MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4657–4666
0270-7306/86/124657-10$02.00/0

Vol. 6, No. 12

Conserved Sequences and Transcription of the hsp70 Gene Family in *Trypanosoma brucei*

DAVID J. GLASS, RAMONA I. POLVERE, AND LEX H. T. VAN DER PLOEG*

Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, New York, New York 10032

Received 26 June 1986/Accepted 5 September 1986

Five *Trypanosoma brucei* 70-kilodalton heat shock protein-encoding genes (hsp70 genes) were found to be arranged in a tandem array. These hsp70 genes are separated by highly conserved intergenic region sequences of 200 base pairs for one gene and 234 base pairs for the other four genes. This intergenic region sequence is also present in front of the first gene of the tandem array, though at a further distance. All five conserved intergenic regions have sequences that are homologous to the eucaryotic control elements, essential for temperature-induced initiation of transcription by polymerase II. In addition, there is a T-rich region at the 3′ end of the hsp70 genes which is homologous to the site of transcription termination of mini-exon genes. Immediately 3′ of a putative TATA box, a branch point consensus sequence and six sequences homologous to known trypanosome 3′ splice sites were found. It is therefore possible that a *Pol*II promoter is present in the intergenic region sequence. Addition of the 35-nucleotide mini-exon to the hsp70 transcript could thus be mediated by bimolecular splicing. The importance of temperature control for development was illustrated by the response of variant surface glycoprotein-encoding genes to heat shock.

All living organisms examined to date respond to significant temperature increases by activating a specific set of heat shock genes. The same genes are also activated by stress signals, such as anoxia, cell transformation, glucose deprivation, and chemicals that interfere with oxidative phosphorylation (for review, see reference 1). In addition, heat shock genes are part of the normal developmental pathway; they are temporarily activated in erythropoiesis, mouse embryogenesis, yeast spore development, *Volvox* spp. sexual differentiation, and *Leishmania* spp. and *Trypanosoma brucei* differentiation (4, 27, 30, 32, 50).

Three different families of heat shock genes have been identified for *Drosophila* spp.: genes encoding heat shock proteins of about 83 and 70 kilodaltons and genes encoding small heat shock proteins of about 23 kilodaltons. Conserved 14-basepair (bp) DNA sequences (heat shock elements [HSEs]) were first found upstream of the TATA box of the *Drosophila* 70-kilodalton heat shock protein (hsp70)-encoding genes (40). Subsequently, homologous sequences were found associated with the other heat shock protein-encoding genes (as summarized in Pelham [40] and Ayme et al. [3]). The presence of a single HSE is sufficient to regulate temperature-inducible transcription by polymerase II (PolII) (38, 46). A common feature of the HSE is a diad nucleotide symmetry. All eucaryotic species tested to date, from plants to man, have HSEs, with at least five of eight consensus nucleotides conserved. Moreover, heat shock transcription factors bind to the HSE and confer temperature-sensitive transcription to the gene (46).

Parasitic protozoa, such as *T. brucei* and *Leishmania major*, are subjected to a heat shock when they are transferred from the 25°C ambient temperature of their insect vector to the 37°C temperature of their mammalian host (31, 50). In *L. major*, this temperature shift is sufficient to induce in vitro differentiation from the insect-adapted promastigote to the amastigote life stage found in an infected host. Because laboratory infections of tsetse flies with trypanosomes are not successful when the flies are artificially maintained at 30°C, it is clear that the life cycle of the parasite is dependent on habitats of different temperatures (19, 26).

Transcription of protein-encoding genes in trypanosomes is not well understood. First, all mRNA in *T. brucei* consists of two exons, a 35-nucleotide (nt) mini-exon and the main coding exon of the gene. Second, these two exons are not found contiguously in the genome and can even be encoded on separate chromosomes (10, 13, 16, 35, 39, 49, 51, 53). The mechanism of the joining of mini-exon and main exon is unclear, and two models have been proposed: (i) joining of the exons by either bimolecular splicing or RNA ligation or (ii) transcription initiation with the mini-exon as a primer (for a review, see reference 11). It has been impossible to distinguish between these models, because promoter sequences have remained unidentified even for mini-exon genes, for which the transcription unit is only 140 nt.

The *T. brucei* hsp70 locus was analyzed to identify its regulatory signals and determine the role of hsp70 transcription in *T. brucei* differentiation. The evolutionary conservation of heat shock genes and their promoters should facilitate identification of a trypanosome PolII promoter. Thus, we want to determine whether trypanosome PolII promoters are like other eucaryotic promoters or whether they have mini-exon homologous sequences necessary for priming. There is also the possibility that protein-encoding genes are transcribed as polycistronic mRNAs which are subsequently spliced. If that was true, the immediate upstream regions of the genes would not be expected to contain promoter sequences. This has been suggested for some of the genes encoding glycolytic enzymes, tubulin genes, calmodulin genes, and a repetitive gene from *T. cruzi* (14, 20, 36, 45, 47).

MATERIALS AND METHODS

DNA analysis, plasmid isolation, and Southern transfer. *T. brucei* variant 118b (33, 34) was grown in white male Wistar rats and purified free from blood elements as described by Fairlamb et al. (18). Nuclear DNA was isolated from

* Corresponding author.

