Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Molecular & Biochemical Parasitology 139 (2005) 163–172
MOLECULAR & BIOCHEMICAL PARASITOLOGY

The transcriptomes of *Trypanosoma brucei* Lister 427 and TREU927 bloodstream and procyclic trypomastigotes

Stefanie Brems[a], D. Lys Guilbride[b], Drifa Gundlesdodjir-Planck[b], Christian Busold[a], Van-Duc Luu[b], Michaela Schanne[a], Jörg Hoheisel[a], Christine Clayton[b,*]

[a] *DKFZ, Im Neuenheimer Feld 580, D 69120 Heidelberg, Germany*
[b] *ZMBH, Im Neuenheimer Feld 282, D 69120 Heidelberg, Germany*

Received 2 September 2004; received in revised form 27 October 2004; accepted 3 November 2004
Available online 28 November 2004

Abstract

We describe developmentally regulated genes in two strains of *Trypanosoma brucei*: the monomorphic strain Lister 427 and the pleomorphic strain TREU927. Expression patterns were obtained using an array of 24,567 genomic fragments. Probes were prepared from bloodstream-form or procyclic-form trypanosomes. Fourteen procyclic-specific and 77 bloodstream-specific signals were obtained from sequences matching variant surface glycoprotein or associated genes, and a further 17 regulated sequences were repetitive or transposable-element-related. Two hundred and eighty-six regulated spots corresponded to mRNAs from other protein-coding genes; these spots represent 191 different proteins. Regulation of 113 different genes (79 from procyclic forms, 34 from bloodstream-forms) was supported by at least two independent experiments or criteria; of these, about 60 were novel. Only two genes – encoding HSP83 and an importin-related protein – appeared to be regulated in the TREU927 strain only. Our results confirmed previous estimates that 2% of trypanosome genes show developmental regulation at the mRNA level.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Trypanosoma; Differentiation; Transcript; Regulation; Array

1. Introduction

The African trypanosome *Trypanosoma brucei* has two major replicative stages: the bloodstream-form, which grows in the blood and tissue fluids of the mammalian host, and the procyclic form found in the midgut of the Tsetse fly vector [1]. The two environments differ in temperature and nutrient availability: the abundance of glucose in the mammalian tissue fluids enables the parasites to synthesise ATP exclusively by glycolysis, whereas in the Tsetse fly amino acids are the principal energy source and mitochondrial metabolism is highly developed [2]. Host defences also differ: in the mammal, African trypanosomes evade the adaptive immune response by variation of the major surface protein, the variant surface glycoprotein (VSG), whereas in the insect, alternative surface proteins, the EP and GPEET procyclins, serve as a barrier against proteases and other invertebrate defences [3].

The control of gene expression in kinetoplastids is primarily post-transcriptional [4]. Expression of quite a large number of genes has been shown to be controlled at the levels of mRNA stability and translation, and nearly all the regulatory sequences investigated so far were mapped to the 3′-untranslated regions (3′-UTR) of the mature mRNAs [4]. We and others have previously suggested that specific regulatory signals in the 3′-UTRs could act as a code for particular regulatory patterns. Interactions of such sequences with specific protein factors could signal mRNA stabilisation or degradation or influence the translation efficiency. The existence of 3′-UTR signals would allow coordinate post-transcriptional regulation of specific sets of transcripts during the life cycle,

Abbreviations: ESAG, expression-site-associated gene; VSG, variant surface glycoprotein; 3′-UTR, 3′-untranslated region

* Corresponding author. Tel.: +49 6221 546 876; fax: +49 6221 545 894.
E-mail address: cclayton@zmbh.uni-heidelberg.de (C. Clayton).

0166-6851/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2004.11.004

rather as specific promoter elements allow coordinate regulation of transcription in other organisms. The most dramatic example of this so far is a 450-nt element found in the 3′-untranslated region (3′-UTR) of a large number of *Leishmania* mRNAs: this element enhances translation in the amastigote stage of the parasite [5]. In bloodstream-form *T. brucei*, a U-rich element depresses translation and stimulates rapid degradation of three otherwise unrelated mRNAs [6].

Genome-wide surveys will help us to understand the host-parasite relationship and the metabolic adaptations occurring in different life-cycle stages, and will also facilitate the identification of elements responsible for post-transcriptional regulation. We previously described the results of a limited survey of regulated trypanosome mRNAs. Genomic arrays were constructed using 21,000 independent random genomic clones from *T. brucei* TREU927. The arrays were hybridised with oligo d(T)-primed labelled cDNAs, generated using total bloodstream-form or procyclic-form RNA as the template. We found that between 1 and 2% of spots on the array showed regulated mRNA abundance, and described the sequences of 60 clones [7].

T. brucei TREU 927 was adapted relatively recently to in vitro culture and can be transmitted readily by Tsetse [8]; it also shows limited resistance to human serum [9]. We were therefore interested in comparing its gene regulation with that of strain Lister 427, which has been subjected to in vitro culture or animal passage for over 30 years and is a standard strain for genetic manipulation [10]. Since comparisons of mRNAs do not detect translational regulation, we were also interested to find out whether additional information could be obtained from a polysomal RNA fraction.

