

**Caracterización molecular de los genes codificantes de proteínas
con dominio específico ALBA (Pfam01918) en *Trypanosoma
cruzi***



Juan Manuel Matiz González

Trabajo de grado

Presentado como requisito parcial para optar por el título de

Microbiólogo Industrial

Carrera de Microbiología Industrial

Facultad de Ciencias

Pontificia Universidad Javeriana

Bogotá D.C., Colombia

junio, 2019

Nota de Advertencia

“La universidad no se hace responsable por los conceptos emitidos por sus alumnos en sus trabajos de tesis. Solo velará por que no se publique nada contrario al dogma y a la moral católica y por que las tesis no contengan ataques personales contra persona alguna, antes bien se vea en ellas el anhelo de buscar la verdad y la justicia”.

Artículo 23 de la Resolución N° 13 de Julio de 1996

**Caracterización molecular de los genes codificantes de proteínas
con dominio específico ALBA (Pfam01918) en *Trypanosoma
cruzi***

Juan Manuel Matiz González

Concepción Judith Puerta, Ph.D
Decana
Facultad de Ciencias

Marcela Franco Correa, Ph.D
Directora
Carrera, Microbiología Industrial

**Caracterización molecular de los genes codificantes de proteínas
con dominio específico ALBA (Pfam01918) en *Trypanosoma
cruzi***

Juan Manuel Matiz González

Paola Andrea Nocua M, Ph.D
Microbióloga Industrial
Tutora

Claudia L. Cuervo Patiño, Ph.D
Bióloga
Tutora

Wilson Terán Pérez, Ph.D
Bioquímico
Jurado

Tabla de contenido

| | |
|---|----|
| 1. Introducción | 1 |
| 2. Marco Teórico | 2 |
| 2.1. Enfermedad de Chagas..... | 2 |
| 2.2. Tratamiento de la enfermedad..... | 3 |
| 2.3. <i>Trypanosoma cruzi</i> | 3 |
| 2.4. Proteínas de Unión a ARN (PUA) | 5 |
| 2.5. Proteínas de la familia ALBA | 7 |
| 2.5.1. Proteínas ALBA en tripanosomátidos | 7 |
| 2.5.2. Proteínas ALBA en <i>T. cruzi</i> | 8 |
| 3. Objetivos | 9 |
| 3.1. Objetivo general | 9 |
| 3.2. Objetivos específicos | 9 |
| 4. Metodología | 10 |
| 4.1. Identificación de los genes codificantes para proteínas con dominio ALBA en <i>T. cruzi</i> y otros tripanosomátidos. | 10 |
| 4.2. Caracterización de las proteínas ALBA identificadas en <i>T. cruzi</i> | 10 |
| 4.3. Caracterización de la proteína TcCLB.510877.40 | 10 |
| 4.3.1. Obtención de ADN del parásito | 10 |
| 4.3.2. Amplificación del gen codificante de TcCLB.510877.40..... | 11 |
| 4.3.3. Caracterización in silico de la proteína TcCLB.510877.40 | 11 |
| 4.4. Análisis comparativo entre proteínas ALBA de <i>T. cruzi</i> y otros tripanosomátidos..... | 12 |
| 5. Resultados | 13 |
| 5.1. Secuencias identificadas en <i>T. cruzi</i> y otros tripanosomátidos..... | 13 |
| 5.2. Características génicas de las proteínas ALBA identificadas..... | 14 |
| 5.3. Características de las proteínas ALBA en <i>T. cruzi</i> | 22 |
| 5.3.1. Propiedades fisicoquímicas..... | 22 |
| 5.3.2. Dominios y/o motivos identificados..... | 23 |
| 5.4. Obtención del modelo tridimensional de la proteína TcY (TcALBA3)..... | 25 |
| 5.5. Análisis comparativo de las proteínas ALBA de <i>T. cruzi</i> y otros tripanosomátidos..... | 27 |
| 6. Discusión | 30 |