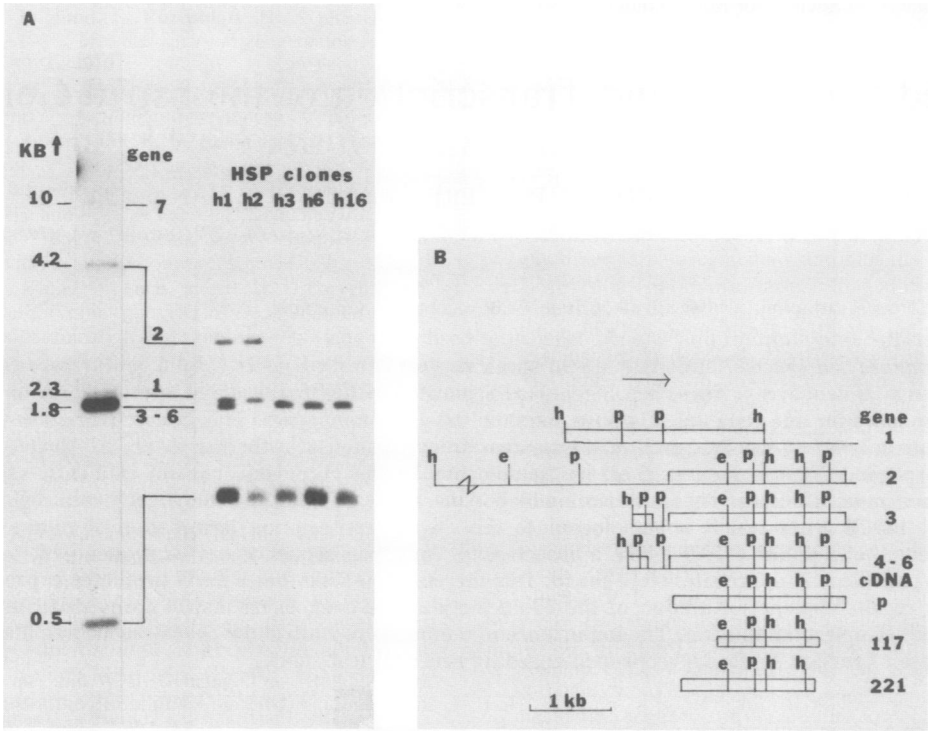


FIG. 1. Analysis of the *T. brucei* hsp70 loci. (A) Hybridization gel. The left lane is a hybridization of genomic *T. brucei* DNA digested with *Hin*dIII and probed with the *T. brucei* hsp70 coding sequence (clones 10 and 11, Fig. 2). Molecular weight standards are shown at left. Because two clones which were labeled at unequal specific activities were used in the hybridization, the 0.5- and 1.8-kb bands hybridized in a nonstoichiometric manner. A comparison of the 2.3- and 1.8-kb bands in gels hybridized with the 5′ coding region of clone 10 only showed a 1-to-4 intensity ratio as determined by densitometric scanning (data not shown). This confirms the presence of four identical genes in the tandem array. To the right of this lane are given the numbers of the genes that correspond to the genes in Fig. 1B and 2. The genomic clones used to map six of the seven genes are shown in the right hand panel, which was hybridized with a *Drosophila* hsp70 probe-containing coding sequence only (1.0-kb *Bam*HI-*Sal*I fragment from clone 229.1 [23]). Posthybridizational washes were at 65°C with 0.1 × SSC–0.1% sodium dodecyl sulfate. (B) Physical maps of the six hsp70 genes and three different cDNA clones. Physical maps were made from subclones of each of the genes as indicated in Fig. 2. The open boxes of genes 2 to 6 indicate the coding sequence (for gene 1, this is tentative). The three cDNA clones that are shown were derived from libraries made with procyclic RNA, variant 117 RNA, and variant 221 RNA (designated p, 117, and 221, respectively). The arrow indicates the direction of transcription. The direction of transcription of gene 1 is unclear. h, *Hin*dIII; p, *Pst*I; e, *Eco*RI; cDNA p, a cDNA clone derived from procyclic trypanosomes.

trypanosomes as described by Van der Ploeg et al. (48) and digested with a fourfold excess of restriction endonuclease, and the products were resolved on 0.8% agarose gels. Plasmids were isolated as described by Birnboim and Doly (8). After transfer to nitrocellulose (44), the filters were prehybridized and subsequently hybridized with ^{32}P-labeled probes as previously described (25, 41). DNA was treated with 0.25 M HCl for 20 min before Southern transfer. After hybridization, filters were washed to a final stringency of 0.1 × SSC (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate)- or 3 × SSC–0.1% sodium dodecyl sulfate at 65°C, as indicated for individual experiments.

Construction and screening of genomic and cDNA libraries. Nuclear DNA of variant 118b was digested with the restriction enzyme *Hin*dIII. In the library made with partially digested DNA, a fraction of 15 kilobases (kb) and larger was used for ligation. Nuclear DNA was ligated into the plasmid vectors pAT153 (library I) and pBR322 (library II) as described by Van der Ploeg et al. (52). The ligation mixtures were used to transform *Escherichia coli* HB101. About 10^5 transformants were screened from each library with a ^{32}P-labeled probe of *Drosophila* hsp70 (1.0-kb *Bam*HI-*Sal*I fragment; clone 229.1) (23). The positively hybridizing clones were purified and then used for subsequent characterizations. The cDNA clone banks were a gift from C. Clayton; they were propagated in *E. coli* C600.

RNA preparation and analysis. Total RNA was isolated by LiCl precipitation as described by Auffray and Rougeon (2). Northern analysis of RNA was performed as described by Boedtker (9). S1 nuclease protection analysis of RNA was carried out essentially as described by Dudler and Travers (17).

DNA sequence analysis. Sequence analysis was performed by the dideoxy chain termination method (43). Three approaches were used to generate suitable DNA fragments for subcloning into the replicative forms of bacteriophage M13, mp18 and mp19. In method 1, DNAs were cleaved with restriction enzymes and ligated into the appropriate M13 vector. In methods 2 and 3, DNA subclones were treated with BAL 31 exonuclease or exonuclease III for various times at room temperature to generate a series of overlapping clones (21); these were subsequently ligated into the M13 vectors. Sequence analysis was performed on hsp70 subclones 10 and 11 (see Fig. 2).

Sandwich hybridization. In this procedure, 5 μg of hsp70 clone 10, 0.5 μg of cDNA clone TcV118-2, and 1 μg of nuclear DNA were digested with *Hin*dIII, *Pst*I, and *Pvu*II, respectively. The DNA was size separated and transferred

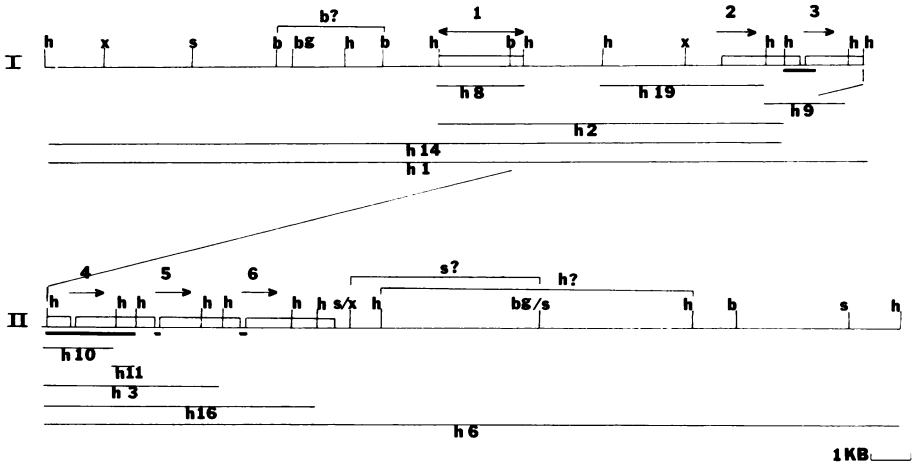


FIG. 2. Physical map of the hsp70 locus in *T. brucei*. Arrays I and II show the physical maps of the hsp70 clones. The open boxes numbered 1 through 6 indicate the hsp70 coding sequences. Arrows give the direction of transcription. We do not know the direction of transcription of gene 1. The hsp70 genes in arrays I and II are most likely linked (as shown by the diagonal line), because single 35-kb *Bgl*II and *Bam*HI fragments and a 30-kb *Bam*HI-*Bgl*II fragment hybridized with the hsp70 probe in nuclear DNA. The thick bars below the physical maps indicate the different areas for which nucleotide sequence was obtained. The overlapping clones, numbered h1 to h19 are shown below the physical maps. b, *Bam*HI; bg, *Bgl*II; h, *Hin*dIII; s, *Sal*I; x, *Xba*I. Question marks indicate that the restriction map has not been determined for the area specified.

to nitrocellulose. One filter was prehybridized in S1 hybridization buffer, supplemented with 50 μg of tRNA per ml, and hybridized in the same buffer with 500 μg of total RNA from variant 118 per ml for 15 h at 42°C. After hybridization, nonspecifically bound RNA was removed by washing the filter for 60 min at 65°C and 3 × SSC. A filter prehybridized with variant 118 RNA and one without RNA were then hybridized with a ^{32}P-labeled 22-mer, complementary to the mini-exon, as described by De Lange et al. (16).

