The Nucleus and Gene Expression

A Focus Topic at BioScience2005, held at SECC Glasgow, U.K., 17–21 July 2005. Edited by W. Bickmore (Edinburgh, U.K.), R. Borts (Leicester, U.K.), J. Cáceres (Edinburgh, U.K.), A. Maxwell (John Innes Centre, Norwich, U.K.), S. Newbury (Newcastle upon Tyne, U.K.), D. Wigley (Cancer Research London, U.K.) and A. Willis (Nottingham, U.K.).

An *in silico* analysis of trypanosomatid RNA polymerases: insights into their unusual transcription

S. Kelly, B. Wickstead and K. Gull[1]

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, U.K.

Abstract

African trypanosomes employ both Pol I (RNA polymerase I) and Pol II to transcribe protein-coding genes in large polycistronic units of up to 50 genes. Subsequent processing produces mature capped mRNAs. Evidence suggests that regulation of gene expression is primarily exerted post-transcriptionally. Here, we use the recently completed genome sequences of three trypanosomatids, *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*, in an *in silico* analysis of their fundamental RNA polymerase complexes. The core complement of Pol II subunits, including those that are shared with Pol I and Pol III are present. However, both Pol I and Pol III complexes are missing members of the rpoE-rpoF subunit groups. Out of the five shared subunits, both RPB5 and RPB6 have two isoforms in the three trypanosomes. One represents the canonical polymerase subunit and the other differs by insertion or deletion of stretches of charged residues. We propose that these alternative isoforms function in distinct polymerase complexes, and may influence recruitment of the trypanosome RPB4–RPB7 heterodimer.

Introduction

African trypanosomes are transmitted between mammalian hosts by tsetse flies. This mode of transmission is relatively inefficient and requires the development of a chronic infection to give the parasites sufficient opportunity for reuptake by the tsetse vector. Therefore the trypanosomes must survive and grow for long periods of time inside their mammalian hosts. As the entire life cycle of the African trypanosome is represented by extracellular stages, the parasite must employ a robust strategy to evade host immune responses. Antigenic variation, the method by which the bloodstream form trypanosome changes its VSG (variant surface glycoprotein) coat, is a paradigm for evasion of host immune responses in many parasitic organisms [1].

In *Trypanosoma brucei*, the active VSG gene is transcribed from a long polycistronic transcription unit termed an expression site (ES). These ESs are located at telomeres and there are approx. 20 such sites, each capable of expressing their encoded VSG. The availability of several ESs adds a degree of complexity to antigenic variation since at any time only one ES is active and only one VSG can be transcribed. Unusually, VSG genes are transcribed by Pol I (RNA polymerase I). This is the only documented example of protein-coding gene transcription by Pol I and it occurs in a nucleoplasmic location separate from the nucleolus [2]. This presents the cell with the challenge of linking the Pol II mRNA processing machinery with the Pol I complex and raises the question as to whether there are differences in the core transcription machinery in *T. brucei*.

Eukaryotic RNA polymerases consist of a large number of proteins of which only a subset can be considered core components fundamental to transcription. Although most of what is known about eukaryotic transcription is confined to yeast and human cells, core polymerase subunits can be defined based on sequence, structural and (if possible) functional conservation across Archaea and Eukarya. Here, we use iterative profile-based searching methods and the recently completed genome of *T. brucei* along with those of two related trypanosomatids, *Trypanosoma cruzi* and *Leishmania major*, to

Key words: antigenic variation, *in silico* analysis, RNA polymerase, transcription, trypanosomatid, variant surface glycoprotein (VSG).

Abbreviations used: ES, expression site; Pol I, RNA polymerase I; VSG, variant surface glycoprotein.

[1]To whom correspondence should be addressed (email keith.gull@path.ox.ac.uk).

