion events with free UTP as a substrate; non-base-paired uridylates in the pre-mRNA become deleted (Seiwert and Stuart 1994; Frech and Simpson 1996; Kable et al. 1996; Seiwert et al. 1996). By utilizing different gRNAs, alternative editing events can occur, which contribute to generate protein diversity within the mitochondria of the parasites (Ochsenreiter and Hajduk 2006; Ochsenreiter et al. 2008a, b).

Early experimental evidence already suggested that RNA editing involves mitochondrial protein components and that the reaction is likely catalyzed by high-molecular-mass protein complexes (Pollard et al. 1992; Göringer et al. 1994; Köller et al. 1994; Corell et al. 1996; Rusché et al. 1997; Peris et al. 1997). In analogy to ribosomes and spliceosomes the complexes have been termed editosomes and they function as a reaction platform for the catalysis of the processing reaction in a series of enzyme-driven steps (Blum et al. 1990; Frech and Simpson 1996; Sabatini and Hajduk 1995). In recent years, the protein inventory of the editosome has been studied in detail and polypeptide candidates for every step of the minimal reaction cycle have been identified (reviewed in Carnes and Stuart 2008; Hajduk and Ochsenreiter 2010).

In this article, I summarize the current knowledge with respect to the individual reaction steps of the RNA editing cycle and I provide a summary of the molecular

components of the editing machinery including accessory factors that contribute to the process. The discussion is limited to the situation in African trypanosomes (*Trypanosoma brucei*). A comparison to other kinetoplastid organisms has been published by Lukeš et al. (2005). Special emphasis is drawn towards the recently described three-dimensional (3D) structures of *T. brucei* RNA editing complexes derived from electron microscopy (EM)-based single particle studies (Golas et al. 2009). I discuss the prominent structural landmarks of the different complexes and analyze their interconversion and functionality in the context of published biochemical and genetic data.

7.2 The Phenomenon

Editing creates translatable mRNAs by three basal modalities: (1) by correcting frameshifts, (2) by generating translational start and/or stop codons, and (3) by creating entire open reading frames (ORF). Edited mRNAs encode subunits of the oxidative phosphorylation system such as the NADH–ubiquinone oxidoreductase (complex I), cytochrome bc_1 (complex III), cytochrome oxidase (complex IV), and ATP synthase (complex V). Of the 20 mitochondrial genes in *Trypanosoma brucei*, 12 pre-mRNAs are substrates of the processing reaction (Hajduk and Ochsenreiter 2010 and references therein), however, the degree of editing can vary considerably. The most modest form of editing exists in the case of COII: only four uridylate residues are inserted. Other examples of "moderate" editing are Cyb (34 U insertions) and MURF2 (maxicircle unidentified reading frame) with 26 U's inserted and 4 U's deleted. COIII on the other hand represents an extensively edited pre-mRNA (a phenomenon also known as "pan-editing"): 547 U residues are inserted and 41 U's deleted. This affects approximately 60% of the nucleotide content of the final mRNA. Other pan-edited pre-mRNAs are ND7 (553 U insertions/89 U deletions), A6 (447/28), ND9 (345/20), and CR4 (325/40) (summarized in Table 7.1).

7.3 The Machinery

Native editosomes of African trypanosomes have been enriched from the endogenous steady-state pool of mitochondrial editing complexes by a variety of biochemical protocols (for a review, see Panigrahi et al. 2007). Starting materials are usually insect-stage trypanosomes, which rely on fully developed mitochondria for their energy consumption and thus are expected to have maximal RNA editing activity. Nonsynchronized cells are harvested at midlog growth conditions and are used to isolate mitochondrial vesicles by various cell disruption and subcellular fractionation methods (Göringer et al. 1994; Rusché et al. 1997; Hauser et al. 1996). Mitochondrial vesicle preparations are converted into low-salt detergent lysates using nonionic detergents followed by isokinetic density gradient centrifugation techniques or

Table 7.1 Extent of RNA editing in *Trypanosoma brucei*

Mitochondrial transcript	Respiratory complex/ function	No. of U insertions/ U-deletions	Length of edited mRNA (nt)
ND1	Complex I	Not edited	
ND3	Complex I	210/13	452
ND4	Complex I	Not edited	
ND5	Complex I	Not edited	
ND7	Complex I	553/89	1,238
ND8	Complex I	259/46	574
ND9	Complex I	345/20	649
Cyb	Complex III	34/none	1,151
COI	Complex IV	Not edited	
COII	Complex IV	4/none	663
COIII	Complex IV	547/41	969
A6	Complex V	447/28	811
S12	Ribosomal protein S12	132/28	325
MURF1	Unknown function	Not edited	
MURF2	Unknown function	26/4	1,111
MURF5	Unknown function	Not edited	
CR3	Unknown function	148/13	299
CR4	Unknown function	325/40	567
9S rRNA	SSU ribosomal RNA	3′ oligo uridylation	
12S rRNA	LSU ribosomal RNA	3′ oligo uridylation	

Abbreviations: *ND* NADH-ubiquinone oxidoreductase subunits 1-9; *Cyb*: apocytochrome b; *CO*: cytochrome oxidase subunits I–III; A6: *ATP* synthase subunit 6; S12: small subunit ribosomal protein 12; *MURF*: maxicircle unidentified reading frame; *CR*: G versus C-strand biased genes no. 3 and 4; *SSU*: small subunit; *LSU*: large subunit

column purification schemes. Following these protocols, active editing complexes harboring as little as 7 (Rusché et al. 1997), 13 (Golas et al. 2009; Aphasizhev et al. 2003a), or up to 20 polypeptides (Panigrahi et al. 2001) have been described. However, all enrichment protocols generate low yields of complexes, either suggesting low steady-state concentrations or low kinetic/thermodynamic and/or chemical (i.e., redox) stabilities of the complexes. Therefore, more recent purification schemes have applied near native enrichment conditions mainly following the tandem affinity purification (TAP) protocol as developed by Rigaut et al. (1999). The procedure relies on transgenic trypanosomes that conditionally express TAP-tagged versions of different editosomal proteins (reviewed in Panigrahi et al. 2007). The TAP-tag contains protein A and calmodulin binding domains separated by a tobacco etch virus (TEV) protease cleavage site, which allows for chemically moderate, i.e., "native-like" chromatographic separation and elution conditions.

TAP-tagged editosome preparations have been visualized by transmission electron microscopy (TEM) and by cryo-EM (Kastner et al. 2008; Golas et al. 2009). Raw EM images display monodisperse populations of two classes of high-molecular-mass assemblies in addition to some high-molecular-mass aggregates. The two classes consist of "large," asymmetric complexes of up to 26 nm in diameter and "smaller," elongated complexes with dimension of 21 × 26 nm.

Both types of complexes are characterized by well-defined, compact shapes with distinct structural features including surface areas of different electron density. In line with previous experimental data (Pollard et al. 1992; Corell et al. 1996), sedimentation analysis characterized the two particle classes as high-molecular-mass assemblies with apparent S-values of 20S and 35–40S (Golas et al. 2009). Importantly, 35–40S complexes have been shown to be associated with endogenous RNA including pre-mRNA and gRNA (Pollard et al. 1992; Corell et al. 1996; Golas et al. 2009) and thus likely embody the steady-state population of editing complexes actively engaged in the processing reaction. 20S complexes are "protein-only" assemblies (Rusché et al. 1997; Golas et al. 2009) that consist of 13 polypeptides: TbMP100, TbMP99, TbMP90, TbMP67, TbMP63, TbMP61, TbMP57, TbMP52, TbMP48, TbMP46, TbMP44, TbMP42, and TbMP24 (summarized in Table 7.2). This includes all required core activities of the editing reaction cycle (see below) and thus, 20S complexes are competent to faithfully edit synthetic pre-edited insertion and deletion substrate mRNAs in a gRNA-dependent fashion (Igo et al. 2000; Igo et al. 2002; Carnes and Stuart 2007). By contrast, isolated 35–40S complexes are inactive to bind and process synthetic pre-mRNA molecules likely because their RNA binding site/sites is/are occupied with endogenous RNA (Golas et al. 2009).

7.4 The Mechanism

A minimal RNA editing reaction cycle can be formally divided into three basal reaction steps: pre-mRNA cleavage, U-nucleotide deletion or U-nucleotide insertion, and "re"-ligation of the initially generated pre-mRNA fragments. Although it is not known whether the overall reaction is distributive or processive, available in vitro RNA editing data suggest that the reaction is carried out within a multifunctional reaction center that mediates the three partial reaction steps consecutively (Fig. 7.1).

The endonucleolytic cleavage of the pre-mRNA occurs at the first unpaired nucleotide upstream within the pre-mRNA/gRNA duplex substrate. Three related proteins, TbMP90, TbMP63, and TbMP67, which all contain U1-like Zn-finger motifs, an RNaseIII domain, and a dsRBM sequence possess endonuclease activity. TbMP90 is specific for the cleavage at deletion sites, while TbMP63 cleaves at insertion sites (Carnes et al. 2005; Trotter et al. 2005). The function of TbMP67 is not yet fully understood. At a deletion site the endonucleolytic cleavage is followed by the removal of unpaired U's. U-nucleotide-specific ribonuclease (exoUase) activity was identified for TbMP99 and TbMP100, two related proteins with N-terminal 5′/3′ exonuclease and C-terminal endo/exo/phosphatase motifs (Kang et al. 2005; Rogers et al. 2007). In addition, these proteins possess nucleotidyl phosphatase activity (Niemann et al. 2009). TbMP42, a protein with two C_2H_2-Zn-finger domains and a putative oligonucleotide/oligosaccharide binding (OB)-fold executes endo- and exoribonuclease activity in vitro, likely following a two-metal-

Table 7.2 *Trypanosoma brucei* proteins involved in RNA editing

Protein designation	Conserved structural motif(s)	Identified or proposed function
MP18	OB-fold	gRNA binding
MP19	OB-fold	Interaction
MP24	OB-fold	gRNA binding
MP41	U1-like	Interaction
MP42	Zn-finger (2×), OB-fold	Endo/exonuclease (in vitro)
MP44	U1-like, RNaseIII, Pum	Editosome integrity
MP46	U1-like, RNaseIII, Pum	Editosome integrity
MP47	U1-like	Interaction
MP48/REL2	Ligase, tau, K	RNA ligase
MP49	U1-like	Interaction
MP52/REL1	Ligase, tau, K	RNA ligase
MP57/RET2	PAP cat/assoc.	TUTase
MP61/REN2	U1-like, RNaseIII, dsRBM	U insertion-specific endonuclease
mHel61p	DEAD-box RNA helicase	RNA helicase, RNPase
MP63	Zn-finger (2×), OB-fold	Interaction
MP67/REN3	U1-like, RNaseIII, dsRBM	COXII-specific endonuclease
MP81	Zn-finger (2×), OB-fold	Interaction
MP90/REN1	U1-like, RNaseIII, dsRBM	U deletion-specific endonuclease
MP99/REX2	5′–3′ Exo, EEP	Nuclease/nucleotidyl phosphatase
MP100/REX1	5′–3′ Exo, EEP	Nuclease/nucleotidyl phosphatase
Accessory proteins		
gBP21	R-rich, Whirly-fold	Matchmaking-type RNA annealing
gBP25	R-rich, Whirly-fold	Matchmaking-type RNA annealing
RBP16	CSD	Interaction
TbRGG1	RGG-repeats	mRNA stabilization
KRET1	Zn-finger, PAP cat/assoc.	TUTase

Proteins are abbreviated following the nomenclature of Panigrahi et al. 2001. Alternative names and acronyms have been suggested by Stuart et al. (2005) and Simpson et al. (2010). *OB-fold*: oligosaccharide/oligonucleotide binding-fold, *U1-like*: U1-like Zn-finger domain, *Pum*: Pumilio domain, *tau*: microtubule-associated tau motif; *k*: kinesin light chain domain, *PAP*: poly(A) polymerase catalytic and associated domains, *dsRBM*: dsRNA binding motif, *5′3′ Exo*: 5′3′ exoribonuclease domain; *Zn-finger*: C2H2-type Zn-finger domain, *EEP*: endonuclease-exonuclease-phosphatase domain; *CSG* cold shock domain; *RGG* RGG RNA-binding repeats; *Whirly-fold*: Whirly transcription factor-fold; *TUTase*: terminal uridylyl transferase

ion reaction mechanism (Brecht et al. 2005; Niemann et al. 2008). Following endonucleolytic cleavage at an insertion site, U's are added to the 3′-end of the 5′-cleavage product of the mRNA in a gRNA-dependent fashion. This step of the reaction is catalyzed by TbMP57, a protein with a catalytic poly(A) polymerase (PAP) domain that executes terminal uridylyltransferase (TUTase) activity (Ernst et al. 2003). At the end of the processing cycle the two editing-specific RNA ligases TbMP48 and TbMP52 rejoin the two processed mRNA fragments (McManus et al. 2001; Rusché et al. 2001; Gao and Simpson 2003; Deng et al. 2004).

Aside from these core activities that catalyze the main steps of the reaction cycle, several accessory factors are required (for reviews see Carnes and Stuart 2008 and Göringer et al. 2008). These proteins presumably bind only temporarily

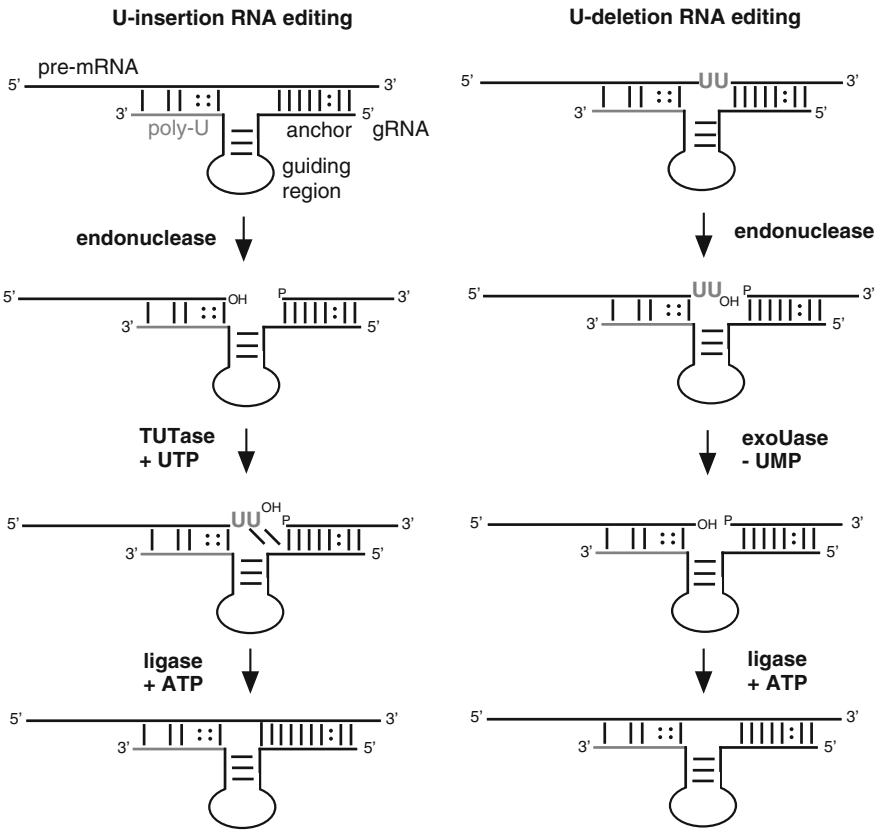


Fig. 7.1 Mechanistic outline of the basal reaction steps of U-insertion (left) and U-deletion (right) RNA editing in *Trypanosoma brucei*. Guide RNA molecules interact with cognate pre-edited mRNAs via antiparallel base pairing, thereby forming a three-helix junction RNA hybrid. The interaction relies on Watson/Crick-type base-pairing (*solid lines*) as well as noncanonical G:U bp (:). Endonucleolytic cleavage of the pre-mRNA occurs at the first non-base-paired nucleotide (nt) upstream of the so-called anchor duplex. U-insertion editing continues by adding U nt (from UTP) to the 3-end of the 5′ pre-mRNA cleavage fragment. The reaction is catalyzed by a terminal uridylyl transferase (TUTase). During deletion editing, U nt are removed from the 3′-end of the 5′ cleavage fragment, using the activity of a 3′ exonuclease (exoUase). In both cases, the number of U's (inserted or deleted) is specified by the "guiding" sequence of the gRNA. Resultant pre-mRNA cleavage fragments are finally ligated by an RNA ligase. Several editing cycles must occur until all editing sites specified by one gRNA are processed. Complete editing of a pre-mRNA requires in most cases multiple gRNAs. The reaction proceeds with a 3′–5′ directionality on the pre-mRNA. Measurements as to the fidelity of the reaction have not been published. However, substantial misediting (Decker and Sollner-Webb 1990; Maslov et al. 1994) as well as aberrant ligation events has been identified in vivo and in vitro (Blum et al. 1991; Koslowsky et al. 1992; Blum and Simpson 1992; Schmid et al. 1996). The involvement of a nucleotidyl phosphatase activity has been suggested by Niemann et al. (2009)

to the editosome to contribute additional functionality. This includes the matchmaking-type RNA/RNA annealing factors gBP21 and gBP25 (Köller et al. 1997; Allen et al. 1998; Müller et al. 2001; Müller and Göringer 2002; Blom et al. 2001; Aphasizhev et al. 2003b; Schumacher et al. 2006), the proteins RBP16 (Hayman and Read 1999; Pelletier et al. 2000; Miller and Read 2003; Pelletier and Read 2003), TbRGG1 (Vanhamme et al. 1998; Hashimi et al. 2008), REAP1 (Madison-Antenucci et al. 1998; Madison-Antenucci and Hajduk 2001; Hans et al. 2007), and RBP38 (Sbicego et al. 2003). It also includes the 3′-end-specific TUTase KRET1 (Aphasizhev et al. 2002, 2003c; Aphasizheva and Aphasizhev 2010) and the mitochondrial DExH/D protein mHel61p (Missel and Göringer 1994; Missel et al. 1997; Göringer et al. 2008; Li et al. 2011). Although the presence of accessory editing factors is unquestioned, no knowledge with respect to the dynamic interplay of these proteins with the editosome exists.

7.5 Structure of the 20S Editosome

Raw EM images of purified 20S editosomes show a monodisperse population of elongated slightly bent particles with dimensions of up to 21–26 nm. The consensus structure of the complex is characterized by an elongated, slightly bent appearance (Fig. 7.2), which results in a concave/convex shape, displaying one concave and one convex contour on opposite sides. The particle is composed of two quasiglobular domains roughly equal in size. Both subdomains interact extensively in an interface region where a protruding arm extends on one side and a triangular protrusion emerges from the opposite side. The two subdomains differ in their structural details, indicating that 20S editosomes are no homodimers. The surface representation of the 20S particle was estimated to enclose a molecular mass of 800 $\pm$ 80 kDa. The value is consistent with the sum of all proteins identified in the biochemical analysis (790 kDa), assuming that every protein is present in single copy (Golas et al. 2009). A theoretical sedimentation coefficient was calculated to 21–26S, in line with the experimentally determined apparent value of 20–24S (Pollard et al. 1992; Corell et al. 1996).

7.6 Structure of the 35–40S Editosome

The consensus 3D structure of the 35–40S complex (Fig. 7.2) is similarly characterized by defined landmarks: an elongated, straight to slightly convex platform is packed against a semispherical element of variable size. The platform extends on two sides into small head-like protrusions one oriented to the top, the other one to the bottom forming a larger foot-like extension. The semispherical back is packed against the platform as a tight network and the interface between both elements is marked by incisions in the upper and lower part. The semispherical back element is asymmetric in

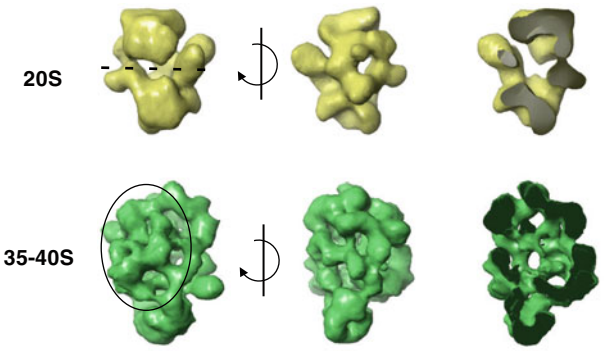


Fig. 7.2 Consensus 3D structures of *Trypanosoma brucei* RNA editing complexes derived from cryo-EM-based single-particle studies (Golas et al. 2009). *Top row*: Front (left) and back (center) surface views of the 20S RNA editing complex (EMDB-1595). The hatched line separates the two nonidentical globular subdomains. The structure to the right represents an inside view into the complex. For a comparison to the *Leishmania tarentolae* L-complex see Li et al. (2009). Bottom row: Front (*left*) and back (*center*) surface views of the 35–40S RNA editing complex (EMDB-1594). The semicircular back domain, which presumably represents bound substrate RNA, is encircled. The structure to the right shows the inside of the complex. Map data and image files can be retrieved at http://www.pdbj.org.emnavi/ (EM Navigator, Protein Data Bank Japan (PDBj)

its appearance displaying on one side a protruding shoulder-type element and on the other side an inclination or proclivity. A structure refinement of the 35–40S complex by cryo-negative staining EM resulted in a resolution of 13–19 nm and the surface of the complex was calculated to enclose a molecular mass of 1.45 ± 0.15MDa (Golas et al. 2009). The data have been used to calculate a theoretical sedimentation coefficient of 35–41S (García de la Torre et al. 2001), which is in agreement with the experimentally derived apparent sedimentation behavior observed in isokinetic glycerol gradients (Pollard et al. 1992; Peris et al. 1997).

7.7 Editosome Interconversion

The structural relationship between 20S and 35–40S editosomes has been investigated by different 3D docking and alignment approaches (Golas et al. 2009). The data suggest that 20S complexes represent a significant part of the platform density of the 35–40S complexes. 20S complexes can be integrated in the head domain, the upper part of the platform and part of the back domain of the 35–40S complex. By contrast, the foot domain, the lower part of the platform density and the majority of the semispherical back appears to be composed of components that have no structural correlate in the 20S particles. Since the main biochemical difference between 20S and 35–40S complexes is their RNA content, it

is tempting to speculate that the semispherical back of the 35–40S complexes represents gRNA and/or pre-mRNA molecules only—although the presence of a small number of additional proteins cannot be excluded. Experimental confirmation of the suggested scenario was gained from interconversion RNase digestion experiments. Treatment of 20S and 35–40S complexes with single- and double strand-specific RNases resulted in a concentration-dependent decrease in the amount of 35–40S editosomes and at the same time in an increase of 20S complexes. Conversely, incubation of isolated 20S editosomes with *T. brucei* mitochondrial RNA, which contains pre- and partially edited mRNAs as well as gRNAs, generated 35–40S complexes in a concentration-dependent fashion (Golas et al. 2009). The substrate RNA/20S interaction is of high affinity with equilibrium dissociation constants (K_d) in the nanomolar concentration range (Golas et al. 2009).

7.8 The Reaction Center

Next to unraveling the protein inventory of the 20S complex, attempts have been published to describe the interconnectivity and general organization of individual proteins within the particle (Schnaufer et al. 2003, 2010). Although the precise architecture is far from being understood, it is undisputed that functional 20S particles can be assembled in the absence of pre-edited mRNA and gRNA (Domingo et al. 2003). Furthermore, biochemical data suggest that insertion and deletion RNA editing may be catalyzed by separate editing subcomplexes or subdomains (for a review, see Carnes and Stuart 2008). This is seemingly supported by the fact that the editing core activities are present in protein pairs or even higher numbers (Stuart et al. 2005; Hajduk and Ochsenreiter 2010). Yeast two-hybrid and coimmunoprecipitation experiments identified possible protein/protein interaction partners among different editosomal proteins, which together suggest that 20S complexes consist of structurally separated insertion and deletion subdomains (Schnaufer et al. 2003). The insertion domain is assembled around a trimeric protein core involving the terminal uridylyltransferase (TUTase) TbMP57, the RNA editing ligase TbMP48, and protein TbMP81. The deletion subdomain relies on the heterotrimeric interaction of the exoUase TbMP99, the RNA editing ligase TbMP52, and protein TbMP63. Other polypeptides associate with both subdomains such as TbMP44 and TbMP24 (Wang et al. 2003; Salavati et al. 2006). TbMP18 likely represents a protein that is located at the interface between both subdomains (Law et al. 2007). Some of these results have been confirmed using TAP-tagged editosome preparations (reviewed in Carnes and Stuart 2008) and have provided insight into the structure/function correlation of various proteins in the 20S complex.

It is tempting to speculate that the compositionally different deletion and insertion subdomains correspond to the bipartite substructures of the 20S consensus model (see Fig. 7.2). The structure displays two nonidentical nearly globular subdomains, which are connected through an interface region in line with the above-described biochemical data. Whether that implies that both subdomains have independent RNA binding sites is not known to date. However, based on the structural details of the 35–40S complexes, in which the semispherical back of the complex was identified as bound substrate RNA, it is more likely that the 20S complex has only one RNA interaction site. The binding site covers a substantial part on one side of the 20S surface and in part reaches into the interface region in-between the two globular subdomains. Thus, a single catalytic core for both editing reactions might be positioned at the interface between the deletion and insertion subdomains representing a "dual mode" reaction center that is triggered by the chemical nature of the bound substrate RNA. Deletional editing substrates activate the deletion subdomain and insertion-type substrate RNAs become processed by the insertion subdomain.

7.9 Compositionally Distinct Editosomes

Evidence for functionally and compositionally distinct 20S complexes has been derived from the analysis of editing complexes isolated from transgenic trypanosome strains that conditionally express TAP-tagged variants of different editosomal proteins. Tandem affinity-purified editosomes possess in some cases only a subcollection of the 20S protein inventory and these compositional differences are mirrored in functional differences (Panigrahi et al. 2006). For instance, TbMP90–TAP editosomes execute only deletional RNA editing, while TbMP61–TAP complexes only process insertional editing substrate RNAs. Protein TbMP100 is only present in TbMP90–TAP editing complexes, while TbMP81, TbMP63, TbMP42, TbMP24, TbMP18, TbMP46, TbMP44, TbMP57, TbMP99, TbMP52, and TbMP48 seem to be ubiquitous. On the contrary, the three proteins TbMP99, TbMP61, and TbMP67 have been shown to be mutually exclusive in certain complexes. Based on these data, Carnes et al. (2008) suggested that at least three compositionally distinct editosomes exist. Whether these partially assembled machineries contribute to the endogenous, steady-state ensemble of editosomes in genetically unaltered, i.e., wildtype trypanosomes is not known. However, the data at least demonstrate that multiple assembly pathways exist for the processing machinery and that heterogeneous populations of 20S complexes are tolerated in in vivo isolates. Furthermore, it cannot be excluded that the composition of editosomes may be in a dynamic equilibrium. Individual components like the above described accessory factors (Göringer et al. 2008) may at different stages of the assembly pathway shuttle in and out of the complex thereby further expanding the structural landscape of the particle.

7.10 Conclusion

Taken together, the described experimental data can be summarized to derive a first picture that correlates some of the prominent structural features of the 20S and 35–40S complexes with functional attributes of the RNA editing reaction cycle. Steady-state editosome preparations in African trypanosomes consist of a mixture of two classes of high-molecular-mass complexes: 35–40S complexes represent the editing machinery that is actively engaged in the processing reaction, while 20S complexes are preassembled precursor complexes that consist of proteins only. The binding of substrate RNAs converts 20S editosomes into 35–40S complexes and depending on the type of RNA, differently shaped 35–40S complexes are formed. Altogether 12 pre-edited mRNAs and literally hundreds of gRNA and partially edited mRNA molecules are present in African trypanosomes. Guide RNAs have molecular masses around 20 kDa, while the involved mRNAs vary between 60 kDa (unedited CR3) and 450 kDa (edited ND7). As a consequence, their hydrodynamic radius must vary and at steady-state conditions, this results in a broad structural landscape of multiple 35–40S complexes with different shapes. However, these particles differ predominantly in only one structural feature: the semispherical back element. This part of the complex represents the RNA interaction site of the editosome, which implies that 20S editosomes have only one RNA binding domain that interacts with a large structural ensemble of differently sized and folded substrate RNAs. The RNA binding site covers a large surface area on one side of the 20S particle and contacts the interface between the two nonidentical, globular subdomains of the 20S editosome.

The protein-only 20S editosome is characterized by a bipartite appearance with two prominent globular subdomains. Both subdomains, however, differ in their structural details indicating that the particle is not a homodimer. In conjunction with the described biochemical data, these substructures likely represent separate insertion and deletion subdomains and they are assembled around different protein core elements in line with the different enzyme requirements that catalyze the two types of editing reactions. The two subdomains connect in an interface region, which is linked to the RNA binding domain of the complex. This suggests a working model in which the editosome reaction center is located at the interface of the insertional and deletional subdomains thereby presenting a catalytic core of bifunctional quality. As a result, editosome-bound substrate RNA can be in physical contact with both catalytic machineries and depending on the RNA editing domain, both, U-insertions and deletions can be executed on the same pre-mRNA. This is further supported by the fact that single gRNAs can mediate insertion as well as deletion editing (Blum et al. 1990; Maslov and Simpson 1992). Understanding the conformational dynamic and adaptive recognition at the catalytic center of the editosome will be the next experimental and perhaps conceptual challenge.

Acknowledgements This work was supported by the Howard Hughes Medical Institute (HHMI), the German Research Foundation (DFG), and the Dr. Illing Foundation. The author would like to thank all members of the laboratory (past and present) for their contributions and their enthusiasm. Mike Roskope is acknowledged for discussions and comments on the manuscript. Holger Stark, Monika M. Golas, and Bjoern Sander are thanked for a productive cooperation.

References

Allen TE, Heidmann S, Reed R, Myler PJ, Göringer HU, Stuart KD (1998) Association of guide RNA binding protein gBP21 with active RNA editing complexes in *Trypanosoma brucei*. Mol Cell Biol 18:6014–6022

Aphasizhev R, Sbicego S, Peris M, Jang SH, Aphasizheva I, Simpson AM, Rivlin A, Simpson L (2002) Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. Cell 108:637–648

Aphasizhev R, Aphasizheva I, Nelson RE, Gao G, Simpson AM, Kang X, Falick AM, Sbicego S, Simpson L (2003a) Isolation of a U-insertion/deletion editing complex from *Leishmania tarentolae* mitochondria. EMBO J 22:913–924

Aphasizhev R, Aphasizheva I, Nelson RE, Simpson L (2003b) A 100-kD complex of two RNA-binding proteins from mitochondria of *Leishmania tarentolae* catalyzes RNA annealing and interacts with several RNA editing components. RNA 9:62–76

Aphasizhev R, Aphasizheva I, Simpson L (2003c) A tale of two TUTases. Proc Natl Acad Sci USA 100:10617–10622

Aphasizheva I, Aphasizhev R (2010) RET1-catalyzed uridylylation shapes the mitochondrial transcriptome in *Trypanosoma brucei*. Mol Cell Biol 30:1555–1567

Benne R, van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC (1986) Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46:819–826

Blom D, van den Burg J, Breek CK, Speijer D, Muijsers AO, Benne R (2001) Cloning and characterization of two guide RNA-binding proteins from mitochondria of *Crithidia fasciculata*: gBP27, a novel protein, and gBP29, the orthologue of *Trypanosoma brucei* gBP21. Nucleic Acids Res 29:2950–2962

Blum B, Simpson L (1990) Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo (U) tail involved in recognition of the preedited region. Cell 62:391–397

Blum B, Simpson L (1992) Formation of guide RNA/messenger RNA chimeric molecules in vitro, the initial step of RNA editing, is dependent on an anchor sequence. Proc Natl Acad Sci USA 89:11944–11948

Blum B, Bakalara N, Simpson L (1990) A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell 60:189–198

Blum B, Sturm NR, Simpson AM, Simpson L (1991) Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. Cell 65:543–550

Brecht M, Niemann M, Schlüter E, Müller UF, Stuart K, Göringer HU (2005) TbMP42, a protein component of the RNA editing complex in African trypanosomes has endo-exoribonuclease activity. Mol Cell 17:621–630

Carnes J, Stuart K (2007) Uridine insertion/deletion editing activities. Methods Enzymol 424:25–54

Carnes J, Stuart K (2008) Working together: the RNA editing machinery in *Trypanosoma brucei*. In: Göringer HU (ed) RNA editing, nucleic acids and molecular biology, vol 20. Springer, Heidelberg, pp 143–164

Carnes J, Trotter JR, Ernst NL, Steinberg A, Stuart K (2005) An essential RNase III insertion editing endonuclease in *Trypanosoma brucei*. Proc Natl Acad Sci USA 102:16614–16619

Carnes J, Trotter JR, Peltan A, Fleck M, Stuart K (2008) RNA editing in *Trypanosoma brucei* requires three different editosomes. Mol Cell Biol 28:122–130

Corell RA, Read LK, Riley GR, Nellissery JK, Allen TE, Kable ML, Wachal MD, Seiwert SD, Myler PJ, Stuart KD (1996) Complexes from *Trypanosoma brucei* that exhibit deletion editing and other editing-associated properties. Mol Cell Biol 16:1410–1418

Decker CJ, Sollner-Webb B (1990) RNA editing involves indiscriminate U changes throughout precisely defined editing domains. Cell 61:1001–1011

Deng J, Schnaufer A, Salavati R, Stuart KD, Hol WG (2004) High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. J Mol Biol 343:601–613

Domingo GJ, Palazzo SS, Wang B, Pannicucci B, Salavati R, Stuart KD (2003) Dyskinetoplastic *Trypanosoma brucei* contains functional editing complexes. Eukaryot Cell 2:569–577

Ernst NL, Panicucci B, Igo RP Jr, Panigrahi AK, Salavati R, Stuart K (2003) TbMP57 is a 3′ terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. Mol Cell 11:1525–1536

Frech GC, Simpson L (1996) Uridine insertion into preedited mRNA by a mitochondrial extract from *Leishmania tarentolae*: stereochemical evidence for the enzyme cascade model. Mol Cell Biol 16:4584–4589

Gao G, Simpson L (2003) Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing and is the REL2 RNA ligase specific for U-insertion editing? J Biol Chem 278:27570–27574

García de la Torre J, Llorca O, Carrascosa JL, Valpuesta JM (2001) HYDROMIC: prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy images. Eur Biophys J 30:457–462

Golas MM, Böhm C, Sander B, Effenberger K, Brecht M, Stark H, Göringer HU (2009) Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages *in vivo*. EMBO J 28:766–778

Göringer HU, Koslowsky DJ, Morales TH, Stuart K (1994) The formation of mitochondrial ribonucleoprotein complexes involving gRNA molecules in *Trypanosoma brucei*. Proc Natl Acad Sci USA 91:1776–1780

Göringer HU, Brecht M, Böhm C, Kruse E (2008) RNA editing accessory factors—the example of mHel61p. In: Göringer HU (ed) RNA editing, nucleic acids and molecular biology, vol 20. Springer, Heidelberg, pp 165–179

Hajduk S, Ochsenreiter T (2010) RNA editing in kinetoplastids. RNA Biol 7:229–236

Hans J, Hajduk SL, Madison-Antenucci S (2007) RNA-editing-associated protein 1 null mutant reveals link to mitochondrial RNA stability. RNA 13:881–889

Hashimi H, Zíková A, Panigrahi AK, Stuart K, Lukeš J (2008) TbRGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex. RNA 14:970–980

Hauser R, Pypaert M, Häusler T, Horn EK, Schneider A (1996) *In vitro* import of proteins into mitochondria of *Trypanosoma brucei* and *Leishmania tarentolae*. J Cell Sci 109:517–523

Hayman ML, Read LK (1999) *Trypanosoma brucei* RBP16 is a mitochondrial Y-box family protein with guide RNA binding activity. J Biol Chem 274:12067–12074

Hermann T, Schmid B, Heumann H, Göringer HU (1997) A three-dimensional working model for a guide RNA from *Trypanosoma brucei*. Nucleic Acids Res 25:2311–2318

Igo RP Jr, Palazzo SS, Burgess ML, Panigrahi AK, Stuart K (2000) Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing. Mol Cell Biol 20:8447–8457

Igo RP Jr, Weston DS, Ernst NL, Panigrahi AK, Salavati R, Stuart K (2002) Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing. Eukaryot Cell 1:112–118

Kable ML, Seiwert SD, Heidmann S, Stuart K (1996) RNA editing: A mechanism for gRNA-specified uridylate insertion into precursor mRNA. Science 273:1189–1195

Kang X, Rogers K, Gao G, Falick AM, Zhou S, Simpson L (2005) Reconstitution of uridine-deletion precleaved RNA editing with two recombinant enzymes. Proc Natl Acad Sci USA 102:1017–1022

Kastner B, Fischer N, Golas MM, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E, Uchtenhagen H, Urlaub H, Herzog F, Peters JM, Poerschke D, Lührmann R, Stark H (2008) GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods 5:53–55

Köller J, Nörskau G, Paul AS, Stuart K, Göringer HU (1994) Different *Trypanosoma brucei* guide RNA molecules associate with an identical complement of mitochondrial proteins *in vitro*. Nucleic Acids Res 22:1988–1995

Köller J, Müller UF, Schmid B, Missel A, Kruft V, Stuart K, Göringer HU (1997) *Trypanosoma brucei* gBP21: An arginine-rich mitochondrial protein that binds to guide RNA with high affinity. J Biol Chem 272:3749–3757

Koslowsky DJ, Göringer HU, Morales TH, Stuart K (1992) *In vitro* guide RNA/mRNA chimaera formation in *Trypanosoma brucei* RNA editing. Nature 356:807–809

Law JA, O'Hearn S, Sollner-Webb B (2007) In *Trypanosoma brucei* RNA editing, TbMP18 (band VII) is critical for editosome integrity and for both insertional and deletional cleavages. Mol Cell Biol 27:777–787

Leung SS, Koslowsky DJ (2001) Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to its cognate gRNA. RNA 7:1803–1816

Li F, Ge P, Hui WH, Atanasov I, Rogers K, Guo Q, Osato D, Falick AM, Zhou ZH, Simpson L (2009) Structure of the core editing complex (L-complex) involved in uridine insertion/deletion RNA editing in trypanosomatid mitochondria. Proc Natl Acad Sci USA 106:12306–12310

Li F, Herrera J, Zhou S, Maslov DA, Simpson L (2011) Trypanosome REH1 is an RNA helicase involved with the 3′-5′ polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. Proc Natl Acad Sci USA 108:3542–3547

Lukeš J, Hashimi H, Zikova A (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. Curr Genet 48:277–299

Madison-Antenucci S, Hajduk SL (2001) RNA editing-associated protein 1 is an RNA binding protein with specificity for preedited mRNA. Mol Cell 7:879–886

Madison-Antenucci S, Sabatini RS, Pollard VW, Hajduk SL (1998) Kinetoplastid RNA-editing-associated protein 1 (REAP-1): a novel editing complex protein with repetitive domains. EMBO J 17:6368–6376

Maslov DA, Simpson L (1992) The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell 70:459–467

Maslov DA, Thiemann O, Simpson L (1994) Editing and misediting of transcripts of the kinetoplast maxicircle G5 (ND3) cryptogene in an old laboratory strain of *Leishmania tarentolae*. Mol Biochem Parasitol 68:155–159

McKean PG (2003) Coordination of cell cycle and cytokinesis in *Trypanosoma brucei*. Curr Opin Microbiol 6:600–607

McManus MT, Shimamura M, Grams J, Hajduk SL (2001) Identification of candidate mitochondrial RNA editing ligases from *Trypanosoma brucei*. RNA 7:167–175

Miller MM, Read LK (2003) *Trypanosoma brucei*: functions of RBP16 cold shock and RGG domains in macromolecular interactions. Exp Parasit 105:140–148

Missel A, Göringer HU (1994) *Trypanosoma brucei* mitochondria contain RNA helicase activity. Nucleic Acids Res 22:4050–4056

Missel A, Souza AE, Nörskau G, Göringer HU (1997) Gene disruption of a mitochondrial DEAD box protein in *Trypanosoma brucei* affects edited mRNAs. Mol Cell Biol 17:4895–4903

Müller UF, Göringer HU (2002) Mechanism of the gBP21-mediated RNA/RNA annealing reaction: matchmaking and charge reduction. Nucleic Acids Res 30:447–455

Müller UF, Lambert L, Göringer HU (2001) Annealing of RNA editing substrates facilitated by guide RNA-binding protein gBP21. EMBO J 20:1394–1404

Niemann M, Brecht M, Schlüter E, Weitzel K, Zacharias M, Göringer HU (2008) TbMP42 is a structure-sensitive ribonuclease that likely follows a metal ion catalysis mechanism. Nucleic Acids Res 36:4465–4473

Niemann M, Kaibel H, Schlüter E, Weitzel K, Brecht M, Göringer HU (2009) Kinetoplastid RNA editing involves a 3′ nucleotidyl phosphatase activity. Nucleic Acids Res 37:1897–1906

Ochsenreiter T, Hajduk SL (2006) Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. EMBO Rep 7:1128–1133

Ochsenreiter T, Anderson S, Wood ZA, Hajduk S (2008a) Alternative RNA editing produces a novel protein involved in mitochondrial DNA maintenance in trypanosomes. Mol Cell Biol 28:5595–5604

Ochsenreiter T, Cipriano M, Hajduk SL (2008b) Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. PLoS One 3:e1566

Panigrahi AK, Schnaufer A, Carmean N, Igo RP Jr, Gygi SP, Ernst NL, Palazzo SS, Weston DS, Aebersold R, Salavati R, Stuart KD (2001) Four related proteins of the *Trypanosoma brucei* RNA editing complex. Mol Cell Biol 21:6833–6840

Panigrahi AK, Ernst NL, Domingo GJ, Fleck M, Salavati R, Stuart KD (2006) Compositionally and functionally distinct editosomes in *Trypanosoma brucei*. RNA 12:1038–1049

Panigrahi AK, Schnaufer A, Stuart KD (2007) Isolation and compositional analysis of trypanosomatid editosomes. Methods Enzymol 424:3–24

Pelletier M, Read LK (2003) RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAs in *Trypanosoma brucei*. RNA 9:457–468

Pelletier M, Miller MM, Read LK (2000) RNA-binding properties of the mitochondrial Y-box protein RBP16. Nucleic Acids Res 28:1266–1275

Peris M, Simpson AM, Grunstein J, Liliental JE, Frech GC, Simpson L (1997) Native gel analysis of ribonucleoprotein complexes from a *Leishmania tarentolae* mitochondrial extract. Mol Biochem Parasitol 85:9–24

Pollard VW, Harris ME, Hajduk SL (1992) Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. EMBO J 11:4429–4438

Reifur L, Koslowsky DJ (2008) *Trypanosoma brucei* ATPase subunit 6 mRNA bound to gA6-14 forms a conserved three-helical structure. RNA 14:2195–2211

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotech 17:1030–1032

Rogers K, Gao G, Simpson L (2007) Uridylate-specific 3′ 5′-exoribonucleases involved in uridylate-deletion RNA editing in trypanosomatid mitochondria. J Biol Chem 282:29073–29080

Rusché LN, Cruz-Reyes J, Piller KJ, Sollner-Webb B (1997) Purification of a functional enzymatic editing complex from *Trypanosoma brucei* mitochondria. EMBO J 16:4069–4081

Rusché LN, Huang CE, Piller KJ, Hemann M, Wirtz E, Sollner-Webb B (2001) The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: cloning the essential band IV gene and identifying the band V gene. Mol Cell Biol 21:979–989

Sabatini R, Hajduk SL (1995) RNA ligase and its involvement in guide RNA/mRNA chimera formation. Evidence for a cleavage-ligation mechanism of *Trypanosoma brucei* mRNA editing. J Biol Chem 270:7233–7240

Salavati R, Ernst NL, O'Rear J, Gilliam T, Tarun S Jr, Stuart K (2006) KREPA4, an RNA binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. RNA 12:819–831

Sbicego S, Alfonzo JD, Estevez AM, Rubio MA, Kang X, Turck CW, Peris M, Simpson L (2003) RBP38, a novel RNA-binding protein from trypanosomatid mitochondria, modulates RNA stability. Eukaryot Cell 2:560–568

Schmid B, Riley GR, Stuart K, Göringer HU (1995) The secondary structure of guide RNA molecules from *Trypanosoma brucei*. Nucleic Acids Res 23:3093–3102

Schmid B, Read LK, Stuart K, Göringer HU (1996) Experimental verification of the secondary structures of guide RNA-pre-mRNA chimaeric molecules in *Trypanosoma brucei*. Eur J Biochem 240:721–731

Schnaufer A, Ernst NL, Palazzo SS, O'Rear J, Salavati R, Stuart K (2003) Separate insertion and deletion subcomplexes of the *Trypanosoma brucei* RNA editing complex. Mol Cell 12:307–319

Schnaufer A, Wu M, Park YJ, Nakai T, Deng J, Proff R, Hol WG, Stuart KD (2010) A protein-protein interaction map of trypanosome 20S editosomes. J Biol Chem 285:5282–5295

Schumacher MA, Karamooz E, Zikova A, Trantirek L, Lukeš J (2006) Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. Cell 126:701–711

Seiwert SD, Stuart K (1994) RNA editing: transfer of genetic information from gRNA to precursor mRNA *in vitro*. Science 266:114–117

Seiwert SD, Heidmann S, Stuart K (1996) Direct visualization of uridylate deletion *in vitro* suggests a mechanism for kinetoplastid RNA editing. Cell 84:831–841

Simpson L, Aphasizhev R, Lukeš J, Cruz-Reyes J (2010) Guide to the nomenclature of kinetoplastid RNA editing: a proposal. Protist 161:2–6

Stuart KD, Schnaufer A, Ernst NL, Panigrahi AK (2005) Complex management: RNA editing in trypanosomes. Trends Biochem Sci 30:97–105

Trotter JR, Ernst NL, Carnes J, Panicucci B, Stuart K (2005) A deletion site editing endonuclease in *Trypanosoma brucei*. Mol Cell 20:403–412

Vanhamme L, Perez-Morga D, Marchal C, Speijer D, Lambert L, Geusken M, Alexandre S, Ismaïli N, Göringer HU, Benne R, Pays E (1998) *Trypanosoma brucei* TBRGG1, a mitochondrial oligo(U)-binding protein that co-localizes with an *in vitro* RNA editing activity. J Biol Chem 273:21825–21833

Wang B, Ernst NL, Palazzo SS, Panigrahi AK, Salavati R, Stuart K (2003) TbMP44 is essential for RNA editing and structural integrity of the editosome in *Trypanosoma brucei*. Eukaryot Cell 2:578–587

Yu LE, Koslowsky DJ (2006) Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of a gRNA bound to its cognate mRNA. RNA 12:1050–1060

Chapter 8
The RNA Interference Pathway in *Trypanosoma brucei*

Elisabetta Ullu, Nikolay G. Kolev, Rebecca L. Barnes, and Christian Tschudi

Abstract In most eukaryotic cells, expression or delivery of long double-stranded RNA (dsRNA) signals the presence of foreign and/or potentially dangerous nucleic acids, such as viruses or transcripts derived from retroposons and transposons. As a consequence cells go on red alert and activate specific defense mechanisms to eliminate the invaders. Among such mechanisms is the RNA interference (RNAi) pathway, which is triggered by long dsRNAs, destroys target transcripts in a sequence-specific manner, and is widespread throughout eukaryotic evolution. Here we summarize our current understanding of the RNAi mechanism in *Trypanosoma brucei*, a protozoan parasite and early divergent eukaryote, and highlight similarities and differences with the RNAi machinery and its function in higher eukaryotes.