RESULTS AND DISCUSSION

Analysis of *T. brucei* hsp70 genes. We determined the presence and organization of hsp70-encoding genes to analyze transcription regulation in *T. brucei*. At a high stringency of hybridization, the *Drosophila* hsp70 coding sequence (50) and the *T. brucei* hsp70 genes (Fig. 1A) recognize several restriction enzyme fragments in *T. brucei* nuclear DNA. From two different *T. brucei* nuclear DNA libraries, we isolated most of the hsp70 genes (Fig. 1B; see Materials and Methods for details). The physical maps of the hsp70-related sequences are shown in Fig. 2. The *T. brucei* hsp70 genes indicated in arrays I and II (Fig. 2) exist in one large locus, as deduced from hybridizations of nuclear DNA digested with the enzymes *Bgl*II and *Bam*HI (data not shown). We identified five genes in tandem (genes 2 to 6) and a single gene (gene 1), which is separated from the tandem array by several kilobase pairs. Because this gene lacks an hsp70-related 3′ end, it may be a pseudogene. A seventh band was detected in the hybridization in Fig. 1A (labeled 7). This band hybridizes only weakly when compared with the bands for the other six genes and might represent a more divergent gene copy.

To determine the transcriptionally active genes and the lengths of the intergenic regions, we isolated cDNA clones from an insect form (procyclic trypanosomes, which are grown at 25°C) and two bloodstream form cDNA libraries of *T. brucei*. The physical maps of these cDNA clones, which share a *Pst*I restriction enzyme site at their 3′ ends that is absent from gene 1, show that at least one of genes 2 to 6 is transcribed (Fig. 1B and 2A).

Sequence analysis of *T. brucei* hsp70 genes. To establish whether transcription of the hsp70 genes in *T. brucei* is regulated by a conserved eucaryotic promoter immediately flanking the gene, we determined the sequence of one of the hsp70 genes (gene 4; Fig. 2). The complete sequence of gene 4 is shown in Fig. 3A. The coding sequence is 68% conserved at the nucleotide level, compared with the sequence of the distal *Drosophila* hsp70 gene from locus 87A7 (24). Homology is restricted to the coding sequence, which has a 660-amino acid open reading frame, with 69% homology to human hsp70 and 65% homology to *Drosophila* hsp70 (Fig. 3B). A relatively high level of nucleotide sequence conservation is found in the comparison of heat shock proteins among other species. The conservation is higher than necessary given the degeneracy of the genetic code. Hunt and Morimoto (22) concluded, therefore, that a function may exist in the nucleotide sequence itself.

To determine the extent of the intergenic regions between these genes, we located the 5′ and the 3′ ends of the mRNA. In Northern analysis, *T. brucei* RNA reveals a 2,300-nt hsp70 transcript (50). Some variants have slightly shorter, less abundant RNA (D. J. Glass and L. H. T. Van der Ploeg, unpublished data). A comparison of the nucleotide sequence of the 3′ end of a procyclic-derived cDNA clone and the genomic clone shows that addition of the poly(A) tail occurred at the end of the sequence TTTTTATTTT(A) (Fig. 4). The polyadenylation signal AAUAAA of higher eucaryotes is absent, as in other *T. brucei* genes (12). Since two single-base substitutions were detected between the cDNA sequence and the genomic sequence, it is clear that the cDNA was derived from mRNA transcribed from another hsp70 gene (gene 3, 5, or 6) (Fig. 2). We determined the 5′ end of the mRNA by S1 nuclease protection experiments with *Bst*EII (Fig. 5) and *Eco*RI ^{32}P-end-labeled fragments (data not shown). The two S1 nuclease-protected fragments shown in Fig. 5, lane 4, are 158 and 825 nt long. The 158-nt RNA is derived from the 5′ end of the gene, whereas the 825-nt fragment is derived from protection of the only internal *Bst*EII fragment. RNA was analyzed with clone 16 (three hsp70 genes) and clone 10 (one repeat unit only) (Fig.

