The EMBO Journal vol.7 no.4 pp.1121 – 1127, 1988

# The genomic organization and transcription of the ubiquitin genes of *Trypanosoma cruzi*

**John Swindle[1,2,5], Janet Ajioka[1,5], Harvey Eisen[1,2], Bishnu Sanwal[2,3], Catherine Jacquemot[2], Zoe Browder[4] and Gregory Buck[2,4]**

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, USA, [2]Unit d'Immunoparasitology, Institut Pasteur, 28, rue du Dr Roux, 75015 Paris, France, [3]Department of Biochemistry, Health Sciences Center, University of Western Ontario, London, Ontario, Canada N6A 5C1, [4]Department of Microbiology and Immunology, Virginia Commonwealth University, Medical College of Virginia, Box 678 MCV Station, Richmond, VA 23298, USA

[5]Present address: Department of Microbiology and Immunology, University of Tennessee – Memphis, Center for the Health Sciences, 858 Madison Avenue, Memphis, TN 38163, USA

Communicated by H.Eisen

**We describe here the organization of the ubiquitin genes of the parasitic protozoan *Trypanosoma cruzi*. *T.cruzi* contains >100 ubiquitin coding sequences all of which are clustered into a 27 kb segment of the genome. Two types of ubiquitin coding sequences were found. There are five fusion genes (*FUS*1 – 5) consisting of a ubiquitin coding sequence fused to a basic non-ubiquitin sequence. The *T.cruzi* ubiquitin fusion protein is 84% homologous to the product of the *UBI* gene of *Saccharomyces cerevisiae*. The non-ubiquitin domains of the two proteins are 67% homologous. There are five polyubiquitin coding genes (*PUB*) each consisting of varying lengths of polyubiquitin coding sequence and terminating with a single copy of the larger fusion gene. Transcription of the ubiquitin genes results in the generation of six major poly(A)$^+$ mRNAs. The pattern of transcription accurately reflects the genomic organization, in that the transcripts consist of either a single copy of the ubiquitin fusion coding sequence or varying lengths of polyubiquitin (up to 52 copies of the ubiquitin coding unit) each ending with a single copy of the ubiquitin fusion sequence. Finally, there are heat shock elements 5′ to the *PUB* genes and transcription patterns are altered under conditions of stress.**

*Key words: Trypanosoma cruzi*/ubiquitin/transcription

## Introduction

Ubiquitin is one of the most conserved proteins in nature and is present in all eukaryotes examined. The function(s) of ubiquitin and the ubiquitin fusion proteins is not yet fully understood. However, two roles involved in protein turnover have been proposed. First, it has been demonstrated that ubiquitination sensitizes some proteins to specific proteases (Ciechanover *et al.*, 1984; Hershko *et al.*, 1984). Secondly, it has recently been proposed that ubiquitin itself is a protease (Fried *et al.*, 1987) and that ubiquitination of surface and nuclear proteins may be a means of targeting the ubiquitin to these organelles.

The genomic organization and expression of the ubiquitin genes has been studied in several organisms (Wiborg *et al.*, 1985; Ozkaynak *et al.*, 1984; Dworkin-Rastl *et al.*, 1984; Bond and Schlesinger, 1985; Guasing and Barkardottir, 1986). In all cases, while the absolute number of genes varies, the overall organization is similar. The majority of the ubiquitin coding sequences exist as direct spacerless repeats organized into one or more polyubiquitin genes. *Saccharomyces cerevisiae* has a single polyubiquitin gene containing six tandemly repeated units (Ozkaynak *et al.*, 1984), while the human genome codes for a polyubiquitin gene containing nine units (Wiborg *et al.*, 1985). Translation of the polyubiquitin gene transcription product results in the generation of a precursor polyubiquitin protein which is subsequently cleaved yielding monomer ubiquitin peptides (Wiborg *et al.*, 1985; Ozkaynak *et al.*, 1984). Ubiquitin genes also exist as dispersed monomers and as a part of a larger coding sequence (Wiborg *et al.*, 1985; Ozkaynak *et al.*, 1987; Lund *et al.*, 1985) consisting of 5′ ubiquitin coding sequence fused to a non-ubiquitin coding sequence.

We describe here the organization of ubiquitin genes in *Trypanosoma cruzi*, an intracellular hemoflagellate that is responsible for Chagas' disease in man. These organisms contain > 100 ubiquitin genes organized into a single cluster containing both polyubiquitin (*PUB*) and ubiquitin fusion (*FUS*) genes. Each *PUB* gene contains a fusion protein coding sequence at its 3′ end. Analysis of the spacer regions preceding the *PUB* genes indicates that while the sequences differ slighly each contain a heat shock element.

## Results

### *The cDNA clone pTC-FUS codes for a ubiquitin fusion protein*

We have isolated a cDNA clone, pTC-FUS, from a *T.cruzi* library by virtue of its homology to the *S.cerevisiae ubi*1 gene and to a probe specific for the 5′ 35 bp spliced leader common to all mRNAs in *T.cruzi* and other trypanosomatids (De Lange *et al.*, 1984; Parsons *et al.*, 1984; Walder *et al.*, 1985). The DNA sequence and potential translation product of pTC-FUS were determined (Figure 1). From this analysis it is clear that this clone codes for a protein which is very similar in length and sequence to the ubiquitin fusion protein encoded by the *UBI*1 and *UBI*2 genes of *S.cerevisiae* (Ozaynak *et al.*, 1987). Overall the two proteins are 84% homologous. The first 76 amino acids of both proteins code for ubiquitin. The remaining 52 residues of the *T.cruzi* polypeptide have 67% amino acid identity with the yeast sequence. Furthermore charge conservation between the two proteins is almost totally complete due to the conservative nature of the amino acid substitutions. Such a high degree of homology between the two proteins must be indicative of a common function. Although no function has been attributed to the fusion protein, one likely possibility is that

