Molecular & Biochemical Parasitology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

Review

Networks of gene expression regulation in *Trypanosoma brucei*

C.E. Clayton*

Universität Heidelberg, Zentrum für Molekulare Biologie der Universität Heidelberg, DKFZ-ZMBH Alliance, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Trypanosoma
mRNA decay
mRNA processing
Splicing
Translation

ABSTRACT

Regulation of gene expression in Kinetoplastids relies mainly on post-transcriptional mechanisms. Recent high-throughput analyses, combined with mathematical modelling, have demonstrated possibilities for transcript-specific regulation at every stage: *trans* splicing, polyadenylation, translation, and degradation of both the precursor and the mature mRNA. Different mRNA degradation pathways result in different types of degradation kinetics.

The original idea that the fate of an mRNA – or even just its degradation kinetics – can be defined by a single "regulatory element" is an over-simplification. It is now clear that every mRNA can bind many different proteins, some of which may compete with each other. Superimposed upon this complexity are the interactions of those proteins with effectors of gene expression. The amount of protein that is made from a gene is therefore determined by a complex network of interactions.

Contents

1. Trypanosome transcription

Gene expression in Kinetoplastids is remarkable because transcription is polycistronic (reviewed in [1]). The 5′ end of each individual mRNA is generated by *trans* splicing of a capped "spliced leader" (SL, 39nt in *Trypanosoma brucei*) at the 5′ end [2]. The *trans* spliceosome is linked, in a way that is not yet understood, to the polyadenylation machinery, such that each splicing reaction directs – and is coupled to – polyadenylation of the preceding mRNA [3]. For individual mRNAs that are synthesised by RNA polymerase II, this set-up results in a complete lack of transcriptional regulation: almost the entire burden of determining how much of each mRNA is present in the cell falls on post-transcriptional processes. Possible points of regulation are *trans* splicing, polyadenylation, degradation of the mRNA within the nucleus, export from the nucleus, and degradation of the mRNA in the cytosol. The amount of protein that is made might also be regulated though control of translation initiation or elongation.

T. brucei grows in the blood and tissue fluids of mammals, and in the midgut, foregut, proventriculus and salivary glands of Tsetse flies [4]. In the mammal, the bloodstream-form trypomastigotes rely on substrate-level phosphorylation for ATP generation, with glucose as the main energy source [5,6]; and upon antigenic variation of variant surface glycoprotein (VSG) for defence against the humoral immune system [7]. The procyclic trypomastigotes in the Tsetse midgut have a much more active mitochondrion [6], and have procyclins on the surface [8]; while epimastigotes in the proventriculus and salivary glands have a third type of

* Tel.: +49 6221 546876; fax: +49 6221 545894.
E-mail address: cclayton@zmbh.uni-heidelberg.de

http://dx.doi.org/10.1016/j.molbiopara.2014.06.005
0166-6851/

Please cite this article in press as: Clayton CE. Networks of gene expression regulation in *Trypanosoma brucei*. Mol Biochem Parasitol (2014), http://dx.doi.org/10.1016/j.molbiopara.2014.06.005

surface protein called BARP [9]. Finally, there are non-dividing forms called metacyclics, with surface VSG, in the salivary glands. A bloodstream-form trypanosome (growing at 37 °C, with a division time of 6–8 h) has about 20,000 mRNAs [10]. (There are about 7000 different open reading frames (ORFs) in the genome [11].) Procyclic forms are grown at 27 °C; they divide slower than bloodstream forms, are about twice the size, and correspondingly have twice as many mRNA molecules [10].

The transcriptomes of bloodstream and procyclic forms have been analysed by several laboratories, initially with the use of microarrays, and later by high-throughout sequencing of cDNA [11–14]. In bloodstream forms, the mRNAs from most genes are present at 1–2 copies per cell, but for some, over 200 are present [14,15]. Genes in the latter category are always present in multiple copies, often as tandem repeats. It appears that the rate of constitutive polymerase II transcription is simply inadequate to supply more than about 50 mRNAs per ORF.

An alternative mechanism to escape the limitations of the constitutive pol II transcription is unique to salivarian trypanosomes: the genes encoding the variant surface glycoprotein (VSG) and the procyclins are transcribed by RNA polymerase I. RNA polymerase I has about ten-fold more active initiation than for polymerase II, and may also show faster elongation [10,16]. In addition, polymerase I initiation can be regulated by epigenetic means, allowing both stage-specific transcription and the exclusive expression of one VSG at a time – the prerequisite for antigenic variation [7].

