

Identificación *in silico*, caracterización molecular y análisis de expresión de la proteína de filamento para flagelar PFR3 del *Trypanosoma brucei*

Silico identification, molecular characterization and expression analysis of the Trypanosoma brucei paraflagellar rod protein PFR3

MORELL M, GARCÍA-PÉREZ JL, THOMAS MC, LÓPEZ MC*

Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina «López Neyra», CSIC. Granada – SPAIN. * Corresponding author e-mail: mclopez@ipb.csic.es

RESUMEN

En el presente artículo se describen la identificación y el aislamiento del gen codificante para la proteína PFR3 del *T. brucei*. La secuencia deducida de aminoácidos produce una proteína de 592 residuos con un punto isoeléctrico de 5,14 y presenta una identidad de secuencia del 68,9% con la proteína PFR3 del *T. cruzi*. Sin embargo, el porcentaje de homología entre la proteína PFR3 de *T. brucei* y otras secuencias disponibles de PFRs de *T. brucei* y *T. cruzi* es inferior al 22%. En contraste con lo descrito para los miembros de la familia de proteínas de filamento paraflagelar, la mayor divergencia entre las proteínas PFR3 de *T. cruzi* y *T. brucei* se encuentra en la región central de la proteína, con una similitud del 38% en 200 aminoácidos. Estimamos que existen dos copias de la proteína PFR3 de *T. brucei* por genoma haploide. El gen se transcribe como mRNA de aproximadamente 3,6 kb de longitud, presente con la misma abundancia en formas parasitarias procíclicas y del torrente sanguíneo.

PALABRAS CLAVE: Número de copias. Gen. *Leishmania*. Caracterización molecular. Proteína del filamento paraflagelar. Identificación *in silico*. *Trypanosoma brucei*. *Trypanosoma cruzi*. Transcripción.

ABSTRACT

In the present paper we describe the identification and isolation of the gene coding for T. brucei PFR3 protein. The deduced amino acid sequence produces a protein of 592 residues with an isoelectric point of 5.14 and shows a 68.9% sequence identity with T. cruzi PFR3 protein. However, the percentage of homology among T. brucei PFR3 and other available PFRs sequences from T. brucei and T. cruzi is lower than 22%. In contrast to that described for members of paraflagellar rod protein family, the highest divergence between T. cruzi and T. brucei PFR3 proteins is located at the central region of the protein with a 38% of similarity over 200 amino acid. We estimate that there exist two copies of the T. brucei PFR3 protein per haploid genome. The gene is transcribed as a mRNA of approximately 3.6 kb in length, equally abundant in both procyclic and bloodstream parasite forms.

KEY WORDS: Copy number. Gene. *Leishmania*. Molecular characterization. Paraflagellar rod protein. *Silico* identification. *Trypanosoma brucei*. *Trypanosoma cruzi*. Transcription.

1. INTRODUCCIÓN

La mayoría de las especies de protozoos de los géneros *Trypanosoma* y *Leishmania* son patógenos humanos y animales que se transmiten mediante vectores de insectos chupadores de sangre. Estos parásitos son los agentes causantes de enfermedades muy difundidas que afectan a millones de personas en países distribuidos por todo el mundo. El *T. brucei* es el agente responsable de la enfermedad del sueño en humanos, así como de diversas enfermedades veterinarias endémicas en gran parte de África. La quimioterapia resulta tóxica e ineficiente en el tratamiento de la enfermedad, y aún no se han desarrollado vacunas. Los aspectos singulares de la biología de este parásito ofrecen objetivos plausibles para nuevas formas de quimioterapia e inmunoterapia en el tratamiento de la infección. Uno de ellos es la estructura del filamento paraxial o paraflagelar (PFR), un complejo entramado de filamentos de 150 nm que discurre paralelo al axonema en la mayoría de los flagelos de los tripanosomátidos, euglenoides y dinoflagelados¹⁻⁴. Se han identificado dos principales proteínas PFRs en diversas especies de tripanosomátidos (revisados en ⁵), conocidas como PFR1 y PFR2. Estas proteínas tienen un alto grado de conservación en las especies y un peso molecular entre 70 y 80 kDa en PFR1/PFRC y entre 68 y 72 kDa en PFR2/PFRA. Además, se han descrito varias proteínas situadas en la misma ubicación asociadas al PFR, aunque su función en la estructura del PFR aún se desconoce⁶. Las mutaciones nulas del PFR han evidenciado que la estructura del PFR es necesaria para la motilidad y viabilidad de tanto de los parásitos de *Leishmania* como de *T. brucei*^{7,8}.

En *T. cruzi* se han descrito dos proteínas PFR adicionales, PFR3 y PFR4⁹. Comparten menos del 15% de la identidad de los aminoácidos con la familia de proteínas PFR1/PFR2 del *T. cruzi* o PFRC/PFRA del *T. brucei*, pero sí comparten características bioquímicas similares, como el peso molecular y la estructura de doble espiral predicha^{5,9}. Estas nuevas proteínas conforman claramente la estructura de filamento paraflagelar, pero su función aún no se ha determinado. Se ha descrito que las proteínas de filamento paraflagelar de *T. cruzi*

1. INTRODUCTION

Most protozoan species from genera *Trypanosoma* and *Leishmania* are human and animal pathogens transmitted by bloodsucking insect vectors. These parasites are the causative agents of widespread diseases that affect millions of people in countries worldwide. *T. brucei* is the responsible agent of human sleeping sickness and diverse veterinary diseases endemic in large part of Africa. The chemotherapy against the sickness results inefficient and toxic and vaccines have still not been developed. Singular aspects of the biology of this parasite offer plausible targets for new immuno- and chemotherapies for infection control. One of them is the paraflagellar or paraxial rod (PFR) structure, a complex lattice of filaments of 150 nm that runs parallel to the axoneme throughout most of the flagellum of Trypanosomatids, Euglenoids, and Dinoflagellates¹⁻⁴. Two major PFR proteins have been identified in several trypanosomatids species (reviewed in⁵) referred as PFR1 and PFR2. These proteins are highly conserved over species and present a molecular weight from 70 to 80 kDa for PFR1/PFRC and from 68 to 72 kDa for PFR2/PFRA. Moreover, several collocated PFR associated proteins have been described, although their role in PFR structure is still unknown⁶. PFR null mutants have evidenced that PFR structure is necessary for motility and viability of both *Leishmania* and *T. brucei* parasites^{7,8}.