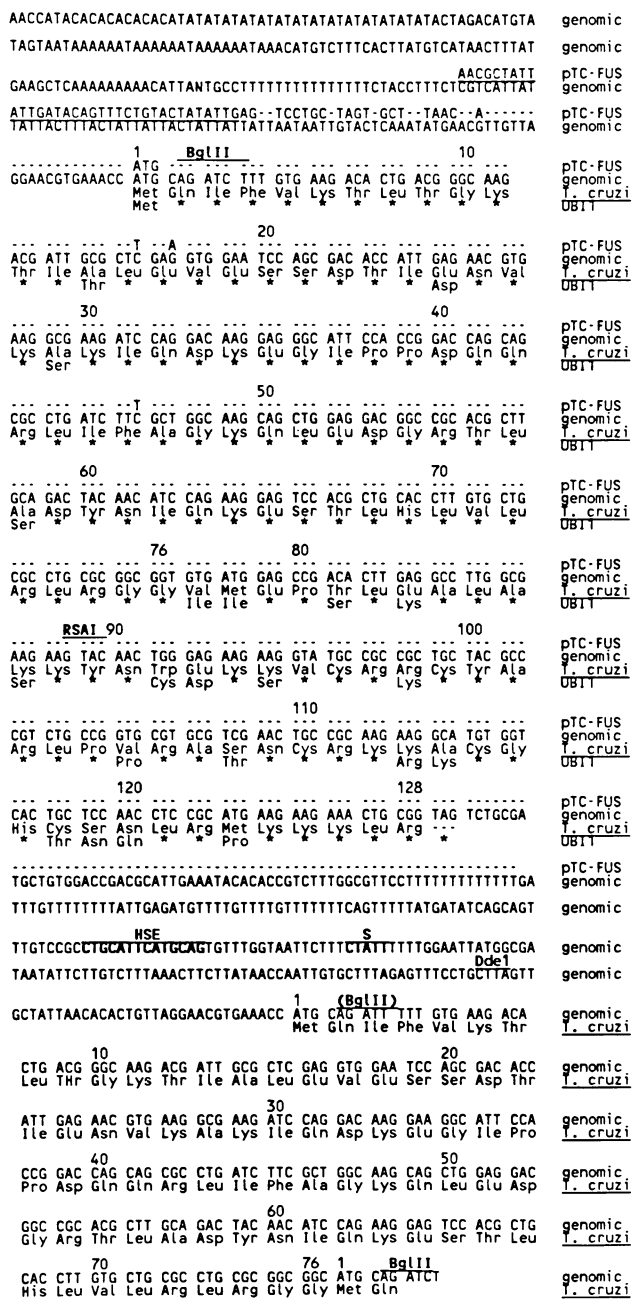

**Fig. 1.** Comparison of the DNA and amino acid sequences of the cDNA (pTC-FUS) and genomic ubiquitin clones of *T.cruzi* with the amino acid sequence of the *UBI1* gene of *S.cerevisiae* (Ozkaynak *et al.*, 1987). pTC-FUS was isolated as described in the text (also see legend to Figure 2). The region which gave rise to the genomic sequence is indicated by the dashed line under pTC-UB1g in Figure 2b. The genomic sequence begins 253 nt before the initial methionine of the first ubiquitin fusion gene (*FUS*1) and continues through the *FUS*1 and intergenic region to the first methionine of the second polyubiquitin unit of *PUB*12.5. The mini-exon of the cDNA is underlined. A dash (−) above the *T.cruzi* genomic sequence indicates a nucleotide identity with the *T.cruzi* cDNA. A star (*) in the UBI1 amino acid sequence indicates amino acid identity with the *T.cruzi* protein. The number 1 marks the position of the first methionine residue of each ubiquitin unit. The non-ubiquitin domain of *FUS*1 begins with residue 77. HSE: position of the homology to the *T.brucei* heat shock element. S: position of the possible CTATT promoter element. The positions of the *Bgl*II and *Dde*I restriction sites are indicated. (*Bgl*II) identifies the BglII recognition site which is eliminated by a DNA polymorphism.

it acts as a DNA binding molecule. This is based on the observation that in both *T.cruzi* and *S.cerevisiae* the non-ubiquitin domain of the fusion protein is highly basic in nature and contains a single zinc binding site both characteristics of known DNA binding proteins (Ozaynak *et al.*, 1987; Miller *et al.*, 1985; Berg, 1985; Harrison, 1986). As expected, analysis of the cDNA sequence revealed the presence of a spliced leader sequence at the 5′ end of the coding strand of pTC-FUS as has been found for all mRNAs in trypanosomes (Figure 1). The cDNA terminates at the 3′ end in a run of oligo(dT)$_n$ rather than the expected oligo(dA)$_n$. We cannot therefore determine the exact 3′ end of the transcript.

