Available online at www.sciencedirect.com
SCIENCE DIRECT
Experimental Parasitology 113 (2006) 112–124
Experimental Parasitology
www.elsevier.com/locate/yexpr

Trypanosoma cruzi: Analysis of the complete PUF RNA-binding protein family

Florence Caro, Natalia Bercovich, Catalina Atorrasagasti, Mariano J. Levin, Martín P. Vázquez *

Laboratorio de Biología Molecular de la Enfermedad de Chagas—INGEBI—CONICET, Departamento de Fisiología y Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires, Vuelta de Obligado 2490 2P, 1428 Buenos Aires, Argentina

Received 31 August 2005; received in revised form 20 December 2005; accepted 21 December 2005
Available online 7 February 2006

Abstract

The members of the PUF family of RNA-binding proteins regulate the fate of mRNAs by binding to their 3′UTR sequence elements in eukaryotes. In trypanosomes, for which gene expression is polycistronic and controlled almost exclusively by post-transcriptional processes, PUF proteins could play a crucial role. We report here the complete analysis of the PUF protein family of *Trypanosoma cruzi* composed of 10 members. In silico analysis predicts the existence of at least three major groups within the *T. cruzi* family, based on their putative binding specificity. Using yeast three hybrid assays, we tested some of these predictions for TcPUF1, TcPUF3, TcPUF5, and TcPUF8 as representatives of these groups. Data mining of the *T. cruzi* genome led us to describe putative binding targets for the TcPUFs of the most conserved group, TcPUF1 and TcPUF2. The targets include genes for mitochondrial proteins and protein kinases. Finally, immunolocalization experiments showed that TcPUF1 is localized in multiple discrete foci in the cytoplasm supporting its proposed function.

Index Descriptors and Abbreviations: Post-transcriptional regulation; Bioinformatics analysis; Yeast three hybrids; Chagas disease; NRE, nanos response element; UTR, untranslated region; HMM, hidden Markov model; ORF, open reading frame; aa, amino acids; TriTryp, *Trypanosoma brucei, Trypanosoma cruzi*, *Leishmania major*

1. Introduction

The Pumilio protein family is large, evolutionary conserved and found exclusively in eukaryotes. These proteins bind 3′-UTR elements of their target mRNAs to reduce expression either by repressing translation or causing mRNA instability. All Pumilio proteins share a domain consisting of eight to nine imperfect Puf repeats of 36–40 amino acids, each one folding into a three helix structure. The α-helices that contact the RNA lie on the concave face of the curved protein structure. On the basis of different studies, a common function established for these proteins is the maintenance of stem cells by promoting proliferation and repressing their differentiation (Wickens et al., 2002). Two analyzed in detail are *Drosophila melanogaster* Pumilio (Pum) and *Caenorhabditis elegans* FBF (fem-3 mRNA-binding factor).

In *Drosophila*, Pum binds to a pair of 32 nucleotide sequences (Nanos respose elements, NREs) within the 3′-UTR of maternal hunchback (hb) mRNA and represses its translation in the posterior portion of the embryo (Murata and Wharton, 1995; Wharton and Struhl, 1991). In *C. elegans*, FBF binds to the 3′-UTR of the fem-3 mRNA, thereby promoting the switch from spermatogenesis to oogenesis (Zhang et al., 1997). Based on structural similarities, FBF and Pumilio were proposed as members of a family of sequence-specific RNA-binding proteins named PUF (for Pumilio and FBF) (Zamore et al., 1997). To date,

* Corresponding author. Fax: +54 11 47868578.
E-mail address: mvazquez@dna.uba.ar (M.P. Vázquez).

0014-4894/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.exppara.2005.12.015

all RNA targets of PUF proteins analyzed contain a UGUR core nucleotide sequence that is critical for binding (Eckmann et al., 2004; Gerber et al., 2004; Jackson et al., 2004; Lamont et al., 2004; Nakahata et al., 2001; Souza et al., 1999; Tadauchi et al., 2001; Wang et al., 2002; Zhang et al., 1997). Moreover, it was shown for human Pumilio1 that each of the eight repeats of the PUF domain makes contacts with a different RNA base via three amino acid side chains at conserved positions. Mutagenesis experiments in these side chains in one repeat altered the sequence specificity of Pumilio1 in a predicted manner. Residues at specific positions make stacking interactions, hydrogen bonds or Van der Waals interactions with the RNA structure. Thus, the high affinity and specificity of the human Pumilio1 for RNA is achieved using multiple copies of the single repeated PUF motif (Wang et al., 2002).

In *Saccharomyces cerevisiae*, DNA microarrays strategies were used to identify a large and functionally related set of specific mRNAs that interact with the five members of PUF proteins. Interestingly, all five PUF localized, predominantly, to multiple discrete foci in the cytoplasm. This raises the possibility of physical clustering of functionally related groups of mRNAs to aid the assembly of complexes and the coordinated control of translation or RNA turnover (Gerber et al., 2004).

In *Trypanosoma cruzi*, where accumulated data provide little or no evidence for regulation at the level of transcription initiation, post-transcriptional regulation is essential (Clayton, 2002). The presence of numerous genes encoding proteins with RNA-binding motifs and domains in the genome of this parasite is consistent with the existence of this type of control of gene expression (Ivens et al., 2005). Several of these proteins containing RNA recognition motives (RRM) and CCCH or CCHC zinc finger motives were studied extensively and implicated in post-transcriptional RNA processing/stability in *T. cruzi* and *T. brucei* (Caro et al., 2005; D'Orso and Frasch, 2002; Hendriks and Matthews, 2005; Hendriks et al., 2003).

A member of the PUF family of proteins was described in *T. brucei* and named TbPUF1. It was demonstrated that TbPUF1 interacts with the product of a VSG expression site associated gene (ESAG8). Although function of ESAG8 is unknown, it localizes in the nucleolus and cytoplasm. Furthermore, genetic perturbations, overexpression or silencing, of TbPUF1 affected growth and differentiation of parasites in culture (Hoek et al., 2002).

Here, we report the identification of 10 members of the Pumilio family present in the genome of *T. cruzi*. An in silico analysis was used to predict their putative binding specificity based on previous data obtained for the human Pumilio1 (Wang et al., 2002). Our results showed at least three major binding groups that we named NRE (for the *D. melanogaster* PUF binding element), UGUR (for the conserved core binding sequence in NRE) and Unknown (for unknown binding specificity). We tested some of these predictions for TcPUF1, TcPUF3, TcPUF5, and TcPUF8 by yeast three hybrid experiments. Bioinformatics analyses and immunolocalization experiments led us to describe the existence of putative binding targets in the *T. cruzi* genome for the TcPUFs NRE group and a foci distribution in the cytoplasm for one of the members of the NRE group, respectively.

2. Materials and methods

2.1. Data mining and bioinformatics analyses

The PFam HMM profile (pfam_ls) for PUF (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00806) was first converted to a GCG profile and then used as input in the MAST algorithm (Motif Alignment and Search Tool, http://meme.sdsc.edu/) to search a licensed copy of the complete WGS individual reads of the *T. cruzi* genome (http://www.tigr.org/tdb/e2k1/tca1/) (Results A, Fig. 1). All coding sequences of PUF proteins were translated and aligned with Align X (Vector NTI 8 package, Informax). The alignment was used as a training set in HMMER 2.3 (http://hmmer.wustl.edu/) to build a *T. cruzi* HMM specific profile that was re-introduced in MAST (Results B, Fig. 1). The coding sequences were also used as input in MEME (Multiple Em for Motif Elicitation, http://meme.sdsc.edu/) for motif discovery. These motives were later converted into a HMM compatible file using the tool Meta-MEME which combines motif models from MEME into a hidden Markov model framework for use in searching sequence databases and re-introduced in MAST (Results C, Fig. 1). The align X analysis was exported as MSF file and used as input in CLUSTAL W (http://www.ebi.ac.uk/clustalw/) to produce the correct output files for tree construction in NJ plot (http://pbil.univ-lyon1.fr/software/njplot.html). The secondary structure analysis of Puf domains was performed online using PSIPRED v2.4 (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html). Pairwise alignment of TriTryp PUF proteins was performed online using the BLAST 2 sequences tool with program blastp without the filter option checked (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html).