2. Methods

2.1. Microarray analysis and Northern blotting

Microarray construction and hybridisation conditions have been previously described [7]. Briefly, independent random genomic clones from *T. brucei* 927 were used to create PCR products which were spotted onto glass slides. Seventy-five percent of insert sizes were within 1.8–2.5 kb. In the original experiments with the 927 strain RNA, 21,024 PCR products were used [7]. In the new experiments reported here, with 427 RNA, the arrays contained 24,567 spots. This array would be expected to cover at least 80% of the trypanosome genome if all fragments were present at equal probability.

Lister 427 procyclic- and bloodstream-form trypanosomes were harvested in the exponential phase of growth, and we prepared both total RNA and RNA from a 75,000 × *g* pellet. The RNA pellet fraction was made as follows, in a protocol adapted from [11]. After addition of cycloheximide (100 μg^{-1}), which freezes translating ribosomes on the mRNA, cells were collected by centrifugation at 2000 × *g* for 5 min at 4 °C and washed twice in ice-cold polysome buffer (120 mM NaCl, 20 mM Tris–HCl, pH 7.5, 2 mM $MgCl_2$, 1 mM DTT, 10 μg ml^{-1} leupeptin, 100 μg^{-1} cycloheximide). Cells were then resuspended in 0.5 ml of polysome buffer and lysed by the addition of NP-40. The cell suspension was homogenized and the lysate was cleared by centrifugation at 10,000 × *g* for 4 min at 4 °C. The supernatant was centrifuged again at 75,000 × *g* for 35 min at 4 °C in an ultracentrifuge. The pellet was resuspended in 1 ml Trizol per 10^6 cells and RNA was extracted following the manufacturer's protocol. Examination of the RNA after this procedure revealed that nearly all of the ribosomal RNA was recovered in the pellet fraction.

Fluorescent, oligo d(T)-primed cDNAs were prepared from bloodstream and procyclic-forms, mixed, and hybridised to the slides as before [7]. Exchanging of the dyes resulted in no significant differences. For total RNA, we obtained six slides giving a strong signal-to background ratio and for the pelleted RNA, four slides were used. Results were analysed using MCHIPS [12,13]. All ratios were recorded. To choose regulated clones for sequence analysis we selected only those with the following parameters—fitted intensities: at least 200,000; ratio: 2.5; min/max-separation: 0.2. All clones satisfying these criteria were sent for sequencing from one end.

For 73 of the regulated clones, a PCR product (pUC plasmids, M13 forward and reverse primers) of the entire insert was gel purified, labelled with ^{32}P dCTP (Amersham, Braunschweig) by random priming (Prime It II, Stratagene) and hybridised to Northern blots of total RNA. 15 μg per lane total RNA from bloodstream (cell culture densities from 5×10^5 to 1×10^6 cells ml^{-1}) and procyclic (cell culture densities from (2 to 4) × 10^6 cells ml^{-1}) was subjected to electrophoresis on formaldehyde gels and blotted onto Nytran membranes. Hybridization to the signal recognition particle (SRP) RNA [14] was used as a loading control. Blots were exposed to phosphorimaging plates for 48–72 h. Signals were measured using a phosphorimager and MacBas v2.0 software. Hybridisations were done only once unless otherwise stated.

2.2. Annotation

From the sequences obtained we identified all RNAs which would hybridise with the corresponding array spot. Prior to completion of the genome the sequences were used to search both GeneDB and the EMBL databases; some of the sequences which clearly matched *VSG*, *ESAG*s, repetitive elements or ribosomal RNA were not further analysed. Upon genome completion, the remaining sequences were used to search the trypanosome genome (mainly release no. 3) in GeneDB using the blastn algorithm. Matches of 96% or more were assumed to be identical. The sequence of the entire cloned fragment was then deduced, assuming a size of 2 kb unless the insert size was known, and predicted RNAs overlapping by at least 100 bp were annotated. To assign the likely position of polyadenylation sites, we estimated a distance of 200 bp between the polyadenylation site and the next downstream open reading frame (Benz et al., manuscript in

preparation). The detailed assignments for all clones, as well as the hybridisation signals, are presented in searchable tabular form as Supplementary Material.

3. Results

3.1. Overview

In all experiments, RNAs from bloodstream and procyclic trypanosomes were used as templates for the generation of fluorescent cDNA probes, using an oligo d(T) primer, and the resulting probes were hybridised to the genomic arrays. We conducted three hybridisation series, comparing six RNA populations:

1. A comparison of bloodstream-form and procyclic-form RNA from the TREU927 strain. Details of this experiment were reported previously [7]; it included six hybridisations. The previous publication described only a third of the differentially regulated clones; the remainder have now been sequenced and the results are included here.
2. A comparison of bloodstream-form and procyclic-form total RNA from the Lister 427 strain. Six slides with good signal-to-noise ratios were used for further analysis.
3. A comparison of ribosome-associated RNA from bloodstream and procyclic-forms of the 427 strain. The ribosome-associated fraction was made by centrifuging a cytoplasmic fraction at 75,000 × *g* for 35 min. Four slides with good signal-to-noise ratios were obtained.