### *The genomic organization of T.cruzi ubiquitin genes*

Genomic Southern analysis using probes specific for both the ubiquitin and non-ubiquitin domains of the cDNA clone pTC-FUS revealed the presence of multiple copies of the *FUS* gene. Digestion of genomic DNA with the restriction endonuclease *Dde*I, which has no recognition sites in the insert of pTC-FUS, indicated that there are at least nine copies of the *FUS* gene represented by the presence of nine bands which hybridized with the non-ubiquitin domain of the cDNA (Figure 3a, lane 11). The pattern of hybridization to *Bgl*II (which has a single recognition site within the ubiquitin coding sequence of pTC-FUS) digested genomic DNA indicated that most of the sequences homologous to the cDNA exist as a 0.65 kb fragment (Figure 3a, lanes 4 and 9). Since all of these hybridized to both the ubiquitin and non-ubiquitin specific probes, each must contain a complete coding sequence for the ubiquitin fusion protein.

To determine the absolute copy number of the ubiquitin coding sequences in *T.cruzi*, we performed quantitative blotting experiments comparing known quantities of cloned ubiquitin and fusion protein coding sequences with varying quantities of genomic DNA (Figure 3b). Assuming that the DNA content of *T.cruzi* is $4.7 \times 10^7$ kb (Castro *et al.*, 1982) these experiments indicate that there are 9−10 copies of the *FUS* gene and that the ubiquitin coding sequence is present in ~100 copies.

As described above, nine of the ubiquitin coding sequences are contained within the *FUS* genes. Therefore there are ~90 additional ubiquitin coding sequences in the genome. These appear to be organized into five polyubiquitin genes, each terminated by a fusion protein coding sequence. Digestion of genomic DNA with the enzyme *Bgl*II, which has a single recognition site within the ubiquitin domain of pTC-FUS, revealed a 0.22 kb band that hybridized uniquely to the ubiquitin specific probe (Figure 3a, compare lanes 4 and 9). The ubiquitin coding sequence is 228 bp in length. Therefore the presence of a repeat unit of 0.22 kb indicates that most of the ubiquitin coding sequences are arranged into direct spacerless repeats. This interpretation was later confirmed through DNA sequence analysis of a portion of the genomic repeat (see below). In addition to the two major bands of hybridization three minor bands are detected. The 800 bp band is the result of a DNA polymorphism in the first ubiquitin unit of *PUB*12.5 that eliminates the *Bgl*II restriction site. The 2.2 kb band represents the 3′ terminal *FUS*5 gene plus flanking sequences. The 1.9 kb band represents the sequence linking *PUB*2.65 with the *FUS*2, see Figure 2a. Furthermore, genomic Southern analysis using the enzyme *Dde*I, which has no recognition sites within