GenBank accession numbers for TcPUF proteins are the following: AY373518–AY373526.

2.2. Cloning of TcPUF genes in the yeast three hybrid system

All four coding sequences of the PUF genes were PCR amplified from 100 ng of total genomic DNA of *T. cruzi* CL-B clone using the proofreading AccuTaq enzyme (Sigma–Aldrich) and the following primers:

Puf 5y3hup 5′AAGCTTATGGTCATGGCCGCGTCAAAC3′,
Puf 5y3h 5′GAGCTCTCACTCATCCATGAGGGG3′,
Puf 3-y3hup 5′AAGCTTAACATTACCCCGGAGGGGTT3′,
Puf 3-down 5′TCAGCTGGAAAGTTGCTGAT3′,
Puf 1-y3h 5′AAGCTTGCTATGAACGGCACAATCAGCC3′,
Puf 1-down 5′CTACGCGTTGCCGTCAGCTT3′,

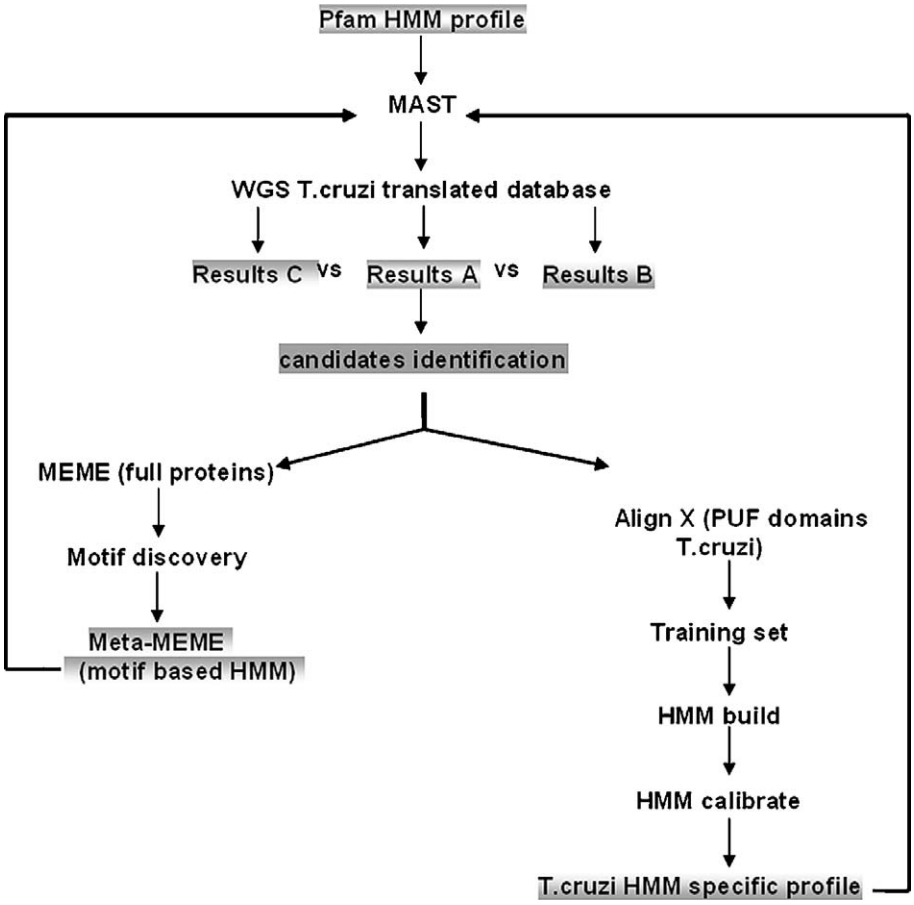


Fig. 1. Data mining strategies of TcPUF proteins in the genome of *T. cruzi*. See Section 2 under the data mining and bioinformatics analyses subheading for detailed description.

Puf 8 Y3Hup 5′AAGCTTATGGTCCAAACGCACA CAAAG3′,
Puf8-down 5′TTACTTTCTGCGTGTTGTCTTC3′

The RNA–protein interaction Hybrid Hunter kit (Invitrogen) was used for the yeast three hybrid experiments. PCR products were cloned into pGEM-T Easy (Promega), subsequently digested with *Hin*dIII and *Eco*RI and subcloned in frame into the pYesTrp3 vector. Plasmids were sequenced on a MegaBACE 500 (Amersham Biosciences) capillary sequencer to verify correct cloning.

The yeast cells L40ura3 strain (MATa, ura3-52, leu2-3112, his3 Δ 200, trp1 Δ1,ade2, LYS2::(LexAop) 4-HIS3, ura3::(LexA-op)8-lacZ) were first transformed with the pHybLexZeo/MS2 plasmid using the lithium acetate method (Sambrook and Russell, 2001). Transformants were selected on plates containing 200 μg/ml zeocin. Thereafter, cells carrying the plasmid were transformed with the pYesTrp3 plasmids containing the fusion of the bait proteins with the activation domain. Double transformants were selected on plates lacking Trp and containing Zeocin (T-Z+). Cells carrying both plasmids were transformed with the plasmid pRH5′ carrying the NRE sequence in the sense and antisense orientations (Cui et al., 2002). The identity and integrity of the pYesTrp3 and pRH5′ fusion products were confirmed by yeast colony PCR.

Activation of *His3* reporter was analyzed by the ability of yeast expressing the corresponding GAL4 fusion proteins to grow on plates lacking histidine (His^-), in the presence of 0.1, 0.5, 1, and 2 mM of 3AT (3-amino-1,2,4-triazole) and incubated at 30 °C for 48 h. The extent of growth on medium minus histidine plus 3AT indicates the strength of the protein–RNA interaction.

2.3. Recombinant protein production and antibody generation

The PUF domain of TcPUF1, from DNA positions 1716 to 3006 (the 430 aa C-terminal domain), was cloned into pGEM-T Easy (Promega), subsequently digested with *Bam*HI and *Not*I and subcloned in frame into the Gateway entry vector pENTR2B (Invitrogen). The ORF was Gateway transferred from pENTR2B to pDEST17 destination vector (Invitrogen) using LR clonase according to manufacturer's instructions. The plasmid was transformed into *Escherichia coli* BL-21 pLys and the recombinant His-tagged protein was expressed by induction with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 12 h at 37 °C. The His-TcPUF1 domain was present in inclusion bodies and it was purified using Ni–NTA columns following the manufacturer's instructions (Qiagen).

A total of 100 µl of a 0.4 µg solution of recombinant protein was mixed with 100 µl of Freund's adjuvant. The solution was injected intraperitoneally into BALB/c mouse three times with two week intervals between each dose. One week after the third injection, blood was collected and serum was obtained.

Western blot analysis was performed as previously described (Vazquez et al., 2003) with TcPUF1 domain mouse serum diluted 1:200.