| | |
|-----------------------|----|
| 7. Conclusiones | 33 |
| 8. Referencias | 34 |

| Índice de figuras | Pag. |
|---|------|
| Fig. 1. Ciclo biológico de <i>Trypanosoma cruzi</i> | 4 |
| Fig. 2. Resumen de la metodología a desarrollar | 12 |
| Fig. 3. Organización de los genes codificantes de proteínas ALBA de <i>T. brucei</i> , <i>T. cruzi</i> y <i>L. infantum</i> descrita en la base de datos TritrypDB | 19 |
| Fig. 4. Electroforesis de PCR miniexón confirmatoria de la DTU de las cepas Y y 058PUJ de <i>T. cruzi</i> | 20 |
| Fig. 5. Información utilizada como base para la amplificación del gen codificante de la proteína TcCLB.510877.40 | 21 |
| Fig. 6. Electroforesis del producto de amplificación del gen codificante de TcCLB.510877.40 de <i>T. cruzi</i> | 21 |
| Fig. 7. Características de las secuencias del gen codificante de TcCLB.510877.40 de la cepa Y y 058PUJ de <i>T. cruzi</i> | 22 |
| Fig. 8. Organización estructural de los dominios y motivos identificados en las proteínas ALBA de <i>T. cruzi</i> | 24 |
| Fig. 9. Predicción de estructura 3D de la proteína TcY de <i>T. cruzi</i> | 26 |
| Fig. 10. Alineamiento de las estructuras del dominio ALBA entre el modelo de la proteína TcY y el de la proteína base para su construcción (Rpp25) | 26 |
| Fig. 11. Árbol filogenético consenso de las proteínas ALBA identificadas, elaborado por el método de Máxima verosimilitud en MEGA 7,0 con Bootstrap de 1000 réplicas | 28 |
| Fig. 12. Árbol filogenético consenso de las proteínas ALBA 3 identificadas, elaborado por el método de Máxima verosimilitud en MEGA 7,0 con Bootstrap de 1000 réplicas | 29 |

Índice de tablas

Pag.

| | |
|---|----|
| Tabla 1. Ejemplos de dominios canónicos de unión a ARN con su respectivo mecanismo ... | 6 |
| Tabla 2. Secuencias proteicas con dominio ALBA identificadas con Blastp realizado con la secuencia de las proteínas TbALBA en las bases de datos TritypDB y NCBI | 13 |
| Tabla 3A. Características de los genes codificantes de proteínas ALBA identificados en algunos tripanosomátidos | 14 |
| Tabla 3B. Características de los genes codificantes de proteínas ALBA identificados en <i>Leishmania</i> spp. | 15 |
| Tabla 3C. Características de los genes codificantes de proteínas ALBA identificados en <i>Trypanosoma</i> spp. | 16 |
| Tabla 4A. Características de proteínas ALBA identificadas en <i>Leishmania</i> spp. y otros tripanosomátidos | 17 |
| Tabla 4B. Características de proteínas ALBA identificadas en <i>Trypanosoma</i> spp. | 18 |
| Tabla 5. Variaciones en la secuencia de proteínas ALBA de la misma clase entre cepas de una misma especie | 19 |
| Tabla 6. Características generales y propiedades fisicoquímicas de las proteínas ALBA en <i>T. cruzi</i> | 23 |
| Tabla 7. Comparación entre las clases definidas de proteínas ALBA | 29 |

1. Introducción

La enfermedad de Chagas es una infección que afecta de 6 a 7 millones de personas alrededor del mundo, la cual es producida por el parásito protozoario *Trypanosoma cruzi* [1]. Para esta enfermedad en la actualidad no se han descrito vacunas y los medicamentos usados para su tratamiento: Benznidazol (2-nitroimidazole) o Nifurtimox (5-nitrofuran) son altamente tóxicos, producen diversos efectos colaterales en el paciente y su administración requiere de periodos prolongados, lo que afecta negativamente la adherencia al tratamiento por parte del paciente [2]. Adicionalmente, el tratamiento presenta poca eficacia en la cura parasitológica durante la fase crónica de la enfermedad [3-5], la cual es la presentación clínica más frecuentemente diagnosticada. En consecuencia, la búsqueda de nuevas alternativas para el tratamiento de la infección con el parásito es prioritaria. En este sentido, la búsqueda e identificación de moléculas esenciales para la viabilidad de *T. cruzi*, que a futuro puedan ser usadas como blancos terapéuticos, constituyen una estrategia prometedora.