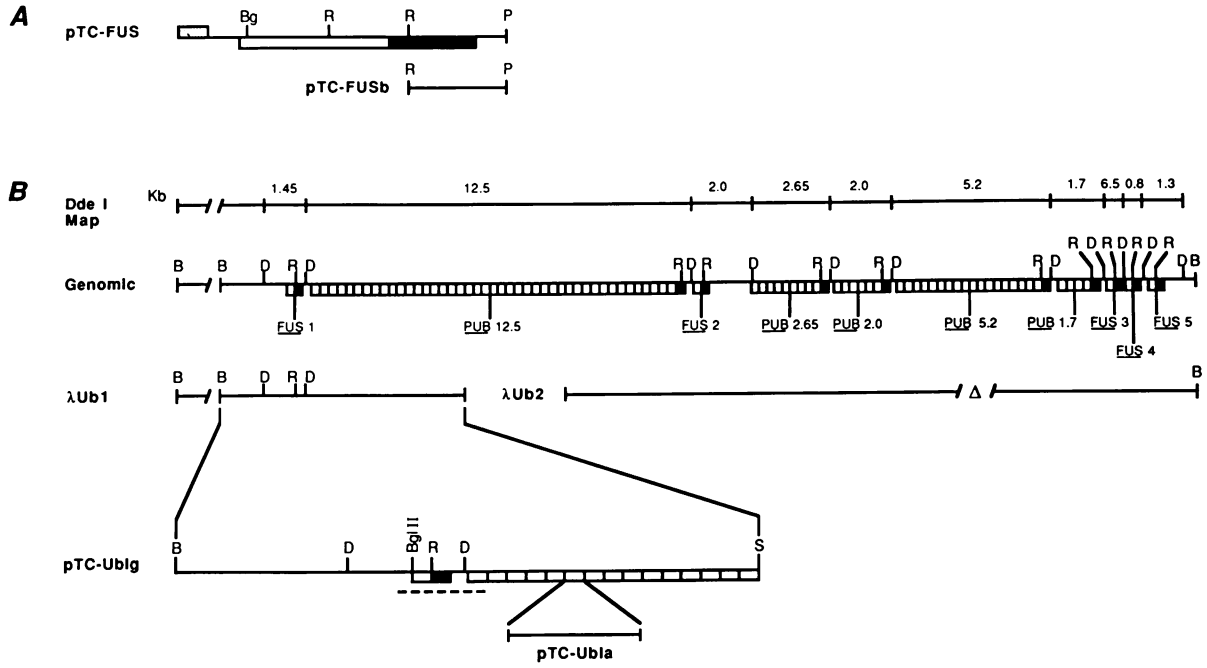


**Fig. 2.** (**A**) Restriction maps of pTC-FUS and pTC-FUSb. pTC-FUS is a 557-bp cDNA inserted into the *Pst*I site of pUC9 (Vieira and Messing, 1982). Stippled box represents the mini-exon sequence (De Lange *et al.*, 1984; Parsons *et al.*, 1984; Walder *et al.*, 1985). The open box represents the ubiquitin domain of the fusion gene and the solid box represents the non-ubiquitin domain of the fusion gene. RNA was isolated and cDNA synthesized from total poly(A)$^+$ RNA as described by Maniatis *et al.* (1982). The 200 bp *Rsa*I–*Pst*I fragment from pTC-FUS was subcloned into Bluescribe (+) (Cloning Vector Systems) to give rise to pTC-FUSb. (**B**) The upper line represents the *Dde*I restriction map of the ubiquitin gene cluster. In the genomic map, solid boxes represent the non-ubiquitin domain of the fusion gene, open boxes represent varying lengths of polyubiquitin (the absolute number of ubiquitin repeats in each length of polyubiquitin has not been determined). The isolation of λUB1 and λUB2 is described in the text. The △ in λUB2 represents a deletion of unknown size. The 6.0 kb *Bam*HI–*Sal*I fragment from UB1 was subcloned into Bluescribe (+) to yield pTC-UB1g. The dashed line under pTC-UB1g indicates the area which was sequenced (see Figure 1, genomic sequence). A single 220 bp *Bgl*II repeat unit from pTC-UBlg was subcloned into the *Bam*HI site of Bluescribe (+) to yield pTC-UB1a. Bg = *Bgl*II, R = *Rsa*I, P = *Pst*I, B = *Bam*HI, D = *Dde*I.

the ubiquitin coding sequence, indicated that there are at least five genes coding for polyubiquitin *PUB*12.5, *PUB*5.2, *PUB*2.6, *PUB*2.0 and *PUB*1.7 (Figure 3a, lane 6). The apparent number of ubiquitin coding sequences within each *PUB* gene increases in direct proportion to the size of the gene. This correlation is evidenced by the fact that the intensity of hybridization increases in direct proportion to the size of the fragment. (In the experiment shown in Figure 3 the DNA was not acid treated before transfer to nitrocellulose so as to optimize transfer and retention of the smaller bands. Thus the 12.5 kb fragment was poorly transferred.) Coupling these results with the gene counting data and the observation that all the ubiquitin sequences are encoded on a single 30 kb *Bam*HI DNA fragment (see below), indicates that in *T.cruzi* the *PUB*12.5 may contain up to 52 direct repeats of the ubiquitin coding sequence.

All sequences homologous to the *FUS* and *PUB* genes are closely clustered in the *T.cruzi* genome. Digestion of genomic DNA with any one of several restriction endonucleases which have a 6 bp recognition sequence but no site within insert of pTC-FUS, resulted in a single band of hybridization (Figure 3a, lanes 3 and 8). Based on the gene counting experiments described above the *FUS* and *PUB* genes would require at least 27 kb of DNA, all of which is contained on the 30 kb *Bam*HI fragment.

The organization of the *FUS* and *PUB* genes were determined through genomic Southern analysis using the restriction endonucleases *Rsa*I and *Dde*I and probes specific for the non-ubiquitin domain of the *FUS* genes or the ubiquitin coding sequence. This analysis indicated that each of the *PUB* genes contains a single copy of the *FUS* gene sequence (Figure 3a, lanes 6 and 11; 1.7, 2.0, 2.6, 5.2 and 12.5 kb *Dde*I fragments). In addition there are five individual *FUS* genes, *FUS*1–5 (Figure 3a, lane 11; 0.65, 0.8, 1.3, 1.45 and 1.7 kb *Dde*I fragments). The 1.7 kb band is a doublet containing both the *PUB*1.7 and the *FUS*1 genes. The same two probes, however, had very different patterns of hybridization to a double digest of *Rsa*I plus *Dde*I. The *FUS* gene specific probe hybridized to a single band of 368 bp (Figure 3a, lane 12) while the ubiquitin specific probe hybridized to bands ranging in size from 1.3 to 12.2 kb (Figure 3a, lane 7). These results indicate that each of the *PUB* genes terminates at the 3′ end with a copy of the *FUS* gene sequence (see the diagram at the base of Figure 3a). The overall organization of the *FUS* and *PUB* genes relative to one another is shown in Figure 2b. This organization was determined as follows. First, genomic clones carrying portions of the ubiquitin gene cluster were isolated from a bacteriophage lambda EMBL3 library (Frischauf *et al.*, 1983). Digestion of DNA from one clone, λUB2, with *Dde*I indicated that the insert DNA contained several *FUS* and *PUB* genes (Figure 2b). The position of the genes within the insert was determined by end labeling the insert DNA at the unique *Bam*HI site and subjecting the purified fragment to partial digestion by either *Dde*I or *Rsa*I (data not shown). The relative position of the *PUB*2.6 and *PUB*2.0 could not be absolutely determined due to the presence of deletions in *PUB*5.2 within the phage.