Keywords Argonaute • Dicer • RNAi genes • siRNA

E. Ullu (✉)
Department of Internal Medicine, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536-0812, USA

Department of Cell Biology, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536-0812, USA
e-mail: elisabetta.ullu@yale.edu

N.G. Kolev • C. Tschudi
Division of Epidemiology of Microbial Diseases, Yale University School of Public Health, 295 Congress Avenue, New Haven, CT 06536-0812, USA
e-mail: nikolay.kolev@yale.edu; christian.tschudi@yale.edu

R.L. Barnes
Department of Internal Medicine, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536-0812, USA
e-mail: rebecca.l.barnes@yale.edu

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_8,

8.1 Overview of RNA Interference

Genetic interference triggered by double-stranded RNA (dsRNA), commonly referred to as RNA interference or RNAi, was first described in 1998 by the laboratories of Andrew Fire and Craig Mello (Fire et al. 1998). The subsequent discovery of a plethora of small 19–30 nucleotide (nt) noncoding RNAs that function in diverse pathways, ranging from mRNA degradation, translational repression, heterochromatin formation, and DNA elimination, has radically changed our understanding of how gene expression is regulated in eukaryotes. In addition, RNAi provided the scientific community with a tool of immense value for the analysis of gene function, especially in organisms not easily amenable to classical genetic approaches.

As surmised by studies in several model organisms, the "natural" or endogenous RNAi pathway functions as a defense mechanism to maintain genome integrity by destroying transcripts derived from molecular parasites, namely viruses, transposons, retroposon, and transgenes, thus limiting viral infection and the catastrophic consequences of transposon/retroposon mobilization (Siomi and Siomi 2009). RNAi is widespread throughout the eukaryotic lineage, from protists to man and based on phylogenetic considerations, it is likely that the ancestral eukaryote was endowed with a functional RNAi pathway (Cerutti and Casas-Mollano 2006). In this review, we focus on the parasitic protozoa *Trypanosoma brucei*, one of the few pathogenic single-cell eukaryotes where the mechanism and functional significance of RNAi has been studied in detail. For more details on RNAi, readers are referred to recent reviews (Siomi and Siomi 2009; Liu and Paroo 2010).

8.2 RNAi in Model Organisms

The minimal set of RNAi components consists of Dicer, an RNase III-family enzyme presiding over the processing of dsRNAs into small interfering RNAs (siRNAs); a "slicer" Argonaute (AGO), an endonuclease with an RNase H-like domain required for target RNA cleavage; and a dsRNA-binding protein (dsRBP), which facilitates Dicer cleavage and assists in transferring siRNAs into AGO (Liu and Paroo 2010). In certain organisms, including *Arabidopsis thaliana*, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe*, a fourth RNAi factor, an RNA-dependent-RNA polymerase (RdRP), amplifies the initial RNAi response, resulting in the production of secondary siRNAs (Cerutti and Casas-Mollano 2006).

In the current model, the RNAi pathway is subdivided into two phases, namely the initiation and the effector steps. In the first phase Dicer, in a complex with the dsRBP and AGO, processes dsRNA into duplex siRNAs with characteristic two nucleotide 3′ extensions, which are then transferred to AGO slicer. AGO cleaves the "passenger" siRNA strand, the strand that is destined to be ejected, while retaining the "guide" siRNA strand. Recent evidence indicates that the process of

siRNA loading into AGO is facilitated by the Hsp90/70 chaperone system, which probably maintains AGO in a conformation suitable for accepting siRNAs (Iwasaki et al. 2010; Miyoshi et al. 2010). The assembly of the AGO-guide siRNA complex leads to the formation of active RISC (RNA-induced silencing complex), of which AGO is the catalytic engine. In the effector step the AGO ribonucleoprotein seeks out and then cleaves homologous target RNA, approximately in the middle of the region of complementarity with the siRNA. The cleaved RNA fragments are then released and degraded by cellular ribonucleases, while RISC is recycled for another round of slicing. It should be noted that the above pathway is based in large parts on studies in higher eukaryotes and considerably less is known in single-cell eukaryotes.

8.3 RNAi in *Trypanosoma brucei*

Ever since the discovery of RNAi in *Trypanosoma brucei* in 1998 (Ngo et al. 1998), a new era began in the study of the biology of this important human pathogen. It is not an exaggeration to say that since the beginning of the year 2000 the way experiments are carried out in *T. brucei* has undergone a major revolution, and gene silencing by RNAi has found wide applications, including forward genetic screens and global analysis of gene function (see Sect. 8.3.10).

To date functional studies have uncovered three components that are essential for RNAi in *T. brucei*: *Tb*AGO1, a member of the Argonaute family of proteins (Durand-Dubief and Bastin 2003; Shi et al. 2004b) and two distinct Dicer-like enzymes, namely *Tb*DCL1 and *Tb*DCL2 (Shi et al. 2006b; Patrick et al. 2009). However, comparative genomics indicates that there are at least two more genes that are candidates for being RNAi factors (see note added in proof), as they are present in RNAi-proficient trypanosomatids, namely the African trypanosomes, *Leishmania Viannia braziliensis* and *Crithidia fasciculata*, but absent in *T. cruzi*, *L. major*, and *L. donovani* (Lye et al. 2010), which are RNAi-negative (Robinson and Beverley 2003; DaRocha et al. 2004). Inspection of the *T. cruzi* and old-world *Leishmania* genomes shows no remnants of the DCL1 and DCL2 genes, whereas in the case of AGO1 a pseudogene is present in *Leishmania* but absent in *T. cruzi*. The loss of the RNAi genes appeared to have occurred twice in the trypanosomatid lineage (Lye et al. 2010), once at some undefined point in the lineage leading to *T. cruzi* and a second time after the divergence of the *Viannia* subgenus from other *Leishmania* species.

A schematic representation of our current view of the RNAi pathway in *T. brucei* is illustrated in Fig. 8.1. *Tb*AGO1, in a complex with single-stranded siRNAs, is found predominantly in the cytoplasm and is essential for mRNA degradation. However, functional evidence supports the model that *Tb*AGO1 is active in the nucleus as well (see Sect. 8.3.3). Notably, the two Dicers are localized in different cellular environments, with *Tb*DCL1 and *Tb*DCL2 predominantly in the cytoplasm

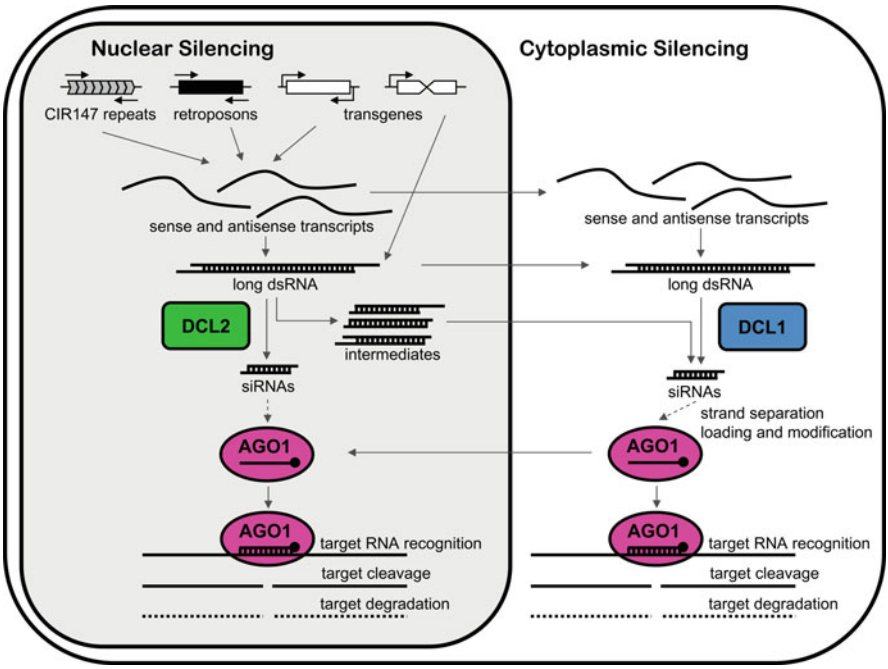


Fig. 8.1 Current working model of RNAi in *T. brucei*. Depicted are the two branches (nuclear and cytoplasmic) of silencing. Sense and antisense (or inverted repeat) RNAs transcribed in the nucleus form long double-stranded RNA (dsRNA) in both cellular compartments. DCL2 and DCL1 process the long RNA duplexes into siRNA duplexes in the nucleus and the cytoplasm, respectively. DCL2 also generates intermediate size dsRNAs, which are converted to siRNAs by DCL1. Argonaute (AGO1) is programmed with single-stranded "guide" siRNA as a result of duplex siRNA strand separation, loading, and modification of the 3′ end of the siRNA. AGO1 finds transcripts complementary to the bound siRNA in both compartments of the cell, and base pairing between the siRNA and the target leads to endonucleolytic cleavage and subsequent degradation of the target RNA

and the nucleus, respectively. In the following sections, we will review our current understanding of the functional significance for the compartmentalization of the two Dicers.

8.3.1 The Cytoplasmic Dicer-Like Protein TbDCL1

A number of early observations predicted the existence of Dicer or Dicer-like protein(s) in *T. brucei*, including the identification of siRNAs (Djikeng et al. 2001), the in vivo RNAi response to transfected long dsRNAs (Ngo et al. 1998), and the detection of an enzymatic activity in trypanosome extracts generating siRNA-size molecules (Best et al. 2005). Dicer and Dicer-like proteins are large multidomain proteins with the canonical signature of two neighboring RNase III

domains, termed RNase IIIa and RNase IIIb (Jaskiewicz and Filipowicz 2008). However, initial homology searches of the *T. brucei* genome with the *Aquifex aeolicus* RNase III domain only returned predicted ORFs with a single RNase III domain (Shi et al. 2006b). Nevertheless, what caught our attention was that two ORFs of 1,648 and 948 amino acids were present in *T. brucei*, the only RNAi-positive trypanosomatid protozoa known at that time, but absent in the RNAi-negative organisms *T. cruzi* and *L. major* (Ullu et al. 2004). Furthermore, as judged by the cellular distribution of GFP fusion proteins, the 1,648 and 948 amino acid-long proteins localized predominantly to the cytoplasm and nucleus, respectively (Shi et al. 2006b). Subsequent refined bioinformatics analysis and the availability of additional genomes of RNAi-positive African trypanosomes allowed the identification of a second RNase III domain in both candidates. Thus, the ORFs of 1,648 and 948 amino acids were named Dicer-like protein 1 or *Tb*DCL1 (Tb927.8.2370) and Dicer-like protein 2 or *Tb*DCL2 (Tb927.3.1230), respectively. *Tb*DCL2 will be discussed in Sect. 8.3.2.

Typically, RNase IIIa and RNase IIIb domains in metazoan and fungal Dicers are adjacent to each other (Jaskiewicz and Filipowicz 2008). However, in *Tb*DCL1 the RNase IIIa domain is close to the amino terminus, and the RNase IIIb domain is located approximately in the center of the molecule. No domains commonly associated with Dicer or Dicer-like proteins, like helicase, PAZ, DUF283, or dsRNA-binding domains, were recognizable along the *Tb*DCL1 polypeptide chain (Shi et al. 2006b). Although the apparent lack of additional domains was surprising at first, the large dataset available to date clearly revealed that Dicer and Dicer-like proteins have diversified substantially throughout eukaryotic evolution, making the two RNase III domains the only universal feature of this class of proteins. It is likely that the diverse domain structure observed in protozoan, fungal, plant, and metazoan proteins reflects organism-specific adaptations.

The RNase IIIa and IIIb domains in *Tb*DCL1 have the potential to form an intramolecular pseudodimer and to build the typical Dicer dsRNA processing center (Jaskiewicz and Filipowicz 2008) composed of two active sites (from the two RNase III domains), each able to perform scission of one of the two strands in the RNA duplex. Indeed, partially purified *Tb*DCL1 catalyzes Mg^{2+}-dependent processing of dsRNA into 24–26 nucleotide-long siRNAs, a size range typical for endogenous siRNAs in *T. brucei* (Shi et al. 2006b). In vivo, the activity of *Tb*DCL1 is responsible, but only in part, for the production of siRNAs from retroposon transcripts, which possibly escape the nucleus, for processing of transfected long dsRNA, and for the cleavage of 35–65 nucleotide dsRNA intermediates generated by *Tb*DCL2 (see Sect. 8.3.2). However, how *Tb*DCL1 recognizes a dsRNA substrate is at present unknown, since PAZ and dsRNA-binding motifs, which are essential for positioning the dsRNA at the catalytic center (MacRae et al. 2007), have so far not been identified.

Although Dicer enzymes have a primary role at the initiation step of RNAi, namely the processing of the dsRNA trigger to siRNAs, there is also evidence for a role in the effector step, namely the cleavage of the target transcript by AGO slicer (Jaskiewicz and Filipowicz 2008). This appears to be the case for *Tb*DCL1 as well,

since a point mutation of a conserved residue outside of the RNase III domains (D966G) does not detectably affect the production of siRNAs, but has a significant dominant negative effect on the cleavage of target mRNAs by *Tb*AGO1 (Shi et al. 2006b). Thus, one prediction would be that *Tb*DCL1 and *Tb*AGO1 interact, yet evidence for such a complex has been elusive so far.

8.3.2 The Nuclear TbDCL2 Initiates the Endogenous RNAi Response

In addition to the bioinformatic analysis described above, experimental evidence provided clues that *T. brucei* contained a second Dicer or Dicer-like enzyme. Firstly, genetic ablation of *Tb*DCL1 did not result in the upregulation of ingi or SLACS retroposon transcripts or transcripts derived from a family of 147-nucleotide-long chromosomal internal repeats, termed CIR147 repeats (Patrick et al. 2009), which are located at putative centromeric regions (Obado et al. 2005). This phenomenon was first observed in *Tb*AGO1 null cells and constitutes the hallmark of RNAi deficiency (Shi et al. 2004b). Secondly, the accumulation of CIR147 siRNAs was independent of *Tb*DCL1, and the siRNAs from ingi and SLACS were only partially depleted in *Tb*DCL1 null cells (Patrick et al. 2009). Thus, we reasoned that a second Dicer must be at work in the trypanosome RNAi pathway and *Tb*DCL2 was the most likely candidate, since sequence homology analysis and three-dimensional structure modeling revealed two RNase III domains and the remnants of a dsRNA-binding domain at the very C terminus (Patrick et al. 2009). Similar to *Tb*DCL1, the two RNase III domains in *Tb*DCL2 are not adjacent, but are found in close proximity to the two ends of the polypeptide, with RNase IIIa at the N terminus and RNase IIIb at the C terminus. One would assume that the two RNase III domains form an intramolecular pseudodimer, since partially purified wild-type *Tb*DCL2 is capable of processing long dsRNA into siRNA-size products in a Mg^{2+}-dependent fashion and single-point mutations in the active site of the RNase IIIb module (K879A and E886A) severely compromised this activity (Patrick et al. 2009). *Tb*DCL1 and *Tb*DCL2 appear to generate duplex siRNA products of slightly different size, as revealed by in vitro dicing assays with *Tb*DCL1 null and *Tb*DCL2 null whole cell extracts. The size range of *Tb*DCL2-generated siRNAs is one nucleotide shorter than the *Tb*DCL1-generated siRNAs; however, the significance of this difference is currently not clear. In addition, *Tb*DCL2 is responsible for the production of intermediate-size RNA molecules (35–65 nt long), which accumulate in the absence of *Tb*DCL1 (Patrick et al. 2009). These molecules are possibly double stranded, since in the case of CIR147 intermediates, both sense and antisense sequences are detected with similar frequency and their distribution is evenly spaced across the entire CIR147 repeat. The intermediate-size RNAs contain a 5′-monophosphate, as indicated by their sensitivity to treatment with Terminator exonuclease (Patrick et al. 2009), which is

consistent with their biogenesis depending on *Tb*DCL2. Although the relevance of the intermediates in the RNAi pathway remains to be investigated, one possibility is that they are side products of *Tb*DCL2 processing of long dsRNA. If *Tb*DCL2 requires substrates with a double-stranded region longer than ~65 bp for binding and processing, then RNA molecules with duplexes shorter than 35–65 bp and single-stranded extensions (5′ or 3′) would not be recognized. These "noncanonical" substrates appear to find their way to *Tb*DCL1, and this Dicer readily accepts them. In vivo, the intermediate RNAs gradually disappear upon increasing the levels of *Tb*DCL1 in the cell (Patrick et al. 2009) with concomitant increase in the amount of corresponding siRNAs, which indicates that the intermediates contain double-stranded regions of at least ~25 bp and serve as a source for siRNA production by *Tb*DCL1.

A *Tb*DCL2-GFP fusion protein localized predominantly to the nucleus, suggesting that the RNAi machinery is present in this compartment. Three lines of evidence support this hypothesis. Firstly, CIR147 transcripts in *Tb*AGO1 null cells are found exclusively in the nucleus, close to or at the nucleolus (Patrick et al. 2009). Secondly, in the absence of *Tb*DCL2, CIR147 siRNAs are below the limit of detection, indicating that their biogenesis is mostly dependent on *Tb*DCL2 (Patrick et al. 2009). Thirdly, work by the group of Michaeli has shown that small nucleolar RNAs (snoRNAs), which as the name indicates are only found within the confines of the nucleus, could be targeted for destruction by expressing homologous dsRNA, a phenomenon termed snoRNAi (Liang et al. 2003). In addition, our two laboratories have recently provided genetic evidence that snoRNAi requires *Tb*DCL2 and *Tb*AGO1, but not cytoplasmic *Tb*DCL1 (Gupta et al. 2010). Taken together, these observations strongly support the existence of a nuclear phase of the RNAi pathway. Since in the absence of *Tb*DCL2 retroposon and CIR147 transcripts accumulate, a phenotype observed in *Tb*AGO1 null but not in *Tb*DCL1 null cells, we proposed that nuclear *Tb*DCL2 initiates the endogenous RNAi response (Patrick et al. 2009). Indeed, *Tb*DCL2 appears to be the sole enzymatic activity producing siRNAs from dsRNAs retained in the nucleus, and its presence is critical for successful RNAi-mediated knockdown approaches targeting nuclear RNAs (Gupta et al. 2010). Notably, knockout of *Tb*DCL2 results in "hypersensitivity" to exogenous dsRNA and synthetic siRNAs, a phenotype that likely reflects decreased competition in *Tb*DCL2 null cells for the formation of active RISC between siRNAs from endogenous and exogenous dsRNA triggers (Patrick et al. 2009). This observation further highlights the crucial role played by *Tb*DCL2 in initiating and sustaining the endogenous RNAi pathway.

Why do trypanosomes have two Dicers? Whereas some of the roles of *Tb*DCL1 and *Tb*DCL2 overlap, namely the generation of siRNAs from retroposons and ensuring *Tb*AGO1 programming against transcripts from these mobile genetic elements to safeguard the trypanosome genome, the two Dicers also have specific functions in the RNAi pathway. *Tb*DCL2 is predominantly nuclear and is the enzyme producing siRNAs from dsRNAs formed by transcripts that do not leave the cell nucleus, e.g., the CIR147 and retroposon RNAs. *Tb*DCL2 appears to have more strict requirements for its substrates, since it leaves behind the intermediate-size

RNAs. *Tb*DCL1 is the second line of defense against dsRNA producing transcripts and accepts a broader arsenal of substrates, including the intermediates. Any RNA that escapes the confines of the nucleus and contains or forms (on its own or with an antisense partner) a double-stranded segment longer than ~25 bp is converted to siRNAs in the cytoplasm by *Tb*DCL1. An additional role for *Tb*DCL1 is in the effector step of the cytoplasmic RNAi branch, since it appears to be required for RISC-mediated cleavage of target transcripts. However, it remains to be determined whether *Tb*DCL2 has a similar function in the nucleus. In summary, trypanosomes have found an evolutionary advantage of having two Dicers with similar properties, but segregated into different cellular compartments, to tackle dsRNA challengers in a "two lines of defense" mechanism and initiate their destruction.

8.3.3 TbAGO1 Is the Argonaute Slicer

The *T. brucei* genome encodes two members of the Argonaute family of proteins, namely *Tb*AGO1, which is essential for the RNAi response (Durand-Dubief and Bastin 2003; Shi et al. 2004b) and *Tb*Piwi-like1 (Durand-Dubief and Bastin 2003), which is present in all trypanosomatid protozoa, irrespective of whether they are RNAi proficient (Ullu et al. 2004). *Tb*Piwi-like 1 is dispensable for the RNAi response (Durand-Dubief and Bastin 2003), but its function is yet to be discovered.

*Tb*AGO1 is required in experimental RNAi for cleavage of mRNA targets in the cytoplasm (Shi et al. 2004b) and snoRNAs in the nucleus (Gupta et al. 2010), and in the endogenous RNAi pathway for cleavage of retroposon- and repeat-derived transcripts, both in the nucleus and in the cytoplasm (Shi et al. 2004b; Patrick et al. 2009). Single-stranded siRNAs form a ribonucleoprotein complex with *Tb*AGO1 (Djikeng et al. 2003), which is highly abundant at 60,000 copies/cell (unpublished observations), as compared to 150–200,000 ribosomes. However, significantly larger *Tb*AGO1-containing complexes are detected in cells ablated of *Tb*DCL1 or *Tb*DCL2 (unpublished observations), possibly representing stalled intermediates in the formation of RISC. Cell fractionation and the microscopic localization of GFP- (Durand-Dubief and Bastin 2003) or epitope-tagged *Tb*AGO1 (Shi et al. 2004b) consistently place the majority of the protein in the cytoplasm. Nevertheless, the requirement of *Tb*AGO1 for downregulation of retroposon- and repeat-derived transcripts and for snoRNAi is consistent with *Tb*AGO1 also functioning in the nucleus, where it presumably travels to pick up siRNAs produced by *Tb*DCL2 and then slices single-stranded and potentially harmful transcripts. How *Tb*AGO1 gets access to the nucleus remains to be addressed experimentally. Lastly, a proportion of *Tb*AGO1 is found associated with polyribosomes (Djikeng et al. 2003; Shi et al. 2009). We have shown that this association is specific for translating ribosomes and appears to be mediated by ionic interactions, as *Tb*AGO1 is released from polyribosomes under high-salt conditions. We proposed that *Tb*AGO1 on polyribosomes functions as a sentinel ready for action, if potentially dangerous retroposon transcripts should become associated with polyribosomes (Shi et al. 2009).

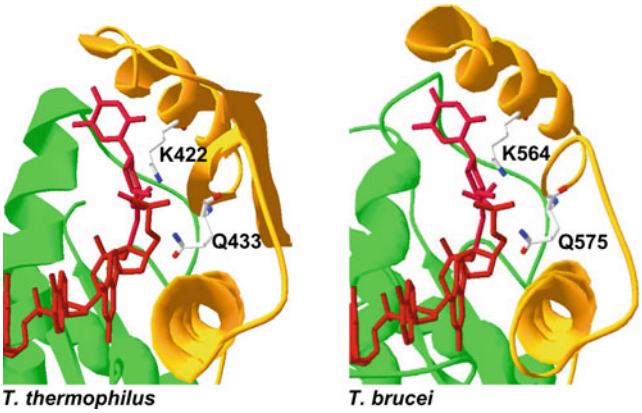


Fig. 8.2 Ribbon (AGO) and stick (DNA) diagrams of the *T. thermophilus* AGO MID domain bound to the 5′-end of a 21-base DNA guide strand (PDB ID 3DLH; *left panel*; (Wang et al. 2008)) and of the predicted model of *Tb*AGO1 with a superimposed DNA guide strand (*right panel*). The positions of two conserved residues in the binding pocket are indicated

8.3.4 Structural and Functional Motifs in Trypanosomatid AGO1

AGO-family proteins are about 100 kDa, highly basic and distinguished by four domains: the N-terminal, PAZ, MID, and PIWI domains, with the PIWI domain showing the highest degree of sequence conservation (Tolia and Joshua-Tor 2007). *Tb*AGO1, as well as its homologues in RNAi-proficient trypanosomatids, conform to the consensus structure of AGO family members, although their primary sequence is highly divergent from that of higher eukaryotic AGOs. The various domains have distinct roles, which have emerged from functional studies and the crystal structures of prokaryotic AGO paralogues (Tolia and Joshua-Tor 2007; Faehnle and Joshua-Tor 2010). The 5′-phosphate and the 3′-end of the guide small RNA are anchored by the MID and PAZ domains binding pockets, respectively, whereas the PIWI domain adopts a fold similar to that of RNase H-type enzymes and confers endonuclease activity to AGO "slicers." Using the crystal structure of *Thermus thermophilus* AGO as a template (Wang et al. 2008), a considerable part of the *Tb*AGO1 C terminus was amenable to three-dimensional modeling, despite low conservation at the primary sequence level (unpublished observation). In particular, the PIWI and part of the MID domain revealed structural similarities to *T. thermophilus* AGO, including the 5′-phosphate-binding pocket made up of the MID domain (Fig. 8.2). In vivo studies further indicated that predicted key amino acid residues in the *Tb*AGO1 5′-phosphate-binding pocket are critical for target RNA cleavage activity (unpublished observation).

The catalytic center of AGO slicers includes the amino acid triad DDH, which is found in all AGO proteins for which slicer activity has been documented (Fig. 8.3; Tolia and Joshua-Tor 2007). However, the presence of these active site residues

Fig. 8.3 Alignment of the motifs containing the catalytic triad DDH in selected Argonaute proteins from humans, fruit fly, and RNAi-proficient trypanosomes. The DDH residues are highlighted in *red* and residues conserved in all of the shown sequences in *blue*

does not guarantee that the AGO protein is endowed with endonuclease activity (Tolia and Joshua-Tor 2007). Of note, the trypanosomatid AGO1 proteins from *T. brucei*, *T. congolense*, *T. vivax*, *L. (V.) braziliensis*, *L. (V.) panamensis*, and *C. fasciculata* are not endowed with the DDH catalytic triad (Fig. 8.3). Whereas the first aspartate residue is conserved in AGO1 of African trypanosomes, it is absent in the *L. Viannia* and *Crithidia fasciculata* counterparts and replaced by valine. Another notable difference between trypanosomatid AGO1s is the presence of a conserved tyrosine at the histidine-equivalent position in the *C. fasciculata* and *L. Viannia* proteins, whereas this position is occupied by alanine in *Tb*AGO1 and by serine in the *T. congolense* and *T. vivax* proteins. Although intriguing, the functional significance of these amino acids substitutions needs to be verified experimentally.

Sequence alignments of the myriad of AGO proteins revealed additional invariant residues, including an arginine residue at position 735 of *Tb*AGO1, just one amino acid upstream from the second aspartate residue of the catalytic triad (Fig. 8.3). Mutation of arginine 735 to alanine in *Tb*AGO1 leads to severe impairment of mRNA cleavage, suggesting that this protein acts as a slicer (Shi et al. 2004c). In support of this model are two additional pieces of evidence. First, upon transfection of α-tubulin dsRNA we observe accumulation of discrete fragments with the size expected for endonucleolytic cleavage of the target mRNA (Shi et al. 2009) and, secondly, we showed that human AGO2 slicer can partially complement the cytoplasmic RNAi response when *Tb*AGO1 is ablated (Shi et al. 2006a). However, the absence in AGO1 of the "classic" catalytic triad of RNase H-type enzymes is puzzling, and this issue can only be resolved by in vitro experiments, which unfortunately have been unsuccessful so far. A possible reason behind our failure to reconstitute *Tb*AGO1 slicing activity in vitro is suggested by the observation that when the pool of newly synthesized *Tb*AGO1 is depleted, the RNAi response declines severely (Shi et al. 2007). We proposed that newly synthesized *Tb*AGO1 is preferentially loaded with newly processed siRNAs, possibly because newly synthesized *Tb*AGO1 is in a conformation that readily accepts siRNAs. We further speculated that the Hsp90 chaperone system may be

involved in facilitating the process of siRNA loading. This prediction has recently gained support by in vitro studies in *Drosophila* and tobacco extracts (Iwasaki et al. 2010; Miyoshi et al. 2010). Interestingly, reconstitution of RISC activity in tobacco could only be achieved by coupling translation of AGO with loading of synthetic siRNAs in a homologous in vitro extract.

8.3.5 Function of the RGG-Rich N Terminus of Tb*AGO1*

One of the striking features of the *T. brucei* AGO1 N-terminal domain is the high representation of arginine-glycine-glycine (RGG) motifs, which make up almost 50% of the first 59 amino acids (Shi et al. 2009). In particular, between positions 9 and 59 there are 10 RGG motifs, which are part of an 11-amino acid repeating unit with the consensus sequence G(Y/R)RGGRGGG(E/F)G. An N-terminal RGG domain is present in all trypanosomatid AGO1 homologues (Shi et al. 2009), indicating a conserved and important function. In addition, certain AGO-family members from higher eukaryotes are also endowed with RGG motifs. Interestingly, the N-termini of certain murine Piwi family proteins have methylated arginines, which guide interactions with Tudor proteins and affect protein stability and localization (Kirino et al. 2009; Vagin et al. 2009).

Through mutagenesis studies we have shown that the RGG domain of *Tb*AGO1, and specifically the arginine residues, are required for efficient cleavage of target mRNA in the cytoplasm and downregulation of retroposon transcripts in the nucleus. A second phenotype observed in cells expressing mutant *Tb*AGO1 carrying partial or complete deletions of the RGG domain or bearing substitutions of all arginine residues by lysine or alanine is the inhibition of the association of a proportion of *Tb*AGO1 with polyribosomes (Shi et al. 2009). A similar phenotype was reported for the *Toxoplasma gondii* AGO protein, which like *Tb*AGO1 contains an RGG-rich N terminus (Braun et al. 2010). Mass spectroscopy analysis, as well as reactivity with antibody Y12, which recognizes symmetric dimethylated arginine (sDMA), support the modification of at least some arginine residues in the *Tb*AGO1 RGG domain to sDMA (Shi et al. 2009). At present we do not know the identity of the specific protein arginine methyltransferases (PRMTs) responsible for modification of *Tb*AGO1. Based on the current model that sDMA is recognized by Tudor domain-containing proteins, we (unpublished observations) and others (Alsford et al. 2010) tested the possibility that *T. brucei* TudorSN, a protein with a recognizable Tudor domain, was required for RNAi, but the results have been negative. Thus, several questions concerning the function of the RGG domain remain to be answered. Which PRMT methylates arginine residues in the RGG domain? Are methylated arginines recognized by a specific, yet to be identified, factor? Does the RGG domain function as a protein and/or an RNA recognition module? Is this domain involved in guide siRNA-target RNA recognition or perhaps in determining the accessibility of the target RNA to the catalytic center?

8.3.6 Endogenous Small Interfering RNAs

siRNAs in a complex with AGO slicer guide recognition of target RNAs. The realization that siRNAs are essential mediators in RNAi was heralded by the seminal discovery in *Arabidopsis* that short 20–25 nt RNAs accumulated in transgenic plants undergoing "cosuppression," namely downregulation of cellular gene expression mediated by ectopic expression of homologous *trans*-genes (Hamilton and Baulcombe 1999). Within a short period of time siRNAs were reported in several different organisms, including *T. brucei* (Djikeng et al. 2001). A decade ago we generated and sequenced the first comprehensive library of endogenous siRNAs (esiRNAs) and discovered that the majority of molecules were derived from the retroposons ingi and SLACS, consistent with the known role of RNAi as a defense mechanism to maintain genome integrity (Djikeng et al. 2001). Once the *T. brucei* genome was completed, our collection of esiRNAs expanded to include the CIR147 siRNAs (Patrick et al. 2009). Recently, deep sequencing of *Tb*AGO1-associated siRNAs has confirmed that retroposons and CIR147 repeats are the major esiRNA-producing loci in *T. brucei*. However, the RNAi machinery processes many different transcripts that have the potential to anneal and form dsRNA (unpublished observations), similarly to what has been described for the esiRNAs in higher eukaryotes (Joshua-Tor and Hannon 2011). Thus, it appears that the nature of esiRNA-producing loci has been conserved throughout eukaryotic evolution. So far, experimental evidence for micro RNAs (miRNAs) is lacking in *T. brucei*.

T. brucei esiRNAs, as well as siRNAs produced from transgenic hairpins, are 24–26 nt in length (Djikeng et al. 2001) making them slightly longer than those described in model organisms, which tend to be around 21–22 nt. As expected from Dicer-mediated cleavage of dsRNA, *T. brucei* siRNAs carry a monophosphate group at the 5′ end (Patrick et al. 2009). The 3′ ends are insensitive to β-elimination, suggesting the presence of a modification on the ribose. Most likely the 2′-OH of the terminal ribose is methylated, similarly to what has been originally described for plant small regulatory RNAs (Li et al. 2005). Addition of the methyl group is carried out by the HEN1 family of methyltransferases and, not surprisingly, the *T. brucei* genome codes for a candidate paralogue of higher eukaryotic HEN1. The current evidence indicates that HEN1 modification protects the small RNA 3′ end from tailing and/or exonucleolytic degradation (Li et al. 2005; Kurth and Mochizuki 2009; Ameres et al. 2010) and is consistent with a model in which the 3′ end of siRNAs is dislodged from the PAZ domain when there is extensive complementarity between the 3′ half of the guide strand and the target. 3′ end methylation would be required to stabilize siRNAs depending on the specific AGO protein, the requirement for pairing between the siRNA and the target, and the rearrangement of the siRNA 3′ end upon binding to the target. Considering that extensive complementarity between siRNAs and their target is required to elicit the RNAi response in *T. brucei* (Aphasizheva et al. 2009) and that siRNAs are modified (Patrick et al. 2009), it is tempting to speculate that the 3′ end of the siRNA is extruded from *Tb*AGO1 after annealing to the target and that 3′ end modification is required to impede siRNA destabilization.

8.3.7 Diversification of the RNAi Mechanism Among Trypanosomatids

A number of observations provocatively suggest that the RNAi mechanism in trypanosomatid protozoa is more diverse than originally anticipated. For instance, the sequence comparison in Fig. 8.3 points to the possibility that the *L. Viannia* and *Crithidia* AGO1s are functionally different from the African trypanosome counterparts. In support of this hypothesis, we found that *L. V. braziliensis* AGO1 was not able to complement RNAi deficiency in *Tb*AGO1 null cells (unpublished observation). Further highlighting a possible mechanistic difference is the observation that in contrast to *T. brucei*, siRNA 3′ ends are not modified in *L. V. braziliensis* and a HEN1 candidate gene is absent in the genome (unpublished observation). Comparative genomic analysis also informed us that whereas DCL1 and DCL2 are found at syntenic loci in African trypanosomes and *L. V. braziliensis*, the AGO1 genes are not. In addition, inspection of the old-world *Leishmania* genomes revealed remnants of an AGO1 gene in a locus syntenic with *L. V. braziliensis* AGO1. Taken together, these as yet unconnected observations might suggest that the ancestor to the trypanosomatid lineage was endowed with two different AGO1 genes and that the African trypanosome and *L. Viannia* lineages did not retain the same gene.

8.3.8 RISC Activation

At the core of RISC activation is the formation of the AGO-guide siRNA complex (Liu and Paroo 2010). In humans, *Drosophila* and *Neurospora crassa* duplex siRNAs are loaded into AGO slicer and the passenger strand is first cleaved by the AGO endonuclease activity and then the resulting fragments are released. This latter step can occur either by destabilization of the passenger strand fragments, which is expected after cleavage, or with the assistance of accessory factors, such as *Nc*QIP (qde2/AGO-interacting protein1), a protein with a 3′–5′ exonuclease domain related to the epsilon exonuclease subunit of prokaryotic DNA Pol III (Maiti et al. 2007) or the heterodimeric endonuclease C3PO (Liu et al. 2009), which has been shown to stimulate RISC activity. Alternatively, AGO itself could unwind the duplex as supported by the observation that human AGO2 can melt a perfect duplex in vitro (Yoda et al. 2010). How does this process take place in *T. brucei?* The current evidence suggests that formation of active RISC in these organisms follows different rules, as *Tb*AGO1 mutants that are severely impaired in target mRNA cleavage are invariably found in a complex with single-stranded siRNAs (Shi et al. 2009). In contrast, active site mutations in *Drosophila*, human or *Neurospora* AGO slicers result in the accumulation of the corresponding AGO loaded with duplex siRNAs (Liu and Paroo 2010). There are at least two scenarios that can reconcile these observations. First, it is possible that in *Tb*AGO1 the passenger strand and

mRNA endonuclease activities are mechanistically distinct, as it has been suggested for human AGO1 and AGO2 from in vitro studies (Yoda et al. 2010). Alternatively, a specific factor may be required to cause double- to single-strand siRNA transition and loading of the guide strand into trypanosome AGO1. A clarification in the mode of RISC activation in *T. brucei* might arise with the aid of comparative genomics, which indicates that there are at least two more candidate RNAi genes, due to their presence in RNAi-proficient, but absence in RNAi-negative trypanosomatids (Lye et al. 2010). In particular, one candidate has been predicted to have a C-terminal domain resembling the DnaQ superfamily of 3′–5′ exonucleases (Zhang et al. 2008) and thus, similarly to *Neurospora* QIP, could play a role in the transition of double- to single-stranded siRNAs (see note added in proof).