2.4. Immunofluorescence assays

Trypanosoma cruzi epimastigotes were fixed in suspension with 3.8% paraformaldehyde and permeabilized with 0.2% Triton X-100. The anti-His-tagged TcPUF1 domain mouse serum was diluted 1:100 in PBS–1% BSA. Bound antibody was detected with FICT-conjugated goat anti-mouse IgG. Images were captured from a Leica DM LB microscope coupled to a CCD camera.

3. Results

3.1. In silico search and characterization of Pumilio homology domains in the T. cruzi genome

The PUF PFam HMM profile (PF00806) was used to search a translated version of the complete WGS individual reads sequences of the *T. cruzi* genome (see Section 2). We found 89 positive clones that were assembled into nine different contigs (Results A, Fig. 1). Each contig contained an ORF with a protein bearing Puf repeats. We used these candidate PUF proteins to build a *T. cruzi* HMM specific profile. The new profile was used to search again the complete WGS individual reads. We found 192 positive clones that were assembled into 10 different contigs (Results B, Fig. 1) each one containing a putative PUF protein, nine of them were coincident with the Results A (Fig. 1). The initial candidates (Results A) were also used in the motif discovery tool MEME (see Section 2). The MEME motives obtained were converted into HMM based profiles using Meta-MEME and subsequently used to search the complete WGS individual reads (Results C, Fig. 1). This strategy identified the same contigs as Results B.

Based on the three search results, we identified 10 different PUF proteins in the haploid genome of *T. cruzi*. The proteins were named TcPUF1 to TcPUF10 according to their highest and lowest scores in the HMM specific profile search and deposited in GenBank (see Section 2 for accession numbers).

Our bioinformatics analyses were done before the initial assembly of *T. cruzi* genome scaffolds. We recently checked our results against the *T. cruzi* genome database v4.0 release (http://www.genedb.org/genedb/tcruzi/) and all the PUF proteins were confirmed. In fact, our naming convention was used in the database.

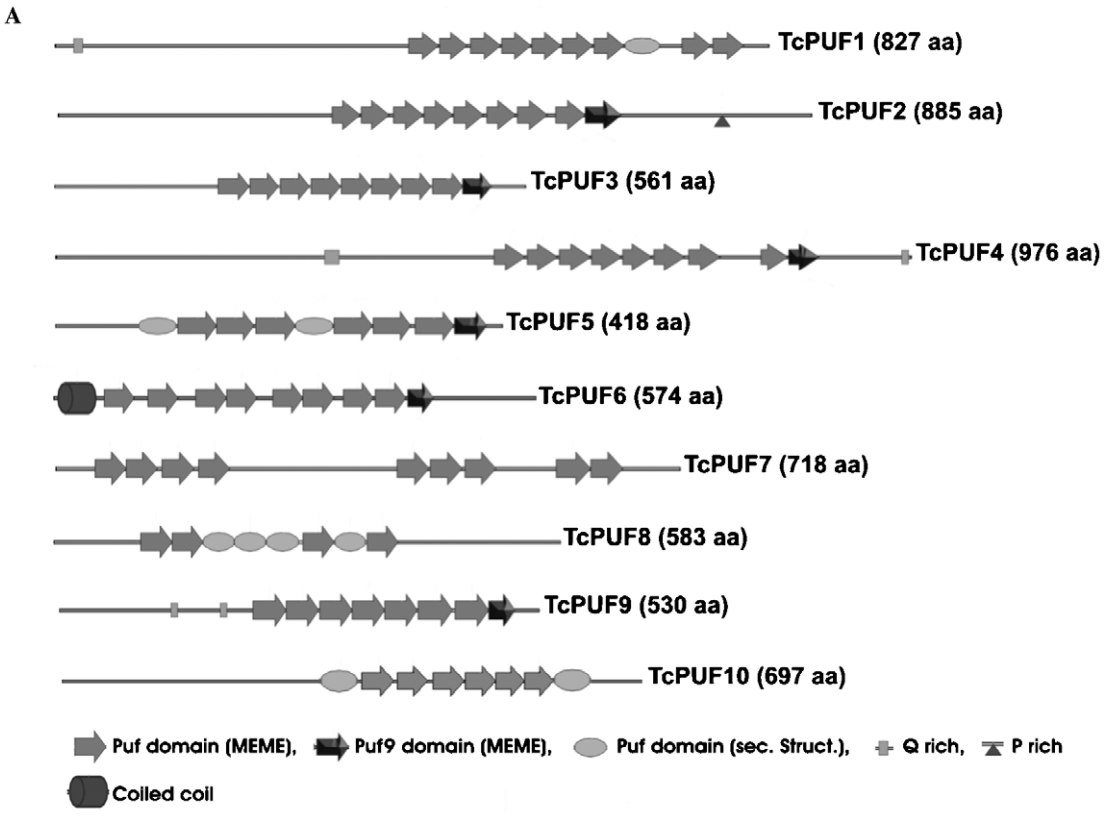
Typically a PUF protein bears eight imperfect Puf repeats of 36–40 amino acids and an alternative divergent ninth repeat that serves as a C-terminal cap of the hydrophobic core (9/c repeat, Edwards et al., 2001). To determine the exact number of repeats present in the TcPUF proteins, we used a combination of bioinformatics tools including as HMM profile searches, MEME discovery tool and secondary structure prediction analysis. This strategy was necessary since the PUF HMM profile failed to identify all the Puf repeats in the TcPUF proteins.

The schematic representation of the 10 TcPUF proteins is depicted in Fig. 2A. The most conserved Puf repeats were detected by HMM profile and MEME analyses. The divergent 9/c like repeat present in TcPUF2, TcPUF3, TcPUF4, TcPUF5, TcPUF6, and TcPUF9 was clearly detected by MEME analysis (Puf 9 domain in Fig. 2A). The less conserved Puf repeats escaped the profile based methods. Instead they were detected only when using secondary structure prediction analysis since they present a typical three helices–two loops structure (Edwards et al., 2001, ovals in Fig. 2A). Various TcPUFs included additional motifs rich in glutamine or proline amino acids that were detected by MEME (Fig. 2A).

The largest PUF protein is TcPUF4 with 976 aa and the shortest is TcPUF5 with 418 aa. Our evolutionary analysis showed that TcPUF1 and TcPUF2 were the closest relatives to the known eukaryotic PUF proteins while TcPUF7, TcPUF8, and TcPUF10 were the most divergent (Fig. 2B).

Analysis of the paralogous TcPUFs was restricted to the PUF domains since the proteins showed no homology outside them. The amino acid sequence comparisons were done using TcPUF1 as a reference. The results indicated that even within the PUF domains the homology was variable (Fig. 2C) due to the divergent nature of some of the Puf repeats. The highest scores were for TcPUF2 (31% identities and 48% positives) while the lowest scores were for TcPUF8 (19% identities and 36% positives).