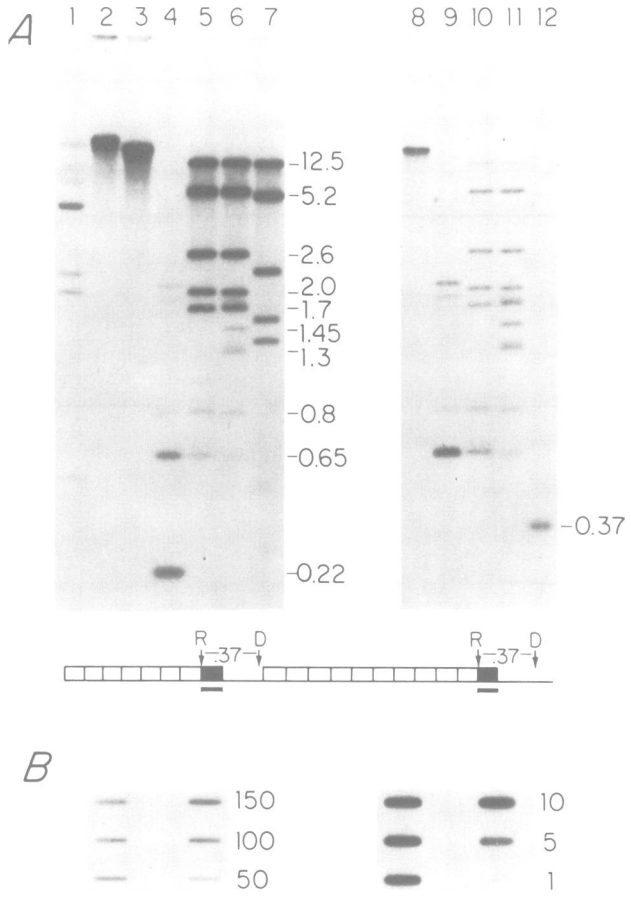


**Fig. 3.** (**A**) Genomic Southern analysis of the ubiquitin genes of *T.cruzi*. Each lane in the Southern blots contains 0.5 μg of genomic *T.cruzi* DNA. **Lanes 1–7** were probed with pTC-UBla and **lanes 8–12** were probed with pTC-FUSb. Lane 1, λ*Hin*dIII size standards; lane 2, uncut *T.cruzi* genomic DNA; lane 3, *Bam*HI; lane 4, *Bgl*II; lane 5, *Rsa*I; lane 6, *Dde*I; lane 7, *Rsa*I + *Dde*I; lane 8, *Bam*HI; lane 9, *Bgl*II; lane 10, *Rsa*I; lane 11, *Dde*I; lane 12, *Rsa*I + *Dde*I. Fragment lengths, given in kb to the right of lane 7, indicate the lengths of the *Dde*I fragments in lanes 6 and 11. The Southern blot of lanes 1–7 was exposed for 12 h and the blot for lanes 8–12 was exposed for 120 h. The restriction map below the Southern blots indicates the relative organization of the polyubiquitin coding sequences and the non-ubiquitin domain of the ubiquitin fusion gene. The open boxes represent ubiquitin coding sequence and the filled boxes represent the non-ubiquitin domain of the fusion gene. The solid line beneath filled boxes indicates the sequences to which the DNA probe used for lanes 8–12 hybridizes. R, *Rsa*I; D, *Dde*I. (**B**) The gene copy number was based on a total genome size of $4 \times 10^7$ bp for *T.cruzi* strain CL (Castro *et al.*, 1982). In both the left and right panels the first lane contains 0.5 μg of genomic DNA in each slot (the equivalent of $1.3 \times 10^7$ genomes). The numbers to the right of the second lane in each panel refer to the equivalent copy number of pTC-UBla and pTC-FUSb respectively (based on a size of 2.9 kb for pTC-UBla and 3.4 kb for pTC-FUSb). Left panel, 50 copies ($2.1 \times 10^{-3}$ μg), 100 copies ($4.2 \times 10^{-3}$ μg), 150 copies ($6.2 \times 10^{-3}$ μg). Right panel, 1 copy ($4.8 \times 10^{-5}$ μg), 5 copies ($2.4 \times 10^{-4}$ μg), 10 copies ($4.8 \times 10^{-4}$ μg). The left panel was hybridized with an insert probe from pTC-UBla and the right panel with an insert probe from pTC-FUSb.

### *DNA sequence analysis of the FUS1 to PUB12.5 region*

We isolated one recombinant phage, λUB1, which contains 10 kb of flanking sequence, the *FUS*1 gene and the first 12 ubiquitin coding units of the *PUB*12.5 gene (Figure 2b). To more precisely determine the relative positions of the *FUS*1 and *PUB*12.5 genes in this clone the 6 kb *Bam*HI–*Sal*I fragment from λUB1 was subcloned into pSP64 (Melton *et al.*, 1984) (pTC-UB1g, Figure 2b). DNA sequence analysis extended from 250 bp 5′ of *FUS*1 through the *Bgl*I site in the second ubiquitin unit of *PUB*12.5 (Figure 1). This analysis indicates that there is a single copy of the *FUS*1 gene which terminates in a T-rich region. Comparison of the DNA sequences of the *FUS*1 gene and the cDNA, pTC-FUS, reveals that while the two encode identical proteins, their nucleic acid sequences differ. Therefore the RNA which gave rise to the cDNA must have resulted from the transcription of another copy of the *FUS* gene and we cannot be sure that the *FUS*1 gene is transcribed. The DNA sequence also confirms that the polyubiquitin coding sequence of *PUB*12.5 consists of direct spacerless repeats of the ubiquitin coding sequence since the first ubiquitin unit terminates with a glycine codon followed immediately by the first methionine codon of the second ubiquitin unit. Partial digestion of pTC-UB1g with *Pvu*II which has a single recognition site in the ubiquitin coding sequence confirms that this clone contains a single copy for the fusion gene followed by at least 12 copies of the unit ubiquitin sequence (data not shown). The organization of the polyubiquitin into direct spacerless repeats is analogous to what has been found in other organisms (Wiborg *et al.*, 1985; Ozkaynak *et al.*, 1984). This indicates that the polyubiquitin in *T.cruzi* is probably translated as a polyprotein which is later proteolytically cleaved to yield the monomer peptide.

```
... ... ... ... ... ... ... ... ... ... ... ... ... ... .......   PUB1.7
CAC TGC TCC AAC CTC CGC ATG AAG AAG AAA CTG CGG TAG TCTGCGA       PUB12.5
His Cys Ser Asn Leu Arg Met Lys Lys Lys Leu Arg END

----------------------------------C--------------A-A-G-CT       PUB1.7
TGCTGTGGACCGACGCATTGAAATACACACCGTCTTTGGCGTTCCTTTTTTTTTTTTTGA     PUB12.5

---T---A--GAGAA--TG-C--G-T-G-----G----TTTTTTTTT------------      PUB1.7
TTTGTTTTTTTTATTGAGATGTTTTGTTTTGTTTTTTT*********CAGTTTTTATGAT     PUB12.5

-CG------------A---------------A----------------C-------        PUB1.7
ATCAGCAGTTTGTCCGCCTGCATTCATGCAGTGTTTGGTAATTCTTTCTATTTTTTGGAA     PUB12.5
                  HSE                          S

----------T------------T---A------C-----------C---------        PUB1.7
TTATGGCGATAATATTCTTGTCTTT*AAACTTCTTATAACCAATTGTGCTTTAGAGTTTC     PUB12.5

---------------------------------C------ --- --- --C --C        PUB1.7
CTGCTTAGTTGCTATTAACACACTGTTAGGAACGTGAAACC ATG CAG ATT TTT        PUB12.5
                                          Met Gln Ile Phe
```