8.3.9 Consequences of Genetic Ablation of the RNAi Pathway

As soon as *Tb*AGO1 was identified, our laboratory (Shi et al. 2004b) as well as Philippe Bastin's group (Durand-Dubief and Bastin 2003) examined the consequences of deleting the RNAi pathway by knocking out the AGO1 gene. RNAi deficiency resulted in a severe growth defect in Lister 427 trypanosomes (Durand-Dubief and Bastin 2003) and in a milder lengthening of the cell cycle in our laboratory procyclic strain derived from human infective *T. b. rhodesiense* (Shi et al. 2004a). Consistent with these observations, we showed that RNAi-deficient trypanosomes, which were selected through resistance to multiple rounds of transfection with α-tubulin dsRNA, displayed a growth defect that increased in severity with the degree of inhibition of the RNAi response and correlated with inhibition in the late stages of cytokinesis (Shi et al. 2004a). On the other hand, Lister 427 AGO1 null trypanosomes had pronounced chromosome segregation alterations (Durand-Dubief and Bastin 2003; Durand-Dubief et al. 2007), a phenotype that could account for their severe growth phenotype. Perhaps the RNAi pathway is involved in establishing pericentric heterochromatin domains, as it has been described in *S. pombe* (Martienssen et al. 2005). However, a direct link between *Tb*AGO1, heterochromatin formation and centromere function has not yet been established. Nevertheless, it is an attractive hypothesis that *Tb*AGO1 function in the nucleus may affect chromosome segregation. Intriguingly, *Tb*AGO1 ablation in two different blood-stream form trypanosome strains adapted to in vitro culture, one a derivative of Lister 427 and the other the pleomorphic strain STIB247, showed no apparent defects in growth, silencing of the variable surface glycoprotein expression sites or life-cycle progression (Janzen et al. 2006). At the present time it is difficult to reconcile these various and somewhat discordant observations into a coherent model. Perhaps, in vitro cultured trypanosome strains have acquired specific adaptations, which may account for the different observed phenotypes.

At the RNA level, *Tb*AGO1 ablation manifests three major phenotypes (1) the siRNAs dramatically decline in abundance (Shi et al. 2004b); (2) there is

upregulation of steady-state transcripts derived from potential mobile elements or their decayed progeny, namely the retroposons ingi and SLACS (Shi et al. 2004b); and (3) transcription of ingi and SLACS retroposons in permeabilized trypanosomes substantially increases (Shi et al. 2004b). Upregulation of retroposon steady-state and newly synthesized transcripts is also observed in in vitro selected RNAi-deficient trypanosomes (Shi et al. 2004a), as well as in *Tb*DCL2 null cells (Patrick et al. 2009), and similar observations apply to the CIR147-derived transcripts (Patrick et al. 2009; unpublished observation). From the above observations we surmise that *Tb*AGO1 and *Tb*DCL2 preside to downregulation of potentially harmful transcripts, which if left intact could promote deleterious hopping of mobile elements. In support of this hypothesis, long-term culturing of *Tb*AGO1 null trypanosomes is accompanied by rearrangements of the SLACS elements (Patrick et al. 2008). The increased transcription of SLACS, ingi, and CIR147 repeats indicate that directly or indirectly the RNAi pathway may regulate the accessibility of these elements to RNA polymerase II transcription, possibly by playing a role in heterochromatin formation, which is a well-established mechanism to silence expression of mobile elements and repeats alike.

8.3.10 RNAi Technology

Soon after its first description in *T. brucei* (Ngo et al. 1998), the power of RNAi has been harnessed for functional genomics studies by generating vectors for heritable and inducible RNAi. This advancement would not have been possible without the efforts of the laboratories of Christine Clayton (Wirtz and Clayton 1995) and George A.M. Cross (Wirtz et al. 1999), who in the meantime had built tetracycline-inducible systems for expression of toxic gene products. The availability of conditional dsRNA expression allowed investigators to apply the RNAi tool to explore the functions of hundreds of genes, both for individual research purposes and for genome-wide screens. Importantly, the rapid validation of the large number of potential drug targets, which are currently under consideration in various centers, would not have been possible without the RNAi tool.

Although the RNAi technology has proven to be invaluable in the analysis of gene function in *T. brucei*, its potential for forward genetics screens has so far not been fully explored on a genome-wide scale. Genomic DNA insert library screens (Morris et al. 2002; Schumann Burkard et al. 2010) rely on the generation of a sufficient number of trypanosome clones (much larger than the actual number of genes) in order to provide several-fold genome coverage. Systematic RNAi screens have been performed only for chromosome I in *T. brucei* (Subramaniam et al. 2006) and were based on knowledge of predicted open reading frames. With the recently obtained genome-wide map of mRNA boundaries through next-generation RNA sequencing (Kolev et al. 2010; Nilsson et al. 2010; Siegel et al. 2010), the stage is set for the systematic design of a genome-wide library producing dsRNA against all identified trypanosome transcripts to significantly reduce the number

of transformant cell lines needed (to ~10,000). Ideally, the genome-wide inducible RNAi library would produce hairpin dsRNA (for its superiority in eliciting an RNAi response) and the vector carrying the inverted repeat would be integrated at a single, defined locus in the genome (Alsford and Horn 2008) to minimize epigenetic influence of surrounding chromatin on dsRNA expression.

8.4 Conclusion

In the last 5 years there has been a tremendous acceleration in all fields of trypanosome biology due to the completion of the genome sequence and the wide application of RNAi. In addition, studies of the RNAi pathway have so far identified three RNAi factors that cooperate to bring about the RNAi response not only in the cytoplasm, but surprisingly also in the nucleus. Nuclear RNAi is fundamental to downregulate expression of transcripts derived from retroposons and putative centromeric repeats and we suspect it may have additional functions to clear away dsRNA originating from other regions of the genome, including the numerous sites of convergent transcription. However, the precise scope of the nuclear RNAi pathway needs to be further investigated in order to understand its impact on trypanosome biology.

Note added in proof Our laboratory recently reported the characterization of two novel and essential RNAi factors, namely *Tb*RIF4, a 3′–5′ exonuclease of the DnaQ superfamily with a critical role in the conversion of duplex siRNAs to the single-stranded form, and *Tb*RIF5, a possible *Tb*DCL1 cofactor (Barnes et al. 2012).

Acknowledgments Research in the author's laboratory was supported by Public Health Service grants AI28798 and AI56333 to E.U and AI43594 to C.T.

References

Alsford S, Horn D (2008) Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei*. Mol Biochem Parasitol 161:76–79

Alsford S, Kemp LE, Kawahara T, Horn D (2010) RNA interference, growth and differentiation appear normal in African trypanosomes lacking Tudor staphylococcal nuclease. Mol Biochem Parasitol 174:70–73

Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, Zamore PD (2010) Target RNA-directed trimming and tailing of small silencing RNAs. Science 328:1534–1539

Aphasizheva I, Ringpis GE, Weng J, Gershon PD, Lathrop RH, Aphasizhev R (2009) Novel TUTase associates with an editosome-like complex in mitochondria of *Trypanosoma brucei*. RNA 15:1322–1337

Barnes RL, Shi H, Kolev NG, Tschudi C, Ullu E (2012) Comparative genomics reveals two novel RNAi factors in *Trypanosoma brucei* and provides insight into the core machinery. PLoS Pathog (in press)

Best A, Handoko L, Schluter E, Goringer HU (2005) In vitro synthesized small interfering RNAs elicit RNA interference in African trypanosomes: an *in vitro* and *in vivo* analysis. J Biol Chem 280:20573–20579

Braun L, Cannella D, Ortet P, Barakat M, Sautel CF, Kieffer S, Garin J, Bastien O, Voinnet O, Hakimi MA (2010) A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. PLoS Pathog 66(5):e1000920. doi:10.1371/journal.ppat.1000920

Cerutti H, Casas-Mollano JA (2006) On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet 50:81–99

DaRocha WD, Otsu K, Teixeira SM, Donelson JE (2004) Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. Mol Biochem Parasitol 133:175–186

Djikeng A, Shi H, Tschudi C, Ullu E (2001) RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA 7:1522–1530

Djikeng A, Shi H, Tschudi C, Shen S, Ullu E (2003) An siRNA ribonucleoprotein is found associated with polyribosomes in *Trypanosoma brucei*. RNA 9:802–808

Durand-Dubief M, Bastin P (2003) TbAGO1, an Argonaute protein required for RNA interference is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. BMC Biol 1:2

Durand-Dubief M, Absalon S, Menzer L, Ngwabyt S, Ersfeld K, Bastin P (2007) The Argonaute protein TbAGO1 contributes to large and mini-chromosome segregation and is required for control of RIME retroposons and RHS pseudogene-associated transcripts. Mol Biochem Parasitol 156:144–153

Faehnle CR, Joshua-Tor L (2010) Argonaute MID domain takes centre stage. EMBO Rep 11:564–565

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Gupta SK, Hury A, Ziporen Y, Shi H, Ullu E, Michaeli S (2010) Small nucleolar RNA interference in *Trypanosoma brucei*: mechanism and utilization for elucidating the function of snoRNAs. Nucleic Acids Res 38:7236–7247

Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950–952

Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell 39:292–299

Janzen CJ, van Deursen F, Shi H, Cross GA, Matthews KR, Ullu E (2006) Expression site silencing and life-cycle progression appear normal in Argonaute1-deficient *Trypanosoma brucei*. Mol Biochem Parasitol 149:102–107

Jaskiewicz L, Filipowicz W (2008) Role of Dicer in posttranscriptional RNA silencing. Curr Top Microbiol Immunol 320:77–97

Joshua-Tor L, Hannon GJ (2011) Ancestral roles of small RNAs: an ago-centric perspective. Cold Spring Harb Perspect Biol 3(10):a003772. doi:10.1101/cshperspect.a003772

Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z (2009) Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat Cell Biol 11:652–658

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 6(9): e1001090. doi:10.1371/journal.ppat.1001090

Kurth HM, Mochizuki K (2009) 2′-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. RNA 15:675–685

Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. Curr Biol 15:1501–1507

Liang XH, Liu Q, Michaeli S (2003) Small nucleolar RNA interference induced by antisense or double-stranded RNA in trypanosomatids. Proc Natl Acad Sci USA 100:7521–7526

Liu Q, Paroo Z (2010) Biochemical principles of small RNA pathways. Annu Rev Biochem 79:295–319

Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science 325:750–753

Lye LF, Owens K, Shi H, Murta SM, Vieira AC, Turco SJ, Tschudi C, Ullu E, Beverley SM (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. PLoS Pathog 6(10):e1001161. doi:10.1371/journal.ppat.1001161

MacRae IJ, Zhou K, Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. Nat Struct Mol Biol 14:934–940

Maiti M, Lee HC, Liu Y (2007) QIP, a putative exonuclease, interacts with the *Neurospora* Argonaute protein and facilitates conversion of duplex siRNA into single strands. Genes Dev 21:590–600

Martienssen RA, Zaratiegui M, Goto DB (2005) RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. Trends Genet 21:450–456

Miyoshi T, Takeuchi A, Siomi H, Siomi MC (2010) A direct role for Hsp90 in pre-RISC formation in Drosophila. Nat Struct Mol Biol 17:1024–1026

Morris JC, Wang Z, Drew ME, Englund PT (2002) Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. EMBO J 21:4429–4438

Ngo H, Tschudi C, Gull K, Ullu E (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc Natl Acad Sci USA 95:14687–14692

Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. PLoS Pathog 6(8):e1001037. doi:10.1371/journal.ppat.1001037

Obado SO, Taylor MC, Wilkinson SR, Bromley EV, Kelly JM (2005) Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. Genome Res 15:36–43

Patrick KL, Luz PM, Ruan JP, Shi H, Ullu E, Tschudi C (2008) Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference-deficient *Trypanosoma brucei*. Mol Microbiol 67:435–447

Patrick KL, Shi H, Kolev NG, Ersfeld K, Tschudi C, Ullu E (2009) Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei*. Proc Natl Acad Sci USA 106:17933–17938

Robinson KA, Beverley SM (2003) Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. Mol Biochem Parasitol 128:217–228

Schumann Burkard G, Jutzi P, Roditi I (2010) Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. Mol Biochem Parasitol 175:91–94

Shi H, Chamond N, Tschudi C, Ullu E (2004a) Selection and characterization of RNA interference-deficient trypanosomes impaired in target mRNA degradation. Eukaryot Cell 3:1445–1453

Shi H, Djikeng A, Tschudi C, Ullu E (2004b) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. Mol Cell Biol 24:420–427

Shi H, Ullu E, Tschudi C (2004c) Function of the trypanosome Argonaute 1 protein in RNA interference requires the N-terminal RGG domain and arginine 735 in the Piwi domain. J Biol Chem 279:49889–49893

Shi H, Tschudi C, Ullu E (2006a) Functional replacement of *Trypanosoma brucei* Argonaute by the human slicer Argonaute2. RNA 12:943–947

Shi H, Tschudi C, Ullu E (2006b) An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. RNA 12:2063–2072

Shi H, Tschudi C, Ullu E (2007) Depletion of newly synthesized Argonaute1 impairs the RNAi response in *Trypanosoma brucei*. RNA 13:1132–1139

Shi H, Chamond N, Djikeng A, Tschudi C, Ullu E (2009) RNA interference in *Trypanosoma brucei*: the role of the amino-terminal RGG domain and the polyribosome association of Argonaute1. J Biol Chem 284:36511–36520

Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GA (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res 38:4946–4957

Siomi H, Siomi MC (2009) On the road to reading the RNA-interference code. Nature 457:396–404

Subramaniam C, Veazey P, Redmond S, Hayes-Sinclair J, Chambers E, Carrington M, Gull K, Matthews K, Horn D, Field MC (2006) Chromosome-wide analysis of gene function by RNA interference in the African trypanosome. Eukaryot Cell 5:1539–1549

Tolia NH, Joshua-Tor L (2007) Slicer and the Argonautes. Nat Chem Biol 3:36–43

Ullu E, Tschudi C, Chakraborty T (2004) RNA interference in protozoan parasites. Cell Microbiol 6:509–519

Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA (2009) Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes Dev 23:1749–1762

Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ (2008) Structure of the guide-strand-containing argonaute silencing complex. Nature 456:209–213

Wirtz E, Clayton C (1995) Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. Science 268:1179–1183

Wirtz E, Leal S, Ochatt C, Cross GA (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

Yoda M, Kawamata T, Paroo Z, Ye X, Iwasaki S, Liu Q, Tomari Y (2010) ATP-dependent human RISC assembly pathways. Nat Struct Mol Biol 17:17–23

Zhang D, Xiong H, Shan J, Xia X, Trudeau VL (2008) Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3′-5′ exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. Biol Direct 3:48

Chapter 9
Translation in Trypanosomatids

Michal Shapira and Alexandra Zinoviev

Abstract Protein synthesis in trypanosomatids plays a key role in shaping the developmental program of gene expression during their complex life cycle. Trypanosomatids are well known for their unusual mechanisms and pathways, including *trans*-splicing of primary mRNAs. The Spliced Leader RNA donates a unique, hypermethylated cap-4 structure to all mRNAs, which requires special adaptations by the cap-binding complex. The cap-binding eIF4F anchors the translation initiation complex in the majority of cases. The trypanosomatid cap-binding proteins show evolutionary divergence, and cannot replace the function of their yeast counterparts. The molecules that comprise the cap-4 binding complex are described and compared to their orthologues from higher eukaryotes, trying to reveal their specific functions. A novel eIF4E-interacting protein is assumed to direct stage-specific translation pathways, adapted to the different environmental conditions. Special attention is given to the unusual finding that differential translation is driven by defined elements in the 3′ UTR, and potential mechanisms are discussed, including structural changes of the RNA. The effects of environmental switches on the translational machinery are also discussed.

9.1 Introduction: Protein Synthesis Pathways in Eukaryotes

Protein synthesis in all living cells is a highly regulated process, both globally and at gene-specific levels. It allows living cells to exercise a rapid response to changing environmental conditions, based on pre-existing intracellular mRNA pools. The process of translation includes three basic steps: initiation, elongation, and termination. The initiation step is predominantly viewed as the limiting step of protein synthesis, thus serving as a target for a multitude of regulatory mechanisms.

M. Shapira (✉) • A. Zinoviev
Department of Life Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel
e-mail: shapiram@bgu.ac.il

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_9,

Translation initiation has been studied intensively in both prokaryotes and eukaryotes, which use fundamentally different mechanisms to target the initiator AUG codon into the active center of the ribosome. Translation initiation in eukaryotes involves a large number of polypeptides, whereas in prokaryotes, this process is less complex.

9.1.1 The Cap-Binding Complex

The general scheme of cap-dependent translation is that a preassembled 43S preinitiation complex (PIC) is targeted to the 5′ end of the mRNA through the cap-binding complex. The 43S PIC consists of the Met-tRNA and a multitude of initiation factors that include eIF1, eIF1A, eIF2, and eIF3, the latter is responsible for recruiting the small ribosomal subunit [reviewed in (Gingras et al. 1999b; Pestova et al. 2007)]. The cap-binding complex, denoted as eIF4F, consists of the cap-binding protein eIF4E, an RNA helicase eIF4A, and the scaffold protein eIF4G that serves as a "hub" to assemble the 43S PIC with the mRNA via its 5′ cap structure and its binding complex (Gross et al. 2003). The human eIF4G is a large protein that holds together the subunits of eIF4F, and binds the eIF3 that is associated with the 40S ribosomal subunit (Yamamoto et al. 2005). In addition, the human eIF4G can interact with the Poly(A) Binding Protein (PABP) that is associated with the poly(A) tail at the 3′ end of the mRNA, causing transient circularization of the translating mRNA during initiation [reviewed in (Cheng and Gallie 2007; Amrani et al. 2008)]. This process explains the strict requirement for the poly(A) tail during translation. The resulting 48S complex scans the 5′ untranslated region (5′ UTR) until it arrives at the initiator AUG codon. At this point the 60S ribosomal subunit joins to form the mature 80S ribosome, with the help of eIF5B (Unbehaun et al. 2007) and eIF6 (Gandin et al. 2008), releasing most of the initiation factors. Once the ribosome advances along the mRNA and evacuates the AUG start codon, a new complex can be formed, creating a polyribosome complex, which assembles on actively translating mRNAs. The scanning process is enabled by helicases which are associated with the initiation complex, with eIF4A serving as the main candidate, although it is not necessarily the only factor involved in this activity (Imataka and Sonenberg 1997; Oberer et al. 2005). The activity of eIF4A is enhanced by an additional factor, eIF4B, which was shown to be essential for the helicase activity (Methot et al. 1994).

9.1.2 Cap-Independent Translation Initiation: IRES Structures

An alternative cap-independent mechanism for translation initiation occurs in certain mRNAs, where the small ribosomal subunit is targeted to a position that is adjacent to the initiator AUG, with the aid of a highly structured RNA element in the 5′ UTR,

denoted as Internal Ribosome Entry Site (IRES) (Hellen and Sarnow 2001). The IRES is usually comprised of a relatively long RNA sequence, which assigns a functional secondary structure, usually with the aid of *trans*-activating factors, such as the Poly-Pyrimidine-Tract-Binding protein (PTB), ITAF45, or the La antigen (Pilipenko et al. 2000). IRES structures are commonly used for translation of viral transcripts, under conditions in which the host cap-dependent translation is inhibited. For example, IRES elements of the Hepatitis C virus do not require any component of the eIF4F complex for binding to the 40S subunit, and rely only on eIF3 and eIF2/eIF5, or eIF5B for this purpose (Pestova et al. 2008; Terenin et al. 2008). A more extreme example is the Cricket Paralysis Virus (CrPV) IRES that is able to assemble the 80S ribosome and advance to the elongation step without the involvement of any translation initiation factor or initiator tRNA (Wilson et al. 2000). IRES elements were also reported for endogenous mRNAs (Holcik et al. 2000; Hellen and Sarnow 2001; Holcik and Sonenberg 2005; Bert et al. 2006); however, they do not account for all cases of cap-independent translation, and other, less understood pathways, may be used (Andreev et al. 2009; Mitchell et al. 2010).

9.1.3 Global Regulation of Translation Initiation by 4E-BP

Exposure of cells to a variety of stress conditions, such as extreme temperatures or amino-acid starvation, leads to a global cessation of protein synthesis in living cells, and cell cycle arrest. This excludes of course proteins that are required by the cell to overcome the damaging effects of the stress. Translation arrest is achieved by dephosphorylation and activation of the 4E-binding protein (4E-BP), which binds to eIF4E in a manner that prevents the interaction between eIF4E and eIF4G. Once this interaction is inhibited, cap-dependent translation is blocked (Lin et al. 1994; Pause et al. 1994). The phosphorylation status of 4E-BP is mediated by the TOR (*T*arget *O*f *R*apamycin) kinase, which is most often controlled by the PI3K/AKT pathway (Lawrence and Abraham 1997). Phosphorylation of 4E-BP by TOR leads to the dissociation of 4E-BP from its target eIF4E [reviewed by (Richter and Sonenberg 2005)]. Stress conditions or amino-acid starvation impede TOR activation, shifting 4E-BP to its dephosphorylated form, and promoting its binding to eIF4E in a manner that prevents the assembly of eIF4F, the cap-binding complex. Thus the TOR proteins are regulators that link metabolism, protein synthesis, and cell cycle (Gingras et al. 1999a; Laplante and Sabatini 2009).

The eukaryote 4E-BP is a highly conserved protein that is expressed in most eukaryotes, except for *C. elegans*. It is interesting to note that a 4E-BP homologue was not identified in the trypanosomatid genomes sequenced to date. Despite the fact that *C. elegans* and trypanosomatids are two remote taxonomical groups, they both use *trans*-splicing for processing the majority (*C. elegans*) or all their mRNAs (trypanosomatids).

9.1.4 Gene-Specific Regulation of Cap-Dependent Translation Initiation

Inhibition of cap-dependent translation can also be achieved in a gene-specific manner, whereby a regulatory protein that is recruited by elements in the 3′ UTR, binds eIF4E in a manner that blocks the access to eIF4G and prevents assembly of the eIF4F complex. During the development of *Drosophila* embryos, specific proteins and mRNAs are distributed unevenly along the body axis, leading to a tightly controlled plan of gene expression. For example, the Cytoplasmatic-Polyadenylation-Element-Binding protein (CPEB) binds to its target sequence in the 3′ UTR, and to Maskin, which is an eIF4E-binding protein. Maskin excludes the formation of the eIF4F complex by competing with eIF4G, preventing translation initiation. CPEB stimulates the formation of the Poly (A) tail, which attracts PABP. PABP interacts with eIF4G, aiding the displacement of Maskin from the relevant residues on eIF4E, promoting cap-dependent translation [reviewed in (Richter and Sonenberg 2005)].

9.1.5 Effects of Heat Shock on Global Translation and Translation Initiation

Exposure to extreme environmental conditions is hazardous to all living cells, and requires adaptive responses. This includes temperature, pH and osmotic switches, UV irradiation, nutrient starvation, oxidative stress, and exposure to a variety of toxic drugs. Stress caused by exposure to extreme temperatures induces cellular damage at multiple levels, and is known to cause a global arrest of protein synthesis, excluding the nascent generation of heat shock proteins (HSPs). In cases where the cell cannot overcome the inflicted damage, apoptosis may be induced (Holcik and Sonenberg 2005). The increased translation of HSPs in higher eukaryotes is controlled by the 5′ UTR, as shown for both HSP70 (Klemenz et al. 1985; McGarry and Lindquist 1985) and HSP90 (Ahmed and Duncan 2004). HSP90 and some variants of HSP70 are expressed also at ambient temperatures, but their synthesis increases dramatically during heat shock.

Despite extensive efforts, no functional motifs that potentially drive preferential translation at elevated temperature were identified. Introduction of highly structured elements into the 5′UTR inhibited translation only at higher temperatures and not under ambient conditions (Hess and Duncan 1996). The 5′ UTR of Hsp90 from *Drosophila* possesses significant secondary structures typical to non-heat shock genes (Ahmed and Duncan 2004). Two modes of Hsp90 translation were postulated: under normal temperatures translation occurs by a typical cap-dependent scanning mechanism, whereas during heat shock, translation shifts to a cap-independent mode. Translation of the human Hsp70 was also examined, and deletion of its 5′ UTR sequences that are complementary to the 3′ hairpin of 18S

rRNA abolished translation during heat shock, suggesting a prokaryotic-like interaction between mRNA and rRNA. However, the Hsp70 5′UTR did not function as an IRES when inserted internally in a bicistronic mRNA (Vivinus et al. 2001). Translation of human Hsp70 can also be explained by a cap-dependent ribosome-shunting mechanism, by which the ribosome is recruited to the 5′ cap, but reaches the first AUG codon by shunting rather than by scanning (Yueh and Schneider 2000). Due to the contradictory data, the mode of preferential translation of HSPs at elevated temperatures in higher eukaryotes remains elusive, although it is generally assumed that it proceeds in a cap-independent manner (Rubtsova et al. 2003).

9.1.6 Global Regulation of Translation by Phosphorylation of eIF2α

Translation initiation of most cellular proteins ceases during temperature stress, whereas synthesis of HSPs preferentially continues. One of the ways to stop global translation is through phosphorylation of eIF2α, which forms a ternary complex with Met-tRNA-GTP. Following the hydrolysis of GTP, activity of eIF2α is regenerated by the eIF2B-mediated exchange of GDP with GTP. Phosphorylation of eIF2α at Ser 51 blocks this exchange, resulting in a global translational arrest (Dever 2002).

9.2 Translation Initiation in Trypanosomatids

9.2.1 Protein-Coding Genes Are Regulated Post-Transcriptionally in Trypanosomatids

Trypanosomatids are ancient eukaryotes that use unusual pathways to control their gene expression. Transcription of protein-coding genes is polycistronic, and the pre-mRNAs are processed by *trans*-splicing and polyadenylation [reviewed in (Clayton 2002; Michaeli 2011)]. Conventional RNA pol II promoters were not identified to date, and there is also no evidence for a process of regulated transcription activation for any of the protein coding genes of trypanosomatids. Thus, mRNA processing, stability, and translation serve as the key mechanisms that direct differential program of gene expression throughout their life cycle [reviewed in (Clayton and Shapira 2007)].

9.2.2 The Trypanosomatid Cap-4 Structure

The trypanosomatid cap-4 structure is donated to the 5′ end of all mRNAs during *trans*-splicing via the SL RNA (Liang et al. 2003). In addition to the consensus m^7GTP, cap-4 consists of 2′-*O* ribose methylations on the first four nucleotides of the SL RNA, and two base methylations on the first adenosine and fourth uridine. These base methylations are unique to trypanosomatids, and are not known in any other group of eukaryotes (Bangs et al. 1992).

9.2.3 The Effect of Ribose Methylation in Cap-4 Nucleotides on Translation

The role of ribose methylation on RNA processing and translation has been examined in *Xenopus* oocytes (Kuge et al. 1998). Individual substitutions of each nucleotide in the cap-4 structure led to impaired utilization of the mutated SL RNA in *trans*-splicing reactions (Mandelboim et al. 2002). However, these elegant studies could not serve to evaluate the role of ribose methylation in translation. The enzymes which are responsible for 2′-*O*-methylation of the ribose moieties in the cap-4 nucleotides were identified (Arhin et al. 2006; Ruan et al. 2007). Efforts to evaluate the requirement for ribose methylation in the cap-4 structure of trypanosomatids were made by testing single and double knock-out mutants of genes encoding one or two of the enzymes responsible for ribose 2′-*O*-methylation at position 1 (TbMTr1), 2 (TbMTr2), 3 and 4 (TbMTr3). The different cell lines showed that preventing the ribose methylations at positions 3 and 4, but not at position 1, caused a decrease in translation rates that were further exacerbated by additional loss of methylation at position 2, without having a dramatic effect on the mRNA levels or cell growth rates. Knock-out of TbMTr1 alone did not cause an inhibitory effect on translation; however, depletion of TbMTr2 or TbMTr3 on TbMTr1-/- background did not yield viable parasites (Zamudio et al. 2006, 2009). Therefore, it was proposed that only a minimal level of mRNA cap ribose methylation is essential for trypanosome viability.

9.2.4 Cap-Binding Proteins in Trypanosomatids

Studies of the translation apparatus in trypanosomatids initiated with genetic annotation, followed by functional analysis. The annotated genes are shown in Table 9.1. Some proteins were subject to a thorough biochemical and cellular analysis, whereas others were only identified based on sequence homology.

eIF4E in higher eukaryotes binds the methylated 5′ cap structure and its association with eIF4G increases the binding affinity. The three-dimensional structure of murine,

Table 9.1 Translation factors in the trypanosomatid genomes

	L. major	*L. infantum*	*L. braziliensis*	*T. brucei*	*T. cruzi**	Sources
eIF4E1	LmjF27.1620	LinJ27_V3.1520	LbrM27_V2.1750	Tb11.18.0004	Tc00.1047053506127.170	Yoffe et al. (2004), Dhalia et al. (2005)
eIF4E2	LmjF19.1500, LmjF19.1480	LinJ19_V3.1520	LbrM19_V2.1750	Tb10.61.0210	Tc00.1047053511353.40	Dhalia et al. (2005), Yoffe et al. (2006)
eIF4E3	LmjF28.2500	LinJ28_V3.2690	LbrM28_V2.2700	Tb11.01.3630	Tc00.1047053504935.10	Dhalia et al. (2005), Yoffe et al. (2006)
eIF4E4	LmjF30.0450	LinJ30_V3.0460	LbrM30_V2.0530	Tb927.6.1870	Tc00.1047053421959.10	Dhalia et al. (2005), Yoffe et al. (2006)
4E-IP	LmjF35.3980	LinJ35_V3.4030	LbrM34_V2.3960	Tb09.211.2380	Tc00.1047053508461.290	Zinoviev et al. (2011)
eIF4G1	LmjF15.0060	LinJ15_V3.0060	LbrM15_V2.0060	Tb927.5.1490	Tc00.1047053506739.10	Dhalia et al. (2005)
eIF4G2	LmjF15.1320	LinJ15_V3.1340	LbrM15_V2.1270	Tb09.160.3980	Tc00.1047053508277.340	Dhalia et al. (2005)
eIF4G3	LmjF16.1600	LinJ16_V31700	LbrM16_V2.1660	Tb927.8.4820	Tc00.1047053508837.130	Dhalia et al. (2005), Yoffe et al. (2009)
eIF4G4	LmjF36.6060	LinJ36_V3.6320	LbrM35_V2.6370	Tb11.01.2330	Tc00.1047053510285.100	Dhalia et al. (2005)
eIF4G5	LmjF10.1080	LinJ10_V3.1160	LbrM10_V2.1180	Tb927.8.4500	Tc00.1047053508989.90	Dhalia et al. (2005)
eIF4G6	LmjF36.5160	LinJ36_V3.5390	LbrM35_V2.5410	Tb11.01.2520	Tc00.1047053506155.40**	Dhalia et al. (2005)
eIF4A1	LmjF01.0780, LmjF01.0770	LinJ01_V3.0800, LinJ01_V3.0790	LbrM01_V2.0740	Tb09.160.3270	Tc00.1047053511585.190	Dhalia et al. (2006)
eIF4A2	LmjF28.1530	LinJ28_V3.1660	LbrM28_V2.1700	Tb11.12.0011	Tc00.1047053511585.190	Dhalia et al. (2006)
eIF4B	None	None	None	None	None	BLAST
ePABP1	LmjF35.5040	LinJ35_V3.5360	LbrM34_V2.4980	Tb09.211.0930	Tc00.1047053506885.70	da Costa Lima et al. (2010)
ePABP2	LmjF35.4130	LinJ35_V3.4200	LbrM34_V2.4130	Tb09.211.2150	Tc00.1047053508461.140**	da Costa Lima et al. (2010)
ePABP3	LmjF25.0080	LinJ25_V3.0080	LbrM25_V2.0080	None	None	da Costa Lima et al. (2010)
eIF1	LmjF24.1210	LinJ24_V3.1240	LbrM24_V2.1230	Tb11.02.3595	Tc00.1047053508515.20	BLAST
eIF1A	LmjF16.0140	LinJ16_V30150	LbrM16_V2.0150	Tb927.8.5880	Tc00.1047053506743.4	BLAST

(continued)

Table 9.1 (continued)

	L. major	*L. infantum*	*L. braziliensis*	*T. brucei*	*T. cruzi**	Sources
eIF2-alpha	LmjF16.0690	LinJ16_V30690	LbrM16_V2.0680	Tb927.5.3450	Tc00.1047053503529.20	BLAST
eIF2-beta	LmjF08.0550	LinJ08_V3.0570	LbrM08_V2.0600	Tb927.5.3120	Tc00.1047053503955.70	BLAST
eIF2-gamma	LmjF09.1070	LinJ09_V3.1130	LbrM09_V2.1130	Tb11.01.4830	Tc00.1047053503819.30	BLAST
eIF3a	LmjF17.0010	LinJ.17.0010	LbrM.17.0020	Tb927.7.6090	Tc00.1047053507975.80	BLAST
eIF3b	LmjF17.1290	LinJ17_V3.1390	LbrM17_V2.1450	Tb927.5.2570	Tc00.1047053511303.60	BLAST
eIF3c	LmjF36.6980	LinJ36_V3.7320	LbrM35_V2.7360	Tb927.10.8290, Tb927.10.8270	Tc00.1047053507611.310	BLAST
eIF3d	LmjF30.3040	LinJ30_V3.3080	LbrM30_V2.3030	Tb927.6.4370	Tc00.1047053506943.160	BLAST
eIF3e	LmjF28.2310	LinJ28_V3.2480	LbrM28_V2.2510	Tb11.01.3420	Tc00.1047053509205.30	BLAST
eIF3f	None	None	None	None	None	BLAST
eIF3g	None	None	None	None	None	BLAST
eIF3h	None	None	None	None	None	BLAST
eIF3i	LmjF36.3880	LinJ36_V3.4070	LbrM35_V2.4110	Tb11.01.1370	Tc00.1047053511589.230	BLAST
eIF3j	None	None	None	None	None	BLAST
eIF3k	LmjF32.2180	LinJ32_V3.2330	LbrM32_V2.2400	Tb11.01.7070	Tc00.1047053509267.40	BLAST
eIF3l	LmjF36.0250	LinJ36_V3.0270	LbrM35_V2.0330	Tb927.10.4640	Tc00.1047053508169.90	BLAST
eIF3m	None	None	None	None	None	BLAST
eIF5	LmjF34.0350	LinJ34_V3.0370	LbrM20_V2.0320	Tb927.10.2770	Tc00.1047053504119.10	BLAST
eIF5A	LmjF25.0720, LmjF25.0730	LinJ25_V3.0740, LinJ25_V3.0750, LinJ25_V3.0760	LbrM25_V2.0580, LbrM25_V2.0590, LbrM25_V2.0600, LbrM25_V2.0610	Tb11.03.0410	Tc00.1047053506925.120, Tc00.1047053506925.130	BLAST
eIF5B	LmjF33.2740	LinJ33_V3.2880	LbrM33_V2.3020	Tb927.2.3780	Tc00.1047053511111.10	BLAST
eIF6	LmjF36.0890	LinJ36_V3.0950	LbrM35_V2.1010	Tb927.10.5300	Tc00.1047053506679.70**	BLAST

eEF1A	LmjF17.0080, LmjF17.0081, LmjF17.0082, LmjF17.0083, LmjF17.0084, LmjF17.0085, LmjF17.0086	LinJ17_V3.0090 LinJ17_V3.0100, LinJ17_V3.0110, LinJ17_V3.0170, LinJ17_V3.0190, LinJ17_V3.0200	LbrM17_V2.0090	Tb927.10.2100, Tb927.10.2110	Tc00.1047053511369.10, Tc00.1047053511369.15, Tc00.1047053511369.20, Tc00.1047053511369.30	Hashimoto et al. (1995), Billaut-Mulot et al. (1996)
eEF1B-alpha	LmjF36.1430	LinJ36_V3.1490	LbrM35_V2.1570	Tb927.10.5840	Tc00.1047053507671.30	Vickers and Fairlamb (2004)
eEF1B-beta	LmjF34.0840, LmjF34.0840	LinJ34_V3.0870, LinJ34_V3.0890	LbrM20_V2.0810, LbrM20_V2.0770	Tb927.4.3570, Tb927.4.3590	Tc00.1047053507671.30	Vickers and Fairlamb (2004)
eEF1B-gamma	LmjF09.0970	LinJ09_V3.1020, LinJ09_V3.1030	LbrM09_V2.1020	Tb11.01.4660, Tb11.01.4750	Tc00.1047053510163.20	Vickers and Fairlamb (2004)
eEF2	LmjF36.0180, LmjF36.0190	LinJ36_V3.0190, LinJ36_V3.0200, LinJ36_V3.0210	LbrM35_V2.0270	Tb927.10.4560, Tb927.10.4570	Tc00.1047053510963.90, Tc00.1047053508169.20	Nakamura et al. (1996)

Homologues of the mammalian consensus translation factors were identified in the genomes of *L. major*, *L. infantum*, *L. braziliensis*, *T. brucei*, and *T. cruzi*. Part of the gene identities were collected from published data, and relevant publications are cited in the "sources" column. For genes that were not explored in the literature, their potential function was inferred from sequence homology based on BLAST analysis

*Esmeraldo-like contigs

**Non-Esmeraldo-like contigs

human, and yeast proteins were formerly determined (Marcotrigiano et al. 1997; Matsuo et al. 1997; Hershey et al. 1999; Gross et al. 2003; Tomoo et al. 2003). While the amino terminus of the protein is flexible and not conserved, the carboxy terminus has a typical structural signature. It acquires the shape of a base-ball glove, with a dedicated structural pocket, which contains a sandwich of three Trp residues (W56, 102, and 166, according to the mouse numbering), for binding of the methylated guanine. The positive charges of Arg 112, 157, and 162 interact with the negatively charged phosphate moieties of the cap structure. eIF4E interacts with eIF4G and with its negative regulators such as 4E-BP through several conserved residues in its carboxy terminus. The conserved structure of the functional region in eIF4E enables proteins from different species to genetically complement each other in yeast (Altmann et al. 1989).

It is reasonable to assume that the cap-4 binding proteins in trypanosomatids have gone through structural changes during their evolution, adjusting their binding to the highly modified cap-4 structure. Gene mining followed by biochemical and cellular analysis revealed that the genomes of trypanosomes and leishmanias encode four isoforms of eIF4E (Dhalia et al. 2005; Yoffe et al. 2006). Homology modeling of these annotated genes and analysis of conserved residues predicted that the cap-binding pocket was conserved. However, none of the four paralogues could substitute for the missing function of the yeast eIF4E, indicating divergence from their higher eukaryotic counterparts (Yoffe et al. 2006). Consequently, assigning roles to the different isoforms is not simple. A synthetic cap-4 analog was prepared (Lewdorowicz et al. 2004) and used to examine the binding specificities of the LeishIF4E paralogs in Trp fluorescence assays. These indicated that isoforms 1 and 4, denoted LeishIF4E-1 and LeishIF4E-4, bound both m^7GTP and cap-4 with comparable kinetics and association constants in the micromolar range. The recombinant LeishIF4E-3 bound mainly to m^7GTP and not to cap-4, and LeishIF4E-2 showed a great preference to cap-4, as compared to m^7GTP (Yoffe et al. 2006). In accordance, the endogenous LeishIF4E-1 and -4 could be eluted from a m^7GTP-Sepharose column. As expected from the binding assays, LeishIF4E-2 could not be eluted from this column, and surprisingly, binding of LeishIF4E-3 was very poor (Yoffe et al. 2006; Freire et al. 2011).

Identification of the subunits of the eIF4F complex by pulldown analysis and evaluation of their migration pattern on sucrose gradients, support the notion that LeishIF4E-4 functions as the conventional eIF4E in promastigotes, and is part of the parasite LeishIF4F cap-4 binding complex. Copurification and pulldown assays showed that LeishIF4E-4 interacts with LeishIF4G-3, a typical MIF4G domain protein (Yoffe et al. 2009; Freire et al. 2011), which is responsible for recruiting LeishIF4A-1 to the complex. A similar interaction was also shown for the *Trypanosoma brucei* orthologues (Freire et al. 2011). These eIF4F subunits comigrate on sucrose density gradients in fractions that are expected to contain the PIC of cap-dependent translation (Yoffe et al. 2009).

Gene silencing experiments were performed on the four *T. brucei* isoforms (Freire et al. 2011). Elimination of TbIF4E-4 inhibits the growth of bloodstream-form cells, but not of procyclic parasites. This result does not coincide with data obtained for the *Leishmania* paralogue, since LeishIF4E-4 binds m^7GTP and LeishIF4G-3 mainly in

promastigotes grown at room temperature, and fails to perform these activities upon exposure to mammalian-like temperatures and in axenic amastigotes (Zinoviev et al. 2011). In addition, the migration profile of LeishIF4E-4 in sucrose gradients shows that in cells exposed to increased temperatures, this protein no longer forms large complexes. Altogether, LeishIF4E-4 seems to be less active under conditions that lead to stage differentiation (Yoffe et al. 2009). The altered experimental design and potential variability in the function of this protein between trypanosomes and leishmanias could explain this difference.

The combined silencing of the *T. brucei* orthologues TbIF4E-4 and TbIF4E-1 led to growth retardation of procyclic cells, suggesting that the two proteins could functionally complement each other. The individual silencing of TbIF4E-4 and TbIF4E-1 was harmful for bloodstream form cells. The *Leishmania* orthologue of TbIF4E-1, LeishIF4E-1, is the only isoform that maintains its expression at elevated temperatures and in axenic amastigotes. Expression of the other three orthologues decreases under these conditions (Yoffe et al. 2006; Zinoviev et al. 2011). In accordance, the 3′ UTR of LeishIF4E-1 contains a conserved 450 nts element (Yoffe et al. 2006), which is typical of transcripts that are preferentially translated in amastigotes (McNicoll et al. 2005). It is therefore possible that LeishIF4E-1 is associated with translation under stress conditions and in amastigotes. TbIF4E-1, as well as its *Leishmania* orthologue, do not interact efficiently with LeishIF4G-3 (Freire et al. 2011; Zinoviev et al. 2011) suggesting that this protein does not generate a typical eIF4F complex and could take part in alternative pathways of translation.