To perform a comparative genomic analysis of PUF proteins of the TriTryp, we searched the *Trypanosoma brucei* and *Leishmania major* genome databases (http://www.genedb.org) by blastp using as query the protein sequences of *T. cruzi*. The collection of protein sequences retrieved were aligned as pairs with each of the TcPUFs using the BLAST 2 sequences tool to produce the results in Table 1. We evaluated the proportions of identical amino acids, positive amino acids (conservative changes), and gaps introduced to maintain maximum homology with the corresponding protein in *T. brucei* and *L. major*. All 10 TcPUFs were found in the genome of *T. brucei* and nine were found in the genome of *L. major* since TcPUF4 was absent (Table 1). This analysis involved the whole protein sequences and not only the Puf domains. Interestingly, the most conserved PUF protein among the TriTryp was PUF8 which is also the most divergent in relation to the other eukaryotes in our analysis (Table 1, Fig. 2B and sections below). In contrast, TcPUF4 was the less conserved and the *L. major* PUF9



C

	TcPUF1			
	%			aa
	I	P	G	L
TcPUF2	31	48	17	380
TcPUF3	20	39	23	310
TcPUF4	24	42	8	381
TcPUF5	29	46	8	223
TcPUF6	21	37	19	275
TcPUF7	22	40	9	162
TcPUF8	19	36	5	256
TcPUF9	22	40	7	299
TcPUF10	22	44	5	159

Fig. 2. (A) Schematic representation of TcPUF proteins. Light arrowheads represent Puf repeats detected by HMM and MEME analysis, dark arrowheads represent the Puf 9/c repeat detected by MEME, ovals represent putative Puf repeats detected by secondary structure prediction. (B) Phylogenetic tree of PUF proteins. MusPUM2, *Mus musculus* PUF; XePUF, *Xenopus laevis* PUF; CePUF(FBF-1), *C. elegans* PUF FBF-1; DmPUF, *D. melanosgaster* Pumilio; DdPUFA, *D. discoideum* Pufa. NRE-FBE group indicates PUF proteins that recognize the NRE or FBE RNA targets sequences of *D. melanogaster* or *C. elegans*, respectively. UGUR group indicates PUF proteins that putatively recognize the UGUR core nucleotide sequence. Unknown group indicates PUF proteins with no obvious target sequence specificity. (C) Analysis of the paralogous family using TcPUF1 as a reference. I, identical residues; P, positive residues; G, gaps; aa, amino acids; L, sequence length in amino acids included in the comparison.

presented a 180 aa N-terminal deletion and a 130 aa C-terminal extension compared to the homologous proteins in *T. cruzi* and *T. brucei*. These changes did not involve the PUF domain. The protein that we here named TcPUF6 was previously published as TbPUF1 in *T. brucei* (Hoek et al., 2002).

3.2. In silico analysis of the binding specificity of TcPUF proteins

It was demonstrated that a high-affinity RNA ligand binds the concave surface of human Pumilio1, where each of the protein's eight repeats makes contact with a different

Table 1
Comparative analysis of PUF proteins in the TriTryp

Trypanosoma cruzi	*Trypanosoma brucei*			*Leishmania major*		
(%)	Identities	Positives	Gaps	Identities	Positives	Gaps
PUF1	64	74	2	46	57	8
PUF2	47	59	8	35	45	22
PUF3	61	72	2	54	69	4
PUF4	38	52	14	n.f	n.f	n.f
PUF5	54	69	2	47	67	5
PUF6	56[a]	72	1	45	65	5
PUF7	54	69	5	44	60	11
PUF8	62	77	3	57	74	1
PUF9	45	57	9	50	68	1[b]
PUF10	56	72	1	38	56	8

n.f, not found.
[a] TbPUF6 was first published as TbPUFI.
[b] LmPUF9 presents a 180 aa N-term deletion and a 130 aa C-term extension.

RNA base via three amino acid side chains at conserved positions. Mutations in one repeat alter the sequence specificity in a predicted manner (Wang et al., 2002). Based on these data, we analyzed the key positions in each repeat in each TcPUF protein as described by Wang et al. (2002) (Table 2A) which are position 10 (van der Waals contact with ribose ring), position 12 (hydrogen bond or van der Waals contact with RNA base), position 13 (stacking with RNA base), position 16 (hydrogen bond with RNA base), position 17 (electrostatic network), and position 20 (electrostatic network). We included in the comparison the human Pumilio1 and the *Dictyostelium discoideum* PufA (Souza et al., 1999). The results are presented in Table 2A, where for each of the positions described, columns show each PUF protein analyzed and rows show each repeat of the PUF domain. The NRE RNA target is shown in the last column on the right (Wang et al., 2002).

From this analysis we speculated that TcPUF1 and TcPUF2 would probably share the same binding specificity with human Pumilio 1 for the NRE (see grey shaded amino acids in Table 2A). Identical amino acids were found in the key positions 12, 13, and 16. Moreover, position 10 presented changes only in repeat 1 and the electrostatic network in positions 17 and 20 were found mostly conserved.

The proteins TcPUF3, TcPUF4, TcPUF5, TcPUF6, and TcPUF9 would be capable of binding at least the UGUR core sequence in NRE (see grey shaded amino acids in Table 2A). Identical or conserved amino acids were found in repeats 6, 7, and 8 in positions 12, 13, and 16. Note that TcPUF4 presented a single non-conservative change (Q to A) in the repeat 6 of position 16 that is involved in binding to the core sequence. This needs to be tested in vitro and in vivo to determine if this change affects binding affinity.

Because of the more divergent nature of TcPUF7, TcPUF8, and TcPUF10, we were unable to determine their putative binding specificity using this approach. In fact, our analysis showed clearly that key positions necessary for binding to the UGUR core were not conserved in these three PUF proteins (see positions 12, 13, and 16 for TcPUF7, TcPUF8, and TcPUF10 in Table 2A). This result is particularly interesting since all PUF proteins described previously bind at least to the UGUR core (Eckmann et al., 2004; Gerber et al., 2004; Jackson et al., 2004; Lamont et al., 2004; Nakahata et al., 2001; Souza et al., 1999; Tadauchi et al., 2001; Zamore et al., 1997; Zhang et al., 1997).

A summary of the putative affinities of TcPUF proteins is presented in Table 2B. Our results indicated the presence of at least three binding groups: the NRE group, the UGUR group, and the Unknown binding group. This was also supported by the evolutionary analysis showed in Fig. 2B where TcPUF proteins essentially grouped in these three main categories.

3.3. Yeast three hybrid analysis of TcPUF binding specificity

To test the binding predictions of our in silico analysis for TcPUFs, we assayed RNA–protein interactions in vivo using the yeast three hybrid system. We used the NRE sequence (Cui et al., 2002) as a target RNA or the antisense NRE sequence as a negative control. These RNAs were produced as hybrid RNAs with the MS2 sequence which binds with high affinity the LexA/MS2 coat fusion protein. We selected representative TcPUF proteins from each group, TcPUF1, TcPUF3, TcPUF5, and TcPUF8 (see Table 2B) to produce fusion proteins with the VP16 activation domain. Activation of *His3* or *LacZ* reporter genes occurs when the RNA–protein complex is established (see Fig. 3A and Section 2). The affinity of the interaction is directly related to the level of reporter gene expression (Bernstein et al., 2005).

We evaluated the activation of *His3* reporter in the presence of 1 or 2 mM 3AT, a competitive inhibitor of the *His3*-encoded enzyme.

The results indicated that TcPUF1 specifically interacted with NRE as predicted since no interaction was detected with the NRE antisense (Fig. 3B), although the yeasts did not grow well at concentrations beyond 2 mM 3AT suggesting that the affinity for NRE is lower than that of the human Pumilio1 (Wang et al., 2002).

Neither TcPUF3/TcPUF5 (UGUR group) nor TcPUF8 (Unknown binding group) recognized the NRE sequence (Fig. 3B), although TcPUF3 yeasts showed weak growth at concentrations below 1 mM 3AT (not shown), suggesting that the UGUR core of NRE was bound but the overall affinity was very low.