The results from experiments 2 and 3 were analysed using MCHIPS software. To choose regulated clones, we applied filters which selected for regulation of at least 2.5-fold, and excluded low signals and results with excessive variation (see Section 2). Single-pass sequences from all the selected clones were mapped on to the *T. brucei* genome. We determined which mRNAs were likely to hybridise with the complete insert and noted the details of the open reading frames within those mRNAs (see Section 2). All the results are presented as a supplementary table.

3.2. Unique genes and large gene families

Table 1 shows an overall summary of the results. We sorted the genes with potential mRNA matches into two classes: "unique" genes and members of large gene families. The "unique" genes are all protein-coding genes which did not fit into the other categories; they include not just single-copy genes, but also genes present in several identical or closely related copies. This class included 286 spots, representing 191 different proteins.

The "large gene family" class included all retroposons (e.g. retrotransposon hotspot, *RIME*, *INGI*) as well as the *VSG* genes and associated genes. The VSGs of *T. brucei* are expressed from telomeric expression sites containing up to 10 additional co-transcribed genes called "expression-site-associated" genes (*ESAG*s) [15]. In addition to the *ESAG*s and VSGs at the telomeres, the *T. brucei* genome contains a large number of chromosome-internal or sub-telomeric *ESAG* and *VSG* genes and pseudogenes, often alternating with repetitive elements [16–18]. Transcripts from the telomeric expression sites in the bloodstream-form are abundant, and will cross-hybridise with similar genes and pseudogenes from elsewhere in the genome. In addition, some genes related to *ESAG*s are independently expressed from chromosomal-internal locations, or specifically expressed in procyclic-

Table 1
Statistics for spots corresponding to the regulated transcripts

(1) Class	(2) # of spots[a]	(3) % of all regulated spots[b]	(4) # of spots supported[c]	(5) % supported[d]
Unique[e]	286	54	208	73
VSG and *ESAG*-related, up-regulated in bloodstream-forms	77	15	40	52
VSG and *ESAG*-related, up-regulated in procyclic-forms	14	3	5	36
RIME, *INGI*, retroposon hotspot	17	3	7	41
Non-coding[f]	16	3	3	19
Ribosomal RNA	42	8	13	31
[g]Not in GeneDB*	40	8	18	45
No insert	6	1	1	17
No sequence obtained	32	6	14	44
Total	530	100	309	58
Total on array	21024			

[a] Gives the overall numbers of regulated spots in the different sequence classes.
[b] Shows the percentage of spots in each sequence class, relative to the total number of clones which were found to be regulated.
[c] Shows how many spots in the individual classes showed regulation which was supported by any of the criteria in Section 3.3.
[d] Gives the percentage of spots with independent support in each individual class (column (4) as a fraction of column (2)).
[e] Unique genes are those which do not belong to the large gene families (*VSG*, *ESAG* and retroposon).
[f] Non-coding sequences came from regions which were not predicted to contain either open reading frames, or 3′- or 5′-untranslated regions. Two of the confirmed spots came from regions in which the transcription direction appeared to change.
[g] Sequences which are not in the genome database may originate from telomeres or minichromosomes, both of which contain VSG-like sequences; 75% of these were up-regulated in bloodstream-forms.

Table 2
Number of spots giving a differential signal from the bloodstream-specific *ISG65* and *MSP* loci and the procyclic-specific *EP-GPEET-PAG* loci

Chromosome or locus	Genes	Source of RNA used for probe						
		927T	427T	427P	927T, 427T	927T, 427P	427T, 427P	ALL
Chromosomes 5 and 2	*ISG65*	1 (0)	3 (2)		1 (0)	1 (1)	4 (6)	1 (2)
Chromosomes 6 and 10	*EP-PAG* loci		3 (0)	2 (1)	2 (1)	1 (0)	4 (8)	1 (3)
Tb08.29O9.350-380	*MSP*	2 (0)	3 (0)		1 (0)		1 (4)	0 (3)

Sources of RNA are as follows: 927T, 927 total RNA; 427T, 427 total RNA; 427P, 427 "polysomal" or ribosome-associated pellet RNA; 927T and 427T, differential signal from both total RNAs; etc. For example, any spot which gave a differential signal with only 427 polysomal RNA will be listed under the column "427P", while a spot which was judged to be differential using each of the probes will be listed under "ALL". The numbers without parentheses are the differential signals which were judged to be reliable after applying filters for statistical significance and signal strength. The numbers in parenthesis include, in addition, all data from the 427 strain hybridisations which indicated a difference of two-fold or more, even if the data were excluded by the filters.

forms (see e.g. [19,20]). Since the similarities within these gene families will lead to extensive cross-hybridisation, the appearance of a differential signal from an individual spot does not imply that that sequence of the DNA on the spot is really transcribed. As expected, the majority of the detected *VSG*/*ESAG*-like spots showed up-regulation in the bloodstream-form.

If we estimate that the large gene families constitute roughly 25% of the genome, then 15–16,000 spots should contain other sequences. Using our relatively stringent criteria, about 2% of these show developmental regulation.