Silencing of TbIF4E-2 had no effect on growth in both life forms of *T. brucei* (Freire et al. 2011). The *Leishmana* orthologue, LeishIF4E-2, was shown to comigrate with heavy polysomes (Yoffe et al. 2006), unlike typical translation initiation factors that are found in the 48S PIC (Hiremath et al. 1989; Rau et al. 1996). A report describing the structure of the 80S ribosome from *T. cruzi* suggested that a distinct factor stabilizes the interaction between the ribosome and mRNAs, which are all protected by the highly modified cap-4 structure (Gao et al. 2005). LeishIF4E-2 may therefore function as a mediator between the ribosome and the parasite mRNAs, but this has yet to be tested experimentally.

The RNAi experiments in *T. brucei* indicated that among the different isoforms, only TbIF4E-3 was essential for survival of both procyclic and bloodstream life forms. Silencing of TbIF4E-3 caused a reduction in incorporation of a radiolabeled amino acid into nascently synthesized polypeptides, and a suggestion that it serves as a translation initiation factor was raised (Freire et al. 2011). However, the reduced translation rates could be due to a downstream effect of other processes that inhibit cell growth in the absence of TbIF4E-3. Furthermore, LeishIF4E-3 has a low affinity to cap-4, not supporting its definition as a typical translation initiation factor. LeishIF4E-3 comigrates with 80S particles, which are nuclease resistant (Yoffe et al. 2006). In accordance, TbIF4E-3 was reported to be part of the RNA granules in *T. brucei* (Cassola et al. 2007; Kramer et al. 2008), and could thus have a crucial role in targeting mRNAs following their export from the nucleus to the cytoplasm, directing them either to the storage bodies, or to actively translating polysomes. TbIF4E-3 interacts with an MIF4G-domain protein, TbIF4G-4, but no

association with other factors, such as eIF2 and eIF3 was shown. Thus the role of the trypanosomatid eIF4E-3 paralogue remains elusive.

9.2.5 *A Novel LeishIF4E-Interacting Protein in Trypanosomatids*

The trypanosomatid genome database does not contain any orthologue of the consensus 4E-BP (~10 kDa). However, pulldown analysis revealed the presence of a novel LeishIF4E-interacting protein, denoted as Leish4E-IP (~80 kDa) (Zinoviev et al. 2011). Leish4E-IP bears no resemblance to 4E-BP from higher eukaryotes, although both proteins are predicted to be unstructured. Leish4E-IP binds mainly to LeishIF4E-1 through a consensus YXXXXLΦ peptide (Zinoviev et al., In preparation), where X is any amino acid, and Φ is a hydrophobic residue, that is present in known eIF4E binding proteins (Mader et al. 1995). Binding of Leish4E-IP to LeishIF4E-1 is observed only in promastigotes, suggesting that this interaction plays a role in directing stage-specific translation of LeishIF4E-1. Preliminary data suggest that the Leish4E-IP phosphorylation status changes during differentiation. It releases LeishIF4E-1 during stress, promoting its function in amastigotes. The regulatory pathway that controls the function of Leish4E-IP is yet to be resolved.

9.2.6 *eIF4G Isoforms*

The genomes of leishmanias and trypanosomes encode several eIF4G candidates that contain the "Middle Portion of eIF4G" (MIF4G) domain. MIF4G consists of several HEAT repeats (Marcotrigiano et al. 2001), typical of all eIF4G proteins in higher eukaryotes (Marintchev and Wagner 2005); it binds eIF3 and eIF4A as well as RNA. The carboxy terminus of the mammalian eIF4G contains a second site for interaction with eIF4A, and its N-terminus (1,600 amino acids) carries binding sites for eIF4E and PABP. Binding to eIF4E is mediated by a consensus YXXXXLΦ motif, which is present in eIF4G, as well as in the eukaryote 4E-BP. Substitution of the conserved Y or LΦ residues in the motif abrogates binding to eIF4E (Mader et al. 1995).

The functional orthologue of eIF4G in *Leishmania* was identified, based on its co-purification and direct interaction with LeishIF4E-4, monitored in yeast two-hybrid assays and reciprocal pulldown experiments (Dhalia et al. 2005; Yoffe et al. 2009; Freire et al. 2011; Zinoviev et al. 2011). LeishIF4G-3 and TbIF4G-3 contain a typical MIF4G domain (Dhalia et al. 2005; Yoffe et al. 2009). However, their N-terminus is short and consists of 50 amino acids; this domain can independently bind the parasite LeishIF4E-4 in a yeast two-hybrid assay. The interaction with LeishIF4E-4 is mediated by a LeishIF4G-3 peptide that is reminiscent of the YXXXXLΦ motif (Mader et al. 1995), but not fully conserved. In addition to the

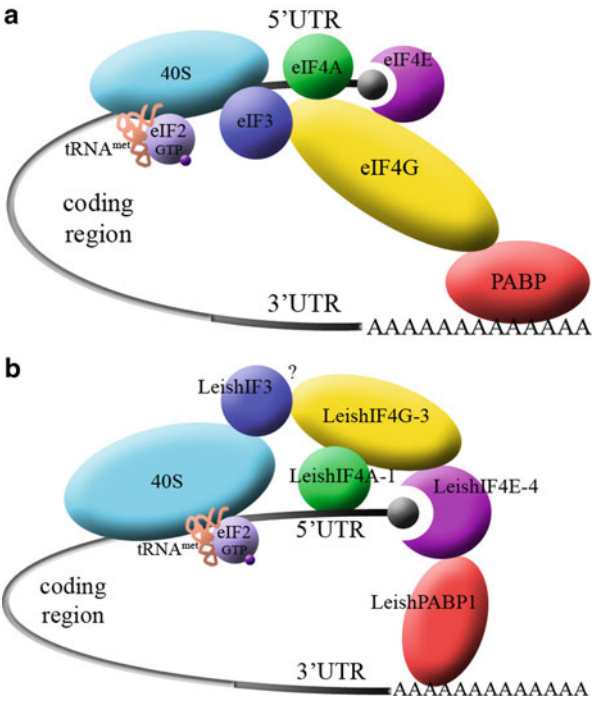


Fig. 9.1 Model for assembly of the eIF4F complex in higher eukaryotes (**a**) and in *Leishmania* promastigotes (**b**). The proteins are not represented in scale

Tyr and Leu at positions 1 and 6 of the peptide, a Phe residue is also required at position 4, and a partial requirement was observed for the Gly and Glu residues at positions 3 and 8 (Yoffe et al. 2009). These alterations emphasize the variability between the trypanosomatid element and the parallel consensus sequence in higher eukaryotes.

The variability between the parasite eIF4G and its human counterpart originates from other domains as well. eIF4G acts as a hub that links the 5′ and 3′ ends, by interacting with both eIF4E and PABP. Binding of the human eIF4G to PABP is mediated by a region in its N-terminus, which is completely missing in the parasite orthologue. In accordance with the short N-terminal domain of LeishIF4G-3, yeast two-hybrid assays exclude its binding with the parasite PABP-1 and PABP-2. Further studies indicate that the mRNA circularization occurs via an interaction between the N-terminus of LeishIF4E-4 and LeishPABP-1. Deletion of this region in LeishIF4E-4 prevents its binding to LeishPABP-1 (Zinoviev et al. 2011). LeishIF4E-4 can therefore be associated directly with translation initiation, along with LeishIF4G-3, which recruits LeishIF4A-1 (Dhalia et al. 2006; Zinoviev et al. 2011); these three polypeptides comprise a typical eIF4F complex in promastigotes (Fig. 9.1).The interaction between LeishIF4G-3 and LeishIF4E-4 is interrupted at elevated temperatures and in axenic amastigotes, suggesting that alternative pathways for translation may be used under these conditions.

Another eIF4E–eIF4G interaction was reported for TbIF4E-3 and TbIF4G-4 (Freire et al. 2011). The authors of this study propose that this complex could be

associated with translation initiation, yet this assumption is contradicted by the fact that it could not be purified over a m^7GTP affinity column (Freire et al. 2011), and that the eIF4E-3 paralogue fails to bind cap-4 in fluorescence titration assays (Yoffe et al. 2006).

9.2.7 eIF4A Isoforms

Scanning of the 5′ UTR by the translation initiation complex requires a potent helicase activity that will unwind secondary RNA structures. This will allow the complex to slide and advance, until it arrives at the initiator AUG codon, where the ribosome will assemble and the actual process of translation will commence. eIF4A is an RNA helicase that is part of the cap-binding complex, and is assumed to be involved in scanning of the 5′ UTR. eIF4A is a member of the conserved DEAD-box family of RNA helicases (Linder 2006). DEAD-box proteins participate in a multitude of processes related to transcription, RNA processing, export, and translation. The genome of most eukaryotes encodes for more than a single eIF4A isoform. One of these homologs is eIF4AIII, which does not have an active role in translation. This protein in higher eukaryotes is part of the Exon Junction Complex (EJC) and interacts with components of the nucleopore, promoting mRNA export from the nucleus to the cytoplasm (Ferraiuolo et al. 2004). The genome of *T. brucei* encodes two isoforms of eIF4A. TbIF4A1 is part of the eIF4F cap-binding complex. It is distributed in the cytoplasm and interacts with TbIF4G-3 (Dhalia et al. 2005; Zinoviev et al. 2011). Silencing of TbIF4A1 caused a dramatic decrease in protein synthesis already in the first 24 h following its inhibition. TbEIF4AIII is predominantly nuclear, and the effect of its silencing on cell proliferation was delayed to 72 h following protein elimination, with hardly any effect on *de novo* protein synthesis. Thus, TbEIF4AI is the translation initiation factor that comprises the cap-4 binding complex (Dhalia et al. 2006).

9.2.8 Poly(A) Binding Proteins of Trypanosomatids

PABPs associate with the 3′ ends of eukaryotic mRNAs (Adam et al. 1986) and are involved in a variety of processes related to RNA processing and export, stabilization and translation. In higher eukaryotes PABP associates with the scaffold protein eIF4G, forming a transient mRNA circle during translation initiation (Wells et al. 1998), thus changes in the poly(A) length usually affect translation. This interaction has been targeted by viral proteases that cleave the eIF4G in a manner that prevents mRNA circularization, as part of the viral strategy to take over the cellular translational machinery; PABP too is targeted by viral proteases for degradation, once the cell is taken over by a viral infection (Jackson 2005). PABP also interacts with the eukaryotic Release Factor (eRF), thus playing a role during the final step of

translation (Jacobson 2005; Amrani et al. 2008). In yeast there is a single PABP, whereas metazoan genomes encode several paralogues, with tissue, or embryonic-specific expression (Kahvejian et al. 2005). PABP consists of four RRMs that have different functional features, and a C-terminal domain that is responsible for protein–protein interactions (Deo et al. 1999). The functions of PABP are complex, and can be regulated by PABP- interacting proteins (PAIPs), such as PAIP1 and PAIP2, which are responsible for positive and negative regulations, respectively (Derry et al. 2006).

Orthologues of PABP are encoded by the genomes of various trypanosomatids (Bates et al. 2000). The genomes of *T. brucei* and *T. cruzi* encode two PABP paralogues (Batista et al. 1994; Hotchkiss et al. 1999), both are essential indicating that they are involved in different cellular functions. In addition to these two paralogues, *Leishmania* encode another unique isoform, PABP3 (da Costa Lima et al. 2010). In higher eukaryotes, PABP shuttles between the nucleus and the cytoplasm, but it is mostly cytoplasmic. All three isoforms in *Leishmania* are indeed cytoplasmic, but when transcription is inhibited, PABP-2 and PABP-3 are retained in the nucleus. PABP-1 copurifies with components of the eIF4F complex (da Costa Lima et al. 2010; Zinoviev et al. 2011), suggesting that it is involved in cap-dependent translation. The role of the other two paralogues is yet to be shown.

9.2.9 *Role of 3′ UTR Elements in Differential Translation in* Leishmania *and Trypanosomes*

The role of the upstream and downstream UTRs on mRNA stability and on translation was evaluated in multiple experimental systems, using reporter genes flanked by UTRs derived from genes showing differential expression patterns throughout the life cycle of trypanosomatids. This was shown for the Hsp70 (Quijada et al. 2000; Folgueira et al. 2005) and Hsp83 (Zilka et al. 2001; Larreta et al. 2004) of *Leishmania*. Although it is not easy to differentiate between the contribution of a 3′ UTR to mRNA stability and translation efficiency, deletion analysis of the Hsp83 3′ UTR showed very clearly that preferential translation was uncoupled from stability. Mutations in the 3′ UTR that caused the destabilization of the mRNA during heat shock did not interfere with preferential translation (Zilka et al. 2001). The regulatory region in the Hsp83 3′ UTR of *L. amazonensis* (886 nts) was mapped to an element of 30 nts that mostly consists of polypyrimidines (positions 312–341). The regulatory region was predicted to be associated with a structural arm that was assigned a very high probability, based on the UNAfold algorithm, with Mfold utilities (David et al. 2010). Although elimination of this region prevented the preferential translation of the reporter gene fused to the Hsp83 UTRs, the minimal fragment that could confer a temperature-induced pattern of expression onto a reporter gene was much larger and spanned the proximal half of the 3′ UTR. It was noted that incubation of the wild-type regulatory

element at elevated temperatures (35°C) caused the mRNA fragment to melt. The parallel region containing a mutation that prevented preferential translation of the fused reporter gene under the same conditions, failed to show a similar pattern. Thus, it was assumed that melting of the regulatory region promotes preferential translation of the *Leishmania* HSP83 during stress (David et al. 2010). The sequence of this element is not conserved throughout trypanosomatids, leading to the assumption that RNA structural changes could be involved in this mode of regulation.

Stage-specific translation was also observed for the Amastin genes in *Leishmania*. Bioinformatics analysis combined with genetic assays identified a 450 nts element in the 3′ UTR. Additional mapping experiments narrowed this region down to 150 nts. Unlike the Hsp83 regulatory fragment, the Amastin element that drives increased translation in amastigotes is conserved, and was also found in other amastigote-specific genes, such as Hsp100 (Boucher et al. 2002; McNicoll et al. 2005). However, no defined secondary structure was reported for this RNA element.

9.2.10 Sequences Surrounding the First AUG Initiation Codon

The nucleotides that flank the first AUG affect recognition by the initiator tRNA, and hence translation efficiency (Kozak 1986). In higher eukaryotes, the "Kozak sequence" comprises of a purine at position -3 and a G at position +4, relative to the initiator AUG codon. These nucleotides are required for optimal initiation of protein synthesis (Kozak 1987). To examine whether the nucleotides adjacent to the AUG triplet affect translation in kinetoplastids, a series of mutants was tested in *Leishmania tarentolae* and *Phytomonas serpens*. Expression levels varied over a 20-fold range, but simple conclusions were not possible because the results depended on the downstream coding region (Lukes et al. 2006).

9.3 Translation Elongation

Although less explored, translation elongation provides an additional level of regulation for protein synthesis in higher eukaryotes. This process has to avoid errors during translation, while remaining synchronized with initiation, to prevent the translating ribosomes from stacking on the mRNA (Song et al. 1989; Carr-Schmid et al. 1999). The aminoacyl-tRNAs are recruited by eEF1A and eEF1B, a G-protein and its exchange factor, respectively [reviewed by (Le Sourd et al. 2006)]. Translocation of the nascent polypeptide along the ribosome is driven by eEF2 [reviewed by (Jorgensen et al. 2006)].

eEF1A is one of the most abundant factors in protein synthesis, but it was shown to be involved in a variety of cellular processes other than translation [reviewed by

(Mateyak and Kinzy 2010)]. It has been implicated in nuclear export, and was also shown to participate in proteasome-mediated degradation of proteins, possibly as part of a quality-control process of nascent polypeptides. Several reports also described a role for eEF1A in arrangement of actin filaments (Yang et al. 1990; Murray et al. 1996; Gross and Kinzy 2005) and apoptosis (Soengas and Lowe 2003; Eberle et al. 2007). However, while some of these findings may be attributed to nonspecific interactions of this abundant protein, it is clear that eEF1A has a pivotal role in the cell, and it probably mediates crosstalk between the translational machinery and other cellular processes.

Although the elongation step of translation has not been studied in trypanosomatids, the elongation factors were reported to take part in other processes. For example, *Crithidia fasciculata* contains an unusual trypanothione S-transferase activity that is associated with its eEF1B orthologue**.** Further analysis of EF1B from *L. major* revealed that all its three subunits, expressed as recombinant proteins, were required to conjugate a variety of electrophilic substrates to trypanothione or glutathionyl spermidine, but not to glutathione (Vickers and Fairlamb 2004). Since GST enzymes are known to play a central role in the mechanism of drug resistance [review in (Hayes et al. 2005)], it was suggested that the GST activity of eEF1Bγ would be implicated in translational responses to oxidative stress.

Despite the high conservation of eEF1A from many organisms, biochemical characterization of the *Leishmania* protein highlighted differences when compared to the mammalian orthologue. Most importantly, a 12-amino-acid deletion in the primary sequence of the *Leishmania* protein was identified, which causes a structural change in the protein. As a consequence of this deletion, an exposed region becomes available on the main body of the protein (Nandan et al. 2003). It was shown that targeting of antibodies specifically to this region of LeishEF1A substantially inhibited translation. Therefore, LeishIF1A was suggested as a novel antileishmanial drug target (Lopez et al. 2007).

EF1A from *L. braziliensis* was reported to be important for pathogenesis. The protein is secreted into the macrophage, where it binds and inactivates a specific tyrosine phosphatase (SHP-1), thus blocking the induction of nitric oxide synthase expression in response to infection (Nandan et al. 2002). EF1A in *T. brucei* was also associated with the trypanosome homolog of RACK1 (Regmi et al. 2008), which was initially identified as a signal anchor protein for protein kinase C and its targets (Ron et al. 1994). Silencing of RACK1 causes an increase in the relative level of the unassembled ribosomal subunits and shifting of the polysome profile. RACK1 silencing also increased the sensitivity of cell growth to the translational inhibitor anisomycin. RACK1 interacts with eIF6, the translation initiation factor that recruits the large ribosomal subunit near the AUG start codon. It can therefore be concluded that the trypanosome EF1A is a novel TbRACK1 binding partner and that TbRACK1 is associated with the translational apparatus in trypanosome.

As mentioned above, eEF2 catalyzes the tRNA and mRNA translocation in the 80S ribosome, after formation of the peptide bond [reviewed in (Jorgensen et al. 2006)]. Phosphorylation of eEF2 can transiently inactivate its function, causing a

translational arrest (Redpath et al. 1993). The *T. brucei* EF2 homolog was shown to interact with all the components of the ribosomal P protein complex (stalk) (Smulski et al. 2011), which is involved in the translocation step of protein synthesis (Liljas 1991). Conversely the *T. brucei* EF2 lacks the phosphorylation site known from higher eukaryotes (Smulski et al. 2011). Therefore, the protein probably acts in a manner analogous to the higher eukaryote counterpart, but the regulation of its function is mediated by other mechanisms.

9.4 Stage Differentiation and Changes in Translation

9.4.1 Environmental Switches and Trypanosomatid Life Cycles

Pathogenic trypanosomatids go through a complex digenetic life cycle that involves an invertebrate vector and a mammalian host. A multitude of diseases is therefore caused by organisms of this family, leading to world-wide morbidity and mortality. Subspecies of *T. brucei* are transmitted by tsetse flies in sub-Saharan Africa, threatening both humans and their live stocks. *T. cruzi*, the causative agent of Chagas disease in South America, is transmitted by the Triatom bug, also known as "The Kissing bug." *Leishmania* parasites are transmitted by sand flies, and different species can cause a variety of clinical symptoms, ranging from simple skin lesions to debilitating mucocutaneous infections and to the lethal visceral disease.

Transmission of the parasites between the invertebrate vector and mammalian host is accompanied by switches in the environmental conditions. *T. brucei* parasites remain as extracellular cells in both vector and host, and it is possible to cultivate the vector-specific procyclic form, as well as the blood stream form of this parasite. Unlike *T. brucei*, the leishmanias become obligatory intracellular parasites within their mammalian host, making their cultivation very complex. Promastigotes that reside in the alimentary canal of sand fly vectors are adapted to temperatures of 22–28°C and a mild basic pH. They migrate into the front part of the insect's digestive system, where they go through metacyclogenesis that increases their virulence (Sacks and Perkins 1984). Upon entry into the mammalian host they reside within acidic (pH 5.5) phagolysomal vacuoles and are exposed to temperatures, which are typical for mammalian hosts (33–35°C in the skin, and 37°C in the visceral organs). True amastigotes can be obtained from skin lesions or internal organs of infected animals. Alternatively, they can be extracted from cultured macrophages. In view of the difficulties to obtain intracellular amastigotes, a system for inducing differentiation under axenic conditions was developed. It has been established in former studies that the environmental switches are responsible for the developmental pattern of gene expression, affecting both mRNA stability and translation (Van der Ploeg et al. 1985; Shapira et al. 1988; Boucher et al. 2002; Barak et al. 2005; Rosenzweig et al. 2008).

To induce axenic transformation, promastigotes are exposed to elevated temperatures and to acidic pH, which resemble the conditions that they experience in the phagolysosomal vacuole. As a result of these switches the cells go through major morphological changes, such as adsorbance of the flagellum and rounding up of the cells, until they acquire the shape of amastigotes. A developmental program of gene expression is also induced. It is important to note that axenic transformation was developed for certain species, such as *L. donovani* and *L. infantum*, but not for *L. major* (Bates et al. 1992; Haile and Papadopoulou 2007; Rosenzweig et al. 2008).

9.4.2 Effects of Stress on Translation in Trypanosomatids

9.4.2.1 eIF2α Phosphorylation

The function of eIF2α from higher eukaryotes is regulated by phosphorylation of a conserved Ser51 residue. Phosphorylation of the eIF2α from *T. brucei* occurs on Thr169 (Moraes et al. 2007) and on the Thr 166 residue in the *Leishmania* orthologue (Lahav et al. 2010). The trypanosomatid orthologue has a long N-terminal extension, thus its Thr 169 in *T. brucei* and Thr 166 from *Leishmania* align with Ser51 of the protein from higher eukaryotes. Phosphorylation of eIF2α from *Leishmania* occurs 10–15 h following the differentiation signal and reaches a plateau after 24 h. In trypanosomes though, the use of a mutant strain T169A/– showed that the Thr phosphorylation was not required for downregulation of translation, or for the formation of heat-shock stress granules. Thus, the mechanism that confers a global decrease in translation remains unclear (Kramer et al. 2008).

9.4.2.2 Formation of P-Bodies During Temperature Stress

Exposure to a variety of environmental and metabolic stress conditions has a dramatic effect on mRNAs in eukaryotes. In most cases translation stops, polysomes dissociate, and the mRNAs are stored in RNA stress granules or in processing bodies (P-bodies). In higher eukaryotes, stress granules have been proposed to be associated with mRNA translation and stability, as well as apoptosis and nuclear processes. Once conditions are appropriate, mRNAs can shuttle out of these granules and associate with polysomes (Buchan and Parker 2009). In trypanosomes, heat shock and nutritional stress cause a decrease in polysomes and the appearance of stress granules, in a manner independent of eIF2α phosphorylation (Kramer et al. 2008). Translation is arrested and the mRNAs are stored in atypical granules that do not contain the RNA degradation machinery and are located at the cell periphery (Cassola et al. 2007; Kramer et al. 2008). This is in line with the observation that the insect form of the parasites may experience

shortage in sugars due to the nutritional diet of the vector, leading to a transient arrest of translation.

9.4.3 Translation in Amastigotes Under Axenic Conditions

The availability of axenic transformation systems was utilized for running high-throughput screens of stage-specific gene expression, both at the transcriptome (Haile and Papadopoulou 2007; Peacock et al. 2007) and proteome levels (Rosenzweig et al. 2008a; Rosenzweig et al. 2008b). Combination of the different screens revealed that control of mRNA levels has a clear role early in the differentiation process, while translation and post-translational regulation are more important later on in that process (Lahav et al. 2010).

A systems approach was taken to evaluate the global changes in the transcriptome and the proteome of *Leishmania*, along the process of differentiation from promastigotes to amastigotes, under axenic conditions. During the initial period after signaling for differentiation, translation decreased dramatically. This included the downregulation of ribosomal proteins, phosphorylation of the eIF2α subunit that causes a global arrest of translation (Dever 2002), and a decrease in the general size of polysomes. By following the changes in the transcriptome and proteome, it was shown that mRNA levels were affected mainly at the initial time period of differentiation, while translation regulation was utilized at later stages of differentiation, to achieve altered levels of expression (Lahav et al. 2010).

9.5 Development of a Cell-Free System for Translation

Cell-free (or *in vitro*) translation systems are widely used in translation research, mainly using rabbit reticulocyte lysates (RRL) and wheat germ extracts. These can initiate translation and synthesize full-length polypeptides encoded by mRNAs that are equipped with a m^7GTP cap at their 5′ ends. However, such systems are not fit for testing the role of the trypanosomatid-specific cap-4 structure and its binding proteins in translation, since the latter cannot complement the function of eIF4E from higher eukaryotes and fail to interact with the eukaryotic factors (Yoffe et al. 2004, 2006).

An *in vitro* translation system in *T. cruzi* has been described over 20 years ago; however, it was limited to polypeptides below 42 kDa (Grossi de Sa et al. 1984). Another trial was reported for *T. brucei* bloodstream form cells; this system could translate mainly endogenous mRNAs, and highlighted the requirement for the SL RNA in translating mRNAs (Moreno et al. 1991). An additional attempt to establish an *in vitro* translation system in *T. brucei* bloodstream form cells was described

(Duszenko et al. 1999) but showed ca. 10% efficiency, as compared to RRL. It is important to note that when the *T. brucei* lysates were added to the RRL, a reduction of 80% in the ability to translate external RNAs was observed, suggesting the presence of a low-molecular-mass inhibitory component. However, the inhibitor could not be removed by separation of the protein mix on a sucrose cushion. Despite its low efficiency, this system was able to perform posttranslational modifications of the tested protein (variable surface glycoprotein—VSG). Addition of exogenous mRNA to the RNaseA treated lysates restored translational activity only partially, even when the total RNA was derived from *T. brucei*. This poor restoration was attributed to the low reinitiation rates of the synthesis process.

A more recent attempt that described a new approach to cell-free translation was established, using *L. tarentolae* (Mureev et al. 2009). The RNA was equipped with a structured leader sequence that promotes cap-independent translation initiation, which was shown to be highly efficient in cell-free translation systems derived from different organisms. Translation of endogenous mRNAs could be disrupted by the addition of an oligonucleotide with a sequence complementary to the SL RNA miniexon, which is present in all mRNAs of *L. tarentolae*. This system could serve as a valuable tool for expression of eukaryotic proteins; however, it is not suitable for research of the trypanosomatid translation machinery, since it is cap-independent and contains the anti-SL oligonucleotide.

9.6 Future Prospects

The development of novel molecular and cellular tools for research in trypanosomatids is at a turning point. Until recently, RNAi experiments could not be conducted in *Leishmania*, since most species did not express the components required for this pathway. However, with the novel report on the occurrence of this pathway in *L. braziliensis* (Lye et al. 2010), it should be interesting to see if silencing of the various factors has similar effects in *Leishmania* and in trypanosomes.

Another emerging field of research deals with the effects of secreted parasite molecules on protein synthesis in infected macrophage host cells via the mTOR signaling pathway (Jaramillo et al. 2011). Such secreted molecules are responsible for a dramatic reduction in the ability of host cells to carry on translation. The field of translation is currently attracting much attention, since it reflects metabolic changes that are observed in cancer and disease. Thus the translation machinery provides an exciting target for novel therapeutic approaches (Moerke et al. 2007). Since the parallel system in trypanosomatids shows evolutionary variability, the prospects for its exploitation as a novel target for drug development appears very attractive.

References

Adam SA, Nakagawa T, Swanson MS, Woodruff TK, Dreyfuss G (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Mol Cell Biol 6:2932–2943

Ahmed R, Duncan RF (2004) Translational regulation of Hsp90 mRNA. AUG-proximal 5′-untranslated region elements essential for preferential heat shock translation. J Biol Chem 279:49919–49930

Altmann M, Muller PP, Pelletier J, Sonenberg N, Trachsel H (1989) A mammalian translation initiation factor can substitute for its yeast homologue in vivo. J Biol Chem 264:12145–12147

Amrani N, Ghosh S, Mangus DA, Jacobson A (2008) Translation factors promote the formation of two states of the closed-loop mRNP. Nature 453:1276–1280

Andreev DE, Dmitriev SE, Terenin IM, Prassolov VS, Merrick WC, Shatsky IN (2009) Differential contribution of the m^7G-cap to the 5′ end-dependent translation initiation of mammalian mRNAs. Nucleic Acids Res 37:6135–6147

Arhin GK, Ullu E, Tschudi C (2006) 2′-O-Methylation of position 2 of the trypanosome spliced leader cap 4 is mediated by a 48 kDa protein related to vaccinia virus VP39. Mol Biochem Parasitol 147:137–139

Bangs JD, Crain PF, Hashizume T, Mccloskey JA, Boothroyd JC (1992) Mass spectrometry of mRNA cap-4 from trypanosomatids reveals two novel nucleosides. J Biol Chem 267:9805–9815

Barak E, Amin-Spector S, Gerliak E, Goyard S, Holland N, Zilberstein D (2005) Differentiation of *Leishmania donovani* in host-free system: analysis of signal perception and response. Mol Biochem Parasitol 141:99–108

Bates PA, Cobertson CD, Tetley L, Coombs GH (1992) Axenic cultivation and characterization of *Leishmania mexicana* amastigote-like forms. Parasitology 105:193–202

Bates EJ, Knuepfer E, Smith DF (2000) Poly(A)-binding protein I of *Leishmania*: functional analysis and localisation in trypanosomatid parasites. Nucleic Acids Res 28:1211–1220

Batista JA, Teixeira SM, Donelson JE, Kirchhoff LV, de Sá CM (1994) Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. Mol Biochem Parasitol 67:301–312

Bert AG, Grepin R, Vadas MA, Goodall GJ (2006) Assessing IRES activity in the HIF-1alpha and other cellular 5′ UTRs. RNA 12:1074–1083

Billaut-Mulot O, Fernandez-Gomez R, Loyens M, Ouaissi A (1996) *Trypanosoma cruzi* elongation factor 1-alpha: nuclear localization in parasites undergoing apoptosis. Gene 174:19–26

Boucher N, Wu Y, Dumas C, Dube M, Sereno D, Breton M, Papadopoulou B (2002) A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3′-untranslated region element. J Biol Chem 277:19511–19520

Buchan JR, Parker R (2009) Eukaryotic stress granules: the ins and outs of translation. Mol Cell 36:932–941

Carr-Schmid A, Durko N, Cavallius J, Merrick WC, Kinzy TG (1999) Mutations in a GTP-binding motif of eukaryotic elongation factor 1A reduce both translational fidelity and the requirement for nucleotide exchange. J Biol Chem 274:30297–30302

Cassola A, De Gaudenzi JG, Frasch AC (2007) Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. Mol Microbiol 65:655–670

Cheng S, Gallie DR (2007) eIF4G, eIFiso4G, and eIF4B bind the poly(A)-binding protein through overlapping sites within the RNA recognition motif domains. J Biol Chem 282:25247–25258

Clayton CE (2002) Life without transcriptional control? From fly to man and back again. EMBO J 21:1881–1888

Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156:93–101

da Costa Lima TD, Moura DM, Reis CR, Vasconcelos JR, Ellis L, Carrington M, Figueiredo RC, de Melo Neto OP (2010) Functional characterization of three Leishmania poly(a) binding

protein homologues with distinct binding properties to RNA and protein partners. Eukaryot Cell 9:1484–1494
David M, Gabdank I, Ben-David M, Zilka A, Orr I, Barash D, Shapira M (2010) Preferential translation of Hsp83 in *Leishmania* requires a thermosensitive polypyrimidine-rich element in the 3′ UTR and involves scanning of the 5′ UTR. RNA 16:364–374
Deo RC, Bonanno JB, Sonenberg N, Burley SK (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 98:835–845
Derry MC, Yanagiya A, Martineau Y, Sonenberg N (2006) Regulation of poly(A)-binding protein through PABP-interacting proteins. Cold Spring Harb Symp Quant Biol 71:537–543
Dever TE (2002) Gene-specific regulation by general translation factors. Cell 108:545–556
Dhalia R, Reis CR, Freire ER, Rocha PO, Katz R, Muniz JR, Standart N, de Melo Neto OP (2005) Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues. Mol Biochem Parasitol 140:23–41
Dhalia R, Marinsek N, Reis CR, Katz R, Muniz JR, Standart N, Carrington M, de Melo Neto OP (2006) The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct. Nucleic Acids Res 34:2495–2507
Duszenko M, Kang X, Bohme U, Homke R, Lehner M (1999) In vitro translation in a cell-free system from *Trypanosoma brucei* yields glycosylated and glycosylphosphatidylinositol-anchored proteins. Eur J Biochem 266:789–797
Eberle J, Kurbanov BM, Hossini AM, Trefzer U, Fecker LF (2007) Overcoming apoptosis deficiency of melanoma-hope for new therapeutic approaches. Drug Resist Update 10:218–234
Ferraiuolo MA, Lee CS, Ler LW, Hsu JL, Costa-Mattioli M, Luo MJ, Reed R, Sonenberg N (2004) A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. Proc Natl Acad Sci U S A 101:4118–4123
Folgueira C, Quijada L, Soto M, Abanades DR, Alonso C, Requena JM (2005) The translational efficiencies of the two *Leishmania infantum* HSP70 mRNAs, differing in their 3′-untranslated regions, are affected by shifts in the temperature of growth through different mechanisms. J Biol Chem 280:35172–35183
Freire ER, Dhalia R, Moura DM, da Costa Lima TD, Lima RP, Reis CR, Hughes K, Figueiredo RC, Standart N, Carrington M, de Melo Neto OP (2011) The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. Mol Biochem Parasitol 176:25–36
Gandin V, Miluzio A, Barbieri AM, Beugnet A, Kiyokawa H, Marchisio PC, Biffo S (2008) Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. Nature 455:684–688
Gao H, Ayub MJ, Levin MJ, Frank J (2005) The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. Proc Natl Acad Sci U S A 102:10206–10211
Gingras AC, Gygi SP, Raught B, Polakiewicz RD, Abraham RT, Hoekstra MF, Aebersold R, Sonenberg N (1999a) Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. Genes Dev 13:1422–1437
Gingras AC, Raught B, Sonenberg N (1999b) eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem 68:913–963
Gross SR, Kinzy TG (2005) Translation elongation factor 1A is essential for regulation of the actin cytoskeleton and cell morphology. Nat Struct Mol Biol 12:772–778
Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, Wagner G (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell 115:739–750
Grossi de Sa MF, de Sa C, Pereira de Almeida ER, Barbosa da Cruz W, Filho SA, Gander ES (1984) Optimization of a protein synthesizing lysate system from *Trypanosoma cruzi*. Mol Biochem Parasitol 10:347–354
Haile S, Papadopoulou B (2007) Developmental regulation of gene expression in trypanosomatid parasitic protozoa. Curr Opin Microbiol 10(6):569–577

Hashimoto T, Nakamura Y, Kamaishi T, Adachi J, Nakamura F, Okamoto K, Hasegawa M (1995) Phylogenetic place of kinetoplastid protozoa inferred from a protein phylogeny of elongation factor 1 alpha. Mol Biochem Parasitol 70:181–185

Hayes JD, Flanagan JU, Jowsey IR (2005) Glutathione transferases. Annu Rev Pharmacol Toxicol 45:51–88

Hellen CU, Sarnow P (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev 15:1593–1612

Hershey PE, McWhirter SM, Gross JD, Wagner G, Alber T, Sachs AB (1999) The Cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4GI. J Biol Chem 274:21297–21304

Hess MA, Duncan RF (1996) Sequence and structure determinants of Drosophila Hsp70 mRNA translation: 5′-UTR secondary structure specifically inhibits heat shock protein mRNA translation. Nucleic Acids Res 24:2441–2449

Hiremath LS, Hiremath ST, Rychlik W, Joshi S, Domier LL, Rhoads RE (1989) In vitro synthesis, phosphorylation, and localization on 48S initiation complexes of human protein synthesis initiation factor 4E. J Biol Chem 264:1132–1138

Holcik M, Sonenberg N (2005) Translational control in stress and apoptosis. Nat Rev Mol Cell Biol 6:318–327

Holcik M, Sonenberg N, Korneluk RG (2000) Internal ribosome initiation of translation and the control of cell death. Trends Genet 16:469–473

Hotchkiss TL, Nerantzakis GE, Dills SC, Shang L, Read LK (1999) *Trypanosoma brucei* poly(A) binding protein I cDNA cloning, expression, and binding to 5 untranslated region sequence elements. Mol Biochem Parasitol 98:117–129

Imataka H, Sonenberg N (1997) Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. Mol Cell Biol 17:6940–6947

Jackson RJ (2005) Alternative mechanisms of initiating translation of mammalian mRNAs. Biochem Soc Trans 33:1231–1241

Jacobson A (2005) The end justifies the means. Nat Struct Mol Biol 12:474–475

Jaramillo M, Gomez MA, Larsson O, Shio MT, Topisirovic I, Contreras I, Luxenburg R, Rosnefeld A, Colina R, McMaster RW et al (2011) *Leishmania* repression of host translation through mTOR cleavage is required for parasite survival and infection. Cell Host Microbe 9:331–341

Jorgensen R, Merrill AR, Andersen GR (2006) The life and death of translation elongation factor 2. Biochem Soc Trans 34:1–6

Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN, Sonenberg N (2005) Mammalian poly (A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev 19:104–113

Klemenz R, Hultmark D, Gehring WJ (1985) Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. EMBO J 4:2053–2060

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kramer S, Queiroz R, Ellis L, Webb H, Hoheisel JD, Clayton C, Carrington M (2008) Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169. J Cell Sci 121:3002–3014

Kuge H, Brownlee GG, Gershon PD, Richter JD (1998) Cap ribose methylation of c-mos mRNA stimulates translation and oocyte maturation in Xenopus laevis. Nucleic Acids Res 26:3208–3214

Lahav T, Sivam D, Volpin H, Ronen M, Tsigankov P, Green A, Holland N, Kuzyk M, Borchers C, Zilberstein D et al (2010) Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. FASEB J 25(2):515–25

Laplante M, Sabatini DM (2009) mTOR signaling at a glance. J Cell Sci 122:3589–3594
Larreta R, Soto M, Quijada L, Folgueira C, Abanades DR, Alonso C, Requena JM (2004) The expression of HSP83 genes in *Leishmania infantum* is affected by temperature and by stage-differentiation and is regulated at the levels of mRNA stability and translation. BMC Mol Biol 5:3
Lawrence JC Jr, Abraham RT (1997) PHAS/4E-BPs as regulators of mRNA translation and cell proliferation. Trends Biochem Sci 22:345–349
Le Sourd F, Boulben S, Le Bouffant R, Cormier P, Morales J, Belle R, Mulner-Lorillon O (2006) eEF1B: At the dawn of the 21st century. Biochim Biophys Acta 1759:13–31
Lewdorowicz M, Yoffe Y, Zuberek J, Jemielity J, Stepinskil J, Kierzek R, Stolarskil R, Shapira M, Darzynkiewicz E (2004) Chemical synthesis and binding activity of the trypanosomatid cap-4 structure. RNA 10:1–10
Liang XH, Haritan A, Uliel S, Michaeli S (2003) *Trans-* and *cis-* splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell 2:830–840
Liljas A (1991) Comparative biochemistry and biophysics of ribosomal proteins. Int Rev Cytol 124:103–136
Lin T-A, Kong X, Haystead TAJ, Pause A, Belsham G, Sonenberg N, Lawrence JCJ (1994) PHAS-I as a link between mitogen-activated protein kinase and translation initiation. Science 266:653–56
Linder P (2006) DEAD-box proteins: a family affair–active and passive players in RNP-remodeling. Nucleic Acids Res 34:4168–4180
Lopez M, Cherkasov A, Nandan D (2007) Molecular architecture of *Leishmania* EF-1alpha reveals a novel site that may modulate protein translation: a possible target for drug development. Biochem Biophys Res Commun 356:886–892
Lukes J, Paris Z, Regmi S, Breitling R, Mureev S, Kushnir S, Pyatkov K, Jirku M, Alexandrov KA (2006) Translational initiation in *Leishmania tarentolae* and *Phytomonas serpens* (Kinetoplastida) is strongly influenced by pre-ATG triplet and its 5′ sequence context. Mol Biochem Parasitol 148:125–132
Lye LF, Owens K, Shi H, Murta SM, Vieira AC, Turco SJ, Tschudi C, Ullu E, Beverley SM (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. PLoS Pathog 6:e1001161
Mader S, Lee H, Pause A, Sonenberg N (1995) The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4G and the translational repressors 4E-binding proteins. Mol Cell Biol 15:4990–4997
Mandelboim M, Estrano CL, Tschudi C, Ullu E, Michaeli S (2002) On the role of exon and intron sequences in *trans*-splicing utilization and cap 4 modification of the trypanosomatid *Leptomonas collosoma* SL RNA. J Biol Chem 277:35210–35218
Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1997) Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell 89:951–961
Marcotrigiano J, Lomakin IB, Sonenberg N, Pestova TV, Hellen CU, Burley SK (2001) A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. Mol Cell 7:193–203
Marintchev A, Wagner G (2005) eIF4G and CBP80 share a common origin and similar domain organization: implications for the structure and function of eIF4G. Biochemistry 44:12265–12272
Mateyak MK, Kinzy TG (2010) eEF1A: thinking outside the ribosome. J Biol Chem 285:21209–21213
Matsuo H, Li H, McGuire AM, Fletcher CM, Gingras AC, Sonenberg N, Wagner G (1997) Structure of translation factor eIF4E bound to m^7GDP and interaction with 4E-binding protein. Nat Struct Biol 9:717–724
McGarry TJ, Lindquist S (1985) The preferential translation of *Drosophila* hsp70 mRNA requires sequences in the untranslated leader. Cell 42:903–911