Together these results confirmed some of the predictions observed in silico (Table 2A). Mutations shown to abolish RNA-binding in *D. melanogaster* and human Pumilio proteins (Edwards et al., 2001, oval shaded amino acids in Table 2A) are depicted in the structural model in Fig. 3C (DmPUF). The structural models for the Puf domains of TcPUF1 and TcPUF3 were obtained from the SWISS-MODEL protein homology modeling server using the *D. melanogaster* protein domain as a template. As shown in Fig. 3C, the key residues for RNA-binding were all

Table 2
In silico analysis of binding specificity of TcPUF proteins

	HsPUM	DdPufA	TcPUF1	TcPUF2	TcPUF3	TcPUF4	TcPUF5	TcPUF6	TcPUF7	TcPUF8	TcPUF9	TcPUF10	
(A)													
Position 10 (van der Waals with ribose rings)													
rep1	H	V	E	D	H	H	S	N	A	R	?	R	
rep2	F	F	F	F	Y	Y	T	N	V	Y	E	D	
rep3	Y	Y	Y	Y	H	H	Q	N	F	F	H	I	
rep4	N	N	N	N	H	N	H	N	Y	K	Y	1	
rep5	Y	Y	Y	Y	Q	Q	Q	Q	V	1	Q	N	
rep6	Y	Y	Y	Y	F	Y	Y	F	A	E	F	Y	
rep7	F	F	F	F	F	F	F	F	Q	Y	Y	N	
rep8	Y	Y	F	Y	F	F	F	Y	H	F	F	G	
Position 12 (hydrogen bond or van der Waals with RNA bases)													RNA
rep1	S	S	S	S	C	C	A	L	H	S	?	C	A3
rep2	N	N	N	N	N	N	R	S	N	H	R	N	U
rep3	C	C	C	C	T	N	C	T	N	I	C	V	A
rep4	N	N	N	N	N	G	R	S	Q	Y	R	N	U/C
rep5	C	C	C	C	C	C	C	C	S	Y	C	S	A
rep6	N	N	N	N	N	N	N	N	S	A	N	S	U
rep7	S	S	S	S	S	S	S	S	S	A	S	R	G
rep8	N	N	N	N	N	N	N	N	C	H	N	Q	U5
Position 13 (stacking with RNA bases)													RNA
rep1	R	R	R	R	R	R	E	R	H	R	?	K	A3
rep2	Y	Y	Y	Y	F	F	H	E	T	F	S	F	U
rep3	R	R	R	R	F	Y	H	R	H	E	H	H	A
rep4	H	H	H	H	H	H	R	H	R	P	R	L	U/C
rep5	R	R	R	R	C	C	I	C	P	P	1	L	A
rep6	Y	Y	Y	Y	Y	Y	Y	Y	Y	P	Y	H	U
rep7	N	N	N	N	N	N	N	N	R	P	N	F	G
rep8	Y	Y	Y	Y	Y	Y	F	Y	Y	Q	F	H	U
rep9	h		P	h	g	r	r		r				g A5

Position 16 (hydrogen bonds with RNA bases)

													RNA
rep1	Q	Q	Q	Q	Q	Q	L	M	L	Q	?	Q	A3
rep2	Q	Q	Q	Q	Q	Q	V	Q	Q	T	L	Q	U
rep3	Q	Q	Q	Q	Q	Q	R	Q	V	N	R	V	A
rep4	Q	Q	Q	Q	Q	Q	Q	A	Q	E	Q	Q	U/C
rep5	Q	Q	Q	Q	Q	Q	I	Q	Q	Q	M	R	A
rep6	Q	Q	Q	Q	Q	A	Q	Q	Q	T	Q	G	U
rep7	E	E	E	E	E	E	E	E	Q	A	E	E	G
rep8	Q	Q	Q	Q	Q	Q	Q	Q	Q	L	Q	C	U5

Position 17 (Electrostatic network)

rep1	L	Q	R	R	R	S	L	K	S	S	?	R
rep2	K	K	K	K	K	K	S	N	E	A	A	K
rep3	K	K	K	K	K	K	A	K	K	A	A	K
rep4	K	K	K	K	K	K	A	R	K	E	A	K
rep5	R	R	C	C	R	K	R	K	K	R	R	C
rep6	H	H	H	H	Y	H	C	F	A	T	C	H
rep7	K	K	K	K	K	I	K	K	K	R	K	K
rep8	K	K	R	R	T	K	A	T	A	H	S	R

Position 20 (Electrostatic network)

rep1	E	E	E	E	G	E	E	D	A	K	?	S
rep2	E	E	E	D	D	E	R	G	E	R	R	G
rep3	E	E	E	E	E	D	E	D	E	S	D	V
rep4	E	E	E	E	Q	Q	E	A	E	Q	E	K
rep5	E	E	Q	E	E	D	E	Q	E	A	E	P
rep6	E	E	Q	L	E	R	Q	D	Q	F	E	E
rep7	T	Q	V	V	R	S	W	R	A	D	K	H
rep8	D	D	D	Q	T	S	D	V	R	T	E	L

Affinity for NRE	Affinity for UGUR	Unknown affinity	Most divergent
(B)			
TcPUF1	TcPUF3	TcPUF7	TcPUF8
TcPUF2	TcPUF4	TcPUF8	TcPUFl0
	TcPUF5	TcPUFl0	
	TcPUF6		
	TcPUF9		

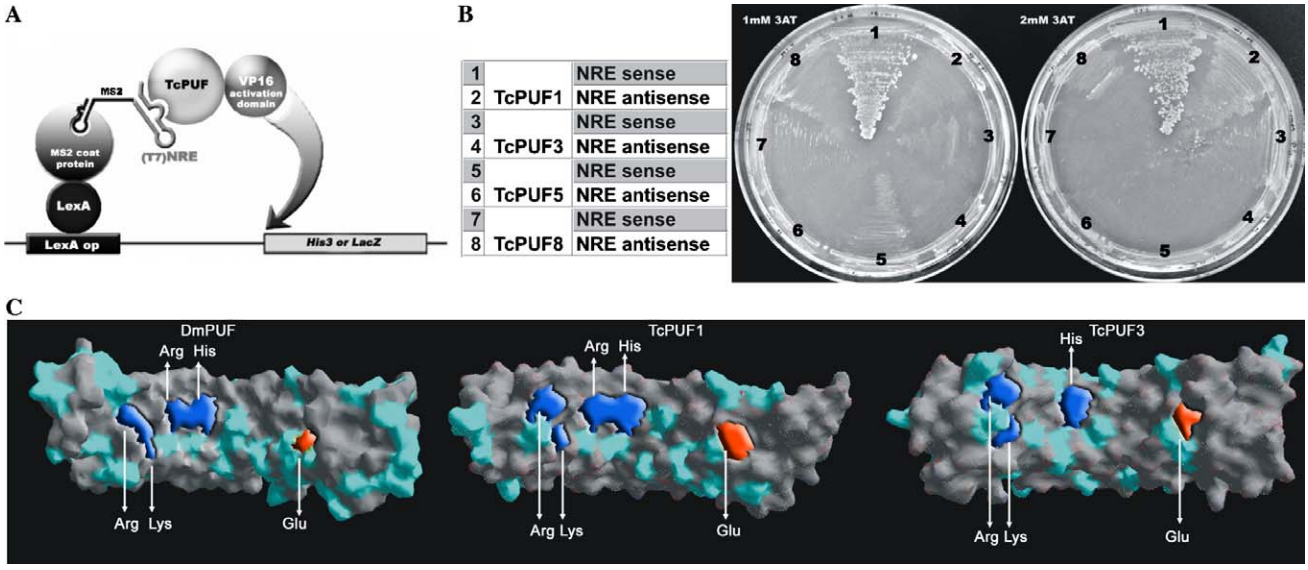


Fig. 3. Yeast three hybrid analysis of RNA binding specificity. (A) Schematic representation of the yeast three hybrid system. (B) Yeast three hybrid analysis of selected TcPUF proteins with NRE sense or antisense as a target. Numbers in the plates correspond to numbers in the table to the left. (C) 3D models of PUF domains, concave surface view. DmPUF, *D. melanogaster* PUF proteins. Arrows point to amino acids known to be critical for RNA-binding in DmPUF.

conserved in TcPUF1 but a single non-conservative change occurred in TcPUF3 (R to F, see Table 2A) that might greatly reduce its affinity for NRE. Other three mutations in relation to TcPUF1 might also account for the difference: S to C in position 12 repeat 1, C to T in position 12 repeat 3 and R to C in position 13 repeat 5. A similar situation occurred with TcPUF5.