**Fig. 4.** Comparison of the intergenic region preceding the *PUB*1.7 and *PUB*12.5 genes. *PUB*1.7 was subcloned into Bluescribe (+) from λUBl as a 1.7 kb *Rsa*I fragment. (–) indicates nucleotide identity between the two sequences, (*) indicates the position of deletions or insertions, nucleotides in the upper sequence indicate single base changes. HSE: position of the heat shock element. S: position of a possible promoter element.

### *The DNA sequences upstream of PUB genes contain a heat shock element*

To identify any sequences which may be important for the expression of the *PUB* genes we sequenced the intergenic region immediately upstream of two *PUB* genes. The result of this analyis is shown in Figure 4. There are multiple single nucleotide changes as well as small insertions between the two intergenic sequences. Both intergenic regions

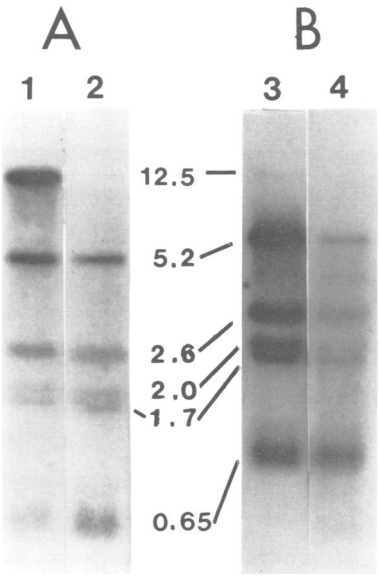


**Fig. 5.** Northern analysis of the ubiquitin transcripts of *T.cruzi*. Lane 1 was treated with 50 mM NaOH to enhance transfer of high mol. wt RNAs. (**A**) Comparison of poly(A)$^+$ and total RNA. **Lane 1**, 0.5 μg of poly(A)$^+$ RNA, **lane 2**, 5.0 μg of total RNA. The blot was hybridized with an insert probe isolated from pTC-FUS containing both the ubiquitin and non-ubiquitin domains of the fusion gene. (**B**) Both lanes 1 and 2 contain 5.0 μg of total RNA. Lanes 1 and 2 were probed with inserts from pTC-UB1a (ubiquitin domain) and pTC-FUSb (non-ubiquitin domain) respectively.

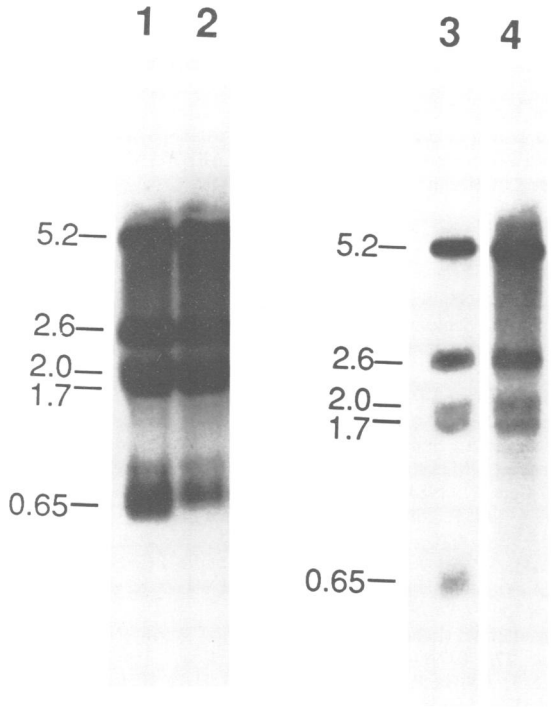


**Fig. 6.** Northern analysis of RNA from heat shock and stationary phase cultures. The blot was hybridized with an insert probe from the cDNA pTC-FUS. RNA was isolated from: **lane 1**, 26°C, **lane 2**, 41°C, **lane 3**, midlog culture; **lane 4**, stationary phase culture.

however have conserved sequences that are strikingly similar to the heat shock element (HSE) identified in front of the major heat shock genes of *T.brucei* (Glass *et al.*, 1986). Both HSEs consist of a sequence of 14 nt containing a direct repeat of the sequence TGCA. The dyad symmetry in the *T.brucei* HSE is limited to this core sequence while the symmetrical region of the *T.cruzi PUB* genes is extended by 2 nt (Figure 4). A putative trypanosome TATA box [CTATT (Campbell *et al.*, 1984)] is located 17 bases 3′ to the HSE (Figure 4). The spacing between the HSE and the TATA sequence of the *PUB* genes is similar to that for other heat shock genes described in other systems (Topol *et al.*, 1985).