3.3. Reliability and reproducibility

We next assessed the validity of the results. Regulation was judged to be "confirmed" if the sequences or the experimental results satisfied one or more of the following criteria:

1. Known developmentally regulated mRNA.
2. Known developmentally regulated protein.
3. Protein involved in a pathway or process that is known to be developmentally regulated: glycolysis in bloodstream-forms, mitochondrial biogenesis and metabolism in procyclic-forms.
4. The same type of differential signal was obtained from two or more different spots containing sequence from the same gene (or family of closely related genes).
5. A significant differential signal (using the MCHIP filters described above) was obtained for a single spot from two or more hybridisation series.
6. The spot showed a significant differential signal in one hybridisation series. In at least one of the other hybridisation series, two-fold or more regulation was seen but the results did not satisfy the significance criteria we had set.
7. The result was confirmed by Northern blot hybridisation or RT-PCR.

Clones which satisfied only criterion 6 or criterion 7 were judged to be only weakly supported, unless the Northern blots had been repeated. The corresponding gene names are in italics in Tables 4 and 5. They will be referred to as "weakly supported" in the discussion below.

The criteria we have used to confirm regulation will exclude some single-copy sequences which are genuinely regulated. A few genes will not be on the arrays at all. Many single-copy genes are likely to be represented by only one spot and we have not performed Northern blots for all clones. The use of two independent hybridisation series for the 427 strain played a major role in confirming the regulation of mRNAs in that strain, but the hybridisation series for the 927 strain has not been repeated. Hence some genes which are developmentally regulated only in the 927 strain have probably been missed.

The effects of applying these criteria are shown in Table 1: most importantly, 73% of genes in the "unique" category were confirmed as regulated; this corresponds to 113 different proteins. A lower level of confirmation was expected for the *VSG*/*ESAG* category as the 927 and 427 strains express different *VSG* and *ESAG* genes, which will cross-hybridise to different-related genes and pseudogenes.

To further compare results from the three hybridisation series we examined the differential signals from several different spots containing cloned DNA from regulated gene families. The results for three gene families are shown in Tables 2 and 3.

The families chosen were the *EP-GPEET-PAG* loci [21], the genes encoding a procyclic-specific major surface protease (MSP) [22] and the bloodstream-specific *ISG65* genes [23]. We found that PCR fragments spanning virtually identical regions gave different results in individual experiments. The original 927 total RNA experiment was less sensitive than the more recent one done using 427 total RNA, possibly because the more recent arrays are of superior quality. Akopyants et al. [24], who compared mRNAs from three life-cycle stages of *Leishmania* using genomic microarrays, reported that the "efficiency" of recognition of a single regulated spot

Table 3
Number of spots recognised in individual hybridisation series for the *ISG65*, *MSP* and *EP-PAG* loci

	ISG65	EP-PAG	MSP	Total
927T	4	4	3	11
427T	9 (10)	9 (11)	5 (7)	23 (28)
427P	6 (9)	8 (11)	1 (7)	15 (27)
Total spots	11	13	7	31

Details as in Table 2. For example, the experiment with 927 total RNA detected four *ISG65* spots, four *EP-PAG* spots, and three *MSP* spots—a total of 11 for the three gene sets.

Table 4
Genes whose expression is up-regulated in procyclic trypanosomes

Gene ID	Description of locus or protein product	# of spots	# of differential signals	427 Northern	Publication
Membrane					
	**EP* loci	13	21 (26)		[30]
Tb10.6k15.3510	CRAM	5	9 (10)	C	[31]
Tb07.25D22.110	Trans-sialidase	2	2 (4)		
Tb08.29O9.350-380	*MSP	7	10 (17)		[22]
Tb05.29K2.440	Trans-sialidase	1	1 (2)	+	
Tb05.3C6.130	META1-like	1	2	+	
Tb09.211.4070	GPI anchored protein	1	2		
Tb927.1.2310	Membrane protein	1	3	+	
Tb927.2.5290	*Mucin-like, 8 copies*	*1*	*1*	+	
Tb05.29K2.440	*Trans-sialidase-like*	*1*	*1* (2)		
Tb11.0310	*Membrane protein*	*1*	*1* (2)		
Transporter					
Tb03.6N20.700	Adenosine transporter	2	2 (3)		
Tb04.3I12.190	Amino acid transporter	4	7 (9)	+	
Tb07.10C21.10/20	13 trans-membrane domains, transporter?	3	6	+(5)	
Tb07.2F2.110-150	*13 trans-membrane domains, transporter?	7	8 (9)		
Tb08.4A8.350	Amino acid transporter	2	3	+	
Tb10.6k15.1350	Pteridine transporter	6	12 (13)	+(2)	
Tb09.160.5480	Nucleoside transporter	2	3		up in short stumpy form (GeneDB annotation)
Tb10.6k15.2020	*THT2A hexose transporter	1	2 (3)		[26]
Tb04.1D20.650	*Amino acid transporter*	*1*	*1* (2)		
Tb927.1.580	*Phosphate-repressible phosphate permease*	*1*	*1* (2)		
Glycosome					
Tb11.02.4150	PPDK pyruvate phosphate dikinase	3	6 (7)		[2]
Tb10.61.0980,	gMDH glycosomal malate dehydrogenase	1	3		[2]
Tb927.2.4210	Phosphoenolpyruvate carboxykinase	1	1 (2)		[2]
Tb03.26J7.860	Fumarate hydratase	3	6 (9)		[32]
Mitochondrion					
Tb07.30D13.350/60	ATP synthase alpha	4	7		[33]
Tb03.1J15.500	ATP synthase beta	2	2		[33]
Tb09.160.4310	Glutamate dehydrogenase	2	3 (4)		
Tb09.160.1820	COX V cytochrome *c* oxidase subunit V	1	2		[2]
Tb03.1J15.440	COX VII cytochrome *c* oxidase subunit VII	1	2		[2]
Tb04.3I12.400	COX VIII cytochrome *c* oxidase subunit VIII	1	2 (3)	+	[2]
Tb11.01.4702	COX X cytochrome *c* oxidase subunit X	2	3 (4)		
Tb08.26A17.490	Cytochrome *c* oxidase copper chaperone domain	1	1	+	[2]
Tb07.13M20.760	NADH-cytochrome b5 reductase	1	1		[2]
Tb09.160.4380	Succinate dehydrogenase	1	1		[2]
Tb09.211.4700	Reiske iron-sulfur protein	1	1		[2]
Tb09.244.2620	NADH-ubiquinone oxidoreductase 39 kDa subunit	1	1		[2]
Tb10.1650	Aconitase	1	1 (2)		[34]