The results for TcPUF8 were expected since we showed that the key residues for NRE binding were all changed (see Table 2A and Fig. 3B).

3.4. In silico search of TcPUF NRE group targets in the *T. cruzi* genome

As we demonstrated that TcPUF1 was capable of binding to NRE, we decided to search for NRE-like sequences in the genome of *T. cruzi*. To do so, we designed an in silico strategy to search for a regular expression based on NRE sequence features: (GTTGT)(.{4, 6})(ATTGTA). In this expression, the UGU core in NRE boxA could be separated by four to six bases from the UGUR core in NRE boxB preceded by AU (Wang et al., 2002). We searched for this regular expression in the complete WGS individual reads sequences of the *T. cruzi* genome and in the *T. cruzi* dbEST database. The initial search found 31 positive clones in the WGS database that matched these criteria which we named Target 1–31. To accurately filter this search, we established two additional criteria: (A) the target sequence must be associated to the 3′ end region downstream the ORF or, if possible, based on data obtained from dbEST, to the 3′-UTR; (B) the 3′ end region is defined as a region laying between the ORF's stop codon and a site located approximately 100–140 nucleotides upstream of the most proximal polypyrimidine tract associated downstream of the ORF. This condition ensures a probable 3′-UTR function of the defined 3′ end region of that ORF.

After applying these filters, we selected five positives from the initial number of 31 targets (Fig. 4A). Within the selected clones, we found the target sequence in the 3′ end region downstream an ORF of unknown function specific of *T. cruzi* and *L. major*; Cox5, a subunit of the mitochondrial cytochrome oxidase; a kinetoplast DNA-associated protein; a repeat motif protein of unknown function and a TriTryp specific putative protein kinase (Fig. 4A). Additionally, EST searches allowed us to determine that the target sequence was indeed contained within the 3′-UTR of Cox5 and the putative protein kinase genes (see Target 2 and Target 8 in Fig. 4A).

Alignment of these target sequences with other known targets of PUF proteins of yeast, *C. elegans* and *D. melanogaster* showed additional conservation of AU or UA dinucleotides downstream the UGUR core as previously determined (Fig. 4B, Bernstein et al., 2005; Gerber et al., 2004). Interestingly, these nucleotides were not included in the original regular expression. Moreover, another target sequence UGUR(3/4)AU was found upstream in Targets 1, 5, 7, and 8 (Fig. 4B).

3.5. Cellular localization of TcPUF1

One important aspect of the PUF function is its subcellular localization. We produced polyclonal antibodies against the His-tagged TcPUF1 domain (430 aa C-terminal sequence) in mice (see Section 2) and performed immunofluorescence assays in *T. cruzi* epimastigotes.

We used the PUF domain instead of the full TcPUF1 protein because bacterial expression of the domain was

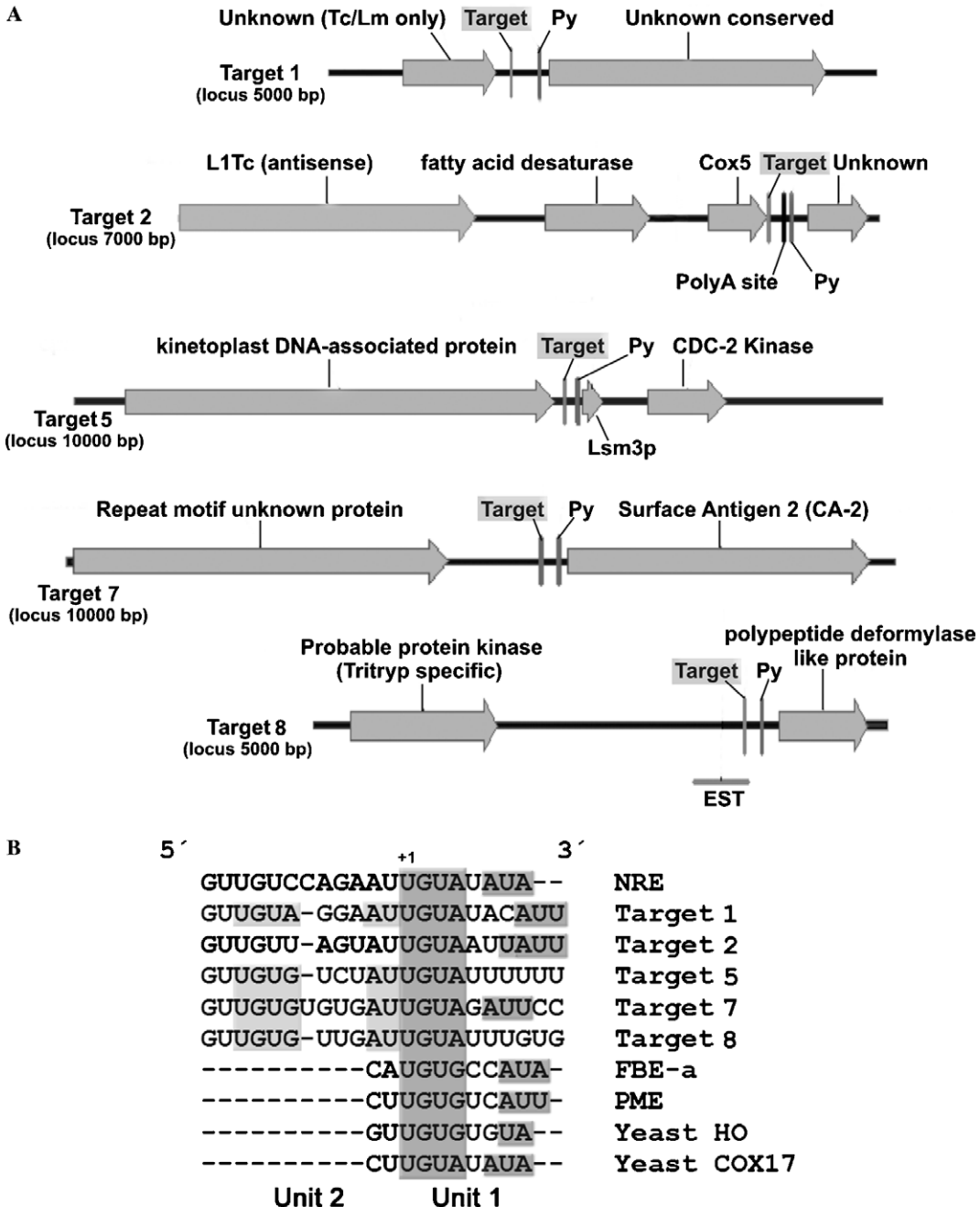


Fig. 4. In silico search of putative targets for the TcPUF NRE group. (A) Schematic representation of five genomic locations with ORFs bearing the regular expression (GTTGT)(.{4, 6})(ATTGTA) in the 3′ non-coding region. Grey shaded target, location of the regular expression. Py, polypyrimidine rich region. Unknown, the function of that ORF is not known at present. (B) Comparison of sequence targets found with the targets of other eukaryotic PUF proteins. NRE, target of *D. melanogaster* Pumilio protein FBF-a, *C. elegans* target of FBF-1; PME, *C. elegans* target of FBF-2; yeast HO, yeast target of PUF5; yeast COX17, yeast target of PUF3. +1 indicates the position of the UGUR core.

more efficient, rendering maximum level in 3 h. Both full-length TcPUF1 and the domain collected in inclusion bodies. (see the 49 kDa band in Fig. 5A).