2. The contributions of nuclear processes to mRNA regulation

Fig. 1 illustrates transcription and processing for an imaginary 10 kb region, in the form of a time-lapse image with the position of RNA pol II indicated at 1-min intervals. The processes that determine the amount of mature mRNA that is made and can be exported to the cytoplasm are as follows:

1) Transcription. The transcription rate has not been measured. We do know, though, that the parasite has to be making enough mRNA all the time in order to both balance mRNA degradation, and to double the amount of mRNA before the cell divides. This means that we can calculate the average rate of RNA synthesis. If we assume that the rate of transcription elongation is similar to average estimates from Opisthokonts – about 20 nt/s – we can calculate how often initiation has to happen. This gives an average distance between pol II elongation complexes of 40 kb [10].
2) *Trans* splicing. Careful measurements on the *PGKB* mRNA in vivo [10], and the *HSP70* [17] and tubulin mRNAs [18] in an in vitro system, indicated that splicing happens 1–2 min after the splice acceptor site was synthesised. For technical and theoretical reasons, it is not possible to measure splicing rates across the entire transcriptome, but results from measurements for a few thousand genes suggested that *trans* splicing half-times might be between 1 and 5 min [14] (which translates into average splicing times of 1.5–7 min). Assuming the elongation rate of 20 nt/s, his would mean that a splice acceptor site will be processed when the pol II complex that made it is 1–6 kb downstream. In the figure, RNAs A. C and D are all *trans* spliced about 1 min after their splice acceptor sites were synthesised, while RNA B is has a poor splicing signal so its splicing is delayed by a further 3 min.
3) Polyadenylation. Since polyadenylation is coupled to *trans* splicing of the next mRNA downstream, it is expected to show the same kinetics [18] – perhaps with a slight delay. In Fig. 1, therefore, RNAs B, C and D are polyadenylated 1–2 min after the poly(A) site was made, but polyadenylation of RNA A is delayed.
4) Instinctively, one would think that if an mRNA were processed really slowly, there should be less of it than if it were processed fast. Indeed, if mRNA synthesis has only just been turned on, the processing rate has a critical influence of the rate at which the mRNA accumulates. However, once steady-state has been reached, the processing rate is irrelevant, since each mRNA always gets made eventually. Fig. 1 shows this: there is just as much of RNA B as there is of RNAs A, C and D. For mRNA processing to have any effect on steady-state abundances, processing has to be competing with degradation. Indeed, degradation does occur in the absence of processing. The only

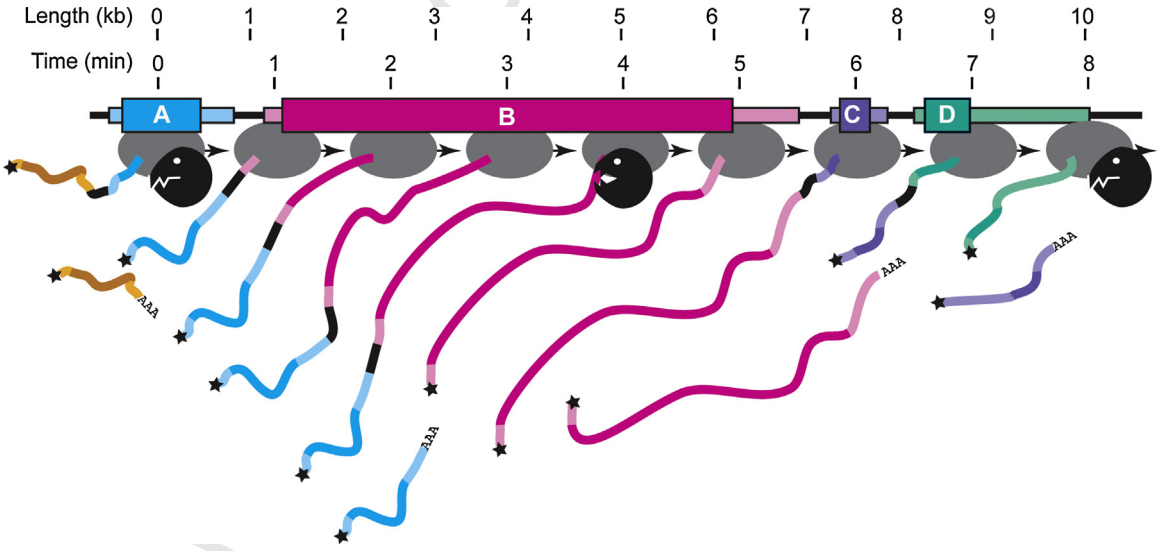

Fig. 1. Pathway of mRNA formation. Four genes – A (blue), B (pink), C (purple) and D (green) – are in a polycistronic transcription unit. They are preceded by another gene that is not shown – mRNA in orange. The coding regions are darker-coloured than the untranslated regions. The figure is a time-lapse snapshot of RNA polymerase II taken at 1-min intervals, assuming an elongation rate of about 20 nt/s. The black star is the spliced leader. In this hypothetical example, 5′-trans splicing of mRNAs A, C, and D happens after 1–2 min of splicing acceptor synthesis, but splicing of B is slower (5 min) – for example, it might have a weak polypyrimidine tract. Note that the slow splicing of "B" does not affect the abundance of RNAs "A" or "B", but causes slow polyadenylation of "A" and thus retention of mRNA "A" at the chromatin. The black "chompers" represent co-transcriptional degradation; the one at the 5 kb mark is currently active. RNAs "A" and "B" are much more likely to be "chomped" than RNAs "C and "D" because they are attached to the chromatin for longer.

Please cite this article in press as: Clayton CE. Networks of gene expression regulation in *Trypanosoma brucei*. Mol Biochem Parasitol (2014), http://dx.doi.org/10.1016/j.molbiopara.2014.06.005