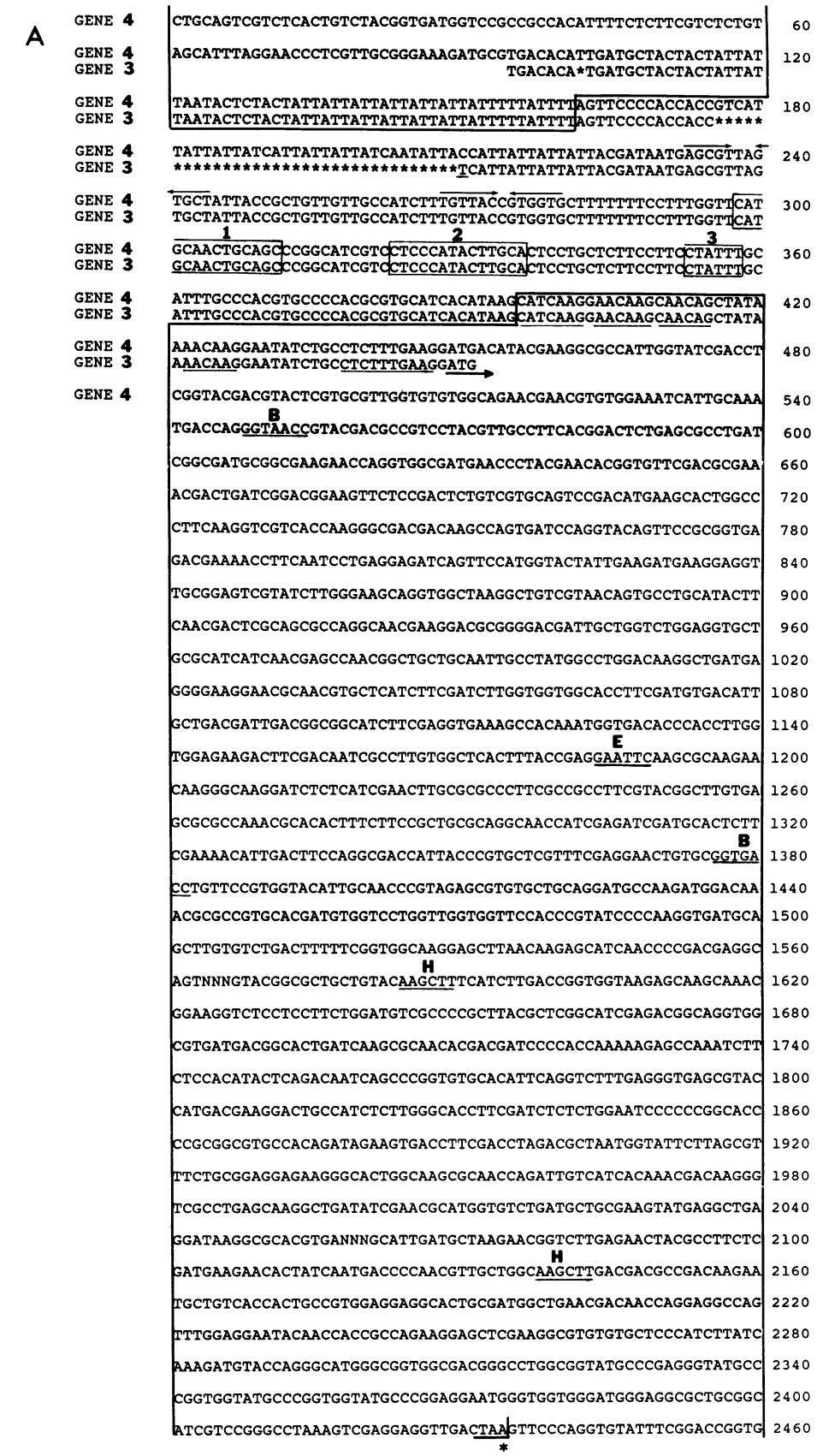

A
GENE 4
GENE 3
1
2
3
B
E
H
H
B
*

FIG. 3. Sequence analysis and protein homologies. (A) Sequence analysis of hsp70 genes in *T. brucei*. Clones h10 and h11 were digested with *Hin*dIII and subcloned into M13. These were treated as described in Materials and Methods to obtain overlapping clones. Since the genes 4, 5, and 6 appear as identical repeats, we do not know from which of these clones 10 and 11 are derived. The sequence is presented starting with the 3′ end of one of the three tandemly arranged genes, arbitrarily numbered 3, spanning nt 1 to 160 (boxed). Making up the intergenic region are nt 161 to 394 (not boxed), as determined by cDNA sequencing and S1 nuclease mapping. Demonstrating the open reading frame and part of the 5′ and 3′ untranslated mRNA extensions are nt 395 to 2434 (boxed). A bold arrow is drawn under the translational start site. The translation stop codon TAA is indicated with an asterisk. The sequence stops at nt 2460, which precedes the *Pst*I site at nt 1 to 6. Box 1 is the first region that shares homology to the HSE of the *Drosophila* hsp70 promoter, with five of eight consensus nucleotides conserved. The box has a strong diad symmetry. Box 2 has 8 of 14 bases in common with the HSE sequence, but only three of eight of the consensus nucleotides are maintained. However, the spacing between the two boxes is identical to that between sites 1 and 2 of the *Drosophila* hsp70 gene (46). Box 3 has homology to the *T. brucei* mini-exon TATA box. It maintains the same spacing to HSE box 2 as in *Drosophila* hsp70. B, *Bst*EII; E, *Eco*RI; H, *Hin*dIII. Next, from positions 96 to 431, sequence data from gene 3 are presented underneath gene 4 to demonstrate that the intergenic region between genes 2 and 3 has a deletion (marked by asterisks) of the region spanning nt 176 to 209, with a base substitution at 210 (nucleotide underlined) and a nucleotide deletion at 102. Finally, opposing arrows at positions 232 to 244 and 268 to 280 highlight a diad symmetry in which G-T base pairs are considered stable. (B) Protein homologies. Boxed areas demonstrate regions of homology between human hsp70 (22) (line 1, 69% homologous), trypanosome hsp70 (line 2), and *Drosophila* hsp70 (24) (line 3, 65% homologous). Gaps were introduced to maximize homology. Four asterisks in the *T. brucei* sequence represent amino acids that could not be determined because of compressions in the DNA sequence.

2). Each of these gave identical S1 nuclease-protected fragments. The intergenic region between genes 3 and 4 is therefore 234 bp in length (Fig. 3A).

We established the extent of conservation between the intergenic regions of the different genes to identify regulatory elements. Since at least one of hsp70 genes 3, 5, or 6 is transcribed, a comparison shows the degree of conservation in the intergenic regions of all potentially functional heat shock genes (Fig. 3A). The intergenic regions from genes 4 to 6 (clone 16) gave identical nucleotide sequences in the

```
HSP 70 GENE   ACATTTTCTC TTCGTCTCTG TAGCATTTAG GAACCCTCGT TGCGGGAAAG
HSP 70 cDNA   ACATTTTCTC TTCGTCTCTG TAGCATTTAG GAACCCTCGT TACGGGAAAG

HSP 70 GENE   ATGCGTGACA CATTGATGCT ACTACTATTA TTAATACTCT ACTATTATTA
HSP 70 cDNA   GTGCGTGACA CATTGATGCT ACTACTATTA TTAATACTCT ACTATTATTA

HSP 70 GENE   TTATTATTAT TATTTTTATT TTAGTTCCCC ACCACCGTCA TTAT
hSP 70 cDNA   TTATTATTAT TATTTTTATT TTAAAAAAAA AAAAAA(A)10
```

FIG. 4. 3′ end of hsp70 cDNA and genomic hsp70 sequences. A sequence comparison is shown for the 3′ end of a hsp70 cDNA clone (indicated in Fig. 1) derived from procyclic trypanosomes, which transcribe hsp70 genes at a low level, and a genomic hsp70 clone. Two sequence deviations are boxed.

region located downstream of the poly(A) addition site (200 nt were compared). Since the physical maps of genes 4 to 6 are also identical, we believe that the intergenic regions of these three genes are highly conserved. Furthermore, the intergenic region between genes 2 and 3 is completely conserved, except for a 34-bp deletion in a TTA repeat and one base substitution (Fig. 3A).

The first gene of the tandem array (gene 2, Fig. 2) has a different restriction enzyme map in its immediate upstream region. However, sequence analysis of its 5′ flanking area showed that the same sequence which is in the intergenic region of genes 3 and 4 is also located upstream of gene 2 but at least 1 kb in front of the gene (data not shown).

If the temperature-sensitive transcription of these heat shock genes results from transcription initiation in front of each gene, we expect to find HSE sequences in the conserved regions that flank these five genes.

Conserved sequence elements of the hsp70 intergenic regions. We looked for putative regulatory signals in the intergenic region of the *T. brucei* hsp70 genes by comparisons of the sequence of this region with the HSE consensus sequence and with the nontranscribed intergenic region sequence of mini-exon genes. Three palindromic sequences can be found in the intergenic regions of the hsp70 genes (Fig. 3A, sequences with arrows and the sequence in box 1). One of these shares five out of eight nucleotides with the HSE consensus sequence (Fig. 3A, box 1; Fig. 6). All base substitutions are such that a perfect palindrome (TGCA) is maintained. The other palindromes, which are also present in front of all five genes, are unrelated to the HSE. A second HSE-related sequence is found 10 bp downstream of the one described. Its homology to the 14-nt HSE sequence of *Drosophila* spp. is 8/14, but its relatedness to the consensus is only 3/8. Since divergence in the HSE sequences exists among different species (Fig. 6), we believe that the most conserved HSE (Fig. 3A, box 1) probably has a function in transcriptional regulation. We are currently testing its function by the transfer of trypanosome hsp70 genes and the analysis of their transcripts in other eucaryotic cells.

That the intergenic region indeed contains a PolII promoter is further supported by the fact that with HeLa cell extracts in vitro, transcription initiates at positions +2 and +10 from a possible TATA box (Fig. 2, box 3). Furthermore, a probe located directly upstream of the 5′ end of the mature mRNA detects a 2,300-nt hsp70 RNA molecule in *T. brucei* total RNA. This RNA may be the precursor to the mature mRNA which carries the mini-exon, and it could have resulted from initiation of transcription at this TATA box (L. H. T. Van der Ploeg, unpublished results).