Table 4 (*Continued*)

Gene ID	Description of locus or protein product	# of spots	# of differential signals	427 Northern	Publication
Tb11.01.1740	2-Oxoglutarate dehydrogenase E1	1	1 (2)	+	[2]
Tb11.01.3550	2-Oxoglutarate dehydrogenase, E2	1	1		[2]
Tb11.02.0290	Succinyl-CoA:3-ketoacid-coenzyme a transferase	1	1 (2)	+	[2]
Tb10.6k15.3250	Succinyl-CoA ligase	1	1		[2]
Tb11.03.0870	Mitochondrial carrier protein	1	1	0	[2]
Tb05.28F8.230	Mitochondrial processing peptidase	2	2		
Tb927.1.1690	MP90 RNA editing complex	1	1		
Tb07.27M11.900	MP57 RNA editing 3′ terminal uridylyl transferase 2	1	1 (2)		
Tb11.03.0940	*Bacterial elongation factor*	*1*	*1*	*0*	
Enzyme					
Tb11.02.2310,	Prostaglandin F synthase	1	1	+	
Tb11.02.5410	Ubiquitin activating enzyme	1	1	+	
Tb927.1.2230	Calpain-like fragment	1	1 (2)	+	
Tb07.26A24.1130	Threonine synthase	1	2		
Tb10.389.1480	Metalloprotease	2	2 (4)		
Tb927.2.2510/20 or Tb927.2.2500	*Conserved hypothetical or serine peptidase*	*1*	*1* (2)		
Tb03.27F10.890	*Protein kinase*	*1*	*1* (2)		
Cytoskeleton					
Tb04.1D20.740	CAP5.5 cytoskeleton-associated protein	3	5 (6)		[35]
Tb08.10J17.170	FLA1 flagellar adhesion protein	2	3		[36]
Tb07.22O10.840	Myosin-like	1	2 (3)		
Tb10.406.0650	Microtubule-associated	1		+	
Tb08.4A8.660	Kinesin	2	2 (4)		
Tb11.01.7880	*CAP17 cytoskeleton-associated protein*	*1*	*1* (2)		
Tb04.5E12.340	*Calcineurin-like phosphoesterase*	*1*	*1* (2)		
Gene expression/sorting					
Tb08.10J17.140	Endonuclease	2	3 (5)	+	
Tb05.26K5.1010	Translation initiation factor	1	3	0	
Tb10.70.4600	Elongation factor	1	2		
Tb10.26.1080	*HSP83	6	6# (8)	0	Very little regulation in 427: array, northern and [27]
Tb07.27M11.780	Importin	3	3#	0	
Tb07.13M20.480,	*ATP-dependent DEAD/H RNA helicase*	*1*	*1* (2)		
Tb11.01.8290	*DEAD/DEAH box helicase*	*1*	*1* (2)		
Tb11.02.2930	*DNA helicase*	*1*	*1* (2)		
Unknown					
Tb08.26N11.180	Hypothetical	2	2	+	
Tb03.1J15.550	Conserved hypothetical, WD-40 repeat	1	1	+	
Tb09.244.2660	Conserved hypothetical	1	3	0	
Tb07.21H15.100	Hypothetical	1	2		
Tb11.01.7880	Hypothetical	1	2		

Table 4 (*Continued*)

Gene ID	Description of locus or protein product	# of spots	# of differential signals	427 Northern	Publication
Tb07.22O10.800	*Possibly GTP-binding*	*1*	*1* (2)		
Tb927.1.2260	*3′-UTR: similar to calpain-like protein*	*1*	*1* (2)		
Tb07.13M20.380	*Fatty acid signaling?*	*1*	*1* (2)		
Tb05.3C6.130	*Conserved hypothetical*	*1*	*1* (2)		
Tb08.10J17.80 or Tb08.10J17.130	*Conserved hypothetical*	*1*	*1* (2)		
	Expression-site	14	24		
	Repeats	7	9		