McNicoll F, Muller M, Cloutier S, Boilard N, Rochette A, Dube M, Papadopoulou B (2005) Distinct 3′-untranslated region elements regulate stage-specific mRNA accumulation and translation in *Leishmania*. J Biol Chem 280:35238–35246

Methot N, Pause A, Hershey JW, Sonenberg N (1994) The translation initiation factor eIF-4B contains an RNA-binding region that is distinct and independent from its ribonucleoprotein consensus sequence. Mol Cell Biol 14:2307–2316

Michaeli S (2011) *Trans*-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. Future Microbiol 6:459–474

Mitchell SF, Walker SE, Algire MA, Park EH, Hinnebusch AG, Lorsch JR (2010) The 5′-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. Mol Cell 39:950–962

Moerke NJ, Aktas H, Chen H, Cantel S, Reibarkh MY, Fahmy A, Gross JD, Degterev A, Yuan J, Chorev M et al (2007) Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. Cell 128:257–267

Moraes MC, Jesus TC, Hashimoto NN, Dey M, Schwartz KJ, Alves VS, Avila CC, Bangs JD, Dever TE, Schenkman S et al (2007) A novel membrane-bound eIF2{alpha} kinase in the flagellar pocket of *Trypanosoma brucei*. Eukaryot Cell 6(11):1979–91

Moreno SN, Ip HS, Cross GA (1991) An mRNA-dependent *in vitro* translation system from *Trypanosoma brucei*. Mol Biochem Parasitol 46:265–274

Mureev S, Kovtun O, Nguyen UT, Alexandrov K (2009) Species-independent translational leaders facilitate cell-free expression. Nat Biotechnol 27:747–752

Murray JW, Edmonds BT, Liu G, Condeelis J (1996) Bundling of actin filaments by elongation factor 1 alpha inhibits polymerization at filament ends. J Cell Biol 135:1309–1321

Nakamura Y, Hashimoto T, Kamaishi T, Adachi J, Nakamura F, Okamoto K, Hasegawa M (1996) Phylogenetic position of kinetoplastid protozoa inferred from the protein phylogenies of elongation factors 1alpha and 2. Journal of biochemistry 119:70–79

Nandan D, Yi T, Lopez M, Lai C, Reiner NE (2002) *Leishmania* EF-1alpha activates the Src homology 2 domain containing tyrosine phosphatase SHP-1 leading to macrophage deactivation. J Biol Chem 277:50190–50197

Nandan D, Cherkasov A, Sabouti R, Yi T, Reiner NE (2003) Molecular cloning, biochemical and structural analysis of elongation factor-1 alpha from *Leishmania donovani:* comparison with the mammalian homologue. Biochem Biophys Res Commun 302:646–652

Oberer M, Marintchev A, Wagner G (2005) Structural basis for the enhancement of eIF4A helicase activity by eIF4G. Genes Dev 19:2212–2223

Pause A, Belsham GJ, Gingras A-C, Donzé O, Lin T-A, Lawrence JCJ (1994) Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5′-cap function. Nature 371:762–7

Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M et al (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat Genet 39:839–847

Pestova TV, Lorch JR, Hellen CHT (2007) The mechanism of translation initiation in Eukaryotes. In: Mathews MB, Sonenberg N, Hershey JWB (eds) Translation control in biology and medicine. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 87–128

Pestova TV, de Breyne S, Pisarev AV, Abaeva IS, Hellen CU (2008) eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. Embo J 27:1060–1072

Pilipenko EV, Pestova TV, Kolupaeva VG, Khitrina EV, Poperechnaya AN, Agol VI, Hellen CU (2000) A cell cycle-dependent protein serves as a template-specific translation initiation factor. Genes Dev 14:2028–2045

Quijada L, Soto M, Alonso C, Requena JM (2000) Identification of a putative regulatory element in the 3′-untranslated region that controls expression of HSP70 in Leishmania infantum. Mol Biochem Parasitol 110:79–91

Rau M, Ohlmann T, Morley SJ, Pain VM (1996) A reevaluation of the cap-binding protein, eIF4E, as a rate-limiting factor for initiation of translation in reticulocyte lysate. J Biol Chem 271:8983–8990

Redpath NT, Price NT, Severinov KV, Proud CG (1993) Regulation of elongation factor-2 by multisite phosphorylation. Eur J Biochem 213:689–699

Regmi S, Rothberg KG, Hubbard JG, Ruben L (2008) The RACK1 signal anchor protein from *Trypanosoma brucei* associates with eukaryotic elongation factor 1A: a role for translational control in cytokinesis. Mol Microbiol 70:724–745

Richter JD, Sonenberg N (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature 433:477–480

Ron D, Chen CH, Caldwell J, Jamieson L, Orr E, Mochly-Rosen D (1994) Cloning of an intracellular receptor for protein kinase C: a homolog of the beta subunit of G proteins. Proc Natl Acad Sci U S A 91:839–843

Rosenzweig D, Smith D, Myler PJ, Olafson RW, Zilberstein D (2008a) Post-translational modification of cellular proteins during *Leishmania donovani* differentiation. Proteomics 8:1843–1850

Rosenzweig D, Smith D, Opperdoes F, Stern S, Olafson RW, Zilberstein D (2008b) Retooling Leishmania metabolism: from sand fly gut to human macrophage. FASEB J 22(2):590–602

Ruan JP, Ullu E, Tschudi C (2007) Characterization of the *Trypanosoma brucei* cap hypermethylase Tgs1. Mol Biochem Parasitol 155:66–69

Rubtsova MP, Sizova DV, Dmitriev SE, Ivanov DS, Prassolov VS, Shatsky IN (2003) Distinctive properties of the 5′-untranslated region of human hsp70 mRNA. J Biol Chem 278:22350–22356

Sacks DL, Perkins PV (1984) Identification of an infective stage of *Leishmania* promastigotes. Science 223:11417–11419

Shapira M, McEwen JG, Jaffe CL (1988) Temperature effects on molecular processes which lead to stage differentiation in *Leishmania*. EMBO J 7:2895–2901

Smulski CR, Longhi SA, Ayub MJ, Edreira MM, Simonetti L, Gomez KA, Basile JN, Chaloin O, Hoebeke J, Levin MJ (2011) Interaction map of the *Trypanosoma cruzi* ribosomal P protein complex (stalk) and the elongation factor 2. J Mol Recognit 24(2):359–70

Soengas MS, Lowe SW (2003) Apoptosis and melanoma chemoresistance. Oncogene 22:3138–3151

Song JM, Picologlou S, Grant CM, Firoozan M, Tuite MF, Liebman S (1989) Elongation factor EF-1 alpha gene dosage alters translational fidelity in *Saccharomyces cerevisiae*. Mol Cell Biol 9:4571–4575

Terenin IM, Dmitriev SE, Andreev DE, Shatsky IN (2008) Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. Nat Struct Mol Biol 15:836–841

Tomoo K, Shen X, Okabe K, Nozoe Y, Fukuhara S, Morino S, Sasaki M, Taniguchi T, Miyagawa H, Kitamura K et al (2003) Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations. J Mol Biol 328:365–383

Unbehaun A, Marintchev A, Lomakin IB, Didenko T, Wagner G, Hellen CU, Pestova TV (2007) Position of eukaryotic initiation factor eIF5B on the 80S ribosome mapped by directed hydroxyl radical probing. Embo J 26:3109–3123

Van der Ploeg LHT, Giannini SH, Cantor CR (1985) Heat shock genes: regulatory role for differentiation in parasitic protozoa. Science 228:1443–1446

Vickers TJ, Fairlamb AH (2004) Trypanothione S-transferase activity in a trypanosomatid ribosomal elongation factor 1B. J Biol Chem 279:27246–27256

Vivinus S, Baulande S, van Zanten M, Campbell F, Topley P, Ellis JH, Dessen P, Coste H (2001) An element within the 5′ untranslated region of human Hsp70 mRNA which acts as a general enhancer of mRNA translation. Eur J Biochem 268:1908–1917

Wells SE, Hillner PE, Vale RD, Sachs AB (1998) Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell 2:135–140

Wilson JE, Pestova TV, Hellen CU, Sarnow P (2000) Initiation of protein synthesis from the A site of the ribosome. Cell 102:511–520

Yamamoto Y, Singh CR, Marintchev A, Hall NS, Hannig EM, Wagner G, Asano K (2005) The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G. Proc Natl Acad Sci U S A 102:16164–16169, Epub 12005 Oct 16127

Yang F, Demma M, Warren V, Dharmawardhane S, Condeelis J (1990) Identification of an actin-binding protein from Dictyostelium as elongation factor 1a. Nature 347:494–496

Yoffe Y, Zuberek J, Lewdorowicz M, Zeira Z, Keasar C, Orr-Dahan I, Jankowska-Anyszka M, Stepinski J, Darzynkiewicz E, Shapira M (2004) Cap-binding activity of an eIF4E homolog from *Leishmania*. RNA 10:1764–1775

Yoffe Y, Zuberek J, Lerer A, Lewdorowicz M, Stepinski J, Altmann M, Darzynkiewicz E, Shapira M (2006) Binding specificities and potential roles of isoforms of Eukaryotic initiation factor 4E in *Leishmania*. Eukaryot Cell 12:1969–1979

Yoffe Y, Leger M, Zinoviev A, Zuberek J, Darzynkiewicz E, Wagner G, Shapira M (2009) Evolutionary changes in the *Leishmania* eIF4F complex involve variations in the eIF4E-eIF4G interactions. Nucleic Acids Res 37:3243–3253

Yueh A, Schneider RJ (2000) Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18 S rRNA. Genes Dev 14:414–421

Zamudio JR, Mittra B, Zeiner GM, Feder M, Bujnicki JM, Sturm NR, Campbell DA (2006) Complete cap 4 formation is not required for viability in *Trypanosoma brucei*. Eukaryot Cell 5:905–915

Zamudio JR, Mittra B, Campbell DA, Sturm NR (2009) Hypermethylated cap 4 maximizes *Trypanosoma brucei* translation. Mol Microbiol 72:1100–1110

Zilka A, Garlapati S, Dahan E, Yaolsky V, Shapira M (2001) Developmental regulation of heat shock protein 83 in *Leishmania*. 3′ processing and mRNA stability control transcript abundance, and translation is directed by a determinant in the 3′-untranslated region. J Biol Chem 276:47922–47929

Zinoviev A, Leger M, Wagner G, Shapira M (2011) A novel 4E-interacting protein in *Leishmania* is involved in stage-specific translation pathways. Nucleic Acids Res 39:8404–8415

Chapter 10
Mitochondrial Translation in Trypanosomatids

Dmitri A. Maslov and Rajendra K. Agrawal

Abstract For a long period of time, the functional properties and the very existence of mitochondrial translation in trypanosomatids remained controversial. The unusual resistance of the system to the known inhibitors of translation, such as chloramphenicol, in conjunction with the extreme hydrophobicity of the translation products appeared to be the main factors that made detection and characterization of this system so difficult. As of today, only two mitochondrial translation products have been reliably identified in *Leishmania tarentolae* and *Trypanosoma brucei*: cytochrome *c* oxidase subunit I (COI) and apocytochrome *b* (Cyb) that are encoded by nonedited and 5′-edited mRNAs, respectively. A large body of circumstantial evidence suggests that the F_1F_0 ATPase subunit A6 and ribosomal protein S12 are also expressed in trypanosomatid mitochondria. The issue of existence of kinetoplast-mitochondrial ribosomes has been addressed lately by the isolation of the 50S monosome particles and reconstruction of their 3D-structure using single-particle cryo-electron microscopy. The overall architecture of these particles strikingly resembles that of eubacterial ribosomes, despite profound differences in the size of ribosomal RNAs and the protein content of these two classes of ribosomes. Evidence begins to accumulate that in order to selectively achieve translation of the fully edited mRNA templates the mitoribosomes are involved in higher-order interactions with mRNA editing and polyadenylation machineries. The pentatricopeptide repeat (PPR) proteins emerge as important participants in these interactions.

D.A. Maslov (✉)
Department of Biology, University of California—Riverside, Riverside, CA 92521, USA
e-mail: maslov@ucr.edu

R.K. Agrawal
Division of Translational Medicine, Wadsworth Center, New York State Department of Health, Albany, NY 12201, USA
e-mail: agrawal@wadsworth.org

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_10,

Keywords Cryo-electron microscopy • Leishmania tarentolae • Mitochondrial translation • Mitoribosome • Pentatricopeptide repeat (PPR) proteins • RNA editing • Trypanosoma brucei

10.1 Introduction

The initial discovery of mitochondrial protein synthesis (or translation) in human beings (McLean et al. 1958) was rapidly followed by several findings of this process in other organisms. The distinct feature of mitochondrial translation that allowed for its clear separation from the translation in the cell cytoplasm was its sensitivity to chloramphenicol and other inhibitors of bacterial protein synthesis, although cytoplasmic translation was insensitive to these inhibitors (Lamb et al. 1968). Conversely, mitochondrial translation was found to be resistant to cycloheximide, an efficient inhibitor of cytoplasmic translation. These observations were consistent with the bacterial origins of mitochondria and were generally regarded as universal. The finding of the separate mitochondrial protein synthesis was rapidly followed by the identification of mitochondrial ribosomes (henceforth referred to as mitoribosomes), first indirectly as membrane-associated ribonucleoprotein granules observed by electron microscopy and later by cosedimentation of ribosomal RNA (rRNA)-containing complexes with nascent polypeptides (reviewed by Curgy 1985). As expected, mitoribosomes proved to be quite different from cytoplasmic ribosomes (O'Brien 2002; O'Brien and Denslow 1996; O'Brien and Kalf 1967a). However, in spite of their likely common origin and the shared sensitivity to certain inhibitors, mitoribosomes apparently diverged from their bacterial counterparts, with changes affecting both their functional properties and the architecture (Borst and Grivell 1971; O'Brien 2003; Pel and Grivell 1994; Sharma et al. 2003; Spremulli et al. 2004). Mitoribosomes are specialized in translation of a limited number of templates (usually under 20) that, with few exceptions, encode integral membrane proteins—subunits of the respiratory chain complexes. The extreme hydrophobicity of these products requires a tight association of mitoribosomes with membranes (Gruschke and Ott 2010; Ott and Herrmann 2010). Apparently in order to accommodate these requirements, the mitochondrial rRNAs, which are invariably encoded by mitochondrial DNA, have undergone a significant reduction in their sizes (Mears et al. 2002, 2006). This loss in rRNA, to some degree, is compensated by a large and diverse set of ribosomal proteins that are encoded in the nucleus, synthesized in the cytoplasm and then imported into mitochondria (O'Brien 2002; Pietromonaco et al. 1991; Smits et al. 2007). Although a large amount of data describing structural and functional aspects of mitochondrial ribosomes from various sources have been accumulated, their detailed structural characterization has been achieved only in a few cases using three-dimensional (3D) cryo-electron microscopy (cryo-EM) (Sharma et al. 2003, 2009). In this chapter, we first present a brief historical account followed by recent developments in the studies of trypanosomatid mitochondrial translation and the machinery involved in the process.

10.2 Protein Synthesis in Kinetoplast-Mitochondria

The early attempts to detect mitochondrial translation in trypanosomatids using the differential sensitivity approach yielded somewhat inconclusive, and often controversial, results. In an insect trypanosomatid, *Crithidia luciliae*, a large amount (up to 50%) of the total protein synthesis was found to be chloramphenicol-sensitive (Laub-Kupersztejn and Thirion 1974). In line with the findings in other systems, this effect was attributed to the mitochondrial protein synthesis. The authors, however, used unusually high (500–800 μg/ml) concentrations of the antibiotic, as in other systems a nearly complete inhibition is achieved usually at much lower concentrations (100 μg/ml). Moreover, the work of Laub-Kupersztejn and Thirion was almost immediately challenged by other group (Kleisen and Borst 1975), which found that only a 2% inhibition of the total protein synthesis in this organism is caused by a similarly high (400 μg/ml) concentration of the drug.

Likewise, in *Trypanosoma brucei*, a strong inhibitory effect of chloramphenicol was observed only at the concentrations exceeding 300 μg/ml; however, this effect was attributed to the secondary effects related to a concomitant inhibition of respiration (Spithill et al. 1981). The lack of a direct inhibitory effect at the lower concentrations was attributed to the membrane permeability barriers; however, even with isolated mitochondria, any significant (~50%) effect on the total protein synthesis was observed only at high concentrations of D-chloramphenicol. Similarly, other bacterial protein-synthesis inhibitors were effective only at high concentrations. The mitochondrial protein synthesis in *T. brucei* was revisited using digitonin-permeabilized cells (Shu and Göringer 1998), and a small (~9%) fraction of the total synthesis was found to be resistant to the effect of cycloheximide. It was further reduced to ~2% by treatment with chloramphenicol at 142 μg/ml, leading to the conclusion that this fraction represented the mitochondrial protein synthesis that was intrinsically sensitive to this antibiotic. This conclusion was reiterated using isolated mitochondria from *T. brucei* that were treated with a low concentration (50 μg/ml) of the inhibitor without additional permeabilization (Nabholz et al. 1999).

The drug permeability issue was also addressed by using digitonin-treated mitochondria from an insect trypanosomatid, *Crithidia fasciculata* (Hill et al. 1975). As is typical for trypanosomatids, this organism was refractory to the presence of chloramphenicol in the medium, although its growth was severely inhibited even at a low (10 μg/ml) concentration of cycloheximide. Accordingly, the incorporation of radiolabeled amino acids in isolated mitochondria was not affected by chloramphenicol, while that in the cytosolic fraction was sensitive to cycloheximide. Remarkably, although the treatment of isolated mitochondria with digitonin stimulated the incorporation of labeled amino acids, this did not cause the organelles to become sensitive to chloramphenicol even at a 250 μg/ml concentration. As a control, the synthesis was shown to be sensitive to puromycin, a structural mimic of the CCA end of an aminoacylated tRNA and the universal inhibitor of translation. The resistance of mitochondrial translation to chloramphenicol in

isolated organelles from *C. fasciculata* was also confirmed later (Tittawella 1998). These results strongly suggested that the lack of sensitivity to chloramphenicol at concentrations effective in other systems was not an artifact, but represented an intrinsic property of the mitoribosomes. This view was further strengthened by the sequence analysis of *T. brucei* and *L. tarentolae* mitochondrial rRNAs (further discussed below) which showed that the peptidyl transferase center in these mitoribosomes was altered in a way that would confer resistance to chloramphenicol (de la Cruz et al. 1985b; Eperon et al. 1983).

Although the contradictory results described above still remain unexplained, the issue of sensitivity or insensitivity of the kinetoplast mitochondrial translation to inhibitors is secondary to the main question about the very existence and uniqueness of the process itself. Because of the failure of the previous approaches to unequivocally address this question, the emphasis over the time has shifted from analysis of the bulk protein synthesis toward the detection and identification of individual proteins encoded by the kinetoplastid mitochondrial genome. Besides the two rRNAs, one each corresponding to two ribosomal subunits, the maxicircles of kDNA in *T. brucei*, *L. tarentolae*, and *C. fasciculata* encode 18 polypeptides. The mRNAs corresponding to 12 of these polypeptides must undergo posttranscriptional editing to various extents ("edited" mRNA) in order to become translation-competent (Arts et al. 1993; Benne et al. 1986; Feagin 2000; Simpson et al. 1998, 2000; Stuart and Feagin 1992). The remaining polypeptides are encoded by mRNAs that do not require editing ("never-edited" or "nonedited" mRNA). A significant body of circumstantial evidence, mainly the fact that edited mRNAs in trypanosomes encode polypeptides homologous to functional mitochondrial proteins in other organisms, indicated that these edited mRNAs are indeed translated (Blum et al. 1990; Feagin et al. 1988; Shaw et al. 1988). Additional evidence included the finding of an immunoreactive polypeptide of the size expected for COII (Breek et al. 1997; Shaw et al. 1989), detection of a subunit with the mass matching that of COII by mass spectrometry analysis of the purified cytochrome *c* oxidase complex (Breek et al. 1997), and induction of an antimycin A resistance mutation in the maxicircle gene encoding apocytochrome *b* (Cyb) mRNA (Schnaufer et al. 2000). However, the direct proof in the form of amino acid sequence was still missing. Furthermore, the anticipated mitochondrially encoded polypeptides could not be detected among subunits of the biochemically purified respiratory complexes (Priest and Hajduk 1992; Speijer et al. 1996, 1997). That was puzzling given the enzymatic activity of the isolated complexes and the crucial role played by their mitochondrially encoded subunits (Berry et al. 2000; Ludwig et al. 2001; Mulkidjanian et al. 2007; Remacle et al. 2008). Therefore, it was hypothesized that an extreme hydrophobicity of the mitochondrial subunits had led to preferential loss of these polypeptides due to aggregation and precipitation, upon disassembly of the respiratory enzyme complexes with SDS (Breek et al. 1997).

The effect of hydrophobicity was employed to identify mitochondrially encoded subunits of cytochrome *c* oxidase and cytochrome bc_1 isolated from mitochondria of *Leishmania tarentolae*. The method is based on the abnormal electrophoretic

mobility of hydrophobic polypeptides in electrophoretic gels; apparently due to a higher SDS content, these polypeptides migrate faster in low-concentration gels as compared to most proteins (Marres and Slater 1977). As a result, in a two-dimensional 9% versus 14% polyacrylamide Tris–glycine–SDS gel, such polypeptides are found as spots located off the main diagonal formed by proteins with normal migration properties. Therefore, even the low-intensity fuzzy spots formed by mitochondrially encoded polypeptides would no longer be obscured by the prominent bands of nuclear-encoded polypeptides of the respiratory enzyme complexes. This approach led to the identification of a series of off-diagonal spots in cytochrome *c* oxidase (Horváth et al. 2000b). The subunit I (COI) polypeptide is encoded by the nonedited mRNA (Simpson et al. 1998). Antibodies raised against an internal peptide inferred from the COI mRNA sequence reacted with some of these spots and this identification was confirmed by partial N-terminal amino acid sequencing. Apparently, a partial oligomerization of the hydrophobic COI polypeptide had occurred upon dissociation of the complex. Another off-diagonal spot was putatively identified by partial sequencing as cytochrome *c* oxidase subunit II (COII). The *L. tarentolae* COII mRNA is minimally edited by insertion of four U-residues in the internal pre-edited site (Shaw et al. 1989) and, therefore, the N-terminal sequence determined is translated from the nonedited segment. A few off-diagonal spots were also detected with cytochrome bc_1 (Horváth et al. 2000a). This complex contained a single, mitochondrially encoded subunit—cytochrome *b* (Cyb), which in *L. tarentolae* is encoded by an mRNA that undergoes 5′-editing by insertion of 39 U-residues at 15 editing sites (Feagin et al. 1988). The anticipated N-terminal amino acid sequence is therefore derived from the edited sequence. The match of the determined Cyb sequence from two of these spots, with the sequence predicted by edited mRNA, was the first direct confirmation of the functionality of RNA editing in trypanosome mitochondria. Again, a partial oligomerization was responsible for multiplicity of spots observed in the gel. Interestingly, in each of the cases described above, the deblocking of the N-terminal methionine was necessary, suggesting that the initiator tRNA was formylated.

The predicted A6 subunit of F_1F_0 ATPase (Bhat et al. 1990) and ribosomal protein S12 (Maslov et al. 1992; Maslov and Simpson 1992), as well as putative subunits of NADH dehydrogenase, still remain to be identified as mitochondrially encoded gene products in trypanosomatids. The latter set represents a more challenging task because the respective complex seems to be relatively low in quantity and high in complexity, as compared to the other respiratory enzymes in trypanosomatids (Čermáková et al. 2007; Fang et al. 2001; Opperdoes and Michels 2008).

The de novo synthesis of COI and Cyb polypeptides was subsequently investigated via labeling of *L. tarentolae* cells with radioactive amino acids in the presence of cycloheximide followed by the 2D gel analysis (Horváth et al. 2002). Several off-diagonal radiolabeled spots were observed, with some of them matching the positions of monomeric COI and Cyb and identified by direct sequencing. The similar set of products was also observed using translation in isolated kinetoplast-mitochondrial fraction. After pulse-chase labeling, radioactive

COI and Cyb polypeptides were incorporated in the respective respiratory complexes, as evidenced by native gel electrophoresis of the detergent-solubilized mitochondrial membranes. Analysis of mitochondrial translation in cycloheximide-inhibited *T. brucei* cells has also revealed several off-diagonal spots (Aphasizheva et al. 2011; Neboháčová et al. 2004). Among the radioactive products observed, the COI product was identified by matching its gel migration with *L. tarentolae* COI and tracing its incorporation in the cytochrome *c* oxidase complex (D. A. Maslov, unpublished observations). The comigration of radiolabeled putative Cyb product with a discrete protein spot, which was identified by mass spectrometry, provided another marker for mitochondrially encoded proteins in *T. brucei* (Neboháčová et al. 2004). As expected, the synthesis of COI and Cyb polypeptides in *L. tarentolae* and *T. brucei* was insensitive to chloramphenicol and other inhibitors at a concentration of 100 μg/ml, but was sensitive to puromycin (Horváth et al. 2002; Neboháčová et al. 2004).

It was noticed that some fully edited mRNA sequences, e.g., pan-edited RPS12 (G6 or CR6) and G4 (CR4) in *T. brucei* (Corell et al. 1994; Read et al. 1992), and G3 in *Phytomonas serpens* (Maslov et al. 1999), contain alternative reading frames encompassing almost the entire mRNA-coding regions but have no sequence homology in the databases. Whether just one or both polypeptides are expressed in each of these cases remains an open question. Nevertheless, it is possible that RNA editing expands the coding capabilities of the trypanosomatid mitochondrial genome, by defining variations in amino acid sequence through alternative editing patterns. Thus, there seem to exist two mature editing patterns for the C terminus of ND3 (CR5) mRNA in *T. brucei* (Read et al. 1994b) and ND9 (G2) in *Leishmania mexicana amazonensis* (Maslov 2010). In *T. brucei*, a partially edited COIII mRNA seems to be translated to produce a protein that is involved in kinetoplast DNA maintenance (Ochsenreiter et al. 2008; Ochsenreiter and Hajduk 2006). However, experimental determination of amino acid sequences of the predicted polypeptides in each of these cases is still missing.

A panel of radiolabeled products and immunoreactive polypeptides with apparently normal gel-migration properties was identified in *Leishmania tropica* to represent several mitochondrial proteins including the components of cytochrome *c* oxidase and cytochrome bc_1 complexes (Goswami et al. 2006; Mukherjee et al. 2007). Given the unusual properties of the *bona-fide* COI and Cyb proteins in the closely related *L. tarentolae*, this identification needs further verification.

10.3 Kinetoplast-Mitochondrial Ribosomes, tRNA, and Translation Factors

Historically, the task of detecting mitoribosomes in other systems was achieved by using their cosedimentation with nascent peptides produced in the presence of cycloheximide (Attardi and Ojala 1971; Brega and Vesco 1971; Greco et al. 1973;

O'Brien and Kalf 1967a; b). The mitochondrial origin of these particles was ascertained by the sensitivity of this labeling to chloramphenicol. In trypanosomatids, however, this approach proved to be unsuccessful, and the nature of mitoribosomes of trypanosomatids remained an open question for a long time. In their pioneer work, Hanas and coworkers (1975) were able to detect 72S particles in a total cell lysate of *T. brucei* that were distinct from the cytoplasmic ~80S ribosomes. The mitochondrial origin of the 72S particles was tentatively deduced from the sensitivity of association of the nascent peptides to chloramphenicol, although a high concentration of the antibiotic was used. In a manner characteristic for mitoribosomes, these particles did not dissociate at low Mg^{2+} concentrations, but their dissociation could be achieved by increasing the K^+ concentration. However, in a later study, at odds with the previous findings, a rather heterogeneous population of the ribonucleoprotein particles, containing the ribosomal small subunit (SSU) 9S and large subunit (LSU) 12S RNAs, was observed in a mitochondrial lysate of *T. brucei* (Shu and Göringer 1998). Among those, the 80S particles were found that apparently contained both rRNAs, and, therefore, they were also the candidate monosomes. These particles dissociated into smaller subcomplexes at a lower Mg^{2+} concentration. Yet in *C. fasciculata*, the ribosomal ribonucleoprotein (RNP) complexes that could be detected by the association with nascent polypeptides were much smaller, about 35S (Tittawella et al. 2003), and, also at odds with the previous report for the same species (Tittawella 1998), this association appeared to be sensitive to chloramphenicol. The reasons for these irreconcilable findings remain unclear.

Besides these results, the earlier reports described the chloramphenicol-sensitive 68S ribosomes in *Crithidia oncopelti* (Zaitseva et al. 1977) and 60S ribosomes in *C. luciliae* (Laub-Kuperszteijn and Thirion 1974). More recently, 75S mitoribosomes were reported from *L. tropica* and *L. donovani* (Maarouf et al. 1995). From their sensitivity to bacterial inhibitors of protein synthesis and the dissociation of their subunits at lower Mg^{2+} concentrations, these ribosomes showed similarity to their bacterial counterparts. In none of these cases, however, the characterization of the isolated particles was sufficiently detailed, while the case of *C. oncopelti* was additionally complicated by the presence of intracellular bacterial endosymbionts (Chang 1974).

In *L. tarentolae*, particles with morphologies reminiscent of the typical monosomes and their small and large subunits were detected by electron microscopic analysis of a mitochondrial lysate (Scheinman et al. 1993). Subsequently, analysis of the sedimentation profiles of the SSU 9S and LSU 12S rRNAs in the gradients was used to detect and isolate ribosomal particles in a detergent lysate of the mitochondria (Maslov et al. 2006). The most abundant of these, termed 45S SSU*, could be purified to near homogeneity. They were shown to contain only the 9S rRNA and nearly 50 proteins, including several universal homologs of small subunit ribosomal proteins, such as S5, S6, S9, S11, S15–S18 and MRPS29 (Maslov et al. 2007). Besides these, there were several pentatricopeptide repeat (PPR)-containing proteins that in other systems are known to participate in various aspects of mRNA maturation (Delannoy et al. 2007; Schmitz-Linneweber and

Small 2008). As determined by electron microscopy analysis, the 45S SSU* particles displayed an unusual bilobed morphology. The current view is that they represent a mixture of homodimers of the 30S SSU and heterodimers of the 30S SSU and an unknown protein complex of about the same size as the 30S SSU. The function of these usual complexes remains unclear. The presence of the PPR proteins among its components suggests that these complexes may participate in sequence-specific mRNA recognition for translation, or in the biogenesis of the ribosomes.

The pool of the free 30S SSU particles was relatively small (Maslov et al. 2006, 2007). On the contrary, the pool of the free LSU was large with the majority of the 12S rRNA found within these 40S particles. The typical LSU morphology of these complexes was easily recognizable by transmission electron microscopy. Both 45S SSU* and 40S LSU complexes tend to dimerize, adding to the multiplicity of the ribosomal RNP complexes present in the mitochondrial lysate.

There was no prominent peak of the monosomes that could be expected to contain both rRNAs with an equimolar ratio. The only candidate for this role was the 50S complex that was obscured in the gradients by the peaks of more abundant 40S LSU, 45S SSU* complexes and their dimers. However, electron microscopy analysis of the 50S material clearly revealed the presence of the characteristic monosome-like particles in this fraction.

The subsequent detailed structural analysis of these particles by single-particle 3D cryo-EM (Sharma et al. 2009) revealed that the overall morphology and dimensions of the *Leishmania* mitoribosome (~245 Å in diameter; henceforth referred to as Lmr) are more similar to those of the 70S eubacterial ribosome (~260 Å) rather than to its mammalian mitochondrial counterpart (~320 Å). The common morphological features, such as the body, the head, and the platform of the small subunit, and the central protuberance and the side-stalks of the large subunit are obvious in Lmr (Fig. 10.1; for a detailed comparison with eubacterial and mammalian mitochondrial ribosomes, see Agrawal et al. 2011). Such a conventional morphology of Lmr was unanticipated given the large differences in the composition of Lmr and bacterial ribosomes. For example, the L1-like stock is obvious in the structure, although there is no protein homologous to eubacterial L1. Moreover, the sizes of the small (9S) and large (12S) rRNAs are greatly (by more than 60%) reduced compared with *E. coli* (de la Cruz et al. 1985a, b; Eperon et al. 1983). In *L. tarentolae* these RNAs are only of 610 and 1,173 nt, respectively, and these are even smaller than the rRNAs in mammalian mitochondria, which are among the smallest (12S for the SSU and 16S for the LSU rRNA). A superimposition of the secondary structures of the *Leishmania* mitochondrial and bacterial rRNA molecules (Figs. 10.2a, b) shows that this reduction in sizes was due to the absence of several stem–loop segments in the former. Thus, the 9S SSU rRNA does not contain 24 out of 45 stem–loops present in the *E. coli* structure. In the 12S LSU rRNA only 42 helical regions are preserved out of 101 such regions of *E. coli*. The entire Domain I is absent, while Domain III is greatly reduced. Domains II, IV, and VI (in particular, the linker to its α-sarcin/ricin stem–loop) are also reduced in size. Only Domain V, which forms the peptidyl-transferase

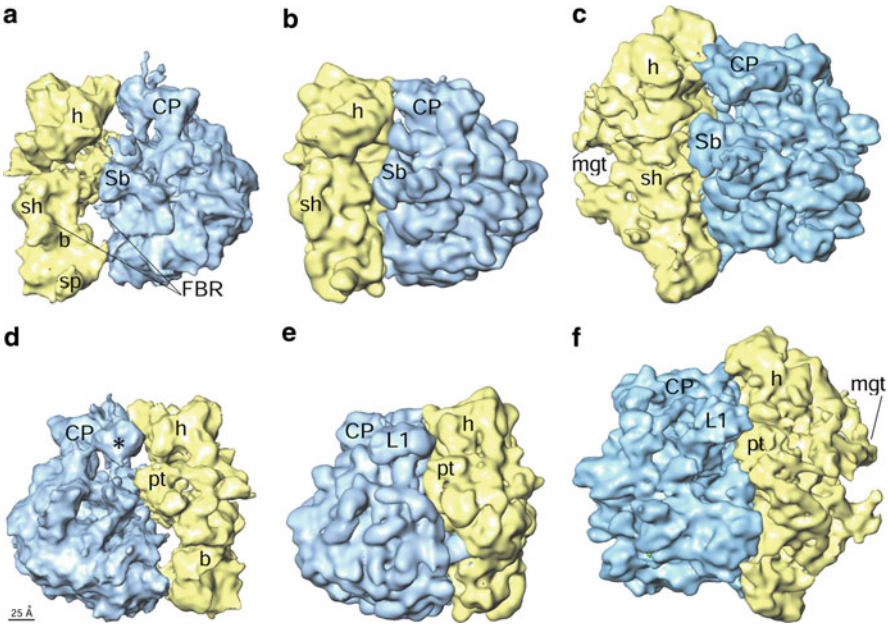


Fig. 10.1 Three-dimensional cryo-electron microscopic map of the *Leishmania* 50S mitoribosome (Lmr) (**a**, **d**) (Sharma et al. 2009) and its comparison with the *E. coli* 70S ribosome (**b**, **e**) (Gabashvili et al. 2000) and the bovine 55S mitoribosome (**c**, **f**) (Sharma et al. 2003). **a**, **b**, **c**—view from the L7/L12 stalk side; **d**, **e**, **f**—view from the L1 side. The small and large ribosomal subunits are depicted in *yellow* and *blue* colors, respectively. Landmarks: *b* body, *CP* central protuberance, *FBR* factor binding region, *h* head, *mgt* mRNA gate (in the bovine mitoribosome), *pt* platform, *Sb* L7/L12 stalk base, *sp* spur, *asterisks* the density in Lmr that substitutes the missing L1 protein

center, is relatively well preserved in the 12S rRNA. With respect to ribosomal proteins, the 50S Lmr also contains 10 proteins that are homologous to eubacterial proteins in the SSU (out of 20 eubacterial proteins) and 21 homologs in the LSU (out of 34 eubacterial proteins). The missing RNA segments (Fig. 10.2) and proteins are only partially (by 50–60%) replaced by the *Leishmania*-specific ribosomal proteins. The result of such an incomplete compensation is a highly porous ribosome structure with several noticeable tunnel-like features transversing the body, the platform, and the head of the SSU, as well as body and the base of central protuberance of the LSU.

Although the recent proteomics analyses of mitoribosomal fractions in *L. tarentolae* (Maslov et al. 2006, 2007) and *T. brucei* (Zíková et al. 2008) failed to detect the predicted ribosomal protein S12 (Maslov et al. 1992; Read et al. 1992), the cryo-EM structure clearly shows the presence of a corresponding protein mass within the SSU of Lmr. A likely reason for the absence of the S12 peptides among the proteolytic digestion products might be a tight association of this protein with the 9S rRNA.

The intersubunit space in Lmr contains a large gap between the SSU and LSU bodies. This gap is mainly due to the absence of a large part of the SSU rRNA helix

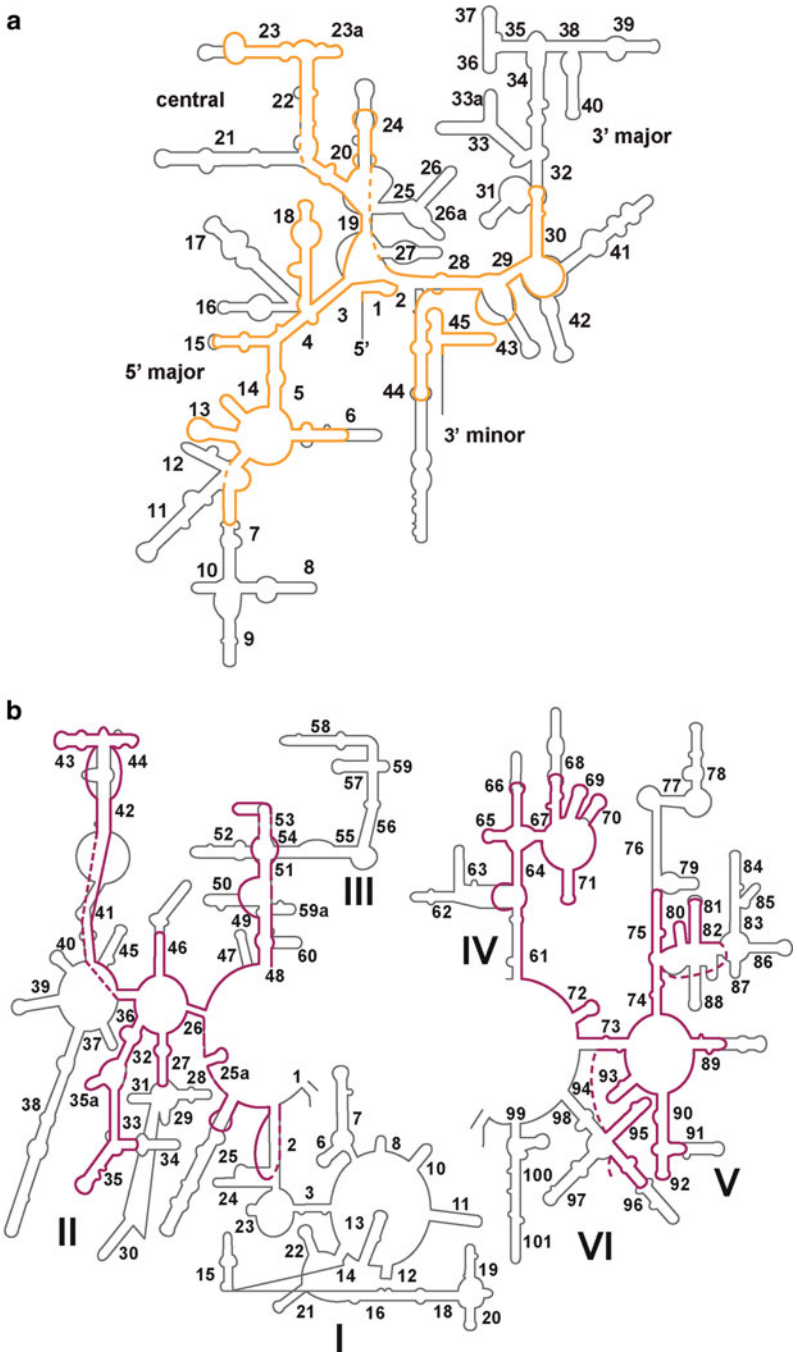


Fig. 10.2 Secondary structure of the trypanosomatid ribosomal RNAs (Sharma et al. 2009). (**a**) Small subunit ribosomal RNA: the structure of *L. tarentolae* 9S rRNA (*orange*) superimposed on the structure of a bacterial 16S rRNA (*gray*). The numbers identify the 16S rRNA helices.

44 (h44) which in eubacterial ribosomes forms much of the SSU:LSU interface in this region. Also, the remaining part of h44 is slightly shifted in Lmr. There are fewer intersubunit bridges in Lmr (nine) compared to eubacterial ribosomes (thirteen) (Yusupov et al. 2001) or mammalian mitochondrial ribosomes (fifteen) (Sharma et al. 2003), and those which are present include only three RNA–RNA contacts, with others being either an RNA–protein or a protein–protein contact.

Several features of the Lmr structure clearly suggest that its *modus operandi* is different from that of eubacterial counterparts. For example, the mRNA channel along the SSU head appears to be lined with Lmr-specific proteins, and the mRNA entrance is rather wide. Although the identities and exact roles of the participating putative proteins remain unknown, it has been suggested that these features reflect the novel modes of recognition of the mRNA by the ribosome, and interactions with specific translation factors (Sharma et al. 2009). It should be noted that the mRNA path in mammalian mitoribosomes is also protein-rich (Sharma et al. 2003).