In *T. cruzi*, two additional PFR proteins have been described, PFR3 and PFR4⁹. They share less than 15% amino acid identity with *T. cruzi* PFR1/PFR2 or *T. brucei* PFRC/PFRA family of proteins, but share similar biochemical characteristics, as molecular weight and predicted coiled-coiled structure^{5,9}. These new proteins clearly conform the paraflagellar rod structure, but their role has not been determined. It has been described that *T. cruzi* paraflagellar rod proteins show a high immunogenicity inducing a protective cellular response^{10,11}. Recently, it has been described that immunization of mice with *T. cruzi* PFR3 and PFR2 recombinant proteins provides protective immunity against *T. cruzi* infection¹². Western blot analysis using *T. brucei* lysates suggests that a PFR3 homologue could exist

presentan una elevada inmunogenicidad que induce una respuesta celular protectora^{10,11}. Recientemente, se ha descrito que la inmunización de ratones con las proteínas recombinantes PFR3 y PFR2 de *T. cruzi* proporciona inmunidad protectora frente a la infección por *T. cruzi*¹². Los análisis Western blot realizados con lisados de *T. brucei* sugieren que podría existir un homólogo del PFR3 en *T. brucei*, dado que se ha observado reactividad cruzada con un anticuerpo monoclonal PFR3 de *T. cruzi*⁹. En el presente artículo se describe el aislamiento del gen codificante de la proteína PFR3 de *T. brucei*, que es homóloga a la PFR3 de *T. cruzi*. También analizamos el número de copias del gen presente en el genoma de *T. brucei* y se muestra el producto de su expresión tanto en formas procíclicas como de torrente sanguíneo del parásito.

2. MATERIALES Y MÉTODOS

Análisis bioinformático: Para investigar la existencia de homólogos de PFR3 en el genoma de *T. brucei* y *Leishmania major*, hemos aprovechado el proyecto del genoma de estos parásitos. Se realizaron búsquedas BLAST en la base de datos del proyecto del genoma del *T. brucei* depositadas en el banco de genes Genbank del NCBI (<http://www.ncbi.nlm.nih.gov>), así como la base de datos GSS del *T. brucei* del EMBL-EBI (<http://www.ebi.ac.uk/blast2/parasites.html>). Ambas bases de datos abarcan más del 50% del genoma del parásito *T. brucei*, basándose en un contenido de ADN de Mb por genoma haploide¹³. Por otra parte, las secuencias de *Leishmania major* depositadas en el EMBL-EBI abarcan más del 50 % del genoma del parásito, basándose en un contenido de ADN de 35 Mb por genoma haploide¹³.

Parásitos: Se utilizaron formas procíclicas y de torrente sanguíneo de la cepa monomórfica 427 de *T. brucei* (Molteno Institute Trypanozoon tipo antigénico 1.2 (MITat 1.2) clon 221a)¹⁴.

Análisis de restricción y aislamiento en *T. brucei* del gen homólogo al PFR3 de *T. cruzi*: El ADN del genoma de *T. bru-*

cei given that cross-reactivity with a *T. cruzi* PFR3 monoclonal antibody has been observed⁹. In the present paper we describe the isolation of the gene coding for *T. brucei* PFR3 protein which is homologous to PFR3 from *T. cruzi*. We also analyze the copy number of the gene present in *T. brucei* genome and show its expression product both in procyclic and bloodstream forms of the parasite.

2. MATERIAL AND METHODS

Bioinformatic analysis: To investigate the existence of PFR3 homologues in the genome of *T. brucei* and *Leishmania major*, we took advantage of the genome project of these parasites. BLAST searches were performed at the *T. brucei* genome project database deposited in Genbank at NCBI (<http://www.ncbi.nlm.nih.gov>) as well as in EMBL-EBI *T. brucei* GSS database (<http://www.ebi.ac.uk/blast2/parasites.html>). Both databases cover over 50% of the *T. brucei* parasite genome, based on a DNA content of 40 Mb per haploid genome¹³. On the other hand, *Leishmania major* sequences deposited at EMBL-EBI covers more than 50 % of the parasite genome, based on a DNA content of 35 Mb per haploid genome¹³.

Parasites: *T. brucei* monomorphic strain 427 procyclic and bloodstream forms (Molteno Institute Trypanozoon antigenic type 1.2 (MITat 1.2) clone 221a) were used¹⁴.

Isolation and restriction analysis of PFR3-like gene from *T. brucei*: *T. brucei* genomic DNA was isolated by standard methods¹⁵. Subsequently, PCR was carried out using *T. brucei* genomic DNA as template and P1Tb5 (5'CAAGGATCCAAAC**ATG**ATCGA-GGTGCAAC3') and P1Tb3 (5'TCTCTAGA-ATTTGCCCT**CT**ACCCACGACCG 3') oligonucleotides, which map, respectively, at the beginning and the end of *T. brucei* PFR3-like coding gene and contain respectively the start and stop codons (bold letters). The DNA amplified fragment was isolated and characterized by restriction mapping and by sequencing after its cloning in pGEM-T vector (Promega).

cei se aisló mediante métodos estándar¹⁵. Seguidamente, se realizó la PCR utilizando ADN del genoma de *T. brucei* como plantilla y oligonucleótidos P1Tb5 (5'CAAGGATCCA-AACATGATCGAGGTGCAAC3') y P1Tb3 (5'TCTCTAGAATTTGCCCTCTACCCA-CGACCG 3'), que se encuentran, respectivamente, al principio y al final del gen codificante para la proteína PFR3 de *T. brucei* y contienen respectivamente los codones de inicio y parada (letras en negrita). El fragmento de ADN amplificado se aisló y se caracterizó mediante el mapa de restricción y secuenciación tras su clonaje en el vector pGEM-T (Promega).

Análisis Northern blot: El ARN total de *T. brucei* se aisló mediante reactivo TRIzol^R (Invitrogene). Se fraccionaron por tamaño 5 µg del ARN total en gel de agarosa-formaldehído al 1% y se transfirieron a una membrana de Z-probe (BioRad) mediante una solución de NaOH de 50 mM. Se amplificó mediante PCR un fragmento de ADN de 1,063 kb correspondiente a los nucleótidos 418 a 1481 del gen *PFR3* de *T. brucei*, se marcó radiactivamente con [α -³²P] dCTP (Feinberg, 1983) y se empleó como sonda. Como control interno se utilizó la sonda 18S, correspondiente a un fragmento de ADN de la subunidad pequeña de ARNr de *T. cruzi* 18S (Marañón et al. 1998). La hibridación para el análisis del nivel de expresión de los ARNs correspondientes se realizó durante 16 horas a 42 °C y 37 °C con, respectivamente, la sonda homóloga y heteróloga en formamida al 50% (v/v), 5x SSC (1x SSC es 0,15M NaCl/0,015 M citrato sódico, pH 7.0), 0,2% SDS, 1x Denhart's, 0,05 M Na₂HPO₄/NaH₂PO₄ tampón y 0,25mg ml⁻¹ de ADN de esperma de arenque. Se realizaron cuatro lavados tras la hibridación en 2x SSC/0,1 % SDS a temperatura ambiente durante 5 minutos y en 0,1x SSC/0,1% SDS (p/v) a 55 °C durante 30 minutos.