The table shows all genes for which there are at least two pieces of evidence for regulated expression. Evidence can be from the arrays, Northern blots, publications, involvement in a regulated pathway or compartmentation in a regulated organelle, as described in Section 3.3. The gene ID and functional annotation are from GeneDB in August 2004; many functions are unconfirmed. Tandemly repeated genes with at least three copies per haploid genome. The third column (# of spots) shows the number of spots recognised which contained the sequence from the gene(s). The number of differential signals (# of differential signals) indicates also the number of independent array series in which each spot gave a "significant" differential signal according to the MCHIPS setting; this will be a maximum three differential signals per spot. The numbers in parenthesis include, in addition, all additional data indicating a difference of two-fold or more even if the MCHIPS criteria of significance were not met. Differential signals which were confirmed only by criterion 6 or 7 (Section 3.3) are in italics. Signals specific to the 927 strain are indicated by #. Northern blots using 427 strain RNA are marked as follows—+, over 1.5-fold regulation in agreement with the array; (+), 1.5×; 0, no regulation; C, regulation conflicting with array. All northerns were done once only unless otherwise stated (number in parentheses).

in individual hybridisations varied from 24 to 76%; for our 427 total RNA, the efficiency of recognition after multiple experiments was 77% for signals which were statistically accepted by our MCHIPS settings, and 90% if all signals were accepted.

Finally we compared the results for 427 total RNA and mRNA from a 75,000 $\times$ g pellet fraction. We expected that RNA from the pellet fraction would be enriched with polysomes, and might therefore be able to be used to detect translational regulation. The results, shown in Tables 2 and 3 and in the supplementary table, did not support this. Overall the "polysomal" RNA seemed merely to mirror the total RNA pattern but with somewhat lower sensitivity. A quantitative comparison for all spots detected by both 427 total RNA and 427 "polysomal" RNA showed that on average the degrees of regulation were similar. We tried sorting the signals and classes in various ways but no significant differences were found. Almost certainly the fractionation procedure – a simple centrifugation – was too crude to give useful results. Nevertheless, the additional data obtained was very helpful in confirming the results from total RNA.

3.4. Genes showing reproducible regulation

The main use of arrays is to reveal differentially regulated genes. In this respect we think our coverage is now likely to be very extensive, although not exhaustive due to the nature of the library used and our inability to obtain sequences for all clones. Tables 4 and 5 are lists of genes whose regulation was judged to be reliable according to at least one of the criteria listed in Section 3.3. In addition to known regulated surface proteins the gene products include glycosomal enzymes – four known to be up-regulated in procyclics and six in bloodstream-forms – and 22 regulated mitochondrial proteins, of which all but one, the alternative oxidase involved in glycerophosphate metabolism, are procyclic-specific. The phosphoglycerate kinase genes *PGKB* (procyclic-specific) and *PGKC* (bloodstream-specific) [25] are too similar to be found using genomic clones, and the bloodstream-specific hexose transporter *THT1*, which is closely related to the oppositely regulated *THT2* [26] was also not detected. The *THT2* gene was in fact found, as one array spot overlapped the 3′-UTR, but not the coding region.

Among the spots corresponding to transcripts for which regulation was independently confirmed, there were 64 genes whose regulation could not have been predicted. Among these, a variety of stage-specific transporters, mostly of unknown substrate specificity, and 20 predicted proteins of unknown function would merit further investigation. Many more novel genes, identified in only one hybridisation series and with no additional supporting evidence, are in the full table (Supplementary Material).

To look for strain-specific mRNAs we restricted our analysis to genes represented at least twice on the array. Specificity for 427 could not be judged as we could no longer access the raw data from the 927 experiments. Six different array spots of *HSP83* were detected as up-regulated in 927 procyclic-forms, and the result was confirmed by RT-PCR [7]. The 427 array results, in contrast, mainly gave results only slightly greater than 1, and Northern blot analysis with 427 RNA also showed less than 1.5-fold regulation ([27] and Robles, ZMBH, unpublished results). Notably, *HSP83* RNA is expressed at higher levels in *Leishmania* promastigotes than amastigotes [28,29]. The only other gene with a 927-strain-specific signal was Tb07.27M11.780, which encodes an importin-like protein.