Furthermore, there appears to be greater contributions of Lmr-specific proteins in the formation of tRNA-binding sites, since several of the eubacterial rRNA segments that form the aminoacyl (A) and peptidyl (P) sites in eubacterial ribosomes are absent in Lmr. The E site seems to be entirely absent in the Lmr. There is also a protein mass that extends from the central protuberance of LSU into the intersubunit space adjacent to A and P sites (referred to as A- and P-site finger or APSF, Fig. 10.3). A similar structure is observed in the mammalian mitoribosome, but there it is smaller, extending only up to the P-site (PSF) (Sharma et al. 2003). It should be noted that, unlike many other mitochondrial systems in which tRNAs are encoded by mitochondrial DNA, all tRNAs in trypanosomatid mitochondria are imported from the cytoplasm and the same tRNAs are also involved in protein synthesis by the cytoplasmic ribosomes (Hancock and Hajduk 1990; Schneider et al. 1994; Shi et al. 1994; Simpson et al. 1989). This includes the mitochondrial initiator tRNA which represents the cytoplasmic elongation Met-tRNA$_m$. When formylated, this tRNA is used for initiation (fMet-tRNA$_i$), and when left unformylated, the same tRNA is used for elongation (Tan et al. 2002). It is conceivable that a set of Lmr-specific proteins has evolved to allow for accommodation of the typical set of eukaryotic tRNAs and the proper discrimination between the two types of Met-tRNA. The homolog of initiation factor IF-2, which brings fMet-tRNA$_i$ to the ribosome, is identifiable in the trypanosomatid genomes (tritrypdb.org/), although it is rather diverged. Homologs of IF-1 and IF-3 are not detectable, further testifying to the peculiarity of the initiation of translation in this system.

Fig. 10.2 (continued) (**b**) Large subunit ribosomal RNA: the *L. tarentolae* 12S rRNA (*purple*) superimposed on a bacterial 23S rRNA (*gray*). The numbers identify the 23S rRNA helices, while the secondary structure domains are identified with *roman numerals*. *Dashed lines* represent the regions with unassigned secondary structure

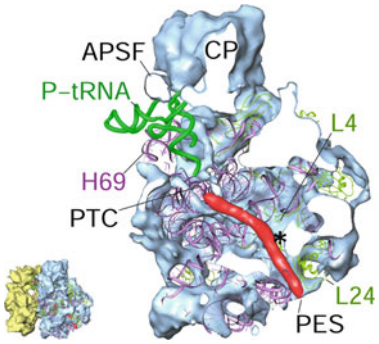


Fig. 10.3 The cut-away view of the 3D cryo-EM structure of the large subunit of the *Leishmania* mitoribosome (Sharma et al. 2009). *White* regions represent the cutting plane. The view is from the L7/L17-stalk side, with the small ribosomal subunit interface to the *left*. Position of the polypeptide-exit tunnel is depicted by docking a low-resolution model of a polypeptide chain (*red*). P-site is *highlighted* by docking a tRNA model (P-tRNA). *APSF* A- and P-site finger, *PES* polypeptide-exit site, *PTC* peptidyltransferase center. The *asterisk* denotes the location of polypeptide accessible site (PAS), which is situated slightly behind the cutting plane. L4 and L24—ribosomal proteins

Besides the canonical AUG initiation codon, which is observed in the majority of cases, trypanosomatids appear to utilize several noncanonical codons for translation initiation. These were predicted initiation codons found in homologous positions for the start codons in multiple alignments. For example, the ORF present in the fully edited RPS12 mRNA sequence from *L. m. amazonensis* contains isoleucine codon Auu (with low case u's added by editing) instead of AUG (Maslov 2010). The ORFs in fully edited sequences of G3 mRNA from *L. amazonensis* and *P. serpens* begin with another isoleucine codon AuA, although there is also an in-frame AuG downstream in the *L. amazonensis* sequence (Maslov et al. 1999; Maslov 2010). There are other cases of predicted noncanonical codons, including the leucine codon UUG (Corell et al. 1994; Read et al. 1994b). In none of such cases, however, was the utilization of the noncanonical codons for initiation verified experimentally. Another codon reassignment is utilization of UGA (along with UGG) for tryptophan, as in many other mitochondrial systems. This is achieved by a C-to-U editing of the CCA anticodon in Trp-tRNA to produce the UCA anticodon that serves to recognize both codons (Alfonzo et al. 1999; Gaston et al. 2007).

The structure corresponding LSU rRNA helix 95 (H95), also known as α-sarcin/ricin stem–loop, is well conserved in the 12S rRNA structure. This rRNA region is known to interact with elongation factors (Moazed et al. 1988), apparently to trigger GTPase activities of the EF-Tu:GTP:aa-tRNA ternary complex, or in the EF-G:GTP complex. However, the spatial position of H95 in Lmr is significantly different from eubacterial ribosome, being shifted by ~30 Å toward the peptidyl-transferase center (Sharma et al. 2009). Perhaps reflecting this, the amino acid sequence of the trypanosomatid mitochondrial EF-Tu (identifiable in the trypanosomatid genomic databases) contains an insertion that might play an accommodating role in its

interactions with Lmr. The homologs of elongation factors EF-Ts and EF-G are also quite distinct in trypanosomatids.

As other mitochondrial ribosomes, Lmr is expected to interact with the inner mitochondrial membrane. The base of the body in the SSU and the nascent polypeptide exit tunnel (PES, polypeptide exit site) of the LSU contain Lmr-specific proteins, reflecting peculiar aspects of these interactions (Fig. 10.3). Yet, the universal features are also observed in trypanosomatids; the homologs of yeast insertase Oxa1, a member of the YidC/Alb3/Oxa1 family of integral membrane proteins interacting with ribosomes close to PES and mediating cotranslational insertion of newly synthesized polypeptides (Stuart 2002), are conserved in trypanosomatids, and so are ribosomal proteins L22, L23, L24, and L40 that participate in forming the PES opening (Nissen et al. 2000; Ott and Herrmann 2010). In addition, the homologs of two membrane-bound proteins Mdm38/LETM1 and Cox11 that mediate ribosome attachment (Herrmann and Neupert 2003; Ott and Herrmann 2010) are also present in trypanosomatid databases. However, the yeast ribosome receptor Mba1 that binds to the LSU next to PES (Ott et al. 2006) is not. An additional interesting feature in Lmr is the extra opening of the nascent peptide tunnel, termed PAS (polypeptide-accessible site). As also observed in mammalian mitoribosomes (Sharma et al. 2003), PAS is located ~25 Å away from the conventional opening (PES) on the solvent side, nearly opposite to the interface with the SSU. The exact function of PAS is unclear, but it is conceivable that this additional opening plays a role in cotranslational insertion of nascent polypeptides into the membrane.

10.4 Higher-Order Interactions of Mitochondrial Ribosomes and the mRNA Recognition Problem

The kinetoplastid mitochondrial ribosome faces a challenging task of recognizing the correct initiation codon in a fully edited and nonedited mRNA while discriminating against relatively abundant pre-edited and partially edited transcripts. In trypanosomes, some initiation codons are created by editing (e.g., pan-edited ND8 in *L. tarentolae*) and others are encoded (e.g., pan-edited G4 mRNA in *L. tarentolae*) (Fig. 10.4) (Gao et al. 2001). Moreover, the correct initiation codon can be preceded by an upstream codon to be avoided, as in the ND8 example above. Finally, the typically short 5′ untranslated regions do not possess discernable universal sequences such as the Shine-Dalgarno element. They are 5′ monophosphorylated and uncapped. Hence, the cap-dependent ribosome recruitment mechanism (Campbell et al. 2003) does not operate in trypanosome mitochondria. The mammalian mitochondria mode of initiation, in which the 5′-untranslated regions are short and the most 5′-proximal AUG (or AUA) serves for initiation (Spremulli et al. 2004) is not directly applicable either. The codon recognition problem may look more similar to that in mRNA of yeast which contain

```
Pan-edited ND8 mRNA (520 nt, encodes 145 aa product)
pre-ed: 5′-UAAACAUAUAUAAUGUAUUAGAUUAAAAGUAA G   G   A GA     G UUUU...
edited: 5′-UAAACAUAUAUAAUGUAUUAGAUUAAAAG*AAuGuuuGuuuAuGAuuuuuGuUUUU...
                                         M  F  V  Y  D  F  C  F   ...

Pan-edited G4 mRNA (537 nt, encodes 169 aa product)
pre-ed: 5′-CAAAUACAUUAAAACACAUAUAUGUAUAAUUGUAAAAG G  G  G     GG        ...
edited: 5′-CAAAUACAUUAAAACACAUAUAUGUAUAAUUGUAAAAGuGuuGuuGuuuuuGGuuuuuuu...
                               M  Y  N  C  K  S  V  V  V  F  G  F  F  ...
```

Fig. 10.4 Partial sequences of fully edited ND8 and G4 mRNA and the inferred amino acid sequences in *L. tarentolae* LEM125. Encoded nucleotides are shown with *upper-case letters*, inserted uridylates with *lower-case* u's and a deleted uridylate with an *asterisk*. The initiation codons are *underlined*. The AUGs *highlighted* in *yellow* point to AUG codons that are to be avoided

long 5′ UTR in which the initiation codon may be preceded by out-of-frame AUGs (Dunstan et al. 1997; Fox 1996; Green-Willms et al. 1998). The logistic requirement to coordinate the initiation at the proper codon with completion of the U-insertion/deletion editing downstream must also be accommodated. That poses a problem of signaling between these two events since editing does not necessarily affect the sequences at the very 5′ end. This is exemplified by the G4 mRNA when the same pre-existing AUG codon needs to be avoided in the pre-edited mRNA, but recognized in the fully edited mRNA (Fig. 10.4).

The initiation codon recognition in yeast mitochondria is aided by highly specific translation activators that interact with the mRNA's 5′ UTR (Green-Willms et al. 2001; Sanchirico et al. 1998; Siep et al. 2000). The COX1 mRNA is recognized by TPR/PPR protein Pet309, the COX2 mRNA by Pet111, and the COB mRNA by a set of several activators. The translation activators represent peripheral or integral membrane proteins that are assembled into a large (~1 MDa) complex tethered to the inner mitochondrial membrane, thereby also bringing the mRNA close to the membrane (Krause et al. 2004; Naithani et al. 2003). However, the yeast mRNA activators are not evolutionally conserved. Therefore, the mRNA recognition factors, if there are any in trypanosome mitochondria, must be trypanosome-specific.

Besides the 5′ end, the *cis*-signals identifying translatable (fully edited and nonedited) mRNA may be located at the 3′ end. Two types of 3′-end tails had been found in trypanosomatid mitochondrial transcripts: the shorter 20–30 nt and, the longer, 200–300 nt tails (Bhat et al. 1991, 1992). The tails' lengths correlate with the editing status of an mRNA: the short tails are found in pre-edited and partially edited mRNA, whereas the long tails are observed in the fully edited type, as well as in non-edited transcripts (Bhat et al. 1992; Etheridge et al. 2008; Militello and Read 1999; Read et al. 1994a). The longer tail represents an extension of the short tail by addition of an A/U heteropolymer upon completion of editing. It was suggested that the long tail represents a "hallmark of translationally competent mRNA" (Etheridge et al. 2008). This hypothesis was verified in a more recent work (Aphasizheva et al. 2011). The long-tail extensions were shown to be produced by the concerted actions of KPAP1 poly(A) polymerase and RET1 terminal uridylyl transferase directed by two PPR proteins termed KPAF1 and 2. Ablation of these

proteins in procyclic trypanosomes resulted in a cessation of the long-tail formation and an inhibition of the de novo protein synthesis. In bloodstream trypanosomes that lack cytochromes and oxidative phosphorylation, the transcripts for subunits cytochrome *c* oxidase and cytochrome bc_1 are devoid of long tails, and de novo synthesis of the corresponding subunits could not be detected. The latter observation has also been reported recently by Richterová and coworkers (2011). However, translation is still expected to occur in bloodstream trypanosomes, at least of the A6 mRNA, which is productively edited and A/U tailed in this stage (Cristodero et al. 2010; Hashimi et al. 2010; Schnaufer et al. 2005).

The results described above have established the long poly(A/U) tails as the general *cis*-elements that identify fully edited mRNAs as templates suitable for translation. The question is still open as to what mediates the interaction between the long-tailed mRNA and the mitoribosomes. Are mRNA-specific factors involved as well? PPR proteins have emerged lately as a class of specific sequence-recognition factors that mediate various aspects of mRNA processing, editing, and translation in plants (Schmitz-Linneweber and Small 2008). Although not so abundant in other organisms, they have also been shown to play similar roles (Delannoy et al. 2007; Lightowlers and Chrzanowska-Lightowlers 2008). In this respect, it is remarkable that trypanosomes contain more than 30 PPR proteins, noticeably more than any other investigated nonplant organism (Mingler et al. 2006; Pusnik et al. 2007). It is conceivable that this abundance is related to a higher complexity of mRNA recognition problem in trypanosomatid mitochondria that is imposed by high rate of RNA editing as compared to that in yeast or mammalian systems. Several PPR proteins were identified in *L. tarentolae* as part of the 45S SSU* complexes (Maslov et al. 2007). It should be noted however that these particles may actually represent a part of a larger complex that is not stable enough to withstand a rigorous isolation procedure. Thus, RNAi-based depletion of several additional PPR proteins in *T. brucei* resulted in a reduction of the 9S or 12S rRNA levels suggesting that these proteins also associate with the ribosomes (Pusnik et al. 2007). This view was further strengthened by the observation that ablation of some of these proteins in turn leads to the degradation of others, as expected, if they are components of the same complex. A relatively gentle affinity isolation procedure using ribosomal proteins S17 and L3 as TAP-fusion "baits" resulted in identification of a large number (nearly 140) of protein components associated with ribosomal particles in *T. brucei*, including 16 PPR proteins (Zíková et al. 2008). This list has recently been extended to 21 ribosomal PPR proteins out of 36 proteins found by database mining and the total list of proteins associated with the mitoribosomes to nearly 300 (Aphasizheva et al. 2011). While the function of the ribosome-associated PPR proteins is unknown, it is conceivable that some of the PPR proteins represent mRNA-specific activators. The long list of mitoribosome-associated proteins suggests that a multitude of tertiary and quaternary protein–protein contacts are made by these particles. However, most of these ~300 proteins do not form integral components of the mitoribosome itself, but may directly or indirectly interact with the mitoribosome as components of the

translation activator complexes, polyadenylation/polyuridylation complexes, or RNA-editing complexes.

The notion that the translating ribosome is substantially larger than the 50S "core" Lmr is further supported by the finding of >70S ribosomal complex in mitochondrial lysates of *T. brucei* (Aphasizheva et al. 2011). This complex represents only a fraction of the total rRNA-containing particles, which are broadly distributed across the gradients with the bulk of them found in the 55–65S range. The >70S complexes stand apart because they cosediment with translatable fully edited mRNA identifiable by the presence of the long 3′-end poly(A/U) tails. Moreover, the same RNP fraction also contains the functional tRNA identifiable as those with repaired CCA-ends (Alfonzo and Soll 2009). The direct association of the tRNA and the large ribosomal RNP was demonstrated by affinity pull-down. However, the physical interaction of the long-tailed mRNA with these complexes was more difficult to demonstrate, as these interactions did not withstand lengthy isolation procedures, and even a relatively gentle TAP method or immunoprecipitation from the gradient fractions resulted in a loss of the associated mRNA (Aphasizheva et al. 2011). However, by applying a rapid cryogenic homogenization followed by affinity pull-downs of the TAP-tagged ribosomes, it was possible to demonstrate a preferential association of the long-tailed mRNA with the small subunits, and that of pre-edited mRNA, as well as components of the edited machinery, with the large subunit (Aphasizheva et al. 2011). Although most of the details of these interactions still need to be uncovered, the emerging view is that of a highly integrated system, in which the process of mRNA maturation is tightly coupled with translation.

Acknowledgments The work in the authors' laboratories has been supported by National Institutes of Health grants AI088292 (to DAM) and GM61576 (to RKA). We thank R. Aphasizhev and L. Simpson for the critical comments on the manuscript.

References

Agrawal RK, Sharma MR, Yassin A, Lahiri I, Spremulli LL (2011) Structure and function of organellar ribosomes as revealed by cryo-EM. In: Rodnina M, Wintermeyer W, Green R (eds) Ribosomes: structure, function, and dynamics. Springer Wien, New York, pp 83–96

Alfonzo JD, Soll D (2009) Mitochondrial tRNA import—the challenge to understand has just begun. Biol Chem 390:717–722

Alfonzo JD, Blanc V, Estevez AM, Rubio MA, Simpson L (1999) C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. EMBO J 18:7056–7062

Aphasizheva I, Maslov DA, Wang X, Huang L, Aphasizhev R (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. Mol Cell 42:106–117

Arts GJ, Van der Spek H, Speijer D, Van den Burg J, Van Steeg H, Sloof P, Benne R (1993) Implications of novel guide RNA features for the mechanism of RNA editing in *Crithidia fasciculata*. EMBO J 12:1523–1532

Attardi G, Ojala D (1971) Mitochondrial ribosomes in HeLa cells. Nature 229:133–135
Benne R, Van den Burg J, Brakenhoff J, Sloof P, Van Boom J, Tromp M (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46:819–826
Berry EA, Guergova-Kuras M, Huang LS, Crofts AR (2000) Structure and function of cytochrome *bc* complexes. Annu Rev Biochem 69:1005–1075
Bhat GJ, Koslowsky DJ, Feagin JE, Smiley BL, Stuart K (1990) An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell 61:885–894
Bhat GJ, Myler PJ, Stuart K (1991) The two ATPase 6 mRNAs of *Leishmania tarentolae* differ at their 3′ ends. Mol Biochem Parasitol 48:139–150
Bhat GJ, Souza AE, Feagin JE, Stuart K (1992) Transcript-specific developmental regulation of polyadenylation in *Trypanosoma brucei* mitochondria. Mol Biochem Parasitol 52:231–240
Blum B, Bakalara N, Simpson L (1990) A model for RNA editing in kinetoplastid mitochondria: "Guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell 60:189–198
Borst P, Grivell LA (1971) Mitochondrial ribosomes. FEBS Lett 13:73–88
Breek CK, Speijer D, Dekker H, Muijsers AO, Benne R (1997) Further evidence for the presence of mitochondrially encoded subunits in cytochrome *c* oxidase of the trypanosomatid *Crithidia fasciculata*. Biol Chem 378:837–841
Brega A, Vesco C (1971) Ribonucleoprotein particles involved in HeLa mitochondrial protein synthesis. Nature 229:136–139
Campbell DA, Thomas S, Sturm NR (2003) Transcription in kinetoplastid protozoa: why be normal? Microbes Infect 5:1231–1240
Čermáková P, Verner Z, Man P, Lukeš J, Horváth A (2007) Characterization of the NADH: ubiquinone oxidoreductase (complex I) in the trypanosomatid *Phytomonas serpens* (Kinetoplastida). FEBS J 274:3150–3158
Chang KP (1974) Ultrastructure of symbiotic bacteria in normal and antibiotic-treated *Blastocrithidia culicis* and *Crithidia oncopelti*. J Protozool 21:699–707
Corell RA, Myler P, Stuart K (1994) *Trypanosoma brucei* mitochondrial CR4 gene encodes an extensively edited mRNA with completely edited sequence only in bloodstream forms. Mol Biochem Parasitol 64:65–74
Cristodero M, Seebeck T, Schneider A (2010) Mitochondrial translation is essential in bloodstream forms of *Trypanosoma brucei*. Mol Microbiol 78:757–769
Curgy J-J (1985) The mitoribosomes. Biol Cell 54:1–38
de la Cruz V, Lake JA, Simpson AM, Simpson L (1985a) A minimal ribosomal RNA: sequence and secondary structure of the 9S kinetoplast ribosomal RNA from *Leishmania tarentolae*. Proc Natl Acad Sci USA 82:1401–1405
de la Cruz V, Simpson A, Lake J, Simpson L (1985b) Primary sequence and partial secondary structure of the 12S kinetoplast (mitochondrial) ribosomal RNA from *Leishmania tarentolae*: conservation of peptidyl-transferase structural elements. Nucleic Acids Res 13:2337–2356
Delannoy E, Stanley WA, Bond CS, Small ID (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochem Soc Trans 35:1643–1647
Dunstan HM, Green-Willms NS, Fox TD (1997) *In vivo* analysis of *Saccharomyces cerevisiae COX2* mRNA 5′-untranslated leader functions in mitochondrial translation initiation and translational activation. Genetics 147:87–100
Eperon I, Janssen J, Hoeijmakers J, Borst P (1983) The major transcripts of the kinetoplast DNA of *T. brucei* are very small ribosomal RNAs. Nucleic Acids Res 11:105–125
Etheridge RD, Aphasizheva I, Gershon PD, Aphasizhev R (2008) 3′ adenylation determines mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. EMBO J 27:1596–1608

Fang J, Wang YD, Beattie DS (2001) Isolation and characterization of complex I, rotenone-sensitive NADH:ubiquinone oxidoreductase, from the procyclic forms of *Trypanosoma brucei*. Eur J Biochem 268:3075–3082

Feagin JE (2000) Mitochondrial genome diversity in parasites. Int J Parasitol 30:371–390

Feagin JE, Shaw JM, Simpson L, Stuart K (1988) Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids. Proc Natl Acad Sci USA 85:539–543

Fox TD (1996) Genetics of mitochondrial translation. In: Hershey JWB, Matthews MB, Sonnenberg N (eds) Translational control. Cold Spring Harbor, New York, pp 733–758

Gabashvili IS, Agrawal RK, Spahn CM, Grassucci RA, Svergun DI, Frank J, Penczek P (2000) Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. Cell 100:537–549

Gao GG, Kapushoc ST, Simpson AM, Thiemann OH, Simpson L (2001) Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: an unexpected complexity. RNA Publ RNA Soc 7:1335–1347

Gaston KW, Rubio MA, Spears JL, Pastar I, Papavasiliou FN, Alfonzo JD (2007) C to U editing at position 32 of the anticodon loop precedes tRNA 5′ leader removal in trypanosomatids. Nucleic Acids Res 35:6740–6749

Goswami S, Dhar G, Mukherjee S, Mahata B, Chatterjee S, Home P, Adhya S (2006) A bifunctional tRNA import receptor from *Leishmania mitochondria*. Proc Natl Acad Sci USA 103:8354–8359

Greco M, Cantatore P, Pepe G, Saccone C (1973) Isolation and characterization of rat-liver mitochondrial ribosomes highly active in poly(U)-directed polyphenylalanine synthesis. Eur J Biochem 37:171–177

Green-Willms NS, Fox TD, Costanzo MC (1998) Functional interactions between yeast mitochondrial ribosomes and mRNA 5′ untranslated leaders. Mol Cell Biol 18:1826–1834

Green-Willms NS, Butler CA, Dunstan HM, Fox TD (2001) Pet111p, an inner membrane-bound translational activator that limits expression of the *Saccharomyces cerevisiae* mitochondrial gene *COX2*. J Biol Chem 276:6392–6397

Gruschke S, Ott M (2010) The polypeptide tunnel exit of the mitochondrial ribosome is tailored to meet the specific requirements of the organelle. Bioessays 32:1050–1057

Hanas J, Linden G, Stuart K (1975) Mitochondrial and cytoplasmic ribosomes and their activity in blood and culture form *Trypanosoma brucei*. J Cell Biol 65:103–111

Hancock K, Hajduk SL (1990) The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. J Biol Chem 265:19208–19215

Hashimi H, Benkovičová V, Čermáková P, Lai DH, Horváth A, Lukeš J (2010) The assembly of F_1F_0-ATP synthase is disrupted upon interference of RNA editing in *Trypanosoma brucei*. Int J Parasitol 40:45–54

Herrmann JM, Neupert W (2003) Protein insertion into the inner membrane of mitochondria. IUBMB Life 55:219–225

Hill RC, Morris CA, Weber MM (1975) Chloramphenicol resistant mitochondrial protein synthesis in *Crithidia fasciculata*. Arch Biochem Biophys 170:392–399

Horváth A, Berry EA, Maslov DA (2000a) Translation of the edited mRNA for cytochrome *b* in trypanosome mitochondria. Science 287:1639–1640

Horváth A, Kingan TG, Maslov DA (2000b) Detection of the mitochondrially encoded cytochrome *c* oxidase subunit I in the trypanosomatid protozoan *Leishmania tarentolae*. J Biol Chem 275:17160–17165

Horváth A, Neboháčová M, Lukeš J, Maslov DA (2002) Unusual polypeptide synthesis in the kinetoplast-mitochondria from *Leishmania tarentolae*. Identification of individual de novo translation products. J Biol Chem 277:7222–7230

Kleisen CM, Borst P (1975) Are 50% of all cellular proteins synthesized on mitochondrial ribosomes in *Crithidia luciliae*? Biochim Biophys Acta 390:78–81

Krause K, de Lopes SR, Roberts DG, Dieckmann CL (2004) The mitochondrial message-specific mRNA protectors Cbp1 and Pet309 are associated in a high-molecular weight complex. Mol Biol Cell 15:2674–2683

Lamb AJ, Clark-Walker GD, Linnane AW (1968) The biogenesis of mitochondria. 4. The differentiation of mitochondrial and cytoplasmic protein synthesizing systems in vitro by antibiotics. Biochim Biophys Acta 161:415–427

Laub-Kuperszteijn R, Thirion J (1974) Existence of two distinct protein synthesis systems in the trypanosomatid *Crithidia luciliae*. Biochim Biophys Acta 340:314–322

Lightowlers RN, Chrzanowska-Lightowlers ZM (2008) PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression. Biochem J 416:e5–e6

Ludwig B, Bender E, Arnold S, Hüttemann M, Lee I, Kadenbach B (2001) Cytochrome *c* oxidase and the regulation of oxidative phosphorylation. Chembiochem 2:403

Maarouf M, Lawrence F, Croft SL, Robert-Gero M (1995) Ribosomes of *Leishmania* are a target for the aminoglycosides. Parasitol Res 81:421–425

Marres CAM, Slater EC (1977) Polypeptide composition of purified QH2:cytochrome c oxidoreductase from beef-heart mitochondria. Biochim Biophys Acta 462:531–548

Maslov DA (2010) Complete set of mitochondrial pan-edited mRNAs in *Leishmania mexicana amazonensis* LV78. Mol Biochem Parasitol 173:107–114

Maslov DA, Simpson L (1992) The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell 70:459–467

Maslov DA, Sturm NR, Niner BM, Gruszynski ES, Peris M, Simpson L (1992) An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. Mol Cell Biol 12:56–67

Maslov DA, Nawathean P, Scheel J (1999) Partial kinetoplast-mitochondrial gene organization and expression in the respiratory deficient plant trypanosomatid *Phytomonas serpens*. Mol Biochem Parasitol 99:207–221

Maslov DA, Sharma MR, Butler E, Falick AM, Gingery M, Agrawal RK, Spremulli LL, Simpson L (2006) Isolation and characterization of mitochondrial ribosomes and ribosomal subunits from *Leishmania tarentolae*. Mol Biochem Parasitol 148:69–78

Maslov DA, Spremulli LL, Sharma MR, Bhargava K, Grasso D, Falick AM, Agrawal RK, Parker CE, Simpson L (2007) Proteomics and electron microscopic characterization of the unusual mitochondrial ribosome-related 45S complex in *Leishmania tarentolae*. Mol Biochem Parasitol 152:203–212

McLean JR, Cohn GL, Brandt IK, Simpson MV (1958) Incorporation of labeled amino acids into the protein of muscle and liver mitochondria. J Biol Chem 233:657–663

Mears JA, Cannone JJ, Stagg SM, Gutell RR, Agrawal RK, Harvey SC (2002) Modeling a minimal ribosome based on comparative sequence analysis. J Mol Biol 321:215–234

Mears JA, Sharma MR, Gutell RR, McCook AS, Richardson PE, Caulfield TR, Agrawal RK, Harvey SC (2006) A structural model for the large subunit of the mammalian mitochondrial ribosome. J Mol Biol 358:193–212

Militello KT, Read LK (1999) Coordination of kRNA editing and polyadenylation in *Trypanosoma brucei* mitochondria: complete editing is not required for long poly(A) tract addition. Nucleic Acids Res 27:1377–1385

Mingler MK, Hingst AM, Clement SL, Yu LE, Reifur L, Koslowsky DJ (2006) Identification of pentatricopeptide repeat proteins in *Trypanosoma brucei*. Mol Biochem Parasitol 150:37–45

Moazed D, Robertson JM, Noller HF (1988) Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. Nature 334:362–364

Mukherjee S, Basu S, Home P, Dhar G, Adhya S (2007) Necessary and sufficient factors for the import of transfer RNA into the kinetoplast mitochondrion. EMBO Rep 8:589–595

Mulkidjanian AY, Makarova KS, Galperin MY, Koonin EV (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. Nat Rev Microbiol 5:892–899

Nabholz CE, Speijer D, Schneider A (1999) Chloramphenicol-sensitive mitochondrial translation in *Trypanosoma brucei*. Parasitol Res 85:779–782

Naithani S, Saracco SA, Butler CA, Fox TD (2003) Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of *Saccharomyces cerevisiae*. Mol Biol Cell 14:324–333

Neboháčová M, Maslov DA, Falick AM, Simpson L (2004) The effect of RNA interference Down-regulation of RNA editing 3′-terminal uridylyl transferase (TUTase) 1 on mitochondrial de novo protein synthesis and stability of respiratory complexes in *Trypanosoma brucei*. J Biol Chem 279:7819–7825

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–930

O'Brien TW (2002) Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. Gene 286:73–79

O'Brien TW (2003) Properties of human mitochondrial ribosomes. IUBMB Life 55:505–513

O'Brien TW, Denslow ND (1996) Bovine mitochondrial ribosomes. In: Attardi GM, Chomyn A (eds) Methods in enzymology, vol 264. Academic Press, San Diego, pp 237–248

O'Brien TW, Kalf GF (1967a) Ribosomes from rat liver mitochondria. II. Partial characterization. J Biol Chem 242:2180–2185

O'Brien TW, Kalf GF (1967b) Ribosomes from rat liver mitochondria. I. Isolation procedure and contamination studies. J Biol Chem 242:2172–2179

Ochsenreiter T, Hajduk SL (2006) Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. EMBO Rep 7:1128–1133

Ochsenreiter T, Cipriano M, Hajduk SL (2008) Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. PLoS One 3:e1566

Opperdoes FR, Michels PA (2008) Complex I of Trypanosomatidae: does it exist? Trends Parasitol 24(7):310–317

Ott M, Herrmann JM (2010) Co-translational membrane insertion of mitochondrially encoded proteins. Biochim Biophys Acta 1803:767–775

Ott M, Prestele M, Bauerschmitt H, Funes S, Bonnefoy N, Herrmann JM (2006) Mba1, a membrane-associated ribosome receptor in mitochondria. EMBO J 25:1603–1610

Pel HJ, Grivell LA (1994) Protein synthesis in mitochondria. Mol Biol Rep 19:183–194

Pietromonaco SF, Denslow ND, O'Brien TW (1991) Proteins of mammalian mitochondrial ribosomes. Biochimie 73:827–836

Priest JW, Hajduk SL (1992) Cytochrome *c* reductase purified from *Crithidia fasciculata* contains an atypical cytochrome c_1. J Biol Chem 267:20188–20195

Pusnik M, Small I, Read LK, Fabbro T, Schneider A (2007) Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. Mol Cell Biol 27:6876–6888

Read LK, Myler PJ, Stuart K (1992) Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei*. J Biol Chem 267:1123–1128

Read LK, Stankey KA, Fish WR, Muthiani AM, Stuart K (1994a) Developmental regulation of RNA editing and polyadenylation in four life cycle stages of *Trypanosoma congolense*. Mol Biochem Parasitol 68:297–306

Read LK, Wilson KD, Myler PJ, Stuart K (1994b) Editing of *Trypanosoma brucei* maxicircle CR5 mRNA generates variable carboxy terminal predicted protein sequences. Nucleic Acids Res 22:1489–1495

Remacle C, Barbieri MR, Cardol P, Hamel PP (2008) Eukaryotic complex I: functional diversity and experimental systems to unravel the assembly process. Mol Genet Genomics 280:93–110

Richterová L, VávrováZ Z, Lukeš J (2011) DEAD-box RNA helicase is dispensable for mitochondrial translation in *Trypanosoma brucei*. Exp Parasitol 127:300–303

Sanchirico ME, Fox TD, Mason TL (1998) Accumulation of mitochondrially synthesized *Saccharomyces cerevisiae* Cox2p and Cox3p depends on targeting information in untranslated portions of their mRNAs. EMBO J 17:5796–5804

Scheinman A, Aguinaldo A-M, Simpson AM, Peris M, Shankweiler G, Simpson L, Lake JA (1993) Reconstitution of a minimal small ribosomal subunit. In: Nierhaus K (ed) The translation apparatus. Plenum Press, New York, pp 719–726

Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 13:663–670

Schnaufer A, Sbicego S, Blum B (2000) Antimycin A resistance in a mutant *Leishmania tarentolae* strain is correlated to a point mutation in the mitochondrial apocytochrome *b* gene. Curr Genet 37:234–241

Schnaufer A, Clark-Walker GD, Steinberg AG, Stuart K (2005) The F_1-ATP synthase complex in bloodstream stage trypanosomes has an unusual and essential function. EMBO J 24:4029–4040

Schneider A, Martin J, Agabian N (1994) A nuclear encoded tRNA of *Trypanosoma brucei* is imported into mitochondria. Mol Cell Biol 14:2317–2322

Sharma MR, Koc EC, Datta PP, Booth TM, Spremulli LL, Agrawal RK (2003) Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. Cell 115:97–108

Sharma MR, Booth TM, Simpson L, Maslov DA, Agrawal RK (2009) Structure of a mitochondrial ribosome with minimal RNA. Proc Natl Acad Sci USA 106:9637–9642

Shaw J, Feagin JE, Stuart K, Simpson L (1988) Editing of mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell 53:401–411

Shaw J, Campbell D, Simpson L (1989) Internal frameshifts within the mitochondrial genes for cytochrome oxidase subunit II and maxicircle unidentified reading frame 3 in *Leishmania tarentolae* are corrected by RNA editing: evidence for translation of the edited cytochrome oxidase subunit II mRNA. Proc Natl Acad Sci USA 86:6220–6224

Shi X, Chen D-HT, Suyama Y (1994) A nuclear tRNA gene cluster in the protozoan *Leishmania tarentolae* and differential distribution of nuclear-encoded tRNAs between the cytosol and mitochondria. Mol Biochem Parasitol 65:23–37

Shu HH, Göringer HU (1998) *Trypanosoma brucei* mitochondrial ribonucleoprotein complexes which contain 12S and 9S ribosomal RNAs. Parasitology 116:157–164

Siep M, van Oosterum K, Neufeglise H, Van der Spek H, Grivell LA (2000) Mss51p, a putative translational activator of cytochrome *c* oxidase subunit-1 (*COX1*) mRNA, is required for synthesis of Cox1p in *Saccharomyces cerevisiae*. Curr Genet 37:213–220

Simpson AM, Suyama Y, Dewes H, Campbell D, Simpson L (1989) Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also small minicircle and maxicircle transcripts of unknown function. Nucleic Acids Res 17:5427–5445

Simpson L, Wang SH, Thiemann OH, Alfonzo JD, Maslov DA, Avila HA (1998) U-insertion/deletion Edited Sequence Database. Nucleic Acids Res 26:170–176

Simpson L, Thiemann OH, Savill NJ, Alfonzo JD, Maslov DA (2000) Evolution of RNA editing in trypanosome mitochondria. Proc Natl Acad Sci USA 97:6986–6993

Smits P, Smeitink JA, Van den Heuvel LP, Huynen MA, Ettema TJ (2007) Reconstructing the evolution of the mitochondrial ribosomal proteome. Nucleic Acids Res 35:4686–4703

Speijer D, Muijsers AO, Dekker H, De Haan A, Breek CKD, Albracht SPJ, Benne R (1996) Purification and characterization of cytochrome *c* oxidase from the insect trypanosomatid *Crithidia fasciculata*. Mol Biochem Parasitol 79:47–59

Speijer D, Breek CKD, Muijsers AO, Hartog AF, Berden JA, Albracht SPJ, Samyn B, Van Beeumen J, Benne R (1997) Characterization of the respiratory chain from cultured *Crithidia fasciculata*. Mol Biochem Parasitol 85:171–186

Spithill T, Shimer S, Hill G (1981) Inhibitory effects of chloramphenicol isomers and other antibiotics on protein synthesis and respiration in procyclic *Trypanosoma brucei brucei*. Mol Biochem Parasitol 2:235–256

Spremulli LL, Coursey A, Navratil T, Hunter SE (2004) Initiation and elongation factors in mammalian mitochondrial protein biosynthesis. Prog Nucleic Acid Res Mol Biol 77:211–261

Stuart R (2002) Insertion of proteins into the inner membrane of mitochondria: the role of the Oxa1 complex. Biochim Biophys Acta 1592:79–87

Stuart K, Feagin JE (1992) Mitochondrial DNA of kinetoplastids. Int Rev Cytol 141:65–88

Tan TH, Bochud-Allemann N, Horn EK, Schneider A (2002) Eukaryotic-type elongator tRNAMet of *Trypanosoma brucei* becomes formylated after import into mitochondria. Proc Natl Acad Sci USA 99:1152–1157

Tittawella I (1998) Protein synthesis in mitochondria isolated from the trypanosomatid protozoan *Crithidia fasciculata*. FEBS Lett 423:351–356

Tittawella I, Yasmin L, Baranov V (2003) Mitochondrial ribosomes in a trypanosome. Biochem Biophys Res Commun 307:578–583

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF (2001) Crystal structure of the ribosome at 5.5 Å resolution. Science 292:883–896

Zaitseva GN, Kolesnikov AA, Shirshov AT (1977) The genetic system of kinetoplasts in trypanosomatids. Mol Cell Biochem 14:47–54

Zíková A, Panigrahi AK, Dalley RA, Acestor N, Anupama A, Ogata Y, Myler PJ, Stuart KD (2008) *Trypanosoma brucei* mitochondrial ribosomes: affinity purification and component identification by mass spectrometry. Mol Cell Proteomics 7:1286–1296

Chapter 11
RNA-Seq Analysis of the Transcriptome of *Trypanosoma brucei*

Jan Mani, Kapila Gunasekera, and Isabel Roditi

Abstract Four recent publications used next generation sequencing to analyse the transcriptome of *Trypanosoma brucei*. These have produced a wealth of data that have refined and, to some extent, redefined our concept of the genome. At the same time, they provide a welcome opportunity to re-examine older publications and place them in a post-genomic context.

Keywords RNA-Seq • Spliced leader trapping • Digital gene expression

11.1 Introduction

In July 2005, a series of research articles reporting the genome sequences of the trypanosomatid parasites *Trypanosoma brucei, Trypanosoma cruzi* and *Leishmania major* made the cover of *Science* (Berriman et al. 2005; El-Sayed et al. 2005; Ivens et al. 2005). The sequencing projects had been initiated more than 10 years earlier (Degrave et al. 2001) and their final publication marked the definitive shift into the post-genomic era for research on these devastating, but fascinating pathogens. The huge impact on the scientific community is reflected by the citation numbers which now range in the hundreds for each article and the fact that new students to the field very often acquaint themselves with the subject by first reading these "genome papers". Even though the first genomic sequences were determined for viruses in the late 1970s, large-scale sequencing efforts were limited for a long time by the lack of suitable sequencing technology. The introduction of the chain-termination method by Frederick Sanger (Sanger et al. 1977) and, most importantly, its later adaptation using dye-terminator dideoxynucleotides and capillary electrophoresis in combination with a high degree of automation, allowed for sequencing

J. Mani • K. Gunasekera • I. Roditi (✉)
Institut für Zellbiologie, Universität Bern, Bern, Switzerland
e-mail: isabel.roditi@izb.unibe.ch

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2_11,

throughputs up to a level appropriate to tackle more complex genomes. However, the basic setup for Sanger sequencing did not undergo fundamental improvements since the introduction of automated sequencers some 20 years ago and the simple (but not necessarily cost-effective) solution to increase throughput was to run many automated sequencing machines in parallel. Therefore, sequencing of whole genomes as well as large-scale cDNA and EST sequencing projects could only be carried out by specialised sequencing centres and was out of reach for the average laboratory. The elucidation of the exact sequence of an organism's genome has captured the imagination of researchers for a long time and can generally be thought of as the first step towards the understanding of the next layers of complexity, namely the transcriptome and the proteome.

The *T. brucei* genome sequence published in 2005 covered 26 Mb on the 11 megabase chromosomes and encompassed 9,068 protein-coding genes (Berriman et al. 2005). Two-thirds of these genes were conserved between the TriTryps and one-third could be assigned tentative functions. Intermediate and mini-chromosomes, which are repositories for large numbers of variant surface glycoprotein (VSG) genes, were not part of the published sequence. Having access to the genome greatly accelerated the pace of gene identification and simplified procedures such as gene knockouts, but there was also a downside. The wealth of new information overshadowed data that had been obtained by studying individual genes in depth or by sequencing a few hundred cDNAs [for example, El-Sayed et al. (1995)]; although these frequently included the complete 5′ and 3′ untranslated regions (UTRs), this information was not incorporated into the genome databases. Thus, for several years, trypanosome genes became synonymous with their open reading frames, despite the known importance of UTRs in regulating RNA stability and translation (Clayton and Shapira 2007; Fernández-Moya and Estévez 2010; Haile and Papadopoulou 2007; Ouellette and Papadopoulou 2009). Four recent publications using next generation sequencing (NGS) to analyse the transcriptome of *T. brucei* have not only identified the 5′ UTRs of most genes and the major 3′ UTRs for several thousand of them, but are also refining and, to some extent, redefining our concept of the genome. At the same time, they provide a welcome opportunity to re-examine older publications and place them in a post-genomic context.