Los organismos pertenecientes a la familia Trypanosomatidae, dentro de los cuales se encuentra *T. cruzi*, se caracterizan por regular la expresión génica, principalmente, a nivel post-transcripcional. En este sentido, las proteínas de unión a ARN (PUA) constituyen un grupo interesante de estudio. Dentro de este grupo, las proteínas pertenecientes a la familia ALBA, las cuales presentan un dominio de unión a ácidos nucleicos llamado ALBA (Pfam01918), han sido descritas en organismos filogenéticamente cercanos a *T. cruzi*, como *Trypanosoma brucei*, en donde se ha sugerido están involucradas en la regulación de la expresión de genes que codifican proteínas necesarias para el adecuado desarrollo del parásito, y participan en procesos como la formación del núcleo o la síntesis del citoesqueleto, puesto que el silenciamiento de la expresión de los genes codificantes para las proteínas de esta familia, generan formas aberrantes del parásito como células sin núcleo o células con morfologías anormales [6].

Teniendo en cuenta lo anterior, en este proyecto se caracterizaron las proteínas con dominio específico ALBA (Pfam01918) en *T. cruzi*, con el propósito de contribuir a dilucidar la función que éstas desempeñan en el parásito y que, a futuro, puedan utilizarse como blanco terapéutico para el desarrollo de nuevos tratamientos contra la enfermedad de Chagas, que sean más efectivos, selectivos y tengan menor toxicidad para el paciente. Para lo cual, se identificaron proteínas ALBA en *T. cruzi* y otros tripanosomátidos mediante análisis Blast de proteínas utilizando como referencia la secuencia de las proteínas de *T. brucei*, así como con la amplificación

y secuenciación de un gen codificante de una proteína ALBA en *T. cruzi*. Adicionalmente, mediante el uso de herramientas bioinformáticas tales como Protparam e Interproscan, se determinaron propiedades fisicoquímicas y estructurales de las proteínas del parásito de estudio, con un especial énfasis en la secuencia proteica obtenida *in vitro*, para la cual se propuso un modelo estereoquímicamente válido, obtenido mediante Phyre2 y analizado por PROCHECK, compatible con estructuras tridimensionales, observadas en Pymol versión 2,1. También se realizó un análisis comparativo entre todas las secuencias obtenidas mediante la realización de árboles filogenéticos o de alineamientos múltiples por medio de las herramientas Mega 7,0 y Clustal Omega, respectivamente, en donde se evidenció la presencia de dos grandes grupos de proteínas ALBA, dentro de los que se formaron clados independientes entre proteínas de *Leishmania* spp. y *Trypanosoma* spp. y sub grupos de proteínas del mismo género, así como se observaron diferencias en el grado de conservación a nivel de número y secuencia entre las proteínas ALBA de los géneros mencionados.