Furthermore, two related 120-bp sequences are present that are homologous to the intergenic regions of the *Trypanosoma vivax* and *Trypanosoma cruzi* mini-exon

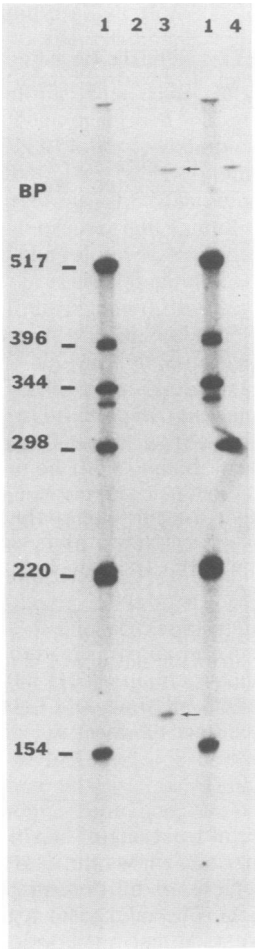


FIG. 5. S1 Protection analysis of the 5′ end of hsp70 mRNA. Clone 10 was digested with *Bst*EII, ^{32}P end labeled, and hybridized with *T. brucei* total RNA. We determined from S1 nuclease protection experiments with *Eco*RI-digested and ^{32}P-end-labeled DNA that the 5′ end of the RNA is located at approximately 825 nt upstream of the single *Eco*RI site of clone 10 (Fig. 1B). The shortest protected fragments in the *Bst*EII digestion, therefore, give the location of the 5′ end of the hsp70 RNA. Lanes: 1, labeled pBR322 *Hin*fI size standard; 2, renaturation control; 3, S1 nuclease-protected fragments obtained from the end-labeled *Bst*EII restriction fragments with total *T. brucei* mRNA made from variant 1.198 (the two bands marked by arrows correspond to the 825-nt internal *Bst*EII fragment and the 158-nt fragment derived from the 5′ end of the hsp70 gene); 4, 825-nt *Bst*EII input fragment, which is a double band with two end-labeled restriction fragments.

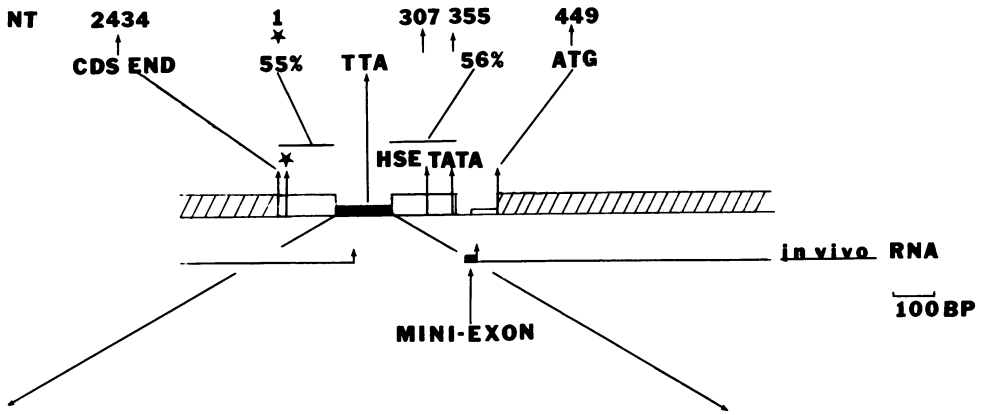
```
T.BRUCEI 1 CATGCAACTGCAGC  5/8
HUMAN      CTGGAATATTCCCG  8/8  Hunt and Morimoto, 1985
CHICKEN    CTGGCAGGTTCCAG  7/8  Morimoto et al., 1986
DROSOPHILA CTCGAATGTTCGCG  8/8  Pelham, 1982
MOUSE      CTGAGAGTTTCCAG  6/8  Hunt and Morimoto, unpublished
HSE        CNNGAANNTTCNNG       Topol et al., 1985

DROSOPHILA CTCGAATGTTCGCG       Topol et al., 1985
           *** **  * **
T.BRUCEI 2 CTCCCATACTTGCA  3/8
```

FIG. 6. Comparison of the *T. brucei* HSE-related sequences with sequences derived from human, chicken, *Drosophila*, and mouse hsp70 genes. The consensus sequence proposed by Topol et al. (46) is marked HSE. The underlined nucleotides constitute the bases of the potential cruciform. Asterisks demonstrate the conserved bases between the *Drosophila* HSE and *T. brucei* box 2. Altogether, 8 of 14 bases are conserved, although only 3 of 8 fall under the consensus.

genes. They are 55% homologous to nt −130 through −280 and −150 through −300 of the *T. vivax* and *T. cruzi* mini-exon genes, respectively (as schematically indicated in Fig. 7). One sequence is located in the 3′ untranslated sequence of the hsp70 mRNA, and one is located immediately upstream of the 5′ end of the hsp70 mRNA start site. These conserved elements are separated by a 120-bp TTA repeat, which is homologous to a region that is located at the 3′ end of the *T. brucei*, *T. vivax*, and *T. cruzi* mini-exon genes (Fig. 7, bottom panel). This stretch of DNA consists of TTA repeats, T-homopolymer stretches, and C+G-rich domains. It may function as a transcription terminator of mini-exon genes (15, 29). The 3′ end of the hsp70 mRNA is also located in this domain. The presence of a sequence that may function as a transcription terminator gives further evidence that the intergenic regions contain a heat shock gene promoter.

***T. brucei* hsp70 mRNA with a mini-exon at its 5′ end.** In vivo, the 5′ end of hsp70 mRNA is located 37 nt downstream of the putative TATA box (Fig. 3A, box 3). However, all mRNAs analyzed in *T. brucei* carry a 35-nt mini-exon at their 5′ end (53). This mini-exon is encoded by another gene and is therefore added in *trans*. If there is a mini-exon at the 5′ end of the hsp70 mRNA, then the S1 nuclease analysis has not mapped the real 5′ end of this mRNA. In addition, the detailed analysis of the hsp70 promoter region may allow us to determine how mini-exon addition occurs. Sequence analysis of the 5′ ends of the hsp70 cDNA clones showed that they all ended 3′ of the site mapped by S1 nuclease protection analysis. One clone ended at location +63 from box 3. Since none of the cDNA clones was full length, we determined the presence of a mini-exon in a sandwich hybridization with *T. brucei* RNA hybridized to a genomic hsp70 clone and subsequently probed with a ^{32}P-labeled synthetic oligonucleotide, complementary to the mini-exon (Fig. 8). Both variant surface glycoprotein 118 (VSG118) and hsp70 mRNAs hybridized with the 22-mer probe of the




```
HSP 70   TATTATTAATACTCTACTATTATTATTATTATTATTTTTATTTTAGTTCCCCACCAC
           ******            ******************

HSP 70   CGTCATATTATTATCATTATTATTATCAATATTACCATTATTATTATTACGATAATGAGC
              ********   *********    *****   ************

HSP 70   GTTAGTGCTATTACCGCTGTTGTTGCCATCTTTCTTACCGTGGTGCTTTTTTTCCTTTGG
           ***      ***

T.VIVAX  AAAACAAATTTTTGTTTTGTATATTTTTTTTTATTTTTTATTATTTATTATTATTTATTAT
                                          ****** ********* *******

T.VIVAX  TATTTTTTTTTATTTGTGTGTTATTATTATTTTTTTTCGTTTTTCCTCCCCGCACACCGCG
           **                 *********
```

FIG. 7. Conserved sequence elements in the hsp70 intergenic regions. The conserved sequence elements are indicated in the top panel. ▨, Coding sequence; ▭, regions that are homologous (55 and 56%) to the *T. vivax* and the *T. cruzi* mini-exon intergenic regions (15); ▬, TTA repeat in which the 3′ end of the mRNA is mapped. This DNA stretch, which consists of TTA repeats, T-homopolymer stretches, and C+G-rich domains, is found with various permutations of these elements at the 3′ ends of *T. brucei*, *T. vivax*, and *T. cruzi* mini-exon genes. The locations of the ends of the RNA are indicated below the map. The nucleotide numbering at the top is identical to the numbering in the nucleotide sequence in Fig. 3A; nt 1 is also indicated by a star. The sequence in the bottom panel compares the conserved elements in the *T. brucei* hsp70 and in the *T. vivax* mini-exon 3′ ends. Asterisks indicate TTA repeats. T-homopolymer stretches are underlined.