11.2 The Next Generation Sequencing Revolution

In recent years, DNA sequencing technology has undergone fundamental changes. The development of NGS platforms by companies such as Roche/454 (http://www.454.com), Illumina (http://www.illumina.com), Applied Biosystems/Life Technologies (http://www.appliedbiosystems.com), and most recently by Helicos (http://www.helicosbio.com), led to an enormous increase in sequencing output while at the same time tremendously lowering the costs per sequenced base. Despite the technological differences in sequencing chemistries, all currently

available platforms depend on the miniaturisation and spatial separation of individual sequencing reactions and subsequent detection of DNA strand extension of hundreds of thousands to millions of these reactions in parallel (Mardis 2008; Metzker 2010). Besides the steep increase in output, the major advantage of all NGS methods is that they no longer rely on cloning of the DNA into vectors prior to sequencing. This makes these procedures far less labour intensive, as well as reducing the bias against DNA fragments that are difficult to clone or that have negative effects on the growth of *E. coli*. The sequencing power of the new machines and the ongoing replacement of Sanger sequencing is reflected by the fact that the newly created sequence read archive (SRA), a depository for raw NGS data, currently holds >500 billion sequence reads consisting of >60 trillion sequenced base pairs (as of mid-September 2010) including far more than 100 fully sequenced human genomes (Leinonen et al. 2010). At the same time, the number of deposited conventional DNA sequence electropherograms at the Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi) is now reaching a plateau after near-exponential increase for the first half or so of the last decade.

Apart from classic genome (re-)sequencing projects, NGS has been very quickly adopted to study transcriptomes and their dynamic properties in an unprecedented manner. Up to this point, genome-wide expression studies mostly relied on microarray technology. Besides the known drawbacks of hybridisation-based methods, such as cross-hybridisation artefacts and a limited dynamic range, commercial microarrays are still only available for widely studied model organisms. Furthermore, a priori knowledge of the transcripts to be examined is a prerequisite for most array designs and therefore represents an important methodological limit. Custom-made arrays for trypanosomes are now available, but expensive, and can come with a surcharge because the sequences are derived from a pathogen! Despite these limitations, useful insights into life-cycle dependent gene expression profiles have been gained from microarray analyses of the TriTryps (Brems et al. 2005; Jensen et al. 2009; Kabani et al. 2009; Minning et al. 2009; Queiroz et al. 2009; Rochette et al. 2008; Srividya et al. 2007). Sequencing and quantification of tags from fragmented cDNA samples can overcome some of the problems associated with microarrays. No a priori knowledge of transcribed regions of a genome is necessary and therefore an unbiased view of a transcriptome is possible. Furthermore, the dynamic range of these methods is limited only by the depth of sequencing. Gene expression is given as a digital output (the number of times a sequence derived from a particular transcript is detected in a sample). This contrasts with microarrays where the readout is given as relative signal intensities that are not truly proportional to transcript abundance. NGS can be used to analyse cDNA samples from any source and can be adapted to exploit organism-specific properties. One study made use of the fact that monocistronic mRNAs in trypanosomatids are processed from large polycistronic precursor RNAs by a coupled *trans*-splicing/polyadenylation reaction that adds a common capped leader sequence to their 5′ ends. The presence of the spliced leader on each mRNA is exploited in the spliced leader trapping (SLT) approach to globally map 5′ splice sites and concomitantly acquire global gene expression profiles (Nilsson et al. 2010). The original form of

expression profiling, digital gene expression (DGE), relied on the use of restriction enzymes to create a site for linker ligation; a limitation of this method is that transcripts lacking the appropriate site do not feature in the analysis. This problem has since been overcome—NGS can now be used to sequence randomly fragmented cDNAs, a method termed RNA sequencing (RNA-Seq). RNA-Seq not only allows quantitative measurements of gene expression, but also provides unbiased insights into which regions of a genome are transcribed and into transcript structures. So far, two publications have examined the transcriptome of *T. b. brucei* by RNA-Seq (Kolev et al. 2010; Siegel et al. 2010), and one has analysed stage-regulation on *T. b. gambiense* by DGE (Veitch et al. 2010). SLT data, as well as RNA-Seq data produced by Nilsson et al. (2010), but not analysed in depth in their publication, can be accessed at http://splicer.unibe.ch. Most of these data are incorporated into TriTrypDB (http://tritrypdb.org).

11.3 Illumina Sequencing

Before going into more detail about SLT and RNA-Seq analysis of *Trypanosoma brucei,* we provide a brief introduction to the sequencing technology by Illumina since this was used for all the publications on trypanosomes. The procedure for sequencing with the Illumina platform consists of three consecutive steps (a) DNA library preparation, (b) generation of clonally amplified DNA clusters and (c) sequencing of these clusters. For cDNA library preparation, either the RNA template or the resulting double-stranded cDNA has to be fragmented. Following the ligation of Illumina or custom-designed adapters, the DNA fragment library is amplified by PCR using adapter-specific primers and size selected by agarose gel electrophoresis. In the next step the DNA library is denatured and one or several channels of a flow cell are seeded with the library at limiting dilutions. The flow cell surface consists of a lawn of covalently bound oligonucleotides complementary to the adapters of the fragment library. Individual DNA molecules attach to the surface by hybridisation with these oligonucleotides. Solid-phase amplification (isothermal bridge amplification) takes place by 3′ extension of bound DNA fragments, leading to covalently attached single molecules. DNA bridges are formed when the free ends of these molecules hybridise with their complementary lawn oligonucleotides. Again 3′ extension is performed, resulting in double-stranded bridges. The process of DNA denaturation, bridge formation and 3′ extension is repeated multiple times to generate spatially separate DNA clusters. Sequencing-by-synthesis takes place by stepwise incorporation of reversible terminator nucleotides carrying nucleotide-specific fluorophores. Depending on how many times this cycle is repeated, read lengths of up to 150 bases are currently possible. Furthermore, bar-coding of sequencing adaptors allows multiple samples to be processed in a single channel of the flow cell, with sequencing reads being assigned to individual DNA libraries after sequencing.

11.4 Different Approaches to Analysing the Transcriptome of Trypanosomes

Although all four laboratories used a common method of sequencing, they used different stocks or subspecies of *T. brucei*, different ways of generating libraries and different ways of analysing and portraying their data. The basic information is summarised in Table 11.1. In the context of transcriptome analysis by RNA-Seq, some cDNA synthesis protocols give information on which strand was transcribed, while others do not. This diversity of approaches has been more fruitful than might be expected at first sight. Not only has it provided a solid core of data that holds for all the strains and sub-species, it has also given us the opportunity to combine data from the individual publications and to consider new concepts of gene regulation.

11.4.1 Spliced Leader Trapping

As mentioned above, SLT exploits the presence of a common 39-nucleotide spliced leader sequence representing the 5′ end of every mRNA in *T. brucei*. *Trans*-splicing of a spliced leader is common to all trypanosomatids and has been found to occur in other organisms, including nematodes, cnidarians, flatworms and ascidians. The SLT approach can therefore be easily adapted for all species relying on similar mechanisms to generate monocistronic mRNAs from large polycistronic precursor RNAs. Furthermore, SLT can be used to selectively analyse the transcriptomes of parasites without the need to purify them from host tissue. This potentially allows investigation of life cycle stages for which no in vitro cultivation system is available, such as for the epimastigote form of *T. brucei*. cDNA library construction for SLT starts with the isolation of total RNA. After enrichment of the samples for polyadenylated RNA by selection on oligo(dT) beads, first-strand cDNA is synthesised using random hexamer primers. Second-strand cDNA synthesis is then primed by a 5′ biotinylated oligonucleotide that serves as the 5′ end adaptor for isothermal bridge amplification and contains a sequence corresponding to the spliced leader. Following second-strand synthesis, 3′ end adaptors are ligated to the double-stranded cDNA and fragments containing the spliced leader are enriched by binding to streptavidin beads. In order to run multiple samples in a single channel of the Illumina flow cell, Nilsson et al. (2010) used custom-designed paired-end 3′ adaptors containing a four-nucleotide barcode. The way in which the cDNA libraries for SLT are constructed and sequenced retains the strand information and the first sequenced base of a read corresponds to the transcript sequence directly downstream of the spliced leader acceptor site in the precursor RNA. For mapping purposes, the SLT method relies on the availability of a fully sequenced reference genome with annotated features. Of the three *T. brucei* life cycle stages analysed by SLT, a total of 4.6 million sequence reads could be assigned to one of the three respective libraries of which 4.5 million could be

Table 11.1 Construction and analysis of cDNA libraries

Species	Stock	Life cycle stage	mRNA enrichment	Library construction	Analysis	References
T. b. brucei	Lister 427; TREU927	PF	poly(A)	oligo(dT) or random hexamer priming	RNA-Seq	Siegel et al. (2010)
	Lister 427 (MITat 1.2, clone 221a)	Long slender BF				
	Lister 427 (MITat 1.2, clone 221a) -derivative "single marker"					
T. b. brucei	Lister 427 (MITat 1.2, clone 221)	Long slender BF	poly(A)	Random hexamer followed by SL priming and affinity isolation of SL-containing cDNAs	SLT	Nilsson et al. (2010)
	AnTat 1.1	Long slender BF				
		Short stumpy BF				
		PF				
	AnTat 1.1	PF	poly(A)	mRNA fragmentation, ligation of 3′ and 5′ DNA/RNA adapters, 3′ adapter-specific priming	RNA-Seq[a]	
T. b. rhodesiense	YTat 1.1	PF	poly(A) ± terminator exonuclease; terminator exonuclease only	oligo(dT) or random hexamer ± SL priming for 5′ enriched libraries; 5′ polyphosphatase treatment followed by DNA/RNA adapter ligation, random hexamer and adapter-specific priming for 5′ triphosphate enriched library	RNA-Seq[b]	Kolev et al. (2010)
T. b. gambiense	STIB 386	BF	poly(A)	oligo(dT) followed by *Nla*III digestion, ligation of adapter 1 (*Mme*I), *Mme*I digestion and ligation of adapter 2	DGE	Veitch et al. (2010)
		PF				

PF procyclic form; *BF* bloodstream form

[a]Strand information preserved

[b]Strand information preserved for 5′-triphosphate enriched library

aligned to the reference genome when allowing for two mismatches per read. Sequencing of SLT libraries to a depth of 1–2 million mappable reads typically detects splicing at about 80% of annotated open reading frames (Nilsson et al. 2010). Since the readout is digital, with each SLT tag representing an associated transcript, expression levels can be displayed as numbers of tags per million. SLT can detect expression levels over a range of at least five orders of magnitude.

11.4.2 RNA-Seq

Although the SLT approach offers some advantages, such as the selective sequencing of parasite transcripts from a mixture of host and parasite RNA, more information can be gained by direct high-throughput sequencing of fragmented cDNA samples by RNA-Seq. Several publications have reviewed the method (Costa et al. 2010; Marguerat and Bahler 2010; Wang et al. 2009; Wilhelm and Landry 2009) and systematic investigation of data generated by RNA-Seq has proven the validity, and in many regards the superiority, of this approach in comparison with hybridisation-based methods for expression profiling (Fu et al. 2009; Marioni et al. 2008; Wilhelm et al. 2008). RNA-Seq not only represents a means of quantifying gene expression, it also offers the possibility of detecting all transcribed regions with a maximal resolution of a single nucleotide. The structure of mRNAs can be determined with great precision, including splice variants and polyadenylation sites (PAS). Since there is overwhelming evidence that many trypanosomal mRNAs are regulated by sequence elements in their 3′ UTRs, exact delineation of transcript boundaries is a prerequisite for genome-wide searches for regulatory sequence motifs. In contrast to other eukaryotes, *cis*-splicing of introns is restricted to two protein-coding genes in *T. brucei* (Mair et al. 2000; Siegel et al. 2010), which greatly facilitates the analysis of RNA-Seq data for this organism. Different strategies were used to prepare cDNA libraries for RNA-Seq. In common with the SLT libraries, the first step consists of enrichment of mRNAs from total RNA. In most instances this is achieved by selection of polyadenylated mRNAs on oligo(dT) beads. For some of their libraries, however, Kolev and colleagues replaced poly(A) selection by treatment of total RNA with terminator exonuclease, an enzyme which degrades RNAs with free 5′ monophosphate ends. A similar approach has been taken by others to analyse the *P. falciparum* transcriptome, where it was further combined with affinity-based depletion of very abundant RNA species (Otto et al. 2010). For library preparation the RNA sample can either be fragmented prior to cDNA synthesis, or the RNA is first converted into cDNA using random or oligo(dT) primers, followed by fragmentation. Most publications used the latter method, even though a bias towards the 3′ end of transcripts has been reported when oligo(dT) is used to prime first-strand cDNA synthesis (Mortazavi et al. 2008; Wang et al. 2009). In common with the SLT method described above, the presence of the spliced leader on all trypanosomatid mRNAs can be exploited to generate 5′ enriched libraries and circumvent this problem, when second-strand cDNA synthesis is primed with an oligonucleotide corresponding to this sequence (Kolev et al. 2010). Construction of sequencing

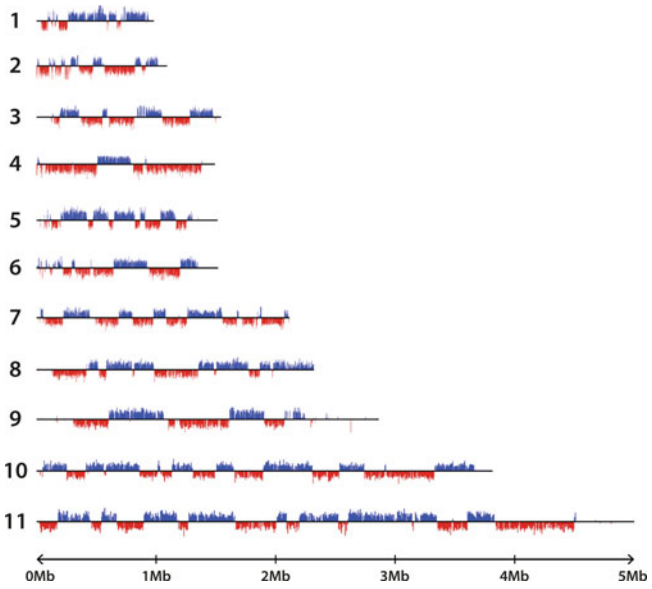


Fig. 11.1 Strand-specific RNA-Seq across the length of the 11 megabase chromosomes. Transcripts corresponding to the *upper* and *lower* strands are depicted in *blue* and *red*, respectively. These data can be accessed at http://splicer.unibe.ch

libraries results in a loss of strand information upon sequencing if no special measures are taken to preserve the strandedness of the sample. So far, strand-specific RNA-Seq data for *T. brucei* has not been published, with one exception: a library enriched for RNAs with 5′-triphosphate ends showed an accumulation of sequence reads at suspected transcription start sites and indicated that transcription initiation by RNA polymerase II at these sites is intrinsically bidirectional (Kolev et al. 2010). An RNA-Seq run, in which strand information was retained, was performed by our laboratory. These data, which are publicly available (http://splicer.unibe.ch), are shown for the 11 megabase chromosomes in Fig. 11.1. The dynamic range of RNA-Seq as observed by Siegel and colleagues in their comparison of expression levels in bloodstream and procyclic forms was close to 10^6. In addition, the technical reproducibility of the method was found to be excellent. The comparison of biological replicates and different approaches for cDNA library construction from the same life cycle stage generally resulted in highly correlated data sets (Kolev et al. 2010; Siegel et al. 2010).

11.5 Splice Site Detection and Mapping of 5′ UTRs

Post-transcriptional regulation of gene expression is very prominent in trypanosomatids, with *cis*-acting elements in UTRs being important determinants of transcript stability and translation (Clayton and Shapira 2007; Fernández-Moya and

Estévez 2010; Haile and Papadopoulou 2007; Ouellette and Papadopoulou 2009). Although most of the regulatory elements that have been identified to date in trypanosomatids are in the 3′ UTRs, elements in 5′ UTRs have also been identified, for example, those governing cell-cycle regulated mRNAs in *Crithidia fasciculata* (Pasion et al. 1996). Delineation of UTRs by experimentally mapping spliced leader addition sites (SAS) and polyadenylation sites (PAS) across the whole *T. brucei* genome is therefore a crucial step before systematic searches for sequence motifs can be performed. Although such searches have been carried out on the *T. brucei* genome, these were limited to computer predictions of the processing sites of mRNAs (Mao et al. 2009).

Sequencing reads derived from RNA-Seq potentially cover transcribed regions along the entire length of a transcript. Mapping of these reads to a reference genome and visualisation of the transcriptional landscape in a genome browser format typically results in alternating peak-and-valley patterns with characteristic sharp changes in numbers of RNA-Seq tags at transcript boundaries. On its own, the observation of rapid changes in RNA-Seq signals can only give a rough estimate of where transcripts actually start or end, with a resolution on the order of tens of base pairs. The identification of end tags that contain poly(A) sequences, greatly facilitates exact mapping of PAS (see below). In contrast to analyses in yeast and other systems, in which the 5′ ends of the majority of transcripts have been determined by a combination of RNA-Seq and 5′ RACE (Nagalakshmi et al. 2008) the spliced leaders at the 5′ ends of trypanosome mRNAs have been exploited to precisely identify SAS from RNA-Seq data. Extraction of reads that cover *trans*-splicing junctions, consisting of sequences corresponding to the conserved spliced leader and the 5′ end of the mRNA to which it is spliced (and therefore also containing strand information), enabled the 5′ boundaries to be mapped for 8,960 and 6,959 transcripts, respectively (Kolev et al. 2010; Siegel et al. 2010). Libraries enriched for 5′ ends, with random priming of first-strand cDNA and a spliced leader primer for synthesis of the second strand were used by Kolev et al. (2010) and Nilsson et al. (2010). Furthermore, the SLT method described above was originally devised for global SAS identification by direct high-throughput sequencing from the spliced leader. SAS for 8,277 genes could be detected in long slender bloodstream forms by the SLT approach and slightly fewer in stumpy bloodstream forms and procyclic forms (Nilsson et al. 2010). The three data sets are in excellent agreement, indicating that strain and sub-species differences do not play a major role. On average there are 2–3 SAS per gene, but in some cases the number is considerably higher. Nilsson et al. therefore assigned an arbitrary cut-off of 60% for major splice sites, below which a transcript was termed alternatively spliced. Based on these criteria, more than 1,200 genes gave rise to alternatively spliced transcripts in the three stages of the life cycle that were analysed by SLT. Reverse transcriptase-PCR (RT-PCR) analysis of a selected number of genes confirmed the presence of alternatively spliced mRNAs under steady-state conditions (Kolev et al. 2010; Nilsson et al. 2010). Investigation of splice acceptor sites confirmed the expected prevalence of the AG splice acceptor dinucleotide at SAS (major SAS: 98% and 94%; minor SAS: 75% and 80%, in the two publications). In two instances

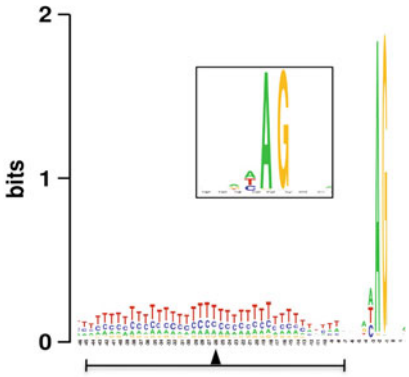


Fig. 11.2 Consensus of the major 3′ splice sites from procyclic form *T. brucei* [reproduced from Nilsson et al. (2010)]. The graphic was generated using Weblogo (Crooks et al. 2004). The *inset* shows a close-up of the conserved AG at the 3′ splice site. The upstream polypyrimidine tract is dominated by T residues starting at an average distance of 26 nucleotides (*black triangle*) upstream of the 3′ splice site

Kolev and colleagues noted that assumed non-AG splice acceptors, based on the *T. brucei* TREU 927 reference genome, were in fact canonical AG acceptors in the strain they used for RNA-Seq. Since most studies that globally mapped SAS relied upon strains other than the genome strain, it cannot be excluded that at least some of the "non-AG" sites that were detected are due to strain differences. There are alternative explanations, however, including sequencing errors or allelic differences. Further assessment of sequence elements surrounding 3′ splice sites in pre-mRNAs confirmed the depletion of G residues at the -3 position in agreement with previous suggestions (Siegel et al. 2005). Polypyrimidine tracts, needed for efficient splicing and described to occur upstream of the SAS, were dominated by T residues and usually separated from the SAS by fewer than 50 bases. Even though the lengths of the polypyrimidine tracts varied, they were typically longer than 10 bases, the minimum length previously reported to be sufficient for *trans*-splicing in trypanosomes (Benz et al. 2005; Huang and Van der Ploeg 1991; Matthews et al. 1994). The predictive power of current bioinformatic tools for SAS detection was found limited in that only 40% of major SAS and 6% of minor SAS were correctly identified when compared with SLT data (Nilsson et al. 2010). The consensus sequence of *T. brucei* 3′ splice sites as determined from SLT data is shown in Fig. 11.2. Calculations of 5′ UTR lengths were different between the three publications with median values being considerably lower in the SLT study (35–42 bases) than by RNA-Seq (~90 bases). To some degree these differences might reflect different approaches to calculating lengths, for example, if only major SAS or all SAS are included. In either case, the detection of multiple SAS for most genes results in heterogenous 5′ UTRs. The shortest 5′ UTRs range from zero, with the start codon immediately downstream of the spliced leader (for example, Tb11.01.5535, encoding an Sm-like protein) to approximately 2.5 kb. An example

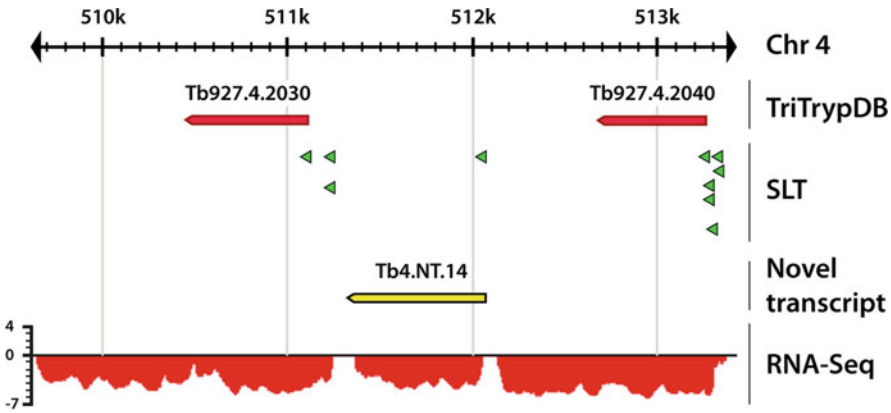


Fig. 11.3 Illustration of a region of the genome showing a novel transcript between genes encoding the related proteins Alba 3 (Tb927.4.2040) and Alba 4 (Tb927.4.2030). The boundaries of the three transcripts are clearly seen in the RNA-Seq trace below. *Arrowheads* mark the positions of SLT tags

of the latter is Tb927.8.7780, which encodes a conserved hypothetical protein. Despite its length, this 5′ UTR does not contain upstream ORFs.

Approximately one thousand potentially misannotated genes were detected by SLT and RNA-Seq. Half of these are spliced within the ORF, and would therefore give rise to shorter proteins than predicted, while the other half potentially start upstream and give rise to larger polypeptides. Some of these were verified experimentally by RT-PCR and Northern blotting or are supported by proteomic data in the case of N-terminal extension of gene products (Kolev et al. 2010; Nilsson et al. 2010; Siegel et al. 2010). Regarding alternative splicing, Nilsson and colleagues proposed four possible functions of transcript variants arising from such events. Accordingly, alternative SAS choice can result in (a) transcripts that are untranslatable because they lack start codons; (b) alterations at the N terminus that change protein localisation, because of a loss or gain of targeting sequences; (c) inclusion or exclusion of regulatory RNA elements such as upstream ORFs or internal ribosome entry sites and (d) the use of alternative ORFs encoded in the transcript (Nilsson et al. 2010). Kolev et al. speculated that some of the heterogeneity in splice site choice in trypanosomatids—at least for those events that do not result in a change of a transcript's coding potential—can be explained on the basis of reduced selective pressure. In contrast, alternative *cis*-splicing would be subjected to a high degree of evolutionary pressure (Kolev et al. 2010). However, alternative *trans*- and *cis*-splicing, leading to distinct localisation and function of proteins, has been previously observed in trypanosomes (Ashibe et al. 2007; Benabdellah et al. 2007; Lopez and Kelleher 2009) and has been proposed to play a role in the dual localisation of some tRNA synthetases (Nilsson et al. 2010; Rettig et al. 2012).

Following the discovery of more than a thousand novel transcripts (NTs) by Kolev et al. (see Sect. 10.8), it was realised that some SLT tags that were designated as alternative splice sites for specific genes were actually derived from unannotated

transcripts upstream of a particular ORF. An example is shown in Fig. 11.3. The situation may be more complicated, however, as alternatively spliced and alternatively polyadenylated mRNAs from a single region may co-exist (Vassella et al. 1994). Integration of SLT and RNA-Seq data and correction of genome annotations will therefore be highly beneficial for our understanding of the transcriptome of *T. brucei*.

11.6 Polyadenylation and Mapping of 3′ UTRs

The polypyrimidine tracts and AG dinucleotides between adjacent genes play a dual role in polyadenylation of the upstream transcript and *trans*-splicing of the downstream transcript (Hug et al. 1994; Matthews et al. 1994; Schürch et al. 1994). In general, it has proven more difficult to predict and validate PAS than SAS. The two publications that used RNA-Seq identified PAS for 2,081 and 5,948 genes, respectively, but the degree of congruence between the two sets of RNA-Seq data was much lower than for that obtained for SAS. As previously predicted from analyses of small data sets (Benz et al. 2005; Schürch et al. 1994), there is a preference for polyadenylation at (small runs of) adenosine residues. Both analyses gave a similar median length for 3′ UTRs in procyclic forms of *T. brucei* (388–400 bases). Once again, however, the range is considerable. There are transcripts with no 3′ UTR, in which the last two bases of the stop codon (UAA) are also the start of the poly(A) tail, e.g. procyclin-associated genes PAG 1 and PAG 2 (Koenig-Martin et al. 1992; Liniger et al. 2001). The longest predicted UTR of 9.7 kb is attributed to an mRNA encoding a hypothetical protein of unknown function (Tb11.01.8270).

Surprisingly, Kolev et al. found that fewer than 60 mRNAs have a unique PAS (Kolev et al. 2010). For the vast majority of transcripts, however, polyadenylation is dispersed. This suggests that there is either imprecise cleavage by the splicing/polyadenylation complex, or a form of "reverse torpedo" effect in which a 3′ to 5′ exonuclease competes with the poly(A) polymerase. In the case of β-tubulin, the same 4 major PAS were identified by the two laboratories employing RNA-Seq, with one site accounting for 63% of the tags. In addition, thirty minor sites were mapped within a region of 76 bases. Is this apparent sloppiness in the choice of PAS deliberate or merely tolerated? What might be the consequences of this heterogeneity, given that elements in 3′ UTRs can affect message stability or translation? Our RNA-fold analyses of the predominant 3′ UTRs of β-tubulin mRNA (data not shown) indicate that these differences in length do not have an impact on the secondary structure. However, it seems quite plausible that trypanosomes might exploit the feature of dispersed polyadenylation to generate subsets of mRNAs with different half-lives or translation efficiencies. As is the case for splicing, some transcripts have alternative PAS, governed by separate polypyrimidine tracts, that can be hundreds of bases apart (Erondu and Donelson 1992; Vassella et al. 1994).

A number of ORFs that are annotated as "hypothetical protein, unlikely", with no evidence of their own spliced leader tags, are included in the 3′ UTRs of *bona fide* mRNAs. In other cases polyadenylation occurs within such ORFs. Transcripts

that are genuinely bicistronic, with both ORFs giving rise to proteins, have not been reported for *T. brucei*, but it is nevertheless possible that they exist. One possible candidate is an mRNA encoding a hypothetical protein (Tb927.4.2550) followed by the ORF for cardiolipin synthetase (CLS; Tb927.4.2560). RNA-Seq detects transcription of this region. All three studies revealed extensive heterogeneity of splicing upstream of Tb927.4.2550, but none of the three studies detected *trans*-splicing of the CLS mRNA in procyclic forms although SLT tags were found on rare occasions at a sequence depth of 40 million (B. Schimanski, K. Gunasekera, T. Ochsenreiter and I. Roditi, unpublished data). Nevertheless, cardiolipin is present in procyclic forms and knockdown of CLS causes growth defects in this stage of the life cycle (Serricchio and Bütikofer 2012). Further experiments are needed to clarify if there might be an internal ribosomal entry site between the two ORFs.

Clusters of protein-coding genes also harbour snoRNAs and a subset of tRNAs in the intergenic regions. Primary transcripts appear to be spliced and polyadenylated before being further processed to their mature forms. In addition, snoRNAs can be embedded in the 3′ UTR or coding region of mRNAs (Kolev et al. 2010). Both snoRNAs and the proteins derived from the mRNAs can be detected. Since there is no evidence that these mRNAs are recycled to produce snoRNAs in the cytoplasm, it would appear likely that a transcript can follow two pathways: further processing in the nucleus to snoRNAs or export to the cytoplasm for translation.

11.7 Transcript Abundance

In contrast to the great degree of consensus on processing sites, calculations of transcript abundance give conflicting results. This may be due to the way in which systems were calibrated, however, rather than to fundamental differences in data. The number of mRNAs per cell has been estimated at 2×10^4 for bloodstream forms and 4×10^4 for procyclic forms of *T. brucei* (Haanstra et al. 2008; Nilsson et al. 2010). Haanstra and colleagues previously calculated that there were 18 ± 2 molecules of phosphoglycerate kinase B (PGK B) mRNA and 72 ± 12 molecules of α-tubulin mRNA per cell in procyclic forms. Estimates of transcript abundance were presented in different ways by the groups that performed RNA-Seq and SLT. Siegel and colleagues used scaled counts (number of tags per kilobase of ORF), without calculating absolute numbers. The largest number of tags was attributed to VSG 221, followed by approximately 20-fold fewer tags for α-tubulin. Kolev et al. (2010) present a supplementary table with numbers of transcript per cell for each gene, and conclude that 75% of transcripts are present at between 1 and 10 copies per cell. These were calibrated according to the number of PGK B transcripts calculated by Haanstra et al. (2008). If transcript numbers are summed up, they total more than 6×10^4 transcripts per cell and the number of α-tubulin mRNAs is extremely high (810 copies per cell). SLT data are presented by Nilsson et al. (2010) as tags per million for all life cycle stages. As expected, VSG mRNAs are the most abundant mRNAs in bloodstream forms (7–11% depending on the strain)

equivalent to 1,400–2,200 molecules per cell. According to SLT data, PGK B mRNA would be present at 2–3 molecules per cell (in procyclic forms), α-tubulin at about 135 molecules per cell, and more than 80% of mRNAs would be present at $\leq$1 copy per cell. In fact, the ratios of α-tubulin to PGK B obtained by the two laboratories are in reasonable agreement—it is the absolute numbers that differ.

11.8 Novel Transcripts

One of the principal aims of transcriptome analysis is to define those parts of a genome that are actively transcribed. Whereas most microarray designs typically rely on annotated features, the introduction of tiling arrays offered a less biased view of an organism's transcriptome. Tiling arrays, however, still have limited resolution for most eukaryotic genomes and are not available for trypanosomes. In their study of the transcriptome of procyclic forms of *T. brucei*, Kolev et al. (2010) found a total of 805 annotated genes to be transcriptionally silent. This could be due to very low expression levels in this stage of the life cycle—below the detection limit of RNA-Seq—but a more likely explanation is that most of these do not really encode proteins. This is supported by the fact that the majority of these non-transcribed genes (64%) are annotated as hypothetical unlikely. The second largest group of non-transcribed genes found by Kolev et al. (2010) consisted of hypothetical genes (23%). This is consistent with the data set from Nilsson et al. (2010), in which hypothetical unlikely genes ranked highest among the annotated genes without SLT tags, followed by hypothetical genes. Balancing this decrease in the number of expressed annotated genes is the discovery of 1,114 *trans*-spliced and polyadenylated novel transcripts (NTs) with no prior annotation in the *T. brucei* genome database (Kolev et al. 2010). Of these, 1,011 are potentially protein-coding, giving rise to polypeptides $\geq$25 amino acids in length, whereas 103 of the NTs have no coding potential at all (assuming a requirement for AUG as the start codon) or would result in extremely short peptides. NTs have been numbered according to the chromosome from which they originate, and then sequentially from left to right (e.g. Tb1.NT.1). We have incorporated all NTs into the Splicer database (http://splicer.unibe.ch). In contrast to annotated genes for which no transcripts were detected, and which are overrepresented on chromosomes 1 and 9, NTs are quite evenly distributed among chromosomes, correlating with chromosome length (Fig. 11.4). NTs range in size from 154 bases to >10 kb, with expression levels resembling those of transcripts from annotated ORFs. Coding NTs (median length: 801 nt; mean length: 1,097 nt) are significantly longer than non-coding NTs (median length: 378 nt; mean length: 460 nt). The majority of coding NTs potentially give rise to proteins $\leq$50 amino acids that would be missed in standard analyses using polyacrylamide gels. Evidence supporting the correctness of some of the newly discovered transcripts comes from different sources. Most convincing is the identification of peptides from an earlier proteomics analysis (Panigrahi et al. 2009), that at the time of publication could not be assigned to any annotated *T.*

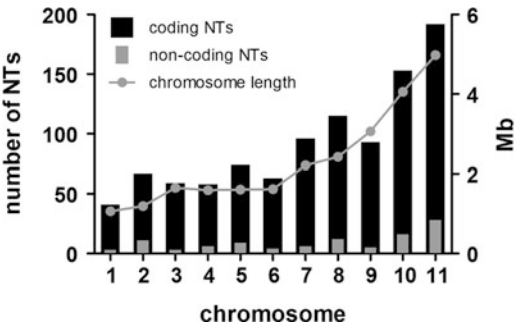


Fig. 11.4 Distribution of novel transcripts (NTs) on different chromosomes in relation to their size

brucei ORF, but were now found to match with 19 predicted proteins encoded by NTs discovered by Kolev et al. (2010). Furthermore, 27 NTs encode ORFs with homology to annotated gene products in the related parasites *T. cruzi* and *L. major*, and another 23 NTs encode potential proteins that show conservation in those species despite the lack of annotation. Five NTs were demonstrated by Northern blot analysis. In the case of Tb8.NT.93-95, the transcripts are somewhat different in size but, interestingly, all seem to code for small highly basic proteins related to human ribosomal protein L41. Expression levels of these three NTs were found to be relatively high (several dozen copies per cell based on calculations using PGK B as a reference) (Haanstra et al. 2008). Furthermore, both SLT data and RNA-Seq data show that their expression is not limited to procyclic forms, but is comparably high in bloodstream forms (Nilsson et al. 2010; Siegel et al. 2010). There are some NTs that appear to encode truncated forms of known proteins, for example Tb11.NT.118 which corresponds to the first 71 amino acids of ribosomal protein L29. Whether such proteins are truly functional, or are merely relics of gene duplication, will need to be established. There are also examples of NTs that contain unannotated ORFs of considerable length, the most striking being Tb9.NT.98, which encodes a repetitive protein of 2,843 amino acids. This overlaps with an ORF on the opposite strand that is annotated as the ISG100 pseudogene (Tb09.v2.0010). There are no SLT tags preceding ISG100, however, whereas there are ample tags for Tb9.NT.98, particularly in bloodstream forms. These findings, taken together with the fact that Tb9.NT.98 is transcribed in the same direction as its flanking genes, strongly supports that it is the correct transcript from this locus.

The significance of defining all transcribed regions of a genome—especially those with no previous annotation—cannot be overemphasised. We have already discussed one of the problems that can arise when SLT tags are assigned to the next annotated gene and interpreted as alternative splicing events, when in reality they derive from NTs. But there are other pitfalls, as illustrated by the example shown in Fig. 11.3. Tb4.NT.14 is a NT that lies between two annotated ORFs on chromosome 4 (Tb927.4.2030, Tb927.4.2040). In this particular case, the two genes flanking Tb4.NT.14 encode proteins of the same family that are highly similar to each other. Given their proximity in the genome and assuming possibly redundant functions, it might be tempting to try a knockout strategy in which both genes are

deleted and replaced by a single resistance cassette in one round of transfection. Under these circumstances the unsuspecting scientist might observe phenotypes that are not caused by deletion of the genes under consideration but rather by removal of the "gene in between".

11.9 Antisense Transcription and Silent Regions

Overlapping sense and antisense transcripts originating from a single locus have been detected in both *T. brucei* (Liniger et al. 2001) and *L. major* (Belli et al. 2003). In the case of *T. brucei*, the transcripts derived from a region where a Pol II transcription unit converged with a Pol I transcription unit (the procyclin expression site on chromosome 10). Both sense and antisense transcripts were *trans*-spliced and polyadenylated. The two publications predated the genome sequences, and at the time it was suggested that this phenomenon might be more widespread. However, SLT and RNA-Seq analyses that contain strand information indicate that antisense transcripts are found almost exclusively at strand-switch regions and that expression levels are often so low that, on average, there would be fewer than one transcript per cell. This would explain why the sense transcripts are not destroyed by RNA interference and why deletion of the argonaute gene Ago1 has little influence on their steady-state levels (Haenni et al. 2009). It should be noted, however, that poly(A)-selected RNA was used for construction of the SLT and RNA-Seq libraries, which biases the detection in favour of processed transcripts. Nuclear run-on analysis of the procyclin expression site and the convergent Pol II transcription unit indicated that overlapping transcription occurs over a region of several kilobases in procyclic forms (Liniger et al. 2001). By mapping tags derived from 5′ triphosphate-enriched RNA, Kolev et al. (2010) demonstrated bidirectional transcription from putative Pol II promoters, which could explain some of these data; however, it is not clear from their publication how far this extends. Interestingly, some of the transcription from the opposite strand to the procyclin transcription unit is resistant to α-amanitin (Liniger et al. 2001), implying that this may be due to Pol I, rather than a continuation of transcription by Pol II.

The procyclin expression site described above consists of procyclin genes and procyclin-associated genes (PAGs). Despite being transcribed from the same Pol I promoter, the steady-state levels of the RNAs differ approximately 100-fold. RNA stability may play a role in this, but there is also evidence that sequences within PAG 1 impede transcription (Haenni et al. 2009), possibly by affecting chromatin conformation since upstream reporter genes are also silenced. Apart from the PAGs in the procyclin expression sites, the genome contains 6 PAG-like genes and $\geq$30 PAG-like sequences, the majority of them at converging strand-switch regions and/or in regions where there is a change of polymerase. Comparing these regions of the genome with the RNA-Seq data obtained from procyclic forms (Nilsson et al. 2010; Siegel et al. 2010) gives a striking picture—these regions are devoid of transcripts for several kilobases (Fig. 11.5). PAG-like sequences are also found in some of the

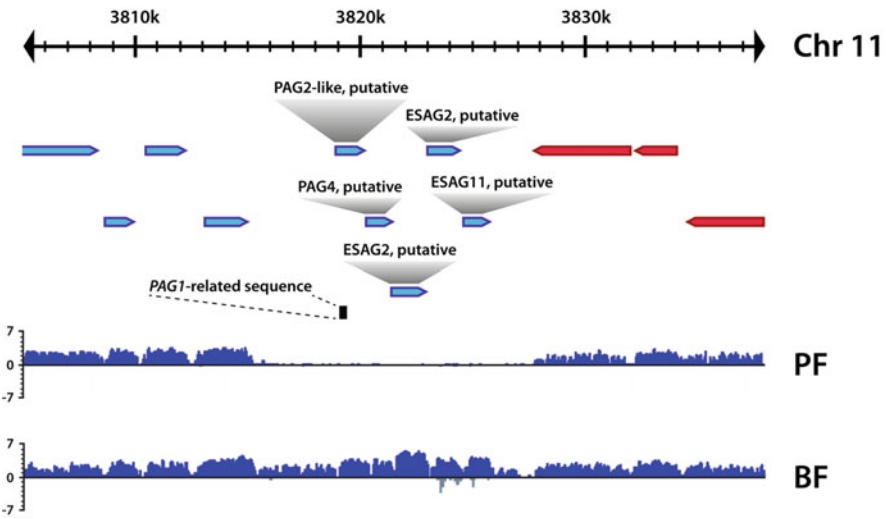


Fig. 11.5 A cluster of genes showing coordinate regulation between procyclic forms (PF) and bloodstream forms (BF). It is striking that procyclin-associated genes (PAGs) and PAG-like sequences are found in regions devoid of transcripts in procyclic forms. RNA-Seq data from Siegel and colleagues (2010) was obtained from http://tritrypdb.org. The PAG-1 related sequence shares 56% identity with a sequence implicated in silencing of transcription (Haenni et al. 2009)

ribosomal spacers; these were already known to be transcriptionally silent (Rudenko et al. 1994), making them preferred integration sites for inducible constructs (Wirtz and Clayton 1995). The possible significance of some of these silent regions is addressed in the section below.

11.10 Stage-Specific Gene Expression

Three of the deep sequencing analyses were performed on two or more life-cycle stages (Nilsson et al. 2010; Siegel et al. 2010; Veitch et al. 2010), enabling stage-regulated genes to be identified. Very different numbers were obtained, however, which is more likely to be due to the different criteria used to identify these genes than to major differences between strains and sub-species. In their comparison of monomorphic (long slender) bloodstream forms and procyclic forms of *T. b. brucei* by RNA-Seq, Siegel et al. (2010) classified 6% of genes as being differentially expressed. For this analysis, paralogous ORFs and ORFs containing repetitive sequences were excluded, with the consequence that some well-known stage-regulated genes, such as the procyclins and CRAM, were overlooked. The criteria for stage-regulation were set at changes in tag abundance >2-fold and $P \leq 0.01$ (one-sided t test). Using DGE profiling to compare bloodstream forms and procyclic forms of *T. b. gambiense*, Veitch et al. (2010) found that 28% showed

≥2.5-fold changes in abundance, with the majority being up-regulated in bloodstream forms. They regarded this threshold as being too arbitrary, however, preferring instead to use a Rank Products analysis that took biological replicates, *P*-values and the false discovery rate (FDR) into account. Using very stringent criteria (FDR 0.2, *P*-value <0.003), 126 genes were found to be up-regulated in bloodstream forms and 63 in procyclic forms. Decreasing the FDR to 0.1 led to the exclusion of some genes that had been proven to be stage-regulated. As mentioned by the authors, not all genes would have been identified. The DGE protocol generates tags from the first *Nla* III site upstream of the poly(A) tail, and as these were only mapped to ORFs, transcripts lacking an *Nla* III or transcripts with an *Nla* III site in the 3′ UTR would not feature in the analysis (once again, procyclin transcripts slip through the net).