### *Transcription products of the FUS and PUB genes*

Northern analysis of total and poly(A)$^+$ RNA isolated from midlog cultures indicated that transcription of the *PUB* and *FUS* genes resulted in the generation of six major poly(A)$^+$ mRNAs ranging in size from 650 bases to 12.5 kb (Figure 5a). Using either pTC-UB1a (homologous to all ubiquitin coding sequences) or pTC-FUSb (specific for the non-ubiquitin domain of the *FUS* genes) as probe revealed the same six RNAs (Figure 5b), indicating that all transcripts code for a ubiquitin fusion peptide. However, while pTC-FUS hybridized to each band equally, the intensity of hybridization of pTC-UB1a to the five largest bands was directly proportional to the size of the RNA molecule. The most probable explanation for this result is that while the 650 bp RNA represents the transcript of one or more of the *FUS* genes, each of the five larger RNA molecules is the transcription product of the *PUB* gene (as defined by *Dde*I sites) of similar size.

### *Transcription of the FUS and PUB genes is altered in response to cellular stress*

The finding that the intergenic regions preceding the *PUB* genes contain HSEs suggested that the genes might be subject to heat shock or stress regulation. Northern analysis of RNA isolated from cultures of *T.cruzi* epimastigotes which have been either subjected to a heat shock or grown to stationary phase reveals an alteration in the pattern of *FUS* and *PUB* gene transcripts. The results of such an analysis are shown in Figure 6. The RNA isolated from a culture which had been heat shocked (see Materials and methods) has approximately the same pattern of *PUB* gene transcripts as seen in control RNA. RNA isolated from stationary phase cultures has a pattern of *FUS* and *PUB* gene transcripts different from that observed in control and heat shocked cultures. Expression of the *PUB* genes is apparently increased in stationary cultures while that of the *FUS* genes is greatly diminished (Figure 6, lane 4). These results indicate that *T.cruzi*'s response to one stress situation involves alteration in the expression of the *FUS* and *PUB* genes. Expression is, however, differentially altered depending upon the nature of the stimulus the cell experiences.

## Discussion

We have presented data describing the organization and expression of the ubiquitin genes of *T.cruzi*. We have shown that *T.cruzi* contains five genes coding for a ubiquitin fusion protein (*FUS* genes) and five polyubiquitin genes (*PUB* genes). The results of DNA sequence analysis and Southern hybridizations indicate that all *FUS* genes are highly homologous. Each would code for a ubiquitin fusion protein of a 76 amino terminal ubiquitin domain fused to an unrelated carboxy-terminal domain of 52 amino acids. The

high degree of homology between the deduced amino acid sequence of the *FUS* gene product and that of the *UBI*1 and *UBI*2 genes of *S.cerevisiae* indicates that the proteins would carry out analogous functions in the two organisms. While the protein products have not been identified, the predicted amino acid sequence of the non-ubiquitin domains raises the possibility that the proteins act as DNA binding molecules (see Results, Ozkaynak *et al.*, 1987). The five *PUB* genes consist of varying numbers of direct spacerless repeats of the ubiquitin coding sequence terminating in each gene with a single copy of the *FUS* gene sequence. This organization has not been seen in other organisms. Each *PUB* gene is ultimately represented in steady state RNA as a single poly(A)$^+$ transcript. Although the nature of the primary transcription product(s) cannot be determined from our data it is likely that under some conditions transcription of the *PUB* and *FUS* genes is independent, since in stationary phase cells transcription of the *PUB* genes is apparently increased while transcription of the *FUS* genes is decreased. A similar phenomenon has been observed in *S.cerevisiae* where transcription of the polyubiquitin gene *UBI*4 is stimulated in stationary phase cultures while transcription of the ubiquitin fusion genes *UBI*1 and *UBI*2 is decreased (Ozkaynak *et al.*, 1987).

The number of ubiquitin coding sequences contained within the *PUB* genes of *T.cruzi* is substantially greater than found in any other organism. This raises intriguing questions as to the role the ubiquitin polypeptide plays in this organism. Ubiquitin's involvement in the elimination of defective proteins from the cell has been well characterized in other organisms (Hershko *et al.*, 1980, 1984; Ciechanover *et al.*, 1984). Also, Finley *et al.* (1987) have shown that the *S.cerevisiae* polyubiquitin coding gene *UBI*1 is essential under conditions of stress. It is thus possible that the large number of ubiquitin coding genes in *T.cruzi* reflects a requirement of the parasite to survive in stressful situations.

*T.cruzi* is an intracellular parasite in the mammalian host and as such must derive its nourishment from the host cell. It would therefore be advantageous to the parasite to maintain a large pool of small peptides and amino acids derived from the host cell's proteins. One way the parasite could do this would be to secrete ubiquitin into the host cytoplasm. It is also possible that parasite glycolysis would result in enough lactate production to cause significant protein denaturation in the host cell cytoplasm. Thus, if ubiquitin availability were limiting for the degradation of denatured proteins, this would increase the rate of degradation. The end result would be to produce adequate amino acids and peptides for the parasite.

The recent identification of ubiquitin as a component of certain cell surface receptors (Siegelman *et al.*, 1986) raises an additional interesting possibility. To establish a successful chronic infection, *T.cruzi* must effectively evade the host's immune system. The parasite could accomplish this by disguising its surface such that it would no longer be recognized as foreign. The extraordinary conservation of ubiquitin throughout evolution makes it a likely candidate to play such a role. If ubiquitin were expressed on the surface, it is possible that the presence of the parasite would not illicit an immune response from the host. Either, or both of the possibilities described would require that the parasite have the capacity to produce large amounts of ubiquitin.