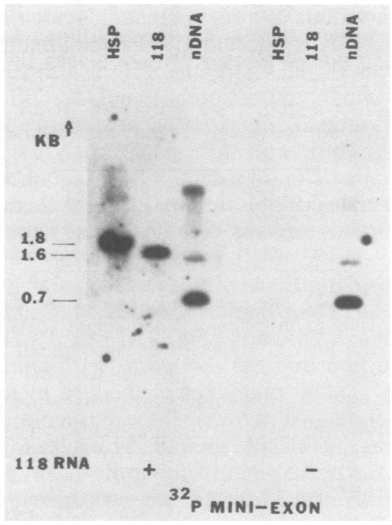


FIG. 8. Mini-exon location on hsp70 mRNA. Because none of the cDNA clones was full length, we determined the presence of a mini-exon in a sandwich hybridization. A 5-μg amount of *T. brucei* clone 10 (1.8-kb *Hin*dIII fragment with hsp70 coding sequence; Fig. 2) was digested with *Hin*dIII; 0.5 μg of cDNA clone Tcv118-2 (6) was digested with *Pst*I, leaving a 1.6-kb insert; and 1 μg of *T. brucei* nuclear DNA was digested with *Pvu*II. All these were size separated in a 0.7% agarose gel and transferred to nitrocellulose filters. Only one filter was hybridized with variant 118 RNA (+), as described in Materials and Methods. The filters were washed at 65°C and 3 × SSC for 60 min after hybridization. Both filters were subsequently hybridized with a ^{32}P-labeled synthetic oligonucleotide complementary to the mini-exon, as indicated in Materials and Methods.

mini-exon specifically, thus proving that a mini-exon is present at the 5′ end of hsp70 mRNA and the VSG118 mRNA. The nuclear DNA, which shows hybridization to the mini-exon repeats only, and the linearized plasmids, which do not hybridize with the 22-mer probe, are controls for the hybridization specificity. The mini-exon is not encoded downstream of the HSE or TATA box. Furthermore, mini-exon homologous sequences are absent from the intergenic regions. A function of the mini-exon in priming is therefore unlikely.

Transcription of VSG genes at 42°C. We wanted to know whether the mRNA levels of other genes are regulated in a temperature-dependent manner. We therefore determined the levels of mature hsp70, VSG118, αβ-tubulin mRNA, and rRNA after a relatively short heat shock of trypanosomes for 120 min at 42°C. The parasite is normally subjected to high temperatures for extended periods of time in the febrile response at peak parasitemias (maximum, 41°C). Northern analysis of the RNAs is shown in Fig. 9. When measured at 37 versus 42°C, VSG118 mRNA was at least 50-fold more abundant, hsp70 mRNA was 10-fold more abundant, and the αβ-tubulin mRNA level was 2-fold more abundant, whereas rRNA remained unaffected (deduced by scanning of the autoradiograms). These data indicate that VSG118 gene mRNA levels may be regulated by heat shock. This may result from enhanced transcription, as for heat shock genes. Given the period of the heat shock (120 min) and the fact that the half-life of VSG118 mRNA is estimated to be 30 min (37), it is not possible that the accumulation of VSG118 mRNA resulted from a diminished rate of mRNA turnover. Only a fourfold increase could maximally have resulted in the absence of mRNA turnover. Thus, both hsp70 PolII and VSG118 gene transcription (presumably transcribed by PolI [28]) may be regulated in a temperature-dependent manner. The data strengthen the importance of heat shock in the development of this parasitic protozoan.

Concluding remarks. Since all mRNAs in trypanosomes, including that of hsp70 (Fig. 8), have a 35-nt mini-exon spliced onto the 5′ end of the mRNA (53), splicing occurs with *T. brucei* hsp70 genes at 37°C. Only two other heat shock genes, a *Drosophila* hsp83-encoding gene and a maize hsp70-encoding gene, are spliced (42). This may reflect the finding that splicing is not efficient at nonbiological extreme heat shock temperatures, whereas it can be rescued by a short prior heat shock (54). We do not know whether heat shock gene transcription in *T. brucei* has a function in splicing. Mini-exon addition occurs between a mini-exon and a main exon, which are not contiguously encoded in the genome (for a review, see reference 11). With hsp70 genes, mini-exons are not encoded in the 2,460-bp hsp70 repeat. Therefore, either the highly conserved intergenic regions of the hsp70 genes must serve in the splicing of the mini-exon and main exon from polycistronic precursor RNAs or they promote transcription initiation of hsp70 genes. The latter could occur by the use of the mini-exon as a primer or by transcription that initiates independent of the mini-exon.

The *T. brucei* hsp70 genes are transcribed by a PolII because its transcription is blocked in the presence of 100 μg of α-amanitine per ml (C. S. Shea and L. H. T. Van der Ploeg, unpublished results). *T. brucei* PolII promoters have not been identified previously because it is not yet possible to transfer genes back into trypanosomes, nor is an in vitro transcription system available. Because of the strict conser-

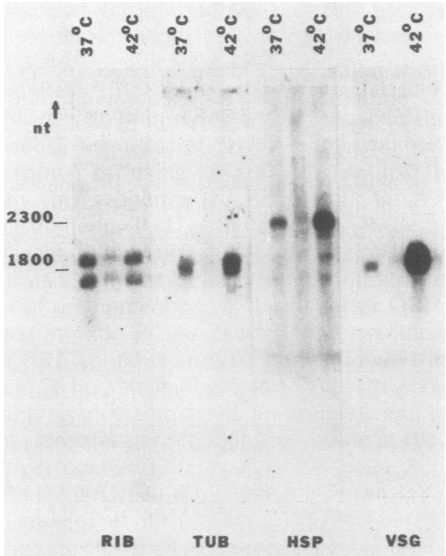


FIG. 9. Analysis of heat-shocked RNAs. Variant 1.8 trypanosomes were grown in a rat and harvested. From half of these trypanosomes, RNA was directly prepared (lanes indicated by 37°C). The remaining trypanosomes were maintained in culture medium at a density of 4 × 10^7/ml at 42°C for 120 min. Culture conditions were as described by Van der Ploeg et al. (50) (IMDM supplemented with 5 × 10^{-5} M 2-mercaptoethanol), after which RNA was made. The same nitrocellulose filter was probed with isolated fragments from a clone containing rRNA genes (RIB), αβ-tubulin genes (TUB), *Drosophila* hsp70 coding sequence (HSP), and a VSG gene 1.8 cDNA fragment from clone Tcv1.8-2 (VSG). Posthybridizational washes were at 65°C at 0.03 × SSC (RIB), 0.1 × SSC (TUB), 3 × SSC (HSP), or 0.1 × SSC (VSG).

vation of hsp promoters among different species, from plants to man, an unambiguous determination of the *T. brucei* promoter elements should be possible. We therefore determined the conserved sequence elements in the five intergenic region sequences of *T. brucei* hsp70 genes.