3. RESULTADOS Y DISCUSIÓN

Búsqueda en la base de datos de PFR3 de *T. brucei*: Se realizó un análisis BLAST de la base de datos de *T. brucei* del NCBI utilizando como sonda la región de

Northern blot analysis: *T. brucei* total RNA was isolated by TRIzol^R reagent (Invitrogene). 5 µg of total RNA were size-fractionated on 1% agarose/formaldehyde gel and transferred to Z-probe membrane (BioRad) using a 50 mM NaOH solution. A 1,063 kb DNA fragment mapping from nucleotides 418 to 1481 of *T. brucei* PFR3-like coding sequence was PCR amplified, [α -³²P] dCTP random labeled (Feinberg, 1983) and employed as a probe. The probe 18S, a DNA fragment coding for the *T. cruzi* 18S small subunit rRNA (Marañón et al. 1998) was used as internal control. Hybridization for RNA analysis was carried out overnight at 42°C and 37°C with, respectively, the homologous and heterologous probe in 50% (v/v) formamide, 5x SSC (1x SSC is 0.15M NaCl/0.015 M sodium citrate, pH 7.0), 0.2% SDS, 1x Denhart's, 0.05 M Na₂HPO₄/NaH₂PO₄ buffer and 0.25mg ml⁻¹ of herring sperm DNA. Post-hybridization washes were performed four times in 2x SSC/0.1 % SDS at room temperature for 5 minutes and in 0.1x SSC/0.1% SDS (w/v) at 55° C for 30 minutes.

3. RESULTS AND DISCUSSION

***T. brucei* PFR3 database search:** BLAST analysis at NCBI was performed against *T. brucei* database using the 1.787 kb coding region from *T. cruzi* PFR3 protein (TcPFR3) protein (accession code AF004380) as a probe. It revealed the existence of a 0.647 kb genomic sequence (accession code AQ950221, GS-1190236) which, at nucleotide level, share 69% identity over 560 bp with *T. cruzi* PFR3 (see figure 1A, grey line). The deduced amino acid sequence from GS-1190236 (AQ950221) shares 72% identity with the deduced amino acid sequence of TcPFR3 protein, and spans from amino acid 586 to 645. However, it only presents 19% of identity with *T. brucei* PFR A/B proteins (figure not shown). In order to isolate the *PFR3* gene by extending the known sequence towards the 5' and 3' ends, BLAST search was redone against *T. brucei* database using GS-1190236 as query sequence (scheme represented in figure 1A). At this step, only retrieved sequences sharing more than 99% identity with the employed

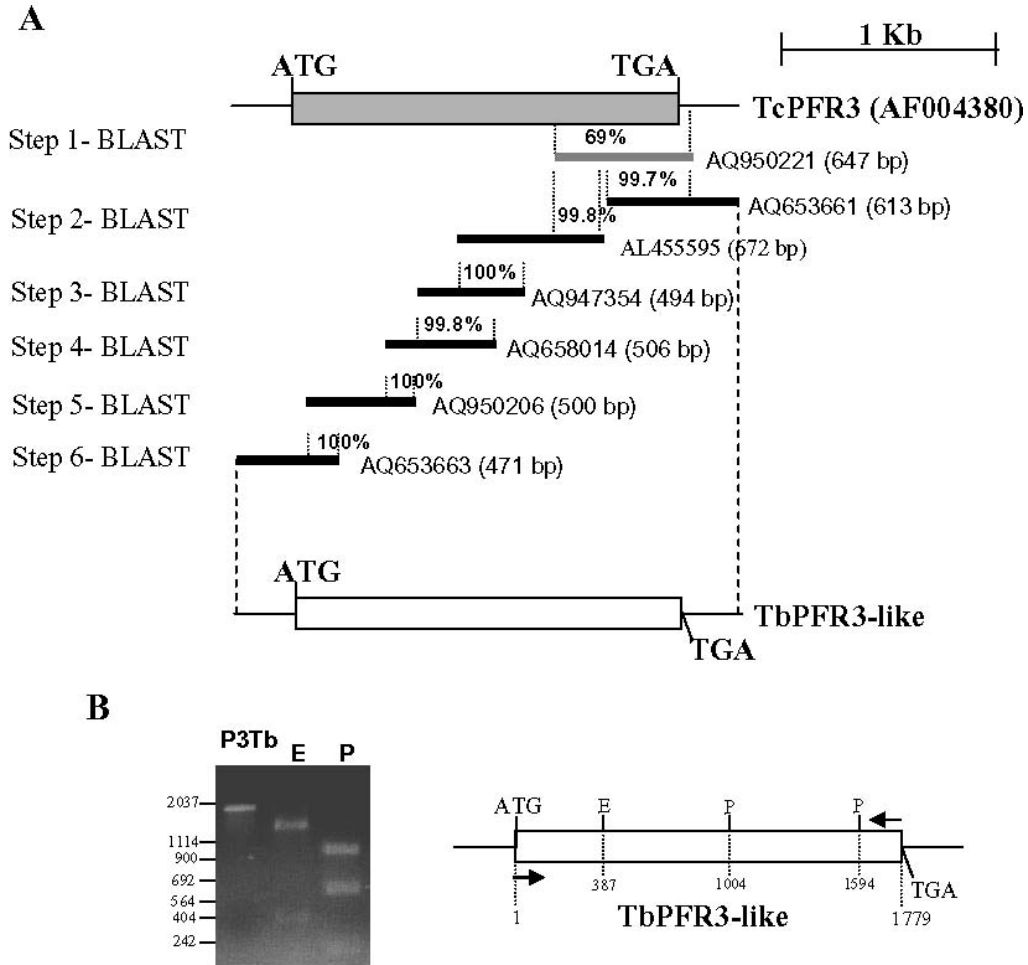
codificación 1,787 kb de la proteína PFR3 de *T. cruzi* (TcPFR3) (código de acceso AF004380). El análisis reveló la existencia de una secuencia genómica de 0,647 kb (código de acceso AQ950221, GSS-1190236) que, en el nivel del nucleótido, comparte un 69% de la identidad en 560 bp con *T. cruzi* PFR3 (véase la figura 1A, línea gris). La secuencia deducida de aminoácidos de GS-1190236 (AQ950221) comparte un 72% de la identidad con la secuencia deducida de aminoácidos de la proteína TcPFR3, y abarca los aminoácidos desde el 586 al 645. Sin embargo, sólo presenta una identidad del 19% con las proteínas PFR A/B del *T. brucei* (figura no mostrada). Para aislar el gen *PFR3* extendiendo la secuencia conocida hacia los extremos 5' y 3', se volvió a realizar la búsqueda BLAST en la base de datos del *T. brucei* utilizando GS-1190236 como secuencia de consulta (esquema representado en la figura 1A). En este paso, sólo se eligieron las secuencias obtenidas que compartían más del 99% de la identidad con la sonda empleada. Tras seis rondas de búsqueda BLAST, se generó una secuencia de ADN de 2,286 bp de *T. brucei*. En el nivel del nucleótido, la secuencia comparte el 62,3% de la identidad con la región de codificación del TcPFR3, lo que indica que este fragmento de ADN corresponde al gen *PFR3* de *T. brucei*, denominándose TkbPFR3-like.