By far the highest number of stage-regulated genes, 40%, was obtained using SLT to compare long slender bloodstream forms, short stumpy bloodstream forms and procyclic forms of a fly-transmissible strain of *T. b. brucei*. In this study, Nilsson et al. (2010) tested genes for significant differences in expression according to Audic and Claverie, which takes both the total number of tags and changes in abundance into account. According to this analysis (setting the thresholds at ≥2-fold change in tags per million, $P < 10^{-5}$) there were 3,214 genes that were differentially expressed, with the greatest differences between long slender forms and procyclic forms. With hindsight, we realise that almost 200 of the changes attributed to annotated ORFs by SLT are in fact due to coding NTs, but this does not change the general picture of stage-regulation. By their design, neither the RNA-Seq analysis of *T. b. brucei* nor the DGE analysis of *T. b. gambiense* identified NTs that were stage-regulated as the analyses were limited to tags mapping to ORFs. Since the raw data were available, the analysis of *T. b. gambiense* was subsequently extended to include the 3′ UTRs and intergenic regions; this information is now accessible at TriTrypDB as "SAGE tags".

There is limited evidence that the site of polyadenylation or the efficiency of RNA processing can be stage-regulated (Erondu and Donelson 1992; Flück et al. 2003), but with the exception of differential splicing of a synthetic luciferase transcript in bloodstream and procyclic forms (Helm et al. 2008), there are no previous reports of stage-regulated splicing. Comparison of the major SAS in three life cycle stages (long slender bloodstream forms, short stumpy bloodstream forms and procyclic forms) indicates that several hundred transcripts are subject to differential splicing (Nilsson et al. 2010). For three transcripts (Tb927.1.790, Tb11.02.2700 and Tb927.6.4240), the SLT data was validated by quantitative RT-PCR. The significance of differential splicing in the context of parasite development remains to be tested. It is currently not clear whether splicing itself is prone to regulation or whether the 5′ UTRs alter RNA stability in a stage-specific manner. The recent establishment of an in vitro *trans*-splicing system may prove valuable for addressing this question (Shaked et al. 2010).

It is generally accepted that there is little regulation of transcription in trypanosomes and that most adjustment to mRNA levels occurs post-transcriptionally. So far, the promoters that show regulated activity through the

life cycle are all specific for Pol I. These include the promoters for the VSG expression sites (both active and silent) and the procyclin promoters. Nilsson and coworkers performed SLT with bloodstream forms of *T. b. brucei* MITat 1.2 (VSG 221); this enabled relative transcript levels to be mapped across the active expression site and, in the process, identified a new expression site-associated gene, ESAG 13, which encodes a polypeptide of 96 amino acids. The VSG expression site provides an excellent example of the interplay between promoter activity and RNA stability. The mRNAs for the VSG mRNA and all ESAGs are derived from the same primary transcript, yet the number of tags detected ranged from zero for ESAG 3-1 and ESAG 8-2 to 69,000 per million for VSG 221. The analysis of *T. b. gambiense* also gives some interesting insights into VSG expression. Two-thirds of the VSG tags obtained in bloodstream forms were assigned to 3 genes (annotated as pseudogenes in the reference genome TREU927), but tags for more than a hundred additional VSG genes were also detected at low levels within this population. The simplest explanation is that the population is heterogeneous, but expression within a cell is monoallelic, as is generally accepted. However, the authors also raise the (currently heretical) possibility that there might be also low-level transcription of multiple VSG genes from internal sites within a single cell (Veitch et al. 2010).

Recently, another publication by Siegel et al. (2009) mapped putative Pol II transcription start sites across the entire genome of the same strain of trypanosomes that they used for RNA-Seq. Two sites were found to be stage specific, one on chromosome 7, which is only detected in bloodstream forms and one on chromosome 11, which is only detected in procyclic forms. When the RNA-Seq profiles of these regions are inspected, however, there are no striking differences between steady-state levels of RNA in the two life cycle stages. Data obtained from the analysis of the transcriptome of *T. b. gambiense*, and consolidated by RNA-Seq, SLT and microarray data from *T. b. brucei* [(Kabani et al. 2009) and K. Matthews, personal communication], suggests that there might be regions of the genome that are coordinately regulated. Veitch and coworkers identified four clusters of genes that were up-regulated in bloodstream forms and one in procyclic forms. All of the bloodstream form clusters contain putative PAGs and/or ESAGs (Veitch et al. 2010). The two clusters containing PAGs are silent regions in procyclic forms (see above). However, it is clear from the comparative RNA-Seq analysis by Siegel et al. that steady-state RNA is derived from these regions in bloodstream forms (Fig. 11.5); furthermore, the RNA is *trans*-spliced, as SLT tags are detected. Given that members of ESAG and PAG families are part of the VSG and procyclin transcription units, which are transcribed by Pol I in a stage-regulated fashion, Veitch and colleagues suggest that this might also be the case for the new clusters. This proposal is extremely tantalising, but further experiments such as nuclear run-on analysis in the presence or absence of α-amanitin are required to clarify this. An alternative interpretation is that PAG-like sequences silence transcription in a stage-specific manner. It is also worth noting that some PAG-containing regions are silent in both bloodstream and procyclic forms, but could conceivably be active in other stages of the life cycle.

In summary, despite the different methods used to identify stage-regulated genes, there is a core set that is found in all cases. These can be regarded as genes that are hard-wired for differential expression, while others may be more susceptible to external factors such as nutrients or temperature. Some of the new core genes encode metabolic enzymes and others are likely to code for membrane proteins, but a substantial proportion of them do not have any annotated features. The next important step will be to devise an intelligent series of high-throughput screens that can provide clues about their functions.

11.11 Future Applications in the Seq Era

The major driving force behind the development of novel high-throughput sequencing technologies was to speed up genome sequencing projects and at the same time cut costs. RNA-Seq and SLT are now sufficiently well established for *T. brucei* that they are cost-effective methods to monitor transcription under a range of conditions. Ongoing projects in various laboratories include RNA-Seq through the entire life cycle, the identification of genes conferring drug resistance (Alsford et al. 2012) and the transcriptome-wide determination of RNA half-lives—to list just a few examples. In recent years, trypanosomal RNA-binding proteins (RBPs) have received increased attention in attempts to define molecular mechanisms of post-transcriptional regulation (reviewed in (Fernández-Moya and Estévez 2010; Kramer and Carrington 2010)). There are examples of microarrays being used to analyse the effect of RBP depletion on the transcriptome of *T. brucei* (Estevez 2008; Stern et al. 2009), but in future these will be replaced by deep sequencing. The identification of mRNA targets for RBPs is another key issue. With the advent of NGS technology, immunoprecipitation of RBPs can be combined with high-throughput sequencing to map RNA-protein interactions at single-nucleotide resolution [HITS-CLIP and iCLIP; (Licatalosi et al. 2008; König et al. 2010)].

Regulation of mRNA stability is generally viewed as the main mechanism to control gene expression in trypanosomatids, but translational control is also likely to play a crucial role (Furger et al. 1997; Hehl et al. 1994; Schürch et al. 1997; Urwyler et al. 2005; Walrad et al. 2009). A method termed ribosome profiling, that globally monitors translation, was recently established for *S. cerevisiae*. Translating ribosomes were stalled on mRNAs by cycloheximide treatment, followed by nuclease digestion of unprotected mRNA fragments and deep sequencing of the remaining ribosome "footprints" (Ingolia 2010; Ingolia et al. 2009). In this manner ribosome positions could be charted at single-codon resolution under different nutrient conditions. The method should easily be adapted for trypanosomes and is certain to provide valuable insights into translational control.

In addition to studies of RNA landscapes, genome-wide mapping of histone variants, the bromodomain factor BDF3 and post-translationally modified histones in *T. brucei* by chromatin immunoprecipitation and deep sequencing (ChIP-Seq) has delivered interesting insights (Siegel et al. 2009; Wright et al. 2010). Trimethylated

H3K4, acetylated H4K10, the histone variants H2AZ and H2BV and BDF3 are enriched at probable RNA Pol II transcription start sites, whereas H3V and H4V preferentially occupy presumed termination sites for RNA Pol II. Similarly, peaks for acetylated histones H3K9/K14 and H4K5/K8/K12/K16, and trimethylated H3K4 are enriched at divergent strand switch regions in *T. cruzi* and acetylated H3K9/K14 is localised preferentially in corresponding regions in *L. major*, marking potential RNA Pol II transcription initiation sites in these parasites (Respuela et al. 2008; Thomas et al. 2009). This suggests a central role for these protein factors and modifications, and therefore for chromatin structure, in the control of transcription initiation and termination. Furthermore, immunoprecipitation of genomic DNA fragments containing the hyper-modified base J (β-D-glucosyl-hydroxymethyluracil) from bloodstream form *T. brucei*, followed by high-throughput sequencing, revealed an enrichment of base J at regions flanking RNA Pol II transcription units (Cliffe et al. 2010).

NGS-based procedures have been devised to gain insights into chromatin structure and organisation and map transcriptional regulatory elements. Genome-wide detection of DNase I hypersensitive genomic sites by DNase-Seq, for instance, has been applied to both *S. cerevisiae* and human primary cells to delineate regions of open chromatin (Boyle et al. 2008; Hesselberth et al. 2009; Song and Crawford 2010). Highly accessible chromatin locations have also been identified on a global scale in HeLa cells starting from crosslinked chromatin which, after sonication and size-fractionation, was analysed by deep sequencing (Sono-Seq; Auerbach et al. 2009). Nucleosome distribution across genomes is non-random and dynamic, and nucleosome occupancy mapping has recently been achieved with unprecedented coverage and resolution. *F*ormaldehyde-*a*ssisted *i*solation of *r*egulatory *e*lements (FAIRE)-Seq and *M*Nase-*a*ssisted *i*solation of *n*ucleosomes (MAINE or MNase)-Seq represent complementary approaches to enrich for nucleosome-free and nucleosome-bound genomic DNA, respectively, followed by high-throughput sequencing (Giresi et al. 2007; Kaplan et al. 2009; Schones et al. 2008; Shivaswamy et al. 2008; Zhou et al. 2011). FAIRE-Seq and MAINE-Seq have been successfully used in combination to chart the nucleosomal landscape in *Plasmodium falciparum* during the erythrocytic cycle, leading the authors to conclude that chromatin dynamics play a more pronounced role in transcriptional regulation and life-cycle progression than was previously envisaged. In particular, nucleosome-free transcription start sites and core promoters were found to correlate with high gene expression in asexual erythrocytic parasites, whereas transcription start sites and promoters that were occupied by nucleosomes in these life stages were associated with genes preferentially expressed in sexual forms of the parasite (Ponts et al. 2010a, b). Comparative investigation of chromatin structure between bloodstream and procyclic forms of *T. brucei*—mostly by electron microscopy—revealed a higher degree of chromatin condensation (heterochromatin) in the bloodstream form, whereas open chromatin (euchromatin) was more prevalent in procyclic form parasites (Rout and Field 2001; Schlimme et al. 1993). Two groups have recently shown that the active VSG expression site in *T. brucei* is depleted of nucleosomes (Figueiredo and Cross 2010; Stanne and Rudenko 2010). Accurate

mapping of nucleosome-free regions or DNase I hypersensitive sites on a genome-wide level, and possibly in different life cycle stages, would greatly ameliorate our currently poor understanding of chromatin structure and dynamics in these parasites.

Cytosine methylation is a well-known epigenetic mechanism that influences gene regulation in many eukaryotes. Methylome analysis has recently experienced a boost by the introduction of genome-wide analysis methods. Different approaches have been used to chart 5-methylcytosine (5MC) positions across different genomes. These include direct high-throughput sequencing of bisulphite-treated DNA (MethylC-Seq or BS-Seq), digestion of genomic DNA with methyl-sensitive restriction enzymes followed by sequencing (Methyl-Seq) and finally sequencing of methylated DNA fragments after enrichment, by a method termed *me*thylated *D*NA *i*mmuno*p*recipitation sequencing (MeDIP-Seq) (Brunner et al. 2009; Cokus et al. 2008; Down et al. 2008; Lister et al. 2008, 2009). Modified bases other than base J and its 5-hydroxymethyluracil precursor were long thought to be absent from the DNA of *T. brucei*, and although the presence of 5MC in *T. cruzi* DNA was reported 20 years ago (Rojas and Galanti 1990), it was identified in *T. brucei* only in 2008 (Militello et al. 2008). ChIP experiments in combination with cloning of precipitated material and conventional Sanger sequencing allowed the authors to identify a limited number of genomic loci containing 5MC. Based on these data and comparison of the role of 5MC in other organisms, it was speculated that one function of DNA methylation in *T. brucei* might be to silence VSG expression sites. To draw more definite conclusions, however, more comprehensive data sets are needed.

11.12 The New Frontier: Third Generation Sequencing

The commonly used NGS platforms are based on sequencing-by-synthesis approaches and rely on an imaging process after incorporation of fluorescently labelled nucleotides or oligonucleotides by polymerases or ligases, or measure bioluminescence signals in the case of pyrosequencing. All these systems use bead-based emulsion PCR or solid-phase amplification of immobilised DNA templates in order to increase signal intensity for the sequencing and imaging processes that follow (Fuller et al. 2009). Amplification-free single-molecule sequencing platforms that reduce the potential bias introduced by the amplification step, and reach even higher throughputs, are currently under development. One of these next-next or third generation sequencing systems (NNGS), the Helicos® Genetic Analysis System (http://www.helicosbio.com) (Harris et al. 2008; Thompson and Steinmann 2010), has already arrived on the market and has shown its utility in a number of studies. These include the sequencing of a human genome, gene expression analysis of yeast and mouse cells from minute amounts of RNA or small numbers of cells, direct sequencing of immunoprecipitated DNA and direct sequencing of RNA to circumvent any bias associated with conversion to cDNA (Goren et al. 2010; Lipson et al. 2009;

Ozsolak et al. 2009, 2010a, b; Pushkarev et al. 2009). The single-molecule real-time NNGS system by Pacific Biosciences (http://www.pacificbiosciences.com), which uses immobilised polymerases rather than immobilised DNA templates, is expected to boost sequencing applications even further (Eid et al. 2009; Korlach et al. 2010). For example, it was recently shown that this system could directly detect DNA methylation without the need for bisulphite conversion of cytosine residues prior to sequencing (Flusberg et al. 2010). Further platforms currently under development include DNA nanoball sequencing by Complete Genomics (http://www.completegenomics.com) and a number of approaches that detect specific features of the DNA bases such as their size, electrical or magnetic properties, or monitor characteristics of the reaction chemistry during base incorporation as the DNA strands are extended. Examples of such are developments by Ion Torrent (http://www.iontorrent.com), Oxford Nanopore Technologies (http://www.nanoporetech.com), ZS Genetics (http://www.zsgenetics.com), Reveo (http://www.reveo.com) and IBM (http://www.ibm.com) (Schadt et al. 2010; Xu et al. 2009).

An interesting and potentially very informative development is the analysis of single cells with regard to their RNA, protein and metabolite content. High-throughput sequencing has already been used to monitor the transcriptome of single mouse blastomeres, and with the advent of single-molecule sequencing and other technologies that further lower the detection limit, can be expected to become more broadly applicable (Tang et al. 2009; Wang and Bodovitz 2010).

Acknowledgements We gratefully acknowledge the financial support of the Swiss National Science Foundation, the Howard Hughes Medical Institute and the Canton of Bern. Annette Macleod, George Cross, Nicolai Siegel and Simon Hänni are thanked for their helpful comments on the manuscript.

References

Alsford S, Eckert S, Baker N, Glover L, Sanchez-Flores A, Leung KF, Turner DJ, Field MC, Berriman M, Horn D (2012) High-throughput decoding of antitrypanosomal drug efficacy and resistance. Nature 482(7384):232–236. doi: 10.1038/nature10771

Ashibe B, Hirai T, Higashi K, Sekimizu K, Motojima K (2007) Dual subcellular localization in the endoplasmic reticulum and peroxisomes and a vital role in protecting against oxidative stress of fatty aldehyde dehydrogenase are achieved by alternative splicing. J Biol Chem 282:20763–20773

Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrancois P, Struhl K, Gerstein M, Snyder M (2009) Mapping accessible chromatin regions using Sono-Seq. Proc Natl Acad Sci USA 106:14926–14931

Belli SI, Monnerat S, Schaff C, Masina S, Noll T, Myler PJ, Stuart K, Fasel N (2003) Sense and antisense transcripts in the histone H1 (HIS-1) locus of *Leishmania major*. Int J Parasitol 33:965–975

Benabdellah K, Gonzalez-Rey E, Gonzalez A (2007) Alternative *trans*-splicing of the *Trypanosoma cruzi* LYT1 gene transcript results in compartmental and functional switch for the encoded protein. Mol Microbiol 65:1559–1567

Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL (2005) Messenger RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol 143:125–134

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B et al (2005) The genome of the African Trypanosome *Trypanosoma brucei*. Science 309:416–422

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132:311–322

Brems S, Guilbride DL, Gundlesdodjir-Planck D, Busold C, Luu VD, Schanne M, Hoheisel J, Clayton C (2005) The transcriptomes of *Trypanosoma brucei* Lister 427 and TREU927 bloodstream and procyclic trypomastigotes. Mol Biochem Parasitol 139:163–172

Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E et al (2009) Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res 19:1044–1056

Bütikofer P, Serricchio M (2012) An essential bacterial-type cardiolipin synthase mediates cardiolipin formation in a eukaryote. Proc Natl Acad Sci USA, epub ahead of print

Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156:93–101

Cliffe LJ, Siegel TN, Marshall M, Cross GA, Sabatini R (2010) Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*. Nucleic Acids Res 38:3923–3935

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature 452:215–219

Costa V, Angelini C, De Feis I, Ciccodicola A (2010) Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol 2010:853916

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

Degrave WM, Melville S, Ivens A, Aslett M (2001) Parasite genome initiatives. Int J Parasitol 31:532–536

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM et al (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26:779–785

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138

El-Sayed NM, Alarcon CM, Beck JC, Sheffield VC, Donelson JE (1995) cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. Mol Biochem Parasitol 73:75–90

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, Ghedin E, Worthey EA, Delcher AL, Blandin G et al (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309:409–415

Erondu NE, Donelson JE (1992) Differential expression of two mRNAs from a single gene encoding an HMG1-like DNA binding protein of African trypanosomes. Mol Biochem Parasitol 51:111–118

Estevez AM (2008) The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. Nucleic Acids Res 36:4573–4586

Fernández-Moya SM, Estévez AM (2010) Posttranscriptional control and the role of RNA-binding proteins in gene regulation in trypanosomatid protozoan parasites. Wiley Interdiscip Rev RNA 1:34–46

Figueiredo LM, Cross GA (2010) Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. Eukaryot Cell 9:148–154

Flück C, Salomone JY, Kurath U, Roditi I (2003) Cycloheximide-mediated accumulation of transcripts from a procyclin expression site depends on the intergenic region. Mol Biochem Parasitol 127:93–97

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7:461–465

Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, Menzel C, Chen W, Li Y, Zeng R et al (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics 10:161

Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, Jovanovich SB, Nelson JR, Schloss JA, Schwartz DC et al (2009) The challenges of sequencing by synthesis. Nat Biotechnol 27:1013–1023

Furger A, Schurch N, Kurath U, Roditi I (1997) Elements in the 3′ untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. Mol Cell Biol 17:4372–4380

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17:877–885

Goren A, Ozsolak F, Shoresh N, Ku M, Adli M, Hart C, Gymrek M, Zuk O, Regev A, Milos PM et al (2010) Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. Nat Methods 7:47–49

Haanstra JR, Stewart M, Luu VD, van Tuijl A, Westerhoff HV, Clayton C, Bakker BM (2008) Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form *Trypanosoma brucei*. J Biol Chem 283:2495–2507

Haenni S, Studer E, Burkard GS, Roditi I (2009) Bidirectional silencing of RNA polymerase I transcription by a strand switch region in *Trypanosoma brucei*. Nucleic Acids Res 37:5007–5018

Haile S, Papadopoulou B (2007) Developmental regulation of gene expression in trypanosomatid parasitic protozoa. Curr Opin Microbiol 10:569–577

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW et al (2008) Single-molecule DNA sequencing of a viral genome. Science 320:106–109

Hehl A, Vassella E, Braun R, Roditi I (1994) A conserved stem-loop structure in the 3′ untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. Proc Natl Acad Sci USA 91:370–374

Helm JR, Wilson ME, Donelson JE (2008) Different trans RNA splicing events in bloodstream and procyclic *Trypanosoma brucei*. Mol Biochem Parasitol 159:134–137

Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS et al (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods 6:283–289

Huang J, Van der Ploeg LH (1991) Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3′ splice acceptor site. EMBO J 10:3877–3885

Hug M, Hotz HR, Hartmann C, Clayton C (1994) Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. Mol Cell Biol 14:7428–7435

Ingolia NT (2010) Genome-wide translational profiling by ribosome footprinting. Methods Enzymol 470:119–142

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream M-A, Adlem E, Aert R et al (2005) The genome of the kinetoplastid parasite, *Leishmania major*. Science 309:436–442

Jensen BC, Sivam D, Kifer CT, Myler PJ, Parsons M (2009) Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. BMC Genomics 10:482

Kabani S, Fenn K, Ross A, Ivens A, Smith TK, Ghazal P, Matthews K (2009) Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*. BMC Genomics 10:427

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J et al (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458:362–366

Koenig-Martin E, Yamage M, Roditi I (1992) A procyclin-associated gene in *Trypanosoma brucei* encodes a polypeptide related to ESAG 6 and 7 proteins. Mol Biochem Parasitol 55:135–145

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 6(9):pii: e1001090

König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17:909–915

Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J, Turner SW (2010) Real-time DNA sequencing from single polymerase molecules. Methods Enzymol 472:431–455

Kramer S, Carrington M (2010) *Trans*-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. Trends Parasitol 27:23–30

Leinonen R, Sugawara H, Shumway M (2010) The sequence read archive. Nucleic Acids Res 39 (Database issue):D19–D21

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469

Liniger M, Bodenmuller K, Pays E, Gallati S, Roditi I (2001) Overlapping sense and antisense transcription units in *Trypanosoma brucei*. Mol Microbiol 40:869–878

Lipson D, Raz T, Kieu A, Jones DR, Giladi E, Thayer E, Thompson JF, Letovsky S, Milos P, Causey M (2009) Quantification of the yeast transcriptome by single-molecule sequencing. Nat Biotechnol 27:652–658

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133:523–536

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322

Lopez V, Kelleher SL (2009) Zinc transporter-2 (ZnT2) variants are localized to distinct subcellular compartments and functionally transport zinc. Biochem J 422:43–52

Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, Falcone FH, Gavrilescu C, Montgomery JL, Santori MI et al (2000) A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. RNA 6:163–169

Mao Y, Najafabadi HS, Salavati R (2009) Genome-wide computational identification of functional RNA elements in *Trypanosoma brucei*. BMC Genomics 10:355

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402

Marguerat S, Bahler J (2010) RNA-seq: from technology to biology. Cell Mol Life Sci 67:569–579

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–1517

Matthews KR, Tschudi C, 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

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46

Militello KT, Wang P, Jayakar SK, Pietrasik RL, Dupont CD, Dodd K, King AM, Valenti PR (2008) African trypanosomes contain 5-methylcytosine in nuclear DNA. Eukaryot Cell 7:2012–2016

Minning TA, Weatherly DB, Atwood J 3rd, Orlando R, Tarleton RL (2009) The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. BMC Genomics 10:370

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. PLoS Pathog 6(8):pii: e1001037

Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, Lemieux J, Barrell B, Pain A, Berriman M et al (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. Mol Microbiol 76:12–24

Ouellette M, Papadopoulou B (2009) Coordinated gene expression by post-transcriptional regulons in African trypanosomes. J Biol 8:100

Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM (2009) Direct RNA sequencing. Nature 461:814–818

Ozsolak F, Goren A, Gymrek M, Guttman M, Regev A, Bernstein BE, Milos PM (2010a) Digital transcriptome profiling from attomole-level RNA samples. Genome Res 20:519–525

Ozsolak F, Ting DT, Wittner BS, Brannigan BW, Paul S, Bardeesy N, Ramaswamy S, Milos PM, Haber DA (2010b) Amplification-free digital gene expression profiling from minute cell quantities. Nat Methods 7:619–621

Panigrahi AK, Ogata Y, Zikova A, Anupama A, Dalley RA, Acestor N, Myler PJ, Stuart KD (2009) A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. Proteomics 9:434–450

Pasion SG, Hines JC, Ou X, Mahmood R, Ray DS (1996) Sequences within the 5′ untranslated region regulate the levels of a kinetoplast DNA topoisomerase mRNA during the cell cycle. Mol Cell Biol 16:6724–6735

Ponts N, Harris EY, Lonardi S, Le Roch KG (2010a) Nucleosome occupancy at transcription start sites in the human malaria parasite: a hard-wired evolution of virulence? Infect Genet Evol 11 (4):716–724

Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG (2010b) Nucleosome landscape and control of transcription in the human malaria parasite. Genome Res 20:228–238

Pushkarev D, Neff NF, Quake SR (2009) Single-molecule sequencing of an individual human genome. Nat Biotechnol 27:847–852

Rettig J, Wang Y, Schneider A, Ochsenreiter T (2012) Dual targeting of isoleucyl-tRNA synthetase in Trypanosoma brucei is mediated through alternative trans-splicing. Nucleic Acids Res 4(3):1299–1306

Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics 10:495

Respuela P, Ferella M, Rada-Iglesias A, Aslund L (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. J Biol Chem 283:15884–15892

Rochette A, Raymond F, Ubeda JM, Smith M, Messier N, Boisvert S, Rigault P, Corbeil J, Ouellette M, Papadopoulou B (2008) Genome-wide gene expression profiling analysis of *Leishmania major* and *Leishmania infantum* developmental stages reveals substantial differences between the two species. BMC Genomics 9:255

Rojas MV, Galanti N (1990) DNA methylation in *Trypanosoma cruzi*. FEBS Lett 263:113–116
Rout MP, Field MC (2001) Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component. J Biol Chem 276:38261–38271
Rudenko G, Blundell PA, Taylor MC, Kieft R, Borst P (1994) VSG gene expression site control in insect form *Trypanosoma brucei*. EMBO J 13:5470–5482
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19:R227–R240
Schlimme W, Burri M, Bender K, Betschart B, Hecker H (1993) *Trypanosoma brucei brucei*: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. Parasitology 107(Pt 3):237–247
Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell 132:887–898
Schürch N, Hehl A, Vassella E, Braun R, Roditi I (1994) Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. Mol Cell Biol 14:3668–3675
Schürch N, Furger A, Kurath U, Roditi I (1997) Contributions of the procyclin 3′ untranslated region and coding region to the regulation of expression in bloodstream forms of *Trypanosoma brucei*. Mol Biochem Parasitol 89:109–121
Shaked H, Wachtel C, Tulinski P, Yahia NH, Barda O, Darzynkiewicz E, Nilsen TW, Michaeli S (2010) Establishment of an in vitro trans-splicing system in *Trypanosoma brucei* that requires endogenous spliced leader RNA. Nucleic Acids Res 38:e114
Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol 6:e65
Siegel TN, Tan KS, Cross GA (2005) Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. Mol Cell Biol 25:9586–9594
Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GA (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. Genes Dev 23:1063–1076
Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GA (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res 38:4946–4957
Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc 2010 (2):pdb prot5384
Srividya G, Duncan R, Sharma P, Raju BV, Nakhasi HL, Salotra P (2007) Transcriptome analysis during the process of in vitro differentiation of *Leishmania donovani* using genomic microarrays. Parasitology 134:1527–1539
Stanne TM, Rudenko G (2010) Active VSG expression sites in *Trypanosoma brucei* are depleted of nucleosomes. Eukaryot Cell 9:136–147
Stern MZ, Gupta SK, Salmon-Divon M, Haham T, Barda O, Levi S, Wachtel C, Nilsen TW, Michaeli S (2009) Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. RNA 15:648–665
Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382
Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ (2009) Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. BMC Genomics 10:152
Thompson JF, Steinmann KE (2010) Single molecule sequencing with a HeliScope genetic analysis system. Curr Protoc Mol Biol Chapter 7:Unit7.10

Urwyler S, Vassella E, Van Den Abbeele J, Renggli CK, Blundell P, Barry JD, Roditi I (2005) Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. PLoS Pathog 1:e22

Vassella E, Braun R, Roditi I (1994) Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site: a dual role for polypyrimidine tracts in trypanosomes? Nucleic Acids Res 22:1359–1364

Veitch NJ, Johnson PC, Trivedi U, Terry S, Wildridge D, MacLeod A (2010) Digital gene expression analysis of two life cycle stages of the human-infective parasite, *Trypanosoma brucei gambiense* reveals differentially expressed clusters of co-regulated genes. BMC Genomics 11:124

Walrad P, Paterou A, Acosta-Serrano A, Matthews KR (2009) Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. PLoS Pathog 5:e1000317

Wang D, Bodovitz S (2010) Single cell analysis: the new frontier in 'omics'. Trends Biotechnol 28:281–290

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48:249–257

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453:1239–1243

Wirtz E, Clayton C (1995) Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. Science 268:1179–1183

Wright JR, Siegel TN, Cross GA (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. Mol Biochem Parasitol 172:141–144

Xu M, Fujita D, Hanagata N (2009) Perspectives and challenges of emerging single-molecule DNA sequencing technologies. Small 5:2638–2649

Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 12:7–18

Index

A
A and P sites, 225
Abnormal electrophoretic mobility, 218, 219
Abundant trypanosome snoRNAs, 130
African trypanosome, 149
AGO. *See* Argonaute (AGO)
Alternative editing, 150, 220
Alternatively spliced transcripts, 245
Alternative reading frames, 220
Alternative splicing, 69, 247
Amastigotes, 197–199, 202, 204–206
Amastin, 202
Antisense transcription, 252
Apocytochrome b (Cyb), 218
APSF, 225
Argonaute (AGO), 168
Argonaute slicer, 174
A6 subunit, 219
A/U heteropolymer, 228
Axenic transformation, 204–205

B
Base J, 257
Base methylation, 39–40
BDF3. *See* Bromodomain containing factor 3 (BDF3)
B double prime 1 (BDP1), 5, 19
BDP1. *See* B double prime 1 (BDP1)
β–elimination, 178
Bicistronic, 249
Bidirectional transcription, 252
Biogenesis, 65
Bloodstream form, 15
Bromodomain containing factor 3 (BDF3), 4, 12, 13

C
Cajal bodies, 136
Cap-4, 187, 192
Cap 1 2'-*O*-ribose methylation, 34–36
CBP20 protein, 40
CCCH zinc finger, 91–92
cDNA libraries, 242
C/D snoRNA, 130
C/D snoRNA and parasite development, 132
Cell cycle, 91
Cell-free system, 206–207
ChIP-seq, 12, 256
Chloramphenicol, 217
Chromatin, 12, 13, 17
Chromosome, 6, 15
Chromosome segregation, 180
CIR147 repeat, 172
cis splicing, 50
Class I transcription factor A (CITFA), 5, 8, 15–18
COI, 219
COII, 219
Cosuppression, 178
Cotranscriptional processing, 31–33
Cox11, 227
Crithidia
  C. fasciculata, 217
  C. luciliae, 217
  C. oncopelti, 221
Cryo-electron microscopy (cryo-EM), 152
Cycloheximide, 217
Cytochrome b (Cyb), 219
Cytochrome bc1, 219
Cytochrome c oxidase, 219
Cytosolic trafficking model, 36

A. Bindereif (ed.), *RNA Metabolism in Trypanosomes*, Nucleic Acids and Molecular Biology 28, DOI 10.1007/978-3-642-28687-2,

D
DAC3, 17
DEAD-box proteins, 200
Deadenylase, 83
Deaminase, 106–107
Decapping, 83–84
Developmental regulation, 80
Dicer, 168
Dicer-like protein(s), 170
Differential splicing, 253
Differentiation, 80, 206
Digitonin-permeabilized cells, 217
Diplonema papillatum, 30
Discovery of mitochondrial protein synthesis, 216
DNA
  maxicircles, 150
  minicircle, 150
DnaQ superfamily, 180
dsRNA-binding protein (dsRBP), 168
dsRNA intermediates, 171
Dynein, 17, 18

E
4E-BP, 189, 198
Editing, 99, 100, 105–109
Editosome, 149
eEF2, 203, 204
eEF1A, 203
eEF1B, 203
eIF2α, 191, 205
eIF4A, 200
eIF4E, 190
eIF4F, 187, 196, 199
eIF4G, 189, 198–200
Endogenous siRNAs (esiRNAs), 178
3′-End tails, 228
EP procyclin, 88
esiRNAs. *See* Endogenous siRNAs (esiRNAs)
E site, 225
Exoribonuclease, 84
Exosome, 84–85
Exportin 1 (Xpo1), 36
Expression sites, 254, 257

F
Facilitates chromatin transcription (FACT), 13, 14, 17
FeS, 109
F1F0 ATPase, 219

G
Genomic organization
  of eukaryotic snoRNAs, 134
  of kinetoplastid SL RNA, 30–31
  of trypanosome snoRNAs, 134
Guide RNAs, 149

H
H/ACA binding proteins, 133
H/ACA snoRNA, 132
HAT1, 17
Head-to-tail tandem arrays, 30, 31
Heat shock proteins, 190
HEN1, 178
Histone, 12, 13, 18
HSP70, 190–191, 201
HSP83, 201–202
Human AGO2, 176

I
Identification of mitochondrial ribosomes, 216
Illumina sequencing, 240
Importin α, 36–37
Inducible RNAi, 181
Ingi, 172, 178, 181
Initiation codon recognition, 228
Inosine, 105, 107
Insertase Oxa1, 227
Intersubunit space in Lmr, 223, 225
In vitro translation, 206–207
ISWI, 17

K
kDNA, 150
Kinase/cyclin-dependent kinase (CDK)/CRK, 7, 10, 11
Kinetoplastids, 29–42
  mitochondrial genome, 218
  protozoa, 149
Kozak sequence, 202

L
Large (12S) rRNAs, 222
Leish4E-IP, 187, 198
LeishIF4A-1, 199
LeishIF4E-1, 196–198
LeishIF4E-2, 96, 198
LeishIF4E-3, 187–198

LeishIF4E-4, 196–197
LeishIF4G-3, 196–199
Leishmania, 191–207
  L. mitoribosome, 222
  L. tarentolae, 31, 33, 39, 218
  L. Viannia braziliensis, 169
Long poly(A/U) tails, 229
LSm ring, 61

M
$m^{6,6}A$ dimethyltransferases, 40
Mdm38/LETM1, 227
Mediator (MED-T), 8, 11
Megabase chromosomes, 244
5-Methylcytosine (5MC), 258
m^7G (Cap 0) formation, 31–33
m^7GTP, 192, 196
Microarray, 12
MID domain, 175
MIF4G domain, 196–197
Misannotated genes, 247
Mitochondria, 149
Mitochondrial genome, 150
Mitochondrially encoded polypeptides, 218
mRNA
  channel, 225
  half-life, 81, 82
  splicing, 52
  turnover, 79–92
MRP, 127
m^3U methylation, 40

N
Next generation sequencing, 238
Noncanonical codons, 226
Noncoding RNAs, 150, 168
Novel transcripts (NTs), 247, 250, 251
Nuclease, 102–104, 114
Nucleolus, 124

O
OB-fold, 153

P
PABP. *See* Poly(A) binding protein (PABP)
PAG. *See* Procyclin-associated genes (PAG)
PAS. *See* Polypeptide-accessible site (PAS)
PAZ domains, 175
P-bodies, 205–206
Pentatricopeptide repeat (PPR)-containing proteins, 221
Peptidyl-transferase center, 223, 226
PIWI domain, 175
Plypyrimidines, 201
Polyadenylation, 248, 254
Poly(A) binding protein (PABP), 198–201
Polypeptide-accessible site (PAS), 227
Polypeptide exit site (PES), 227
Polypyrimidine tracts, 246, 248
Polyribosomes, 174, 177
PPR proteins, 229
Pre-edited mRNAs, 160
PRMTs. *See* Protein arginine methyltransferases (PRMTs)
Processing body, 86–87
Procyclic form, 14, 17
Procyclin, 3, 8, 14, 15, 17
Procyclin-associated genes (PAG), 248, 252, 253
Promoter, 1, 2, 6–20
Protein arginine methyltransferases (PRMTs), 177
Protein complex, 3, 7, 11, 18
Protein factors involved in rRNA processing, 137
Protein synthesis, 187–207
Pseudouridine (Ψ), 34
Pseudouridylation, 34
PsiScan, 133
Pumilio, 90–91
Puromycin, 217, 220

R
RdRP. *See* RNA dependent RNA polymerase (RdRP)
Resistance to chloramphenicol, 218
Retroposon, 168
RGG domain, 177
2′-*O*-Ribose methylations, 192
Ribosomal protein, 219
Ribosomal protein S12, 223
Ribosomal RNA (rRNA), 1, 2, 9, 16, 18, 19
Ribosome recruitment mechanism, 227
RISC. *See* RNA-induced silencing complex (RISC)
RISC activation, 179
RNA
  editing, 149
  interaction, 159

RNA (*cont.*)
interference, 168
processing, 149
RNA dependent RNA polymerase (RdRP), 168
RNAi, 168
deficient, 180
factors, 182
technology, 181
RNA-induced silencing complex (RISC), 169
RNA polymerase (RNA pol)/CTD/CTD phosphorylation, 1–20
RNA recognition motif (RRM), 89
RNase H, 175
RNase H-like domain, 168
RNase III domains, 170, 171
RNase III-family, 168
RNA sequencing (RNA-Seq), 79, 237, 243
RRM. *See* RNA recognition motif (RRM)
rRNA processing in yeast, 124

S
Sand flies, 204
α–Sarcin/ricin stem-loop, 222, 226
Silencing, 196–197
Silent regions, 252
Single hairpin AGA snoRNAs, 133
Single-particle 3D cryo-EM, 222
siRNAs, 168
SLA1, 136
SLACS, 178, 181
SLACS retroposon, 172
SL RNA/SL RNA gene, 2, 3, 6–11, 13–15, 18, 19, 59, 192
40S LSU, 222
SLT. *See* Spliced leader trapping (SLT)
Small (9S), 222
Small nuclear ribonucleoproteins, 53
Small nuclear RNA/snRNA/snRNA gene, 6–8, 11, 18, 19, 53
Sm core variation, 62
Sm-dependent Cap 2′-*O*-ribose methylation, 38–39
SMN. *See* Survival of motor neurons (SMN)
Sm proteins, 36–37, 57
SNAPc, 4, 7–9, 12, 15, 18, 19
snoRNAi, 137, 173
snoRNAs, 249
conservation, 136
involved in processing, 125
polyadenylation, 135
processing, 135
snR30 in yeast, 126
snRNA incomplete 3' processing (SNIP), 39
snRNAs, 53
snRNPs, 53
Splice acceptor sites, 245, 246
Spliced leader RNA, 59
Spliced leader (SL) RNA, 29–42
Spliced leader trapping (SLT), 239, 241
Spliceosome, 53
Splice site, 246
Splice site detection, 244
Splicing regulation, 66
60S processing, 142
40S processing in yeast and trypanosomes, 138
30S SSU, 222
45S SSU, 221
Stage-specific, 253
Strand-specific RNA-Seq, 244
Strand switch region (SSR), 4, 12
Stress granules, 86–87, 205–206
Structure/crystal structure/EM structure/NMR structure, 2, 9–12, 14, 19
Survival of motor neurons (SMN), 65

T
TAP-tagged ribosomes, 230
TB11Cs2C1 snoRNA, 129
TB11Cs2C2 snoRNA, 129
TbIF4A-1, 200
TbIF4E-1, 197
TbIF4E-2, 197
TbIF4E-3, 197–198
TbIF4E-4, 196–197
TbMTr1, 34, 35
TbMTr2 and TbMTr3, 38
TBP-related factor 4 (TRF4)/TBP, 4, 5, 7–10, 12, 13, 19
TEM. *See* Transmission electron microscopy (TEM)
Terminator exonuclease, 172
TFIIA, 4, 7–10, 19
TFIIB, 4, 7–11, 19
TFIIE, 7, 10
TFIIF, 7, 20
TFIIH, 4, 7, 8, 10, 11, 20
TFIIIB, 5, 9, 19
TFIIS, 1, 13, 20
Toxoplasma gondii AGO, 177
Transcript abundance, 249
Transcription, 31–33, 191

initiation, 1, 2, 7–13, 17
start sites, 254
termination, 2
Transcription factor/Basal transcription factor (BTF), 1–20
Transcriptome, 237
Translation, 41, 187–207
cap-dependent, 188
cap-independent, 188–189
gene specific, 190
during heat shock, 190–191
Transmission electron microscopy (TEM), 152
Transposons, 168
Trans-splicing, 2, 6, 14, 40–41, 50, 241
tRNA/tRNA gene, 2, 3, 8, 12, 14, 18–20
export, 100, 101, 104, 108
import, 102, 104, 108–126
thiolation, 101, 105, 107–109, 111, 116
Trypanosoma, 191–207
T. brucei, 151, 169
Trypanosomatids, 187, 191–207
Trypanosome
C/D snoRNAs, 131
homologues to protein factors involved in ribosome biogenesis, 139–141
rRNA processing, 124, 128
snoRNA with unique processing functions, 129
sR30, 127
U3, 126
TudorSN, 177

U
U14 and TB11Cs2C1, 127
5′ and 3′ Untranslated regions (UTRs), 187, 197, 201–202, 238, 244, 248
U-nucleotide
deletion, 153
insertion, 153
U1 snRNP, 63
U2 snRNP, 60
U5 snRNP, 63
UTR lengths, 246
UTRs. *See* 5′ and 3′ Untranslated regions (UTRs)
U4/U6 di-snRNP, 61

V
Variant surface glycoprotein (VSG), 3, 8, 14–18
VSG expression site, 14