In the following discussion, we arbitrarily denote the nucleotide after the AG dinucleotide of the first 3′ splice site as no. 1. The DNA sequence between −6 and +51 at the 5′ end of the mRNA has six *T. brucei* specific 3′ splice sites: CATAAG, CATCAAG, ACAAG, CAACAG, ACAAG, and CTCTTTGAAG. These sequences are similar to the previously described 3′ splice sites for VSG117 (CATAAG) and VSG221 (CTCTTGAAG) (5, 10). In longer exposures of the S1 nuclease protection experiments, several protected fragments between 100 and 158 nt were detected (data not shown). These may have resulted from the use of alternate splice sites. This is suported by the observation of two sites of mini-exon addition in the calmodulin genes (47). A putative lariat branch point sequence is found at position −18 from the first 3′ splice site. The sequence GCCCCAC matches the consensus (Y)NYY(R)AY (55).

A putative hsp70 TATA box, CTATTT at position −37, is embedded in a 16-nt-long stretch of pyrimidines. This TATA box is similar to the putative TATA box of mini-exon genes, which consists of 8 pyrimidines: CTTATTTTT (13, 15). That this sequence may be the TATA box is supported by in vitro transcription with the hsp70 gene in HeLa cell extracts. Three minor and two major transcripts initiated in the intergenic region. The two major transcripts are mapped in S1 nuclease protection at positions +2 and +10 of this TATA box (S. Bondopadhyay and L. H. T. Van der Ploeg, unpublished data). In addition, a probe located immediately upstream of the first 3′ splice site in the putative intron detected a 2,300-nt RNA molecule in *T. brucei* total RNA, which may be the precursor to the mature mRNA. It could have resulted from the initiation of transcription at this TATA box (L. H. T. Van der Ploeg, unpublished data).

In summary, the conserved sequence elements of the intergenic regions are from 5′ to 3′: sequences that may function as a transcription terminator, two sequence blocks also found in the intergenic regions of mini-exon genes, an HSE at position −84; a putative TATA box at position −37 (Fig. 3A, box 3); a branch point consensus sequence at position −18; and six 3′ splice sites from positions −6 to +51. All of these indicate that transcription of hsp70 genes in *T. brucei* initiates from a normal PolII promoter. Furthermore, Bienz and Pelham (7) have shown that a close proximity of HSE and TATA sequences makes a CCAAT box redundant. This may explain why a CCAAT box is not found in between the HSE and TATA box in *T. brucei*.

We have presented evidence that supports the notion that a PolII promoter exists in the conserved intergenic region sequences. Sequences that are homologous to the mini-exon and that could indicate a function in priming are absent from the intergenic regions. It is, therefore, more likely that the mini-exon is added in *trans* by bimolecular splicing. RNA ligation, followed by splicing, cannot be excluded, but the use of alternate 3′ splice sites for calmodulin (47) and perhaps hsp70 genes renders this less likely.

ACKNOWLEDGMENTS

We thank Mary Gwo-Shu Lee, Cathy Shea, David Misek, Argiris Efstratiadis, and Charles Cantor for critical reading of the manuscript and discussion. We thank Scott Zeitlin and Marcello Soares for technical advice, Hon Ip for advice and help with the computer analysis, and Christine Clayton for a gift of the *T. brucei* cDNA libraries.

This work was supported by Public Health Service grant AI21784 from the National Institutes of Health, a Searle scholarship and Irma T. Hirschl award to L.H.T.V.D.P., and a grant from the John D. and Catherine T. MacArthur Foundation.

LITERATURE CITED

1. **Atkinson, B. G., and D. B. Walden (ed.).** 1985. Changes in eukaryotic gene expression in response to environmental stress. Academic Press, Inc., New York.
2. **Auffray, C., and F. Rougeon.** 1980. Purification of mouse immunoglobulin heavy-chain RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
3. **Ayme, A., R. Southgate, and A. Tissieres.** 1985. Nucleotide sequences responsible for the thermal inducibility of the *Drosophila* small heat-shock protein genes in monkey cos cells. J. Mol. Biol. **182**:469–475.
4. **Bensaude, O., C. Babinet, M. Morange, and F. Jacob.** 1983. Heat shock proteins, first major products of zygotic gene activity in mouse embryo. Nature (London) **305**:331–333.
5. **Bernards, A., J. M. Kooter, and P. Borst.** 1985. Structure and transcription of a telomeric surface antigen gene of *Trypanosoma brucei*. Mol. Cell. Biol. **5**:545–553.
6. **Bernards, A., L. H. T. Van der Ploeg, A. C. C. Frasch, P. Borst, J. C. Boothroyd, S. Coleman, and G. A. M. Cross.** 1981. Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3′ end. Cell **27**:497–505.
7. **Bienz, M., and H. R. B. Pelham.** 1986. Heat shock regulatory elements function as an inducible enhancer in the *Xenopus* Hsp 70 gene and when linked to a heterologous promoter. Cell **45**:753–760.
8. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
9. **Boedtker, H.** 1971. Conformation-independent molecular weight determination of RNA by gel electrophoresis. Biochim. Biophys. Acta **240**:448–453.
10. **Boothroyd, J. C., and G. A. M. Cross.** 1982. Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene **20**:281–289.
11. **Borst, P.** 1986. Discontinuous transcription and antigenic variation in trypanosomes. Annu. Rev. Biochem. **55**:701–732.
12. **Borst, P., and G. A. M. Cross.** 1982. Molecular basis for trypanosome antigenic variation. Cell **29**:291–303.
13. **Campbell, D. A., D. A. Thornton, and J. C. Boothroyd.** 1984. Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature (London) **311**:350–355.
14. **Clayton, C.** 1985. Structure and regulated expression of genes encoding fructose biphosphate aldolase in *Trypanosoma brucei*. EMBO J. **4**:2997–3003.
15. **De Lange, T., T. M. Berkvens, H. J. G. Veerman, A. C. C. Frasch, J. D. Barry, and P. Borst.** 1984. Comparison of the genes coding for the common 5′ terminal sequence of the messenger RNAs in three trypanosome species. Nucleic Acids Res. **12**:4431–4443.
16. **De Lange, T., A. Y. C. Liu, L. H. T. Van der Ploeg, P. Borst, M. C. Tromp, and J. Van Boom.** 1983. Tandem repetition of the 5′ mini-exon genes: a multiple promoter for VSG gene transcription? Cell **34**:891–900.
17. **Dudler, R., and A. A. Travers.** 1984. Upstream elements necessary for optimal function of the Hsp 70 promoter in transformed flies. Cell **38**:391–398.
18. **Fairlamb, A. H., P. O. Weisvogel, J. H. J. Hoeijmakers, and P. Borst.** 1978. Isolation and characterization of kinetoplast DNA from bloodstream form of *Trypanosoma brucei*. J. Cell Biol. **76**:293–309.
19. **Glasgow, J. P.** 1970. The *Glossina* community, p. 348–381. *In* H. W. Mulligan (ed.), The African trypanosomiasis. John Wiley & Sons, Inc., New York.
20. **Gonzalez, A., T. J. Lerner, M. Huecas, B. Sosa-Pineda, N. Nogueira, and P. M. Lizardi.** 1985. Apparent generation of a