Figura 1: Identificación de PFR 1 de *T. brucei* mediante análisis BLAST. (A) Esquema de los análisis BLAST utilizados para componer la región de codificación del PFR3 de *T. brucei* mediante la unión de fragmentos de ADN depositados en las bases de datos del NCBI (<http://www.ncbi.nlm.nih.gov>) y del EMBL-EBI (<http://www.ebi.ac.uk/blast2/parasites.html>). El recuadro gris representa la región de codificación del PFR3 de *T. cruzi* (AF004380) utilizada como sonda en la búsqueda BLAST, paso 1. La línea gris corresponde a la secuencia GS-1190236 (AQ950221) de *T. brucei*, recuperada en la búsqueda BLAST 1, que fue utilizada como sonda en el paso 2 de la búsqueda BLAST. Las líneas negras representan secuencias extraídas de sucesivas búsquedas BLAST, en las que se muestra encima de cada línea el tamaño (en bp), los números de acceso y el porcentaje de homología.

probe were chosen. After six BLAST rounds, a 2,286 bp *T. brucei* DNA sequence was generated. This sequence shares, at nucleotide level, 62.3% identity with TcPFR3 coding region, arguing that this DNA fragment corresponds to the PFR3 homologue contained in the *T. brucei* genome, so it was denominated TbPFR3-like.

Figure 1: *T. brucei* PFR 1 identification by BLAST analyses. (A) Scheme of the BLAST analyses used to compose the PFR3 coding region from *T. brucei* through joining DNA fragments deposited at NCBI (<http://www.ncbi.nlm.nih.gov>) as well as at EMBL-EBI (<http://www.ebi.ac.uk/blast2/parasites.html>) databases. Grey box represents the *T. cruzi* PFR3 coding region (AF004380) used as a probe in BLAST search, step 1. Grey line corresponds to *T. brucei* GS-1190236 sequence (AQ950221), retrieved from BLAST 1 search, which was used as a probe in step 2 of BLAST search. Black lines represent extracted sequences from successive BLAST searches, where size (in bp), accession numbers and the percentage of homology with the probe used in the BLAST analysis is shown above each line. TbPFR3-like sequence composed by the extracted sequences is represented with a white box. (B) PCR amplification of *T. brucei* PFR3 homologue. Oligonucleotides P1Tb5 and P1Tb3 were used to PCR amplify PFR3 sequence from *T. brucei* genomic DNA. White box indicates the TbPFR3-like sequence, and oligonucleotide position is represented as arrows. *EcoRI* and *PstI* restriction sites in the 1779 amplified DNA are shown as E and P respectively. Numbers are relative to the amplified DNA length. It is shown a 1% agarose gel electrophoresis of the amplified 1.8 kb DNA (lane P3Tb), as well as *EcoRI* and *PstI* digestions of the amplified DNA (lanes E and P, respectively).

Isolation and characterization of *T. brucei* PFR3 gene: To isolate the gene coding for PFR3 protein from *T. brucei*, we synthesized two oligonucleotides which map, respectively, at the ATG and TGA codons from the TbPFR3-like nucleotide sequence, and employed them in a PCR assay using *T. brucei* genomic DNA as template. As it can be ob-



logía con la sonda utilizada en el análisis BLAST. La secuencia del TbPFR3-like compuesta por las secuencias extraídas se representa con un recuadro blanco. (B) Amplificación PCR del homólogo PFR3 de *T. brucei*. Para la amplificación PCR de la secuencia del PFR3 del ADN genómico de *T. brucei* se utilizaron oligonucleótidos P1Tb5 y P1Tb3. El recuadro blanco indica la secuencia del TbPFR3-like, mientras que la posición de los oligonucleótidos se representa mediante flechas. Los sitios de restricción de *EcoRI* y *PstI* en el ADN con amplificación 1779 se indican como E y P respectivamente. Los números se refieren a la longitud del ADN amplificado. Se muestra como electroforesis de gel de agarosa al 1% del ADN amplificado a 1,8 kb (línea P3Tb), así como las digestiones de *EcoRI* y *PstI* del ADN amplificado (líneas E y P, respectivamente).

served in Figure 1B, lane P3Tb, a 1.8 kb amplified fragment was generated, confirming the presence of a PFR3 homologue in the *T. brucei* genome. The amplified DNA fragment was analyzed by restriction mapping using *EcoRI* and *PstI* enzymes which cut one and two times, respectively, in the *T. brucei* PFR3 identified sequence (nucleotide 387 for *EcoRI* and nucleotides 1004 and 1594 for *PstI*, see schematic map in figure 1B). The results, shown in figure 1B, lanes E and P, show fragments with the expected size for *EcoRI*- and *PstI*- digested *T. brucei* PFR3 gene. P1Tb fragment was subsequently cloned in pGEM-T (Promega) and sequenced by the dideoxy chain-termination method¹⁶ in a 3100 genetic analyzer (Applied Biosystems). The nucleotide and deduced amino acid sequences are available in the GenBank data base under accession number AY352895). The TbPFR3-like deduced amino acid sequence is a protein of 592 residues with a molecular mass of 68.5 kDa