Table 1 | **Core polymerase subunits identified in *S. cerevisiae*, *Homo sapiens* and the three trypanosomatids *T. brucei*, *T. cruzi* and *L. major***

Subunits are arranged according to size, group and polymerase class. ✓, one homologue present; ✓✓, two homologues present; ×, homologue not detected. Key features are shaded grey.

	S. cerevisiae	*H. sapiens*	*T. brucei*	*T. cruzi*	*L. major*
RPA190	✓	✓	✓	✓	✓
RPB1	✓	✓	✓✓	✓	✓
RPC160	✓	✓	✓	✓	✓
RPA135	✓	✓	✓	✓	✓
RPB2	✓	✓	✓	✓	✓
RPC128	✓	✓	✓	✓	✓
RPC40	✓	✓✓	✓	✓	✓
RPB3	✓	✓	✓	✓	✓
RPA14	✓	×	×	×	×
RPB4	✓	✓	✓	✓	✓
RPC17	✓	✓	×	×	×
RPB5	✓	✓	✓	✓	✓
RPB5z	×	×	✓	✓	✓
RPB6	✓	✓	✓	✓	✓
RPB6z	×	×	✓	✓	✓
RPA43	✓	✓	×	×	×
RPB7	✓	✓	✓	✓	✓
RPC25	✓	✓	✓	✓	✓
RPB8	✓	✓	✓	✓	✓
RPA12	✓	✓	✓	✓	✓
RPB9	✓	✓	✓	✓	✓
RPC11	✓	✓	✓	✓	✓
RPB10	✓	✓	✓	✓	✓
RPC19	✓	✓	✓	✓	✓
RPB11	✓	✓✓	✓	✓	✓
RPB12	✓	✓	✓	✓	✓

begin to decipher the proteome behind this unusual transcription.

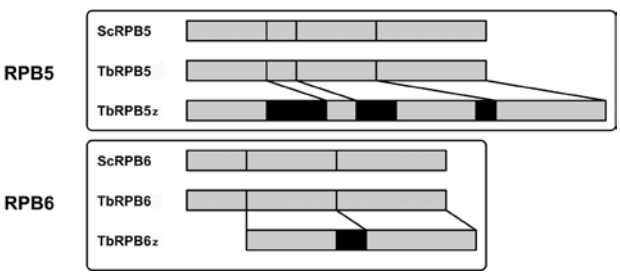
Results and discussion

Iterative profile-based searching techniques are a powerful way of identifying homologues in diverse organisms. Here, simple BLAST searches [3] using a single yeast or archaeal polymerase subunit were performed. Several sequences with low expectation values from closely related organisms were easily identified. These sequences were then aligned using ClustalX [4] and manually edited before being used to generate hidden Markov models [5]. These models were then used to search a large protein sequence database from a diverse set of organisms. The resultant hits were aligned to the model and unconvincing low scoring matches were discarded. A new alignment containing the new sequence matches was then used to regenerate the hidden Markov model and the database was re-searched. This process was repeated until no further satisfactory matches could be detected. These refined models were then used to search the genomes of trypanosomatids *T. brucei*, *T. cruzi* and *L. major*. The final alignments were then trimmed to informative blocks and used to generate maximum likelihood trees [6], from which sequences were assigned to specific protein groups. The results are presented in Table 1.

The core complement of Pol II subunits (RPB1–RPB12) is present in all three trypanosomatids. However, the Pol I and Pol III complexes are missing two and one subunits respectively. The absent Pol I subunits, RPA14 and RPA43, have been demonstrated previously to be paralogues of the Pol II dissociable heterodimer components RPB4 and RPB7 respectively [7,8]. The subunit missing from Pol III, RPC17, is also a paralogue of RPB4 [9]. This therefore raises the possibility that the RPB4–RPB7 heterodimer interacts with and regulates all three classes of RNA polymerase in trypanosomatids.

Figure 1 | **Cartoon of the domain structures of TbRPB5, TbRPB5z, TbRPB6 and TbRPB6z in comparison with the *S. cerevisiae* RPB5 and RPB6 subunits**

The z isoform of each subunit group differs from the canonical subunit by insertion (black blocks) or deletion of stretches of charged residues. Sc, *S. cerevisiae*; Tb, *T. brucei*.