Table 5
Genes whose expression is up-regulated in bloodstream-forms

Gene ID	Description of locus or protein product	# of spots	# of differential signals	427 Northern	Publication
Membrane					
	*ISG75	4	4 (7)		[23]
	*ISG65	11	17 (21)		[23]
Tb11.0360	GP63-3 homologue (MSP)	1	1#		[22]
Tb11.33.0013	GP63-1 homologue (MSP)	1	1#		[22]
Tb927.2.6000	GPI-PLC, glycosylphosphatidylinositol-specific phospholipase C	1	2 (3)		[37]
Tb11.01.3790	Membrane protein	2	2 (4)		
Transporter					
Tb11.01.0725-30	Cation transporter	3	3 (5)		
Tb927.2.6150	P1-type nucleoside transporter	2	3		[38]
Tb03.26J7.120-160	*13 trans-membrane domains	2	3 (4)		
Tb04.3M17.20-69	Amino acid transporter	2	2(4)		
Tb927.2.6200	*TbNT3 adenosine transporter 2*	*1*	*1* (2)		
Glycosome					
Tb10.70.1370	Aldolase	3	3 (6)		[2]
Tb10.70.5800	Hexokinase	2	2 (3)		[2]
Tb03.3K10.320,	ATP-dependent phosphofructokinase	1			[2]
Tb09.211.3540-3580	Glycerol kinase	1			[2]
Tb10.1550	PYK2 pyruvate kinase 2	1	1 (2)		[32]
Tb927.1.3830	PGI glucose-6-phosphate isomerase	1			[2]
Mitochondrion					
Tb10.6k15.3640,	Alternative oxidase	1	1 (2)		[39]
Cytoskeleton					
Tb10.61.1750	Kinesin	2	2 (3)		
Tb11.01.3805	CAP15	2	3 (4)		
Gene expression/sorting					
Tb927.2.3880	HNRNPH	2	2 (3)	(+)	
Tb05.45E22.550	*Histone H4*	*1*	*1*	+	
Tb07.87P12.500	*RBP8*	*1*	*1* (2)		
Other					
Tb07.5F10.450,	*Glycosyl hydrolase*	*1*	*1* (2)		
Tb10.70.5230	*Metacaspase MCA4*	*1*	*1* (3)		
Tb11.02.0730	*Metacaspase*	*1*	*1* (2)		
Unknown					
Tb05.25N21.20	Hypothetical	3	4 (6)		
Tb05.26C7.190,220	Hypothetical	3	3 (6)		
Tb10.6k15.0940	Hypothetical	1	2 (3)		
Tb11.01.3810	Hypothetical	1	2		
Tb08.26A17.40	*Hypothetical*	*1*	*1*	+	
Tb07.26A24.60	*Conserved hypothetical*	*1*	*1* (2)		
Tb10.70.3090	*Hypothetical*	*1*	*1* (2)		
Tb08.26A17.40	*Hypothetical*	*1*	*1* (3)		
	Expression-site	77	82		
	Repeats	10	10		

Details as for Table 2.

3.5. Outlook

Our results indicated that – if the large gene families are excluded – approximately 2% of trypanosome genes show at least two-fold regulation at the level of mRNA abundance. This result agrees with that obtained by Akopyants et al. who used *Leishmania major* genomic arrays similar to ours and looked for two-fold regulation [24]. When EST sequences were used to make the arrays and the boundary was set lower, the percentage of regulated genes appeared higher [28]. Whichever boundary is set, these analyses reveal large classes of regulated genes. The information can be used to investigate mechanisms of regulation and to illuminate the changes in parasite biochemistry and cell biology required for adaptation to different hosts.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinshaft. The *Trypanosoma brucei* genome was sequenced by TIGR and the Sanger Centre with support from the National Institutes of Health (USA) and the Wellcome Trust (UK).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at 10.1016/j.molbiopara.2004.11.004.