Aislamiento y caracterización del gen de *T. brucei* PFR3: Para aislar el gen codificante para la proteína PFR3 de *T. brucei*, se sintetizaron dos oligonucleótidos que corresponden, respectivamente, a los codones ATG y TGA de la secuencia de nucleótidos del TbPFR3-like, y se emplearon en un ensayo de PCR utilizando ADN genómico de *T. brucei* como plantilla. Como se puede observar en la figura 1B, línea P3Tb, se generó un fragmento amplificado de 1.8 kb, lo que confirmó la presencia en el genoma del *T. brucei* de un gen homólogo al PFR3 de *T. cruzi*. El fragmento de ADN amplificado se analizó mediante un mapa de restricción mediante enzimas *EcoRI* y *PstI* que cortan una y dos veces, respectivamente, en la secuencia PFR3 de *T. brucei* identificada (nucleótido 387 en *EcoRI* y nucleótidos 1004 y 1594 en *PstI*, véase el mapa esquemático de la figura 1B). Los resultados, que se muestran en la figura 1B, líneas E y P, muestran fragmentos con el tamaño esperado del gen PFR3 de *T. brucei* digeridos por *EcoRI* y *PstI*. Seguidamente, se clonó el fragmento de P3Tb en pGEM-T (Promega) y se secuenció por el método de terminación de cadena dideoxi¹⁶ en un analizador de genes 3100 (Applied Biosystems). Las secuencias identificadas de nucleótidos y aminoácidos, correspondiente a la proteína PFR3 de *T. brucei*, están disponibles en la base de datos de GenBank a través del número de acceso AY352895. La secuencia deducida de aminoácidos de TbPFR3-like se corresponde con una proteína de 592 residuos con una masa molecular de 68,5 kDa y un punto isoelectrico de 5,14, lo que concuerda con las características acídicas descritas de las PFR de *T. brucei*¹⁷. La alineación de las secuencias de aminoácidos de TbPFR3-like (P3Tb) y *T. cruzi* PFR3 (P3Tc) se muestra en la figura 2A. La comparación de la secuencia deducida de aminoácidos del clon TbPFR3-like con secuencias conocidas de otras proteínas del filamento paraflagelar de *T. brucei* y *T. cruzi* revela un grado de conservación de secuencia significativo con la proteína PFR3 de *T. cruzi*⁹, que alcanza una identidad de secuencia del 68,9%. Estos datos confirman la existencia de un homólogo de PFR3 en *T. brucei*, como se había predicho anteriormente mediante la reactividad cruzada de anticuerpos dirigida

and an isoelectric point of 5.14 which is in agreement with the acidic characteristics described for *T. brucei* PRFs¹⁷. Alignment of TbPFR3-like (P3Tb) and *T. cruzi* PFR3 (P3Tc) deduced amino acid sequences is shown in Figure 2A. Comparison of the TbPFR3-like deduced amino acid sequence with known sequences of *T. brucei* and *T. cruzi* paraflagellar rod proteins reveals a significant sequence conservation degree with *T. cruzi* PFR3 protein⁹, reaching 68.9% sequence identity. These data confirm the existence of a PFR3 homologue in *T. brucei* as it was previously predicted by the cross-reactivity of antibodies directed against *T. cruzi* PFR3 with *T. brucei* cell lysates⁹. The percentage of homology among *T. brucei* PFR3 and other available PRFs sequences from *T. brucei* and *T. cruzi* is lower than 22% (Figure 2B). As it occurs with other members of paraflagellar rod protein families, divergences between *T. brucei* and *T. cruzi* PFR3 proteins accumulate at the amino and carboxyl ends sharing only 3/10 aa at the NH₂ terminus and 1/11 aa at the COOH terminus (Fig. 2A). Moreover, it is observed that *T. brucei* PFR3 protein contains an isoleucine at position two that is not present in the *T. cruzi* homologous protein and lacks the last four amino acid of the carboxyl-terminal *T. cruzi* PFR3 protein. However, in contrast to that described for members of paraflagellar rod protein family⁹, and as it is shown in figure 2A, high divergence occurs at the central region of the protein with a 38% similarity over 200 amino acid. Since this central region has been described to conform a coiled-coil rod structure, common and conserved motif in several cytoskeletal filaments, including the described paraflagellar rod proteins^{5,9}, it is supposed that *T. brucei* PFR3 protein could be provided with a specific functional characteristic. On the other hand, it is interesting to note that the *T. cruzi* PFR3₆₄₋₁₇₁ peptide described as a good target for specific cytotoxic T lymphocytes induced in response to parasite-infected cells¹⁸ is conserved in the *T. brucei* PFR3 (underlined in Figure 2A).

contra PFR3 de *T. cruzi* con lisados de células de *T. brucei*⁹. El porcentaje de homología entre la proteína PFR3 de *T. brucei* y otras secuencias de PRF disponibles de *T. brucei* y *T. cruzi* es inferior al 22% (figura 2B). Al igual que ocurre con otros miembros de las familias de proteínas de filamento paraflagellar, las divergencias entre las proteínas PFR3 de *T. brucei* y *T. cruzi* se acumulan en los extremos amino y carboxilo, compartiendo sólo 3/10 aa en el término NH₂ y 1/11 aa en el término extremo carboxilo (fig. 2A). Además, se observa que la proteína PFR3 de *T. brucei* contiene una isoleucina en la posición dos que no está presente en la proteína homóloga de *T. cruzi* y que carece de los cuatro últimos aminoácidos del extremo carboxilo de la proteína PFR3 de *T. cruzi*. Sin embargo, en contraste con lo descrito para los miembros de la familia de proteínas de filamento paraflagellar⁹, y como se muestra en la figura 2A, se produce una gran divergencia en la región central de la proteína, con una similitud del 38% en 200 aminoácidos. Como se ha descrito que esta región central conforma una es-

Figure 2: Deduced amino acid sequence of the *T. brucei* PFR3 protein. (A) Alignment of the deduced amino acid sequence of *T. cruzi* (P3Tc) and *T. brucei* (P3Tb) PFR3 proteins. Black shade indicates identical residues and grey shade indicates conservative amino acids. Amino acid position is enumerated for each sequence at the left-hand side. Underlined sequence in P3Tc indicates the sequence of the *T. cruzi* PFR3₁₆₄₋₁₇₁ cytotoxic peptide¹⁸. (B) Comparison of deduced amino acid sequences among *T. cruzi* and *T. brucei* paraflagellar rod protein family members. Identity of P3Tb with others PFR proteins is indicated on light grey boxes, and homology between P3Tb and P3Tc is shown on a dark grey box with white letters.