Western blot analysis with PUF domain polyclonal antibodies recognized the recombinant His-PUF domain (Fig. 5B, lane 1) as expected. Analysis in *T.cruzi* epimastigote extracts was done using the soluble cytoplasmic fraction (Fig. 5B, lane 2) and the nuclear fraction (Fig. 5B, lane 3). The results indicated that a protein was recognized only in the cytoplasmic fraction. Interestingly, two bands were detected of about 62 and 30 kDa which together sum up to a protein of 92 kDa which is the approximate size of TcPUF1. This result was obtained consistently in the presence of classical anti-proteolitic cocktails.

Immunofluorescence assays to detect cellular localization showed that TcPUF1 was restricted to the cytoplasm of the parasite and it was clearly excluded from the nucleus (Fig. 5C). These results are in accordance with that of the

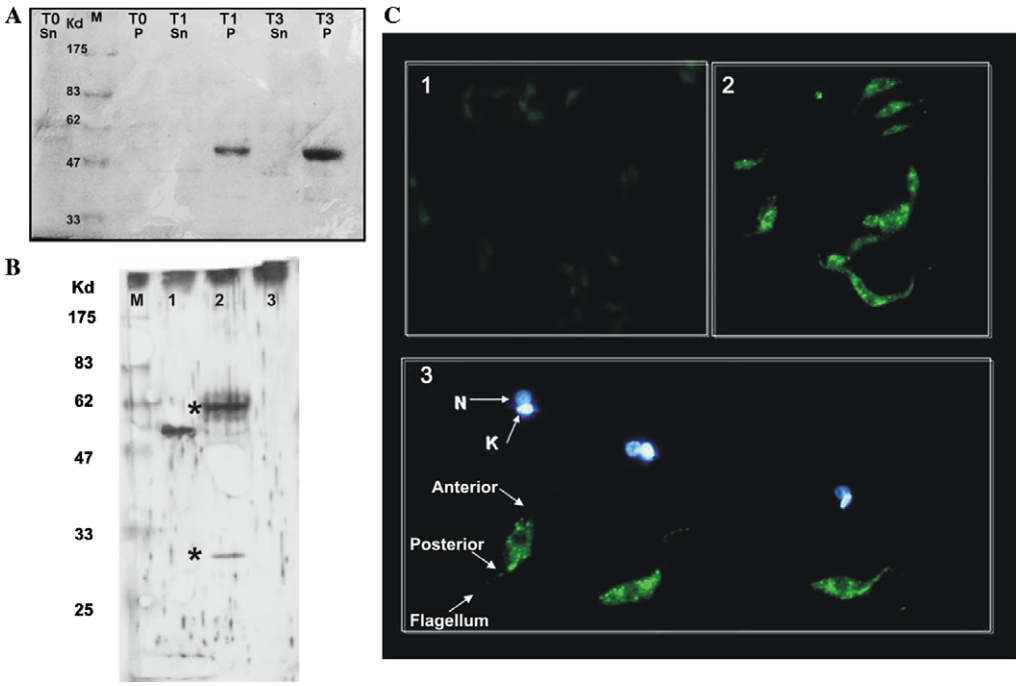


Fig. 5. Cellular localization of TcPUF1 in *T. cruzi* epimastigote cells detected by indirect immunofluorescence assays and Western blot. (A) Bacterial protein expression and purification of His-PUF domain of TcPUF1. Induction was performed at three timepoints: 0, 1, and 3 h (T0, T1, and T3). Soluble (Sn) and pellet (P) fractions were purified by Ni–NTA columns and visualized in PAGE–SDS stained with Coomassie blue. (B) Western blot analysis using the His-PUF1 domain mouse polyclonal antibodies. Lane M, molecular weight marker. Lane 1, recombinant His-PUF1 domain. Lane 2, *T. cruzi* epimastigote extract (cytoplasmic fraction). Lane 3, *T. cruzi* epimastigote extract (nuclear fraction). Asterisks indicate the two putative proteolytic products of complete TcPUF1. (C) immunofluorescence assays. Panel 1, fixed epimastigotes incubated with a pre-immune mouse serum. Panel 2, fixed epimastigotes incubated with anti-His-tagged TcPUF1 mouse serum. Panel 3, same as in panel 2, plus DAPI staining of DNA. N, nucleus; K, kinetoplast. Images were obtained from a Leica DM LB microscope coupled to a CCD camera and processed by Zeiss LSM Image Browser 3.0 and Adobe Photoshop 6.0 software packages.

Western blots (Fig. 5B, lanes 2 and 3). Moreover, the cytoplasmic localization was not diffuse but rather concentrated in multiple discrete foci that predominantly excluded the apical zone, the flagellar pocket and the flagellum itself (Fig. 5C, panel 3).

Our results are in accordance with PUF subcellular localizations described in previous reports for other eukaryotes such as yeast (Gerber et al., 2004), *T. brucei* (PUF6, previously reported as TbPUF1, Hoek et al., 2002), and *T. cruzi* (TcPUF6, Dallagiovanna et al., 2005).

4. Discussion

A combination of in silico strategies led us to determine that the haploid genome of *T. cruzi* encoded ten different PUF RNA-binding proteins. The same group of proteins was present in the genome of *T. brucei* but the orthologue of TcPUF4 was lost in the genome of *L. major*. Interestingly, the most divergent PUF protein (TcPUF8) compared to other eukaryotes was the most conserved among the TriTryp.

Since the TriTryp concentrate gene regulation almost exclusively in post-transcriptional events, one would expect to find an expanded population of RNA-binding proteins. Indeed, the TriTryp genomes contain more than 100 proteins with RRM domains, more than 40 proteins with CCCH zinc finger motives and around 20 proteins with CCHC zinc knuckle motives (Ivens et al., 2005). The PUF proteins reported here are double the number of PUF proteins present in yeasts (Gerber et al., 2004), five times the number present in mammals and other parasites such as *Plasmodium falciparum* and are comparable with the number present in *C. elegans* (Cui et al., 2002; Wickens et al., 2002), although the TriTryp are unicellular organisms.

PUF proteins were implicated in numerous events of RNA metabolism including repression of mRNA translation, enhanced mRNA turnover and mRNA localization (Wickens et al., 2002). It was proposed that yeast PUF proteins bind to functionally related families of mRNAs and localized them to specific cytoplasmic loci. This physical clustering could aid the assembly of complexes and the coordinated control of translation or mRNA turnover (Gerber et al., 2004). This suggestion is compatible with the foci localization of PUF proteins. Interestingly, three TriTryp PUF proteins also showed foci localization (PUF6 in *T. brucei* and *T. cruzi* and PUF1 in *T. cruzi*, Dallagiovanna et al., 2005; Hoek et al., 2002; Fig. 5). The *T. brucei* PUF6 interacts with ESAG8, a nucleolar protein (Hoek et al., 2002). It is tempting to speculate that this PUF protein could aid in the assembly of snoRNPs in discrete foci in the cytoplasm as it happens with the yeast PUF4 (Gerber et al., 2004).

As suggested by Gerber et al. (2004) combinatorial binding of mRNAs by specific proteins, linking their post-transcriptional regulation or specific signal transduction

pathways, could allow rapid and efficient reprogramming of gene expression during development or in response to changing physiological conditions. This is exactly the case for the TriTryp where the control of gene expression is shifted to post-transcriptional events. In this context, the ten PUF proteins could have a fundamental role in these parasites.