segmented mRNA from two separate tandem gene families in *Trypanosoma cruzi*. Nucleic Acids Res. **13**:5789–5804.
21. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
22. **Hunt, C., and R. Morimoto.** 1985. Conserved features of eukaryotic Hsp70 genes revealed by comparison with the nucleotide sequence of human Hsp70. Proc. Natl. Acad. Sci. USA **82**:6455–6459.
23. **Ingolia, T. D., and E. A. Craig.** 1982. *Drosophila* gene related to the major heat shock-induced gene is transcribed at normal temperatures and not induced by heat shock. Proc. Natl. Acad. Sci. USA **79**:525–529.
24. **Ingolia, T. D., E. A. Craig, and B. J. McCarthy.** 1980. Sequence of three copies of the gene for the major *Drosophila* heat shock induced protein and their flanking regions. Cell **21**:669–679.
25. **Jeffreys, A. J., and R. A. Flavell.** 1977. A physical map of the DNA regions flanking the rabbit β globin gene. Cell **12**:429–449.
26. **Killick-Kendrick, R.** 1976. Biology of *Leishmania* in phlebotomine sandflies, p. 394–460. *In* W. H. R. Lumsden and D. E. Evans (ed.), Biology of the kinetoplastida, vol. 2. Academic Press, Inc. (London), Ltd., London.
27. **Kirk, D. L., and M. M. Kirk.** 1986. Heat shock elicits production of sexual inducer in Volvox. Science **231**:51–54.
28. **Kooter, J. M., and P. Borst.** 1984. α-Amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucleic Acids Res. **12**:9457–9472.
29. **Kooter, J. M., T. De Lange, and P. Borst.** 1984. Discontinuous synthesis of mRNA in trypanosomes. EMBO J. **3**:2387–2392.
30. **Kurtz, S., J. Rossi, L. Petko, and S. Lindquist.** 1986. An ancient developmental induction: heat-shock proteins induced in sporulation and oogenesis. Science **231**:1154–1157.
31. **Lawrence, F., and M. Robert-Gero.** 1985. Induction of heat shock and stress proteins in promastigotes of three *Leishmania* species. Proc. Natl. Acad. Sci. USA **82**:4414–4417.
32. **Mandaleswahr, K. S., and J. Yu.** 1984. Accumulation of a heat shock-like protein during differentiation of human erythroid cell line K562. Nature (London) **309**:631–633.
33. **Michels, P. A. M., A. Y. C. Liu, A. Bernards, P. Sloof, M. M. W. Van der Bijl, A. H. Schinkel, H. H. Menke, P. Borst, G. H. Veeneman, M. C. Tromp, and J. H. Van Boom.** 1983. Activation of the genes for variant surface glycoproteins 117 and 118 in *Trypanosoma brucei*. J. Mol. Biol. **166**:537–556.
34. **Michels, P. A. M., L. H. T. Van der Ploeg, A. Y. C. Liu, and P. Borst.** 1984. The activation and reactivation of an expression-linked gene copy of a variant surface glycoprotein in *Trypanosoma brucei*. EMBO J. **3**:1345–1351.
34a. **Morimoto, R. I., C. Hunt, S. Y. Huang, K. L. Berg, and S. S. Banerji.** 1986. Organization, nucleotide sequence and transcription of the chicken Hsp 70 gene. J. Biol. Chem. **261**:12692–12699.
35. **Nelson, R. G., M. Parsons, P. J. Barr, K. Stuart, M. Selkirk, and N. Agabian.** 1983. Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in *Trypanosoma brucei*. Cell **34**:901–909.
36. **Osinga, K. A., B. W. Swinkels, W. C. Gibson, P. Borst, G. H. Veeneman, J. H. Van Boom, P. A. M. Michels, and F. Opperdoes.** 1985. Topogenesis of microbody enzymes: a sequence comparison of the genes for the glycosomal (microbody) and cystolic phosphoglycerate kinases of *Trypanosoma brucei*. EMBO J. **4**:3811–3817.
37. **Overath, P., J. Czichos, and C. Nonnengaesser.** 1983. Repression of glycoprotein synthesis and release of surface coat during transformation of *Trypanosoma brucei*. EMBO J. **2**:1721–1728.
38. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an Hsp 70 gene. Cell **37**:273–283.
39. **Parsons, M., R. G. Nelson, K. P. Watkins, and N. Agabian.** 1984. Trypanosome mRNAs share a common 5′ spliced leader sequence. Cell **38**:309–316.
40. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the *Drosophila* Hsp 70 heat-shock gene. Cell **30**:517–528.
41. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling DNA to high specific activity in vitro by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
42. **Rochester, D. E., J. A. Winer, and D. M. Shah.** 1986. The structure and expression of maize genes encoding the major heat shock protein, Hsp 70. EMBO J. **5**:451–458.
43. **Sanger, F., A. R. Coulsen, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA-sequencing. J. Mol. Biol. **143**:161–178.
44. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
45. **Thomashow, L. S., M. Milhausen, W. J. Rutter, and N. Agabian.** 1983. Tubulin genes are tandemly linked and clustered in the genome of *Trypanosoma brucei*. Cell **32**:35–43.
46. **Topol, J., D. M. Ruden, and C. S. Parker.** 1985. Sequences required for in vitro transcriptional activation of a *Drosophila* hsp 70 gene. Cell **42**:527–537.
47. **Tschudi, C., A. S. Young, L. Ruben, C. L. Patton, and F. F. Richards.** 1985. Calmodulin genes in trypanosomes are tandemly repeated and produce multiple mRNAs with a common 5′ leader sequence. Proc. Natl. Acad. Sci. USA **82**:3998–4002.
48. **Van der Ploeg, L. H. T., A. Bernards, F. A. M. Rijsewijk, and P. Borst.** 1982. Characterization of the DNA duplication-transposition that controls the expression of two genes for variant surface glycoproteins in *Trypanosoma brucei*. Nucleic Acids Res. **10**:593–609.
49. **Van der Ploeg, L. H. T., A. W. C. A. Cornelissen, P. A. M. Michels, and P. Borst.** 1984. Chromosome rearrangements in *Trypanosoma brucei*. Cell **39**:213–221.
50. **Van der Ploeg, L. H. T., S. H. Giannini, and C. R. Cantor.** 1985. Heat shock genes: regulatory role for differentiation in parasitic protozoa. Science **228**:1443–1446.
51. **Van der Ploeg, L. H. T., A. Y. C. Liu, P. A. M. Michels, T. De Lange, P. Borst, H. K. Majumder, H. Weber, G. H. Veeneman, and J. Van Boom.** 1982. RNA splicing is required to make the mRNA for a variant surface antigen in trypanosomes. Nucleic Acids Res. **10**:3591–3604.
52. **Van der Ploeg, L. H. T., D. Valerio, T. De Lange, A. Bernards, P. Borst, and F. G. Grosveld.** 1982. An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res. **10**:5905–5923.
53. **Walder, J. A., P. S. Eder, D. M. Engman, S. T. Brentano, R. Y. Walder, D. S. Knutzon, D. M. Dorfman, and J. E. Donelson.** 1986. The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNA's. Science **233**:569–571.
54. **Yost, J. H., and S. Lindquist.** 1986. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell **45**:185–193.
55. **Zeitlin, S., and A. Efstratiadis.** 1984. In vivo splicing products of the rabbit β-globin pre-mRNA. Cell **39**:589–602.