Copy number estimation of PFR3 gene in *T. brucei* genome: Using a similar strategy to that used for copy number estimation of a retroposon in the *T. cruzi* genome¹⁹, we determined the copy number of the *T. brucei* PFR3 gene in the genome of the parasite. For this purpose, and although *T. brucei* database

A

| | | |
|------|-----|---|
| P3Tb | 1 | MIEVQLNS FNL LPP EYPRR LVO EEE ENHRA VAAL IEL AET AIATAENY SGYVDSRL LPLS |
| P3Tc | 1 | M.TVYHEQ FALVPPQYPRR AVQ. EA ENHRA LAALVELVENAIATAENY VAYTEGRLVPLS |
| P3Tb | 61 | SRGFDLRS AAS EMCERY KHEAPCGWTE EKVM RFCEQEVTR EKI AELL SRQPLDV VAVESI |
| P3Tc | 59 | SRGSEL LASKELRERY RHEAPCGWTE QKVMQHCEQEVTR VEEMAEELL LRPPLDV AGIRSI |
| P3Tb | 121 | LSTLRDS TVEF PRGRFLL LRENLS LKPHOP RDL I RDLS EL CGALYEIQ FVDL LGKMR EE |
| P3Tc | 119 | LRTLQETRTQL PRGRFLL MRDNL SA LKPHOP PDLARDLS DV CGALYEIQ FVDL MAELRAE |
| P3Tb | 181 | LA EFD PPL DAAK KVDAL EAHK RAV VG GDV LEVER THRQLIAARYE LVEVCAQR LKAF L |
| P3Tc | 179 | LSELDGER DKVQE K LKDAL HLYE RAV AK GDV VEVER AHRQLIAARYE FVNA CA KL MH QI L |
| P3Tb | 241 | SQN LENOEKGF IAE LQAL ERD SL EA LQQFTDYHSNQ RDVIQE DIK KCS D KLLQE GATHEK |
| P3Tc | 239 | GED MASQESGF AAE MEAL RRD AA DS LSRFAE ALHER RQAFRN DLHNC DK KRLLEDGNHQK |
| P3Tb | 301 | CLE EYKVK ERE MKD NL YA VI SA KQKLVEE IQR KAAELVQ LTA KQREI VDE QIK AKKEEER |
| P3Tc | 299 | CLE VYNAE ERE TAA QI GQ IV EQ KKLVEE LRQ KARELRD I TLKQKE MVEA QVRAKRAEE E |
| P3Tb | 361 | RITTYNEF ANMGTQQK QRLK KCL EYCE RLVLS VVPG IK SYV GEMV KRLP WQKLR DARNE VN |
| P3Tc | 359 | RVTAYNEF VNMEEQQK HRLLRCLAYFDGMEE LTAD LR SYV DEMV TR LPQQNLR QVLDQ LN |
| P3Tb | 421 | DTEAE CFMEAYDP FVECCGELTVKKMHR HDTL SRQ SRL VEHNRS SSMESLDPNM NNYSR AE |
| P3Tc | 419 | DIEAE VFMNAYGGFVSCCGELTVKKMHR LDTL ERQARL LEHNRS SSMESLDPNM SNYSR LE |
| P3Tb | 481 | LEEL IEQMKGV TGVINALNATQDAGEQLF LSVEK AI LE KYER AATPFVHPLQ VHGVKSVE |
| P3Tc | 479 | LDDI IEQMKGV SGVINALNATQDAGEQLF QSVEK GVLAKYER SGTPTFVHPLQ EYGLKSVE |
| P3Tb | 541 | ERDRFVNRSM QVVE DEERKVLK KSVL ERMR RAVE GEETA VELAIRD PVGR A*~*~* |
| P3Tc | 539 | ERNRFVDRSM H VVE NEERKVLK RNVL NRMR QAVE EDEA FE SAIR N LNEEP A APEY* |

B

| | | | | | | | | |
|------------------|------|-------|-------|-------|-------|-------|-------|------|
| <i>T. brucei</i> | PFR1 | PFR1 | | | | | | |
| | PFR2 | 63.3% | PFR2 | | | | | |
| <i>T. cruzi</i> | PFR3 | 19.3% | 19.5% | PFR3 | | | | |
| | PFR3 | 19.8% | 20.3% | 68.9% | PFR3 | | | |
| | PFR2 | 62.6% | 90.2% | 19.6% | 19.9% | PFR2 | | |
| | PFR1 | 87.6% | 65.4% | 20.4% | 20.2% | 65.9% | PFR1 | |
| | PFR4 | 23.6% | 21.2% | 22% | 21.1% | 21.2% | 22.5% | PFR4 |

estructura de filamento de doble espiral, un motivo común que se conserva en diversos filamentos citoesqueleto, entre los que se incluyen las proteínas de filamento paraflagelar descritas^{5,9}, se supone que la proteína PFR3 de *T. brucei* podría tener una característica funcional específica. Por otra parte, es interesante observar que el péptido PFR3₆₄₋₁₇₁ de *T. cruzi*, descrito como un buen objetivo para los linfocitos T citotóxicos específicos inducidos como respuesta a las células infectadas por parásitos¹⁸, se conserva en la PFR3 de *T. brucei* (subrayado en la figura 2A).

Figura 2: Secuencia de aminoácidos deducida de la proteína PFR3 de *T. brucei*. (A) Alineación de la secuencia de aminoácidos deducida de las proteínas PFR3 de *T. cruzi* (P3Tc) y *T. brucei* (P3Tb). El sombreado negro indica residuos idénticos y el sombreado gris aminoácidos conservadores. En la parte izquierda se enumera la posición de los aminoácidos en cada secuencia. La secuencia subrayada de P3Tc indica la secuencia del péptido citotóxico¹⁸ del PFR3₁₆₄₋₁₇₁ de *T. cruzi*. (B) Comparación de las secuencias deducidas de aminoácidos de diferentes miembros de la familia de proteínas de filamento paraflagelar de *T. cruzi* y *T. brucei*. La identidad de P3Tb con otras proteínas PFR se indica en recuadros gris claro, y la homología entre P3Tb y P3Tc se muestra en un recuadro gris oscuro con letras blancas.

Estimación de número de copias del gen PFR3 en el genoma de *T. brucei*: Mediante una estrategia similar a la utilizada para la estimación del número de copias de un retroposon en el genoma de *T. cruzi*¹⁹, determinamos el número de copias del gen PFR3 de *T. brucei* en el genoma del parásito.

does not cover the whole genome, we used the GSS database for estimating the copy number of a repeated sequence²⁰. PFR1 and PFR2 proteins, are characterized for being arranged in a tandem manner in the genome of *T. brucei*, with 4 copies of each PFR gene per haploid genome¹⁷. Copy number estimation per haploid genome (CN) includes the number of GSS (GSS) homologous to the probe, the size of the probe (GS), the size of the haploid genome (HGS=40 Mb), and the number of GSS contained in the library (TGSS=47854): $CN = (GSS \times HGS) / (TGSS \times GS)$ ^{19,21}. BLAST analysis at NCBI against *T. brucei* database using *T. brucei* PFR3 coding sequence as a probe, indicated that the genome of *T. brucei* contains two copies per haploid genome (2 +/- 0.33). These data contrast with the ten copies described for *T. cruzi* PFR3 gene (Fouts et al., 1998) but is in accordance to those present for other *T. brucei* PFR proteins (four copies for PFR1/B and PFR2/D, respectively)¹⁷ or for the *T. cruzi* PFR4 gene for which there is one or two copies per haploid genome⁹.

