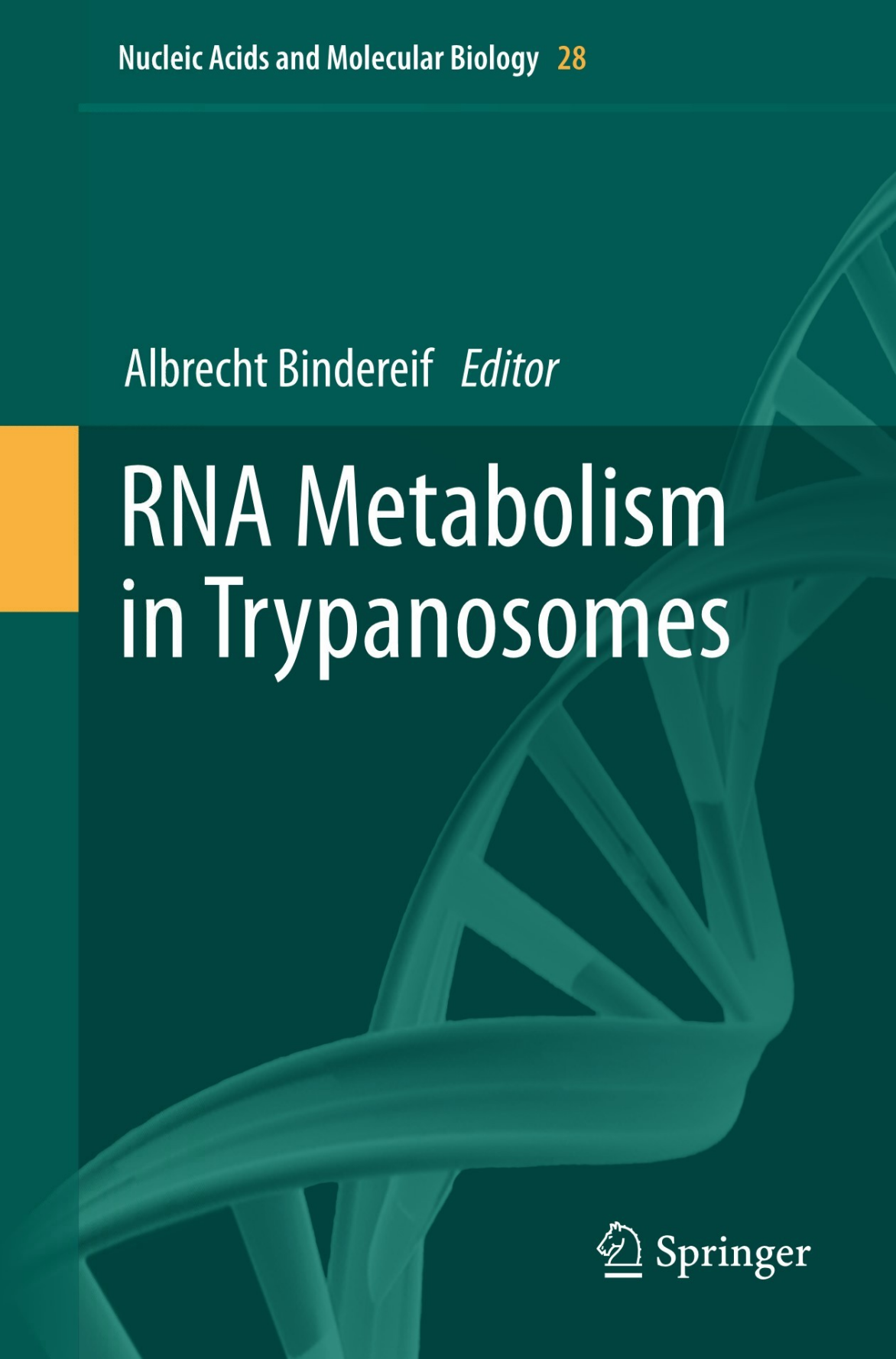
Nucleic Acids and Molecular Biology 28

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

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).

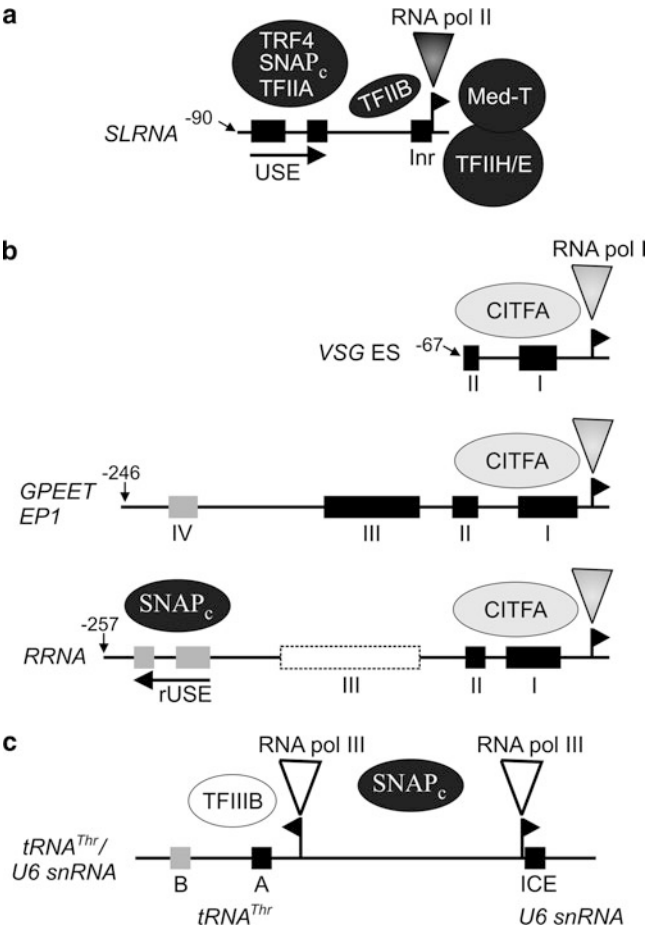


Fig. 1.1 Schematic depiction of *T. brucei* promoters and transcription factors. Transcription initiation sites (TISs) are represented by *flags* and essential promoter elements by *black* boxes. *Gray* boxes indicate promoter elements that are important only in vivo. Furthermore, RNA pols are drawn 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

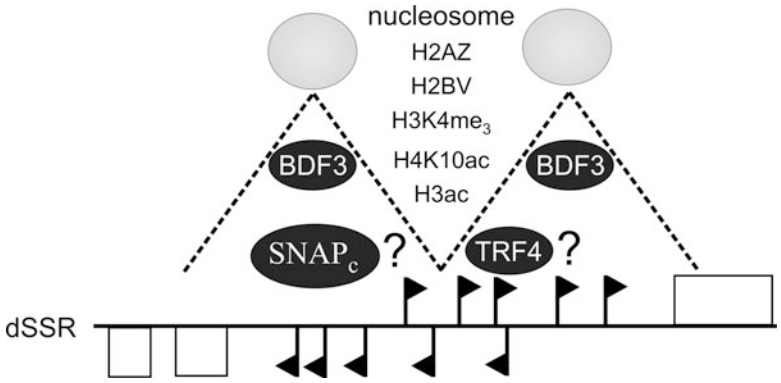


Fig. 1.2 Schematic drawing of a divergent strand switch region (dSSR). In ChIP-seq experiments, two broad peaks of open chromatin marks (*dashed lines*) are present within a dSSR, and there are multiple TISs for transcription in both directions over a broad region. Factor BDF3 appears to be linked to 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,