## Materials and methods

### *Culture conditions*

*T.cruzi* strain CL epimastigotes were cultured in brain heart infusion (BHI; Difco) media (35 g/l) supplemented with 20 μg/ml hemin and 10% bovine calf serum at 26°C. To heat shock cells, cultures grown at 26°C were pelleted by centrifugation at 1500 *g* and resuspended in pre-warmed BHI media (41°C). The cultures were incubated at the elevated temperature for 1 h followed by the addition of an equal volume of frozen BHI media. The chilled cultures were pelleted and RNA prepared as described below.

### *Construction of plasmids*

Plasmid constructions are described in the text and figure legends.

### *DNA sequencing analysis*

DNA sequences were determined by the dideoxy chain termination method (Sanger *et al.*, 1977) using synthetic oligonucleotides homologous to various portions of the clones as primers for elongation.

### *Southern hybridizations*

DNA blot analyses were carried out using Gene Screen Plus. Hybridizations were carried out at 42°C in 50% formamide, 5 × SSC, 5 × Denhardt's, 1% SDS and 100 μg/ml carrier tRNA. All filters were washed at 65°C in 0.1 × SSC, 1.0% SDS for 60 min.

### *Isolation of RNA and Northern hybridizations*

Total cellular RNA was isolated by the guanidinium/cesium chloride method of Maniatis *et al.* (1982). Poly(A)$^+$ was isolated after three cycles through oligo(dT)–cellulose as described (Maniatis *et al.*, 1982). RNA was size fractionated on 1.1% agarose gels containing 1 M formaldehyde (Maniatis *et al.*, 1982). All Northern hybridizations were performed after blotting the size fractionated RNA overnight in 20 × SSC to Gene Screen Plus. Hybridization and wash conditions were precisely as described above for Southern blot analysis.

## Acknowledgements

We wish to thank Alexander Varshavsky for the communication of unpublished data. This work was supported by grants from the French Ministry of Research and Technology, the New Development Fund from the Fred Hutchinson Cancer Research Center and the American Cancer Society. G.B. was supported by a grant from the Jeffress Memorial Trust J81. J.S. was supported by postdoctoral fellowship 5 F32 AI06965 02 from the National Institutes of Health.

## References

Berg,J.M. (1985) *Science*, **232**, 485–487.
Bond,U. and Schlesinger,M.J. (1985) *Mol. Cell. Biol.*, **5**, 949–956.
Campbell,D.A., Thorton,D.A. and Boothroyd,J.C. (1984) *Nature*, **311**, 350–355.
Castro,C., Craig,S.P. and Castaned,M. (1982) *Mol. Biochem. Physiol.*, **4**, 273–282.
Ciechanover,A., Finley,D. and Varshavsky,A. (1984) *Cell*, **37**, 55–66.
De Lange,T., Michels,P.A.M., Veerman,H.J.G., Cornelissen,A.W.C.A. and Borst,P. (1984) *Nucleic Acids Res.*, **12**, 3777–3790.
Dworkin-Rastl,E., Shrutkowski,A. and Dworkin,M.B. (1984) *Cell*, **39**, 321–325.
Finley,D., Ozkaynak,E. and Varshavsky,A. (1987) *Cell*, **48**, 1035–1046.
Fried,V.A., Smith,H.T., Hildebrandt,E. and Wiener,K. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3685–3689.
Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.
Glass,D.J., Polvere,R.I. and Van der Ploeg,L.H.T. (1986) *Mol. Cell. Biol.*, **6**, 4657–4666.
Guasing,K. and Barkardottir,R. (1986) *Eur. J. Biochem.*, **158**, 57–62.
Harrison,S. (1986) *Nature*, **322**, 597–598.
Hershko,A., Ciechanover,A., Heller,H., Hass,A.L. and Rose,I. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1783–1786.
Hershko,A., Leshinsky,E., Ganoth,D. and Heller,H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1619–1623.
Lund,P.K., Moats-Staats,B.M., Simmons,J.G., Hoyt,E., D'Ercole,A.J., Martin,F. and Van Wyk,J.J. (1985) *J. Biol. Chem.*, **260**, 7609–7613.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A*

*Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Ozkaynak,E., Finely,D. and Varshavsky,A. (1984) *Nature*, **312**, 663–666.
Ozkaynak,E., Finely,D., Solomon,M.J. and Varshavsky,A. (1987) *EMBO J.*, **6**, 1429–1439.
Parsons,M., Nelson,R.G., Watkins,K.P. and Agabian,N. (1984) *Cell*, **38**, 309–316.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Siegelman,M., Bond,M.W., Gallatin,W.M., St John,T., Smith,H.,T., Fried,V.A. and Weissman,I.L. (1986) *Science*, **231**, 823–829.
Topol,J., Ruden,D.M. and Parker,C.S. (1985) *Cell*, **42**, 527–537.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.
Walder,J.A., Eder,P.S., Engman,D.M., Brentano,S.T., Walder,R.Y., Knutzon,D.S., Dorfman,D.M. and Donelson,J.E. (1985) *Science*, **233**, 569–571.
Wiborg,O., Pederson,M.S., Wind,A., Berglund,E., Marcker,K.A. and Vuust,J. (1985) *EMBO J.*, **4**, 755–759.

*Received on September 21, 1987; revised on January 15, 1988*