Expression of the transcripts of *T. brucei* PFR3 gene: To evaluate the level of RNA transcripts of the *T. brucei* PFR3 gene, total RNA extracted from *T. brucei* procyclic and bloodstream forms was probed with an [α -³²P]-labeled DNA fragment corresponding to 1.06 kb of *T. brucei* PFR3 coding region. Northern blot analysis (Fig. 3), evidenced a single hybridization band of approximately 3.6 kb in length in both forms of the parasite. Moreover, densitometry analysis showed that there were no differences in the level of *T. brucei* PFR3 transcripts from procyclic and bloodstream forms, in contrast to that described for *T. cruzi* PFR3 gene, where PFR3 trans-

Con este propósito, y aunque la base de datos del *T. brucei* no abarca todo el genoma, hemos utilizado la base de datos GSS para estimar el número de copias de una secuencia repetida²⁰. Las proteínas PFR1 y PFR2 se caracterizan por estar distribuidas en forma de tándem en el genoma de *T. brucei*, con cuatro copias de cada gen PFR por genoma haploide¹⁷. La estimación del número de copias por genoma haploide (CN) incluye el número de homólogos GSS (GSS) con la sonda, el tamaño de la sonda (GS), el tamaño del genoma haploide (HGS=40 Mb) y el número de GSS contenido en la biblioteca (TGSS=47854): $CN = (GSS \times HGS) / (TGSS \times GS)^{19,21}$. El análisis BLAST de la base de datos del *T. brucei* del NCBI utilizando la secuencia de codificación de la PFR3 de *T. brucei* como sonda, indicó que el genoma de *T. brucei* contiene dos copias por genoma haploide ($2 \pm 0,33$). Estos datos contrastan con las diez copias descritas del gen *PFR3* del *T. cruzi* (Fouts et al., 1998) pero concuerdan con los disponibles para otras proteínas PFR del *T. brucei* (cuatro copias para *PFR1/B* y *PFR2/D*, respectivamente)¹⁷ o para el gen *PFR4* del *T. cruzi*, en el que sólo hay una o dos copias por genoma haploide⁹.

Expresión de las transcripciones del gen PFR3 de *T. brucei*: Para evaluar el nivel de las transcripciones de ARN del gen *PFR3* de *T. brucei*, el ARN total extraído de formas procíclicas y de torrente sanguíneo de *T. brucei* se analizaron usando como sonda un fragmento de ADN de 1,06 kb correspondiente a la región codificante para la proteína *PFR3* de *T. brucei*, marcada radiactivamente con [$\alpha^{32}P$]. El análisis Northern blot (fig. 3) evidenció una banda de hibridación única de aproximadamente 3,6 kb de longitud en ambas formas del parásito. Además, los análisis de densitometría demostraron que no había diferencias en el nivel de transcripciones de *PFR3* de *T. brucei* en las formas procíclica y de torrente sanguíneo, en contraste con las descritas en el gen *PFR3* de *T. cruzi*, donde las transcripciones de *PFR3* son más abundantes en epimastigotas que en tripomastigotas (Fouts et al., 1998).

Búsqueda de PFR3 en la base de datos de *L. major*: En contraste, los análisis de

criptos are more abundant in epimastigotes than in trypomastigotes (Fouts et al., 1998).

PFR3 search in *L. major* database: In contrast, BLAST analyses carried out against *Leishmania major* genome database at NCBI or EMBL-EBI, using as a probe the *T. cruzi* *PFR3* coding region (AF004380) or the identified *T. brucei* *PFR3* coding region, did not retrieve any sequences with identity higher than 24% at amino acid level. This result can be taken as one of three situations, (i) there is *PFR3* protein in *Leishmania spp.*, but the amount of sequenced *Leishmania spp.* genome at the available databases is not enough for *PFR3* homologous gene identification, (ii) *PFR3* protein from *Leishmania spp.* is highly divergent from the known PFRs, or (iii) there is not a *PFR3* homologous gene in *Leishmania spp.* Since there exists a high degree of conservation between *PFR3* proteins from *T. cruzi* and *T. brucei* it is expected the existence of a homologous protein in *Leishmania spp.* Thus, different approaches should be performed in order to identify whether *Leishmania spp.* genome contains a *PFR3* homologue.

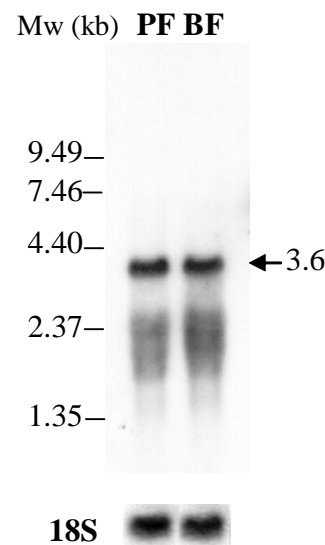
BLAST realizados en la base de datos del genoma de *Leishmania major* del NCBI o el EMBL-EBI, utilizando como sonda la región codificante par la proteína PFR3 de *T. cruzi* (AF004380) o la región de codificación del PFR3 de *T. brucei* identificada, no recuperaron ninguna secuencia con identidad superior al 24% en el nivel del aminoácido. El resultado se puede interpretar de una de las tres formas siguientes: (i) existe proteína PFR3 en *Leishmania spp.*, pero la cantidad de genoma de *Leishmania spp.* secuenciado en las bases de datos disponibles no es suficiente para la identificación de homólogos del gen PFR3; (ii) la proteína PFR3 de *Leishmania spp.* es muy distinta a las PFR conocidas; o (iii) no hay ningún gen homólogo de PFR3 en *Leishmania spp.* Como existe un alto grado de conservación entre las proteínas PFR3 de *T. cruzi* y *T. brucei*, se puede esperar la existencia de una proteína homóloga en *Leishmania spp.* Por tanto, se deben realizar distintos enfoques para identificar si el genoma de *Leishmania spp.* contiene un homólogo de PFR3.