The five subunits that, in *Saccharomyces cerevisiae*, have been shown to be common to all three polymerase classes (RPB5, RPB6, RPB8, RPB10 and RPB12) are present in the three trypanosomatids. It is interesting to note that, out of these five subunits, both RPB5 and RPB6 have more than one isoform. One isoform of each subunit is orthologous to the canonical polymerase subunit, and intriguingly, the other differs from the canonical subunit by insertion or deletion of stretches of charged residues near the N-terminus of the protein (Figure 1). We have termed these variant isoforms of the shared subunits RPB5z and RPB6z. It is possible that these subunits are not common to all three polymerase classes as they are in yeast but are specific to individual complexes. Indeed, recent evidence shows that the TbRPB5z co-purifies with Pol I complexes in *T. brucei* [10].

Many lines of evidence point to a role for RPB5 in interacting with DNA-bound transcription activation factors [11,12]. The position of RPB5 adjacent to the DNA entering the polymerase within the known structure of Pol II [13] is also highly supportive of this role. It will be interesting to see whether the two different RPB5 isoforms in trypanosomes are indeed specific to individual complexes, whether they interact with independent sets of transcription activation factors, or whether they represent alternative transcriptionally permissive states of all three polymerase complexes.

The presence of two RPB6 homologues is interesting since RPB6 is a point of contact between the RPB4–RPB7 subcomplex and the RNA polymerase [14]. As the Pol I and Pol III homologues of RPB4–RPB7 heterodimer are either

missing or incomplete, the choice of RPB6 subunit may directly affect whether or not the RPB4–RPB7 subcomplex is recruited to Pol I or Pol III. This recruitment will certainly have wide ranging effects on the polymerase complexes at both transcription initiation and termination. Whatever their roles, these alternative subunits are likely to be important features in the unusual function of the trypanosome RNA polymerases.

This work was supported by the Wellcome Trust and the EP Abraham Trust.

References

1 Donelson, J.E. (2003) Acta Trop. **85**, 391–404
2 Navarro, M. and Gull, K. (2001) Nature (London) **414**, 759–763
3 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. **25**, 3389–3402
4 Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. **22**, 4673–4680
5 Eddy, S.R. (1998) Bioinformatics **14**, 755–763
6 Huelsenbeck, J.P. and Ronquist, F. (2001) Bioinformatics **17**, 754–755
7 Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M. and Carles, C. (2002) Proc. Natl. Acad. Sci. U.S.A. **99**, 14670–14675
8 Meka, H., Daoust, G., Arnvig, K.B., Werner, F., Brick, P. and Onesti, S. (2003) Nucleic Acids Res. **31**, 4391–4400
9 Siaut, M., Zaros, C., Levivier, E., Ferri, M.L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A. and Conesa, C. (2003) Mol. Cell. Biol. **23**, 195–205
10 Walgraffe, D., Devaux, S., Lecordier, L., Dierick, J.F., Dieu, M., Van den Abbeele, J., Pays, E. and Vanhamme, L. (2005) Mol. Biochem. Parasitol. **139**, 249–260
11 Hayashi, K., Watanabe, T., Tanaka, A., Furumoto, T., Sato-Tsuchiya, C., Kimura, M., Yokoi, M., Ishihama, A., Hanaoka, F. and Ohkuma, Y. (2005) Genes Cells **10**, 207–224
12 Wei, W., Dorjsuren, D., Lin, Y., Qin, W., Nomura, T., Hayashi, N. and Murakami, S. (2001) J. Biol. Chem. **276**, 12266–12273
13 Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R. and Kornberg, R.D. (2000) Science **288**, 640–649
14 Tan, Q., Prysak, M.H. and Woychik, N.A. (2003) Mol. Cell. Biol. **23**, 3329–3338

Received 1 July 2005