References

[1] Matthews KR. Developments in the differentiation of *Trypanosoma brucei*. Parasitol Today 1999;15:76–80.
[2] Hannaert V, Bringaud F, Opperdoes FR, Michels PAM. Evolution of energy metabolism and its compartmentation in Kinetoplastida. Kinetoplastid Biol Dis 2003;2:11.
[3] Roditi I, Liniger M. Dressed for success: the surface coats of insect-borne protozoan parasites. Trends Microbiol 2002;10: 128–34.
[4] Clayton CE. Developmental regulation without transcriptional control? From fly to man and back again. EMBO J 2002;21: 1881–8.
[5] Boucher N, Wu Y, Dumas C, et al. A common mechanism of stage-regulated gene expression in Leishmania mediated by a conserved 3′ UTR element. J Biol Chem 2002;277:19511–20.
[6] Quijada L, Hartmann C, Guerra-Giraldez C, Drozdz M, Irmer H, Clayton CE. Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* induces differentiation-related changes in the abundance of developmentally-regulated mRNAs. Nucl Acids Res 2002;30:1–11.
[7] Diehl S, Diehl F, El-Sayed NM, Clayton CE, Hoheisel JD. Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA microarray. Mol Biochem Parasitol 2002;123:125–33.
[8] van Duersen FJ, Shahi SH, Turner CMR, et al. Characterisation of the growth and differentiation in vivo and in vitro of bloodstream-form *Trypanosoma brucei* strain TREU 927. Mol Biochem Parasitol 2001;112:163–72.
[9] Vanhamme L, Renault H, Lecordier L, Poelvoorde P, Van Den Abbeele J, Pays E. The *Trypanosoma brucei* reference strain TREU927/4 contains *T. brucei rhodesiense*-specific SRA sequences, but displays a distinct phenotype of relative resistance to human serum. Mol Biochem Parasitol 2004;135:39–47.
[10] Wirtz E, Leal S, Ochatt C, Cross GAM. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Mol Biochem Parasitol 1999;99:89–102.
[11] Djikeng A, Shi H, Tschudi C, Shen S, Ullu E. An siRNA ribonucleoprotein is found associated with polyribosomes in *Trypanosoma brucei*. RNA 2003;9:802–8.
[12] Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M. Correspondence analysis applied to microarray data. Proc Natl Acad Sci USA 2001;98:10781–6.
[13] Fellenberg K, Hauser NC, Brors B, Hoheisel JD, Vingron M. Microarray data warehouse allowing for the statistical analysis of experiment annotations. Bioinformatics 2004;18:423–33.
[14] Michaeli S, Podell D, Agabian N, Ullu E. The 7SL RNA homologue of *Trypanosoma brucei* is closely related to mammalian 7SL RNA. Mol Biochem Parasitol 1992;51:55–64.
[15] Pays E, Lips S, Nolan D, Vanhamme L, Perez-Morga D. The VSG expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. Mol Biochem Parasitol 2001;114:1–16.
[16] Berriman M, Hall N, Sheader K, et al. The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. Mol Biochem Parasitol 2002;122:131–40.
[17] Bringaud F, Biteau N, Melville SE, et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. Eukaryot Cell 2002;1:137.
[18] Bringaud F, Biteau N, Zuiderwijk E, et al. The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. Mol Biol Evol 2003.
[19] Berberof M, Pays A, Pays E. A similar gene is shared by both the variant surface glycoprotein and procyclin gene transcription units of *Trypanosoma brucei*. Mol Cell Biol 1991;11:1473–9.
[20] Paindavoine P, Rolin S, van Assel S, et al. A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. Mol Cell Biol 1992;12:1218–25.
[21] Roditi I, Clayton CE. An unambiguous nomenclature for the major surface glycoprotein genes of the procyclic form of *Trypanosoma brucei*. Mol Biochem Parasitol 1999;103:99–100.
[22] LaCount DJ, Gruszynski AE, Grandgenett PM, Bangs JD, Donelson JE. Expression and function of TbMSP (GP63) genes in African trypanosomes. J Biol Chem 2004;278:24658–64.
[23] Ziegelbauer K, Multhaup G, Overath P. Molecular characterisation of two invariant surface glycoproteins specific for the bloodstream stage of *Trypanosoma brucei*. J Biol Chem 1992;267:10797–803.
[24] Akopyants NS, Matlib RS, Bukanova EN, et al. Expression profiling using random genomic DNA microarrays identifies differentially expressed genes associated with three major developmental stages of the protozoan parasite *Leishmania major*. Mol Biochem Parasitol 2004;136:71–86.
[25] Gibson WC, Swinkels BW, Borst P. Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. J Mol Biol 1988;201:315–25.
[26] Bringaud F, Baltz T. Differential regulation of two distinct families of glucose transporter genes in *Trypanosoma brucei*. Mol Cell Biol 1993;13:1146–54.
[27] Glass DJ, Polvere RI, van der Ploeg LHT. Conserved sequences and transcription of the hsp70 gene family in *Trypanosoma brucei*. Mol Cell Biol 1986;6:4657–66.
[28] Almeida R, Gilmartin BJ, McCann SH, et al. Expression profiling of the Leishmania life cycle: cDNA arrays identify developmentally regulated genes present but not annotated in the genome. Mol Biochem Parasitol 2004;136:87–100.
[29] Argaman M, Aly R, Shapira M. Expression of the heat shock protein 83 in Leishmania is regulated post-transcriptionally. Mol Biochem Parasitol 1994;64:95–110.
[30] Roditi I, Furger A, Ruepp S, Schurch N, Butikofer P. Unravelling the procyclin coat of *Trypanosoma brucei*. Mol Biochem Parasitol 1998;91:117–30.
[31] Lee MG-S, Bihain BE, Russell DG, Deckelbaum RJ, Van der Ploeg LH. Characterization of a cDNA encoding a cysteine-rich cell surface protein located in the flagellar pocket of the protozoan *Trypanosoma brucei*. Mol Cell Biol 1990;10:4506–17.
[32] Besteiro S, Biran M, Biteau N, et al. Succinate secreted by *Trypanosoma brucei* is produced by a novel and unique glycosomal enzyme. NADH-dependent fumarate reductase. J Biol Chem 2002;277:38001–12.
[33] Brown BSV, Chi TB, Williams N. The *Trypanosoma brucei* mitochondrial ATP synthase is developmentally regulated at the

level of transcript stability. Mol Biochem Parasitol 2001;115: 177–87.

[34] van Weelden SW, Fast B, Vogt A, et al. Procyclic *Trypanosoma brucei* do not use Krebs cycle activity for energy generation. J Biol Chem 2003;278:12854–63.

[35] Hertz-Fowler C, Ersfeld K, Gull K. CAP5. 5, a life-cycle-regulated, cytoskeleton-associated protein is a member of a novel family of calpain-related proteins in *Trypanosoma brucei*. Mol Biochem Parasitol 2001;116:25–34.

[36] Nozaki T, Haynes PA, Cross GAM. Characterization of the *Trypanosoma brucei* homologue of a *Trypanosoma cruzi* flagellum adhesion protein. Mol Biochem Parasitol 1996;82:245–56.

[37] Mensa-Wilmot K, Hereld D, Englund PT. Genomic organization, chromosomal localization, and developmentally regulated expression of the glycosyl-phosphatidylinositol-specific phospholipase C of *Trypanosoma brucei*. Mol Cell Biol 1990;10: 720–6.

[38] Sanchez MA, Ullman B, Landfear SM, Carter NS. Cloning and functional expression of a gene encoding a P1 type nucleoside transporter from *Trypanosoma brucei*. J Biol Chem 1999;274: 30244–9.

[39] Chaudhuri M, Hill GC. Cloning, sequencing, and functional activity of the *Trypanosoma brucei brucei* alternative oxidase. Mol Biochem Parasitol 1996;83:125–9.