Figura 3: Análisis por Northern blot del transcrito correspondiente a la proteína PFR3 de *T. brucei*. El análisis Northern blot del ARN total de formas procíclicas (PF) y de torrente sanguíneo (BF) de *T. brucei* se realizó con una sonda de ADN de 1,063 kb, correspondiente al fragmento central del gen codificante para la PFR3 de *T. brucei*. El tamaño molecular de la transcripción y el marcador en escalera de ARN (Invitrogen) se indican en kb. El filtro se volvió a hibridar con una sonda de ADNr 18S²² de *T. cruzi* para detectar cualquier diferencia en las cantidades de ARN cargadas en cada carril.

4. AGRADECIMIENTOS

Los datos secuenciales preliminares se obtuvieron del sitio Web del Instituto para la Investigación Genómica (Institute for Genomic Research) en la dirección <http://www.tigr.org>, y estamos sumamente agradecidos al TIGR y al Sanger Institute. La secuenciación está soportada por el National Institute of Allergy and Infectious Diseases,

Figure 3: Northern blot analysis of the *brucei* PFR3 transcript. The Northern blot of total RNA from procyclic (PF) and bloodstream (BF) forms of *T. brucei*, was probed with a 1,063 kb DNA fragment from the central *T. brucei* PFR3 coding region. The molecular size of the transcript and the RNA ladder marker (Invitrogen) are indicated in kb. Filter was rehybridized with a *T. cruzi* 18S rDNA probe²² to detect any differences in loading RNA amounts between tracts.



4. ACKNOWLEDGEMENTS

Preliminary sequence data was obtained from the Institute for Genomic Research website at (<http://www.tigr.org>), and we are extremely grateful to TIGR and Sanger Institute. Sequencing is supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. We would like to thank effort of EMBL-EBI for all the GSS and EST sequences released on the *T. brucei* and *L. major* databases (<http://www.ebi.ac.uk/blast2/parasites.html>). We are grateful to Dr. M. Navarro and Dr. L.M. Ruiz-Pérez (IPBLN-CSIC, Spain) for providing us DNA and RNA from *T. brucei*. This work was supported by Grants PI02/0565 PI02/0862 from FIS, and RICET C03-04 (MSC), Spain.

National Institutes of Health. Descamos agradecer al EMBL-EBI sus esfuerzos para la publicación de todas las secuencias GSS y EST en las bases de datos de *T. brucei* y *L. major* (<http://www.ebi.ac.uk/blast2/parasites.html>). Expresamos nuestro agradecimiento a los Dres. M. Navarro y L. M. Ruiz-Pérez (IPBLN-CSIC, España) por proporcionarnos ADN y ARN de *T. brucei*. Este trabajo ha sido apoyado por las becas PI02/0565 PI02/0862 del FIS y RICET C03-04 (MSC), España.

BIBLIOGRAFÍA/BIBLIOGRAPHY

1. Hyams JS. The Euglena paraflagellar rod: structure, relationship to other flagellar components and preliminary biochemical characterization. *J Cell Sci* 1982;55:199-210.
2. Cosson J, Cachon M, Cachon J, et al. Swimming behavior of the unicellular biflagellate *Oxyrrhis marina*: in vivo and in vitro movement of the two flagella. *Biol Cell* 1988;63:117-26.
3. Saborio JL, Manuel Hernandez J, Narayanswami S, et al. Isolation and characterization of paraflagellar proteins from *Trypanosoma cruzi*. *J Biol Chem* 1989;264:4071-5.
4. Bastin P, Pullen TJ, Moreira-Leite FF, et al. Inside and outside of the trypanosome flagellum: a multifunctional organelle. *Microbes Infect* 2000;2:1865-74.
5. Maga JA, LeBowitz JH. Unravelling the kinetoplastid paraflagellar rod. *Trends Cell Biol* 1999;9: 409-13.
6. Woodward R, Cardew MJ, Gull K. Molecular characterization of a novel, repetitive protein of the paraflagellar rod in *Trypanosoma brucei*. *Mol Biochem Parasitol* 1994;67:31-9.
7. Hunger-Glaser I, Seebeck T. Deletion of the genes for the paraflagellar rod protein PFR-A in *Trypanosoma brucei* is probably lethal. *Mol Biochem Parasitol* 1997;90:347-51.
8. Santrich C, Moore L, Sherwin T, et al. A motility function for the paraflagellar rod of *Leishmania* parasites revealed by PFR-2 gene knockouts. *Mol Biochem Parasitol* 1997;90:95-109.
9. Fouts DL, Stryker GA, Gorski KS, et al. Evidence for four distinct major protein components in the paraflagellar rod of *Trypanosoma cruzi*. *J Biol Chem* 1998;273:21846-55.
10. Wrightsman RA, Miller MJ, Saborio JL, et al. Pure paraflagellar rod protein protects mice against *Trypanosoma cruzi* infection. *Infect Immun* 1995;63:122-5.
11. Miller MJ, Wrightsman RA, Stryker GA, et al. Protection of mice against *Trypanosoma cruzi* by immunization with paraflagellar rod proteins requires T cell, but not B cell, function. *J Immunol* 1997;158:5330-7.
12. Luhrs KA, Fouts DL, Manning JE. Immunization with recombinant paraflagellar rod protein induces protective against *Trypanosoma cruzi* infection. *Vaccine* 2003;21:3058-69.
13. Degraeve WM, Melville S, Ivens A, Aslett M (2001) Parasite genome initiatives. *Int J Parasitol* 31:532-536.
14. Cross GAM. Identification, purification and properties of variant-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*. 1975;71:393-417.
15. Sambroock J, Maniatis T, Fritsh EF (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, U.S.A.
16. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci, USA* 1977;74: 5463-7.
17. Deflorin J, Rudolf M, Seebeck T. The major components of the paraflagellar rod of *Trypanosoma brucei* are two similar, but distinct proteins which are encoded by two different gene loci. *J Biol Chem* 1994;269:28745-51.
18. Wrightsman RA, Luhrs KA, Fouts D, et al. Paraflagellar rod protein-specific CD8+ cytotoxic T lymphocytes target *Trypanosoma cruzi*-infected host cells. *Parasite Immunol* 2002;24:401-12.
19. Bringaud F, Garcia-Perez JL, Heras SR, et al. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 2002;124:73-8.
20. Agüero F, Verdun RE, Frasch AC, et al. A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res* 2000;10:1996-2005.
21. Degraeve WM, Melville S, Ivens A, et al. Parasite genome initiatives. *Int J Parasitol* 2001;31:532-6.
22. Marañón C, Puerta C, Alonso C, et al. Control mechanisms of the H2A genes expression in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 1998;92:313-24.