In relation to the TcPUF binding specificities, our in silico analysis based on the data for human Pumilio1 (Wang et al., 2002), indicated three distinct groups: NRE, UGUR, and Unknown, suggesting a flexible combination of sequences that could be recognized by TcPUFs. Moreover, our analysis suggests that TcPUF7, TcPUF8, and TcPUF10 do not bind the typical UGUR core. These proteins may recognize new sequence or structural elements in the RNA.

The in silico analysis was supported in part by the in vivo yeast three hybrid experiments. We confirmed that TcPUF1 was able to bind to the NRE sequence but representatives of the two other groups were not. In a previous report it was shown that TcPUF6 binds to NRE sequences in vitro in electrophoretic mobility shift assays (EMSA) (Dallagiovanna et al., 2005). Our in silico analysis and in vivo experiments with TcPUF proteins of the same group (TcPUF3 and TcPUF5) did not agree with that observation. Several facts could account for this discrepancy. First, Dallagiovanna et al. (2005) used different techniques to evaluate binding to NRE sequence. Second, the affinity of the interaction between TcPUF6 and NRE was not estimated and could actually be low. Third, we did not directly evaluate binding of TcPUF6 to NRE in yeast three hybrid assays.

Our in silico search for putative binding targets of TcPUF1 or TcPUF2 led to the analysis of five interesting genomic targets. Two of them, Cox5 and a putative protein kinase, were confirmed to bear the target sequence in their mRNA 3′-UTRs. These types of cellular functions are actual targets of PUF proteins in yeast (Gerber et al., 2004). In fact, yeast PUF3 binds to several mitochondrial mRNAs including Cox17, another subunit of Cytochrome Oxidase (Olivas and Parker, 2000).

Our next approach will certainly include the confirmation of these targets in vivo in *T. cruzi* and an extensive analysis of the sequence requirements for binding of each of the TcPUF proteins.

Acknowledgments

We are grateful to Dr. Liwang Cui for generously providing the pRH5′-NRE plasmid. This work was supported by grants from FONCYT—PICT REDES 2003-00300 and UBACyT X-153 (University of Buenos Aires) to M.V., and World Health Organization/Special Program for Research and Training in Tropical Diseases (WHO/TDR) to M.J.L. In addition, M.J.L was supported in part by an International Research Scholar grant from the Howard Hughes Medical Institute, Chevy Chase, MD, USA. F.C. is a fellowship from the University of Buenos Aires; N.B. and C.A. are fellows from CONICET; M.V. and M.J.L. are members of the career of scientific investigator of CONICET, Argentina.

References

Bernstein, D., Hook, B., Hajarnavis, A., Opperman, L., Wickens, M., 2005. Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. RNA 11, 447–458.

Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M., Vazquez, M., 2005. Protein interactions within the TcZFP zinc finger family members of *Trypanosoma cruzi*: implications for their functions. Biochemical and Biophysical Research Communications 333, 1017–1025.

Clayton, C., 2002. Developmental regulation without transcriptional control. From fly to man and back again. The EMBO Journal 21, 1881–1888.

Cui, L., Fan, Q., Li, J., 2002. The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. Nucleic Acids Research 30, 4607–4617.

Dallagiovanna, B., Perez, L., Sotelo-Silveira, J., Smiricich, P., Duhagon, M.A., Garat, B., 2005. *Trypanosoma cruzi*: molecular characterization of TcPUF6, a Pumilio protein. Experimental Parasitology 109, 260–264.

D'Orso, I., Frasch, A.C., 2002. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and poly(A)-binding proteins forming a ribonucleoprotein complex. The Journal of Biological Chemistry 277, 50520–50528.

Eckmann, C.R., Crittenden, S.L., Suh, N., Kimble, J., 2004. GLD-3 and control of the mitosis/meiosis decision in the germline of *Caenorhabditis elegans*. Genetics 168, 147–160.

Edwards, T.A., Pyle, S.E., Wharton, R.P., Aggarwal, A.K., 2001. Structure of Pumilio reveals similarity between RNA and peptide binding motifs. Cell 105, 281–289.

Gerber, A.P., Herschlag, D., Bro, P.O., 2004. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biology 2, 0342–0354.

Hendriks, E.F., Abdul-Razak, A., Matthews, K.R., 2003. tbCPSF30 depletion by RNA interference disrupts polycistronic RNA processing in *Trypanosoma brucei*. The Journal of Biological Chemistry 278, 26870–26878.

Hendriks, E.F., Matthews, K.R., 2005. Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. Molecular Microbiology 57, 706–716.

Hoek, M., Zanders, T., Cross, G.A., 2002. *Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a Pumilio family protein. Molecular and Biochemical Parasitology 120, 269–283.

Ivens, A.C., Peacock, C., Worthey, E., et al., 2005. The genome of the kinetoplastid parasite, *Leishmania major*. Science 309, 436–442.

Jackson, J.S., Houshmandi, S.S., Lopez Leban, F., Olivas, W.M., 2004. Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. RNA 10, 1625–1636.

Lamont, L.B., Crittenden, S.L., Bernstein, D., Wickens, M., Kimble, J., 2004. FBF-1 and FBF-2 regulate the size of the mitotic region in the *C. elegans* germline. Developmental Cell 7, 697–707.

Murata, Y., Wharton, R.P., 1995. Binding of Pumilio to maternal hunchback mRNA is required for posterior patterning in *Drosophila* embryos. Cell 80, 747–756.

Nakahata, S., Katsu, Y., Mita, K., Inoue, K., Nagahama, Y., Yamashita, M., 2001. Biochemical identification of *Xenopus* Pumilio as a sequence-specific cyclin B1 mRNA-binding protein that physically interacts with a nanos homolog, Xcat-2, and a cytoplasmic polyadenylation element-binding protein. The Journal of Biological Chemistry 276, 20945–20953.

Olivas, W., Parker, R., 2000. The puf3 protein is a transcript specific regulator of mRNA degradation in yeast. The European Molecular Biology Organization Journal 19, 6602–6611.

Sambrook, J., Russell, D., 2001. Molecular Cloning: A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Souza, G.M., da Silva, A.M., Kuspa, A., 1999. Starvation promotes *Dictyostelium* development by relieving PufA inhibition of PKA translation through the YakA kinase pathway. Development 126, 3263–3274.

Tadauchi, T., Matsumoto, K., Herskowitz, I., Irie, K., 2001. Posttranscriptional regulation through the HO 3_-UTR by Mpt5, a yeast homolog of Pumilio and FBF. The European Molecular Biology Organization Journal 20, 552–561.

Vazquez, M., Atorrasagasti, C., Bercovich, N., Volcovich, R., Levin, M.J., 2003. Unique features of the *Trypanosoma cruzi* U2AF35 splicing factor. Molecular and Biochemical Parasitology 128, 77–81.

Wang, X., McLachlan, J., Zamore, P.D., Taaka Hal, T.M., 2002. Modular Recognition of RNA by a Human Pumilio-Homology Domain. Cell 110, 501–512.

Wharton, R.P., Struhl, G., 1991. RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen nanos. Cell 67, 955–967.

Wickens, M., Bernstein, D.S., Kimble, J., Parker, R., 2002. A PUF family portrait: 3′UTR regulation as a way of life. Trends in Genetics 18, 150–157.

Zamore, P.D., Williamson, J.R., Lehmann, R., 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA 3, 1421–1433.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., Wickens, M.P., 1997. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature 390, 477–484.