2. Marco Teórico

2.1. Enfermedad de Chagas

La enfermedad de Chagas es una enfermedad tropical causada por el parásito protozoario *Trypanosoma cruzi*. Dicha patología cursa con dos fases, una aguda y otra crónica. La fase aguda se caracteriza por una elevada parasitemia durante las primeras 2-4 semanas postinfección [7,8], en la mayoría de los casos sin la presencia de síntomas, salvo en algunos pocos en los que se produce fiebre, hepatoesplenomegalia, linfocitosis atípica, linfadenopatía o, en el caso de presentar una fase aguda severa (1-5% pacientes), miocarditis aguda, meningoencefalitis y efusión del pericardio [8,9]. Usualmente la respuesta inmune mediada por células es capaz de controlar al parásito, resolviendo los síntomas espontáneamente. Sin embargo, el parásito persiste en el paciente, marcando el comienzo de la fase crónica asintomática la cual está caracterizada por la seropositividad para el agente protozoario con ausencia de signos clínicos [10]. Aunque en la mayoría de los casos los pacientes permanecen en la fase crónica asintomática toda su vida, 10 a 30 años después del desarrollo de la fase aguda, un porcentaje de la población desarrolla la fase crónica sintomática, en la que se producen complicaciones cardíacas (14-45%) e intestinales (10-21%) severas como cardiomiopatía o megavísceras digestivas como megacolon y/o megaesófago [11,12].

Esta enfermedad es considerada como uno de los mayores problemas de salud pública en 21 países latinoamericanos donde la enfermedad es endémica, dentro de los cuales se encuentra Colombia [1,13]. En estos, se concentra gran parte de los cerca de 8 millones de afectados alrededor del mundo, de los cuales 10.000 - 12.000 mueren al año producto del daño cardíaco o gastrointestinal que sufren las personas que desarrollan la fase crónica sintomática de la enfermedad [13,14], lo cual representa un gasto global de alrededor de US\$ 7 - 19 billones y, en Colombia, de \$267 millones de pesos al año en términos del cuidado en la salud de pacientes crónicos sintomáticos [15,16].

2.2. Tratamiento de la enfermedad

Actualmente no existen vacunas para la prevención de la enfermedad de Chagas; no obstante, han sido propuestos dos medicamentos para su tratamiento: el Benznidazol (2-nitroimidazole) y el Nifurtimox (5-nitrofurano). Sin embargo, el tratamiento de la enfermedad tiene poca adherencia, debido a las elevadas concentraciones del medicamento (5-10 o 5-20mg/Kg x día de Benznidazol o Nifurtimox, respectivamente) que deben ser administradas durante prolongados periodos de tiempo (60 – 90 días para ambos) y a los efectos colaterales producto de su consumo (nauseas, pérdida de peso, vómito y diarrea, hasta complicaciones serias como fiebre y septicemia, agranulocitosis y púrpura trombocitopénica) [2,17,18]. Además, el tratamiento durante la fase crónica es poco eficaz en la cura parasitológica de la enfermedad (Fase crónica, 8-20%) [3-5,17,18], siendo esta la fase en la que frecuentemente se diagnostican los individuos. Según Urbina y Docampo [19], esto puede deberse a que dichos compuestos químicos presentan una corta vida media y una limitada capacidad de penetración en los tejidos, a los cuales migra el parásito en esta fase de la enfermedad, lo cual disminuye su actividad frente al microorganismo.

2.3. *Trypanosoma cruzi*

T. cruzi es un parásito protozoario flagelar perteneciente al orden kinetoplastidae y a la familia Trypanosomatidae. Este protozoo presenta una elevada diversidad genética y fenotípica intraespecífica, siendo clasificado en 7 grupos diferentes, denominados Unidades Discretas de Tipificación (DTU, del inglés: *Discrete Typification Unit*), nombrados como TcI al TcVI y Tcbat [20]. Estos grupos presentan diferencias asociadas con la patogénesis, distribución geográfica, respuesta al tratamiento y hospederos principales. Adicionalmente, las características clínicas de la enfermedad, como la generación de cardiomiopatías o megasíndromes viscerales, han sido

asociadas con DTU diferentes. [20-22]. En Colombia, se ha reportado la presencia de todas las DTUs [21,23], siendo TcI y TcII los grupos con mayor prevalencia en el país. [20-25].

Este organismo presenta un ciclo de vida digenético (**Figura 1**), alternando entre un hospedero intermediario invertebrado (artrópodos de la familia Triatominae) y un hospedero definitivo vertebrado (mamíferos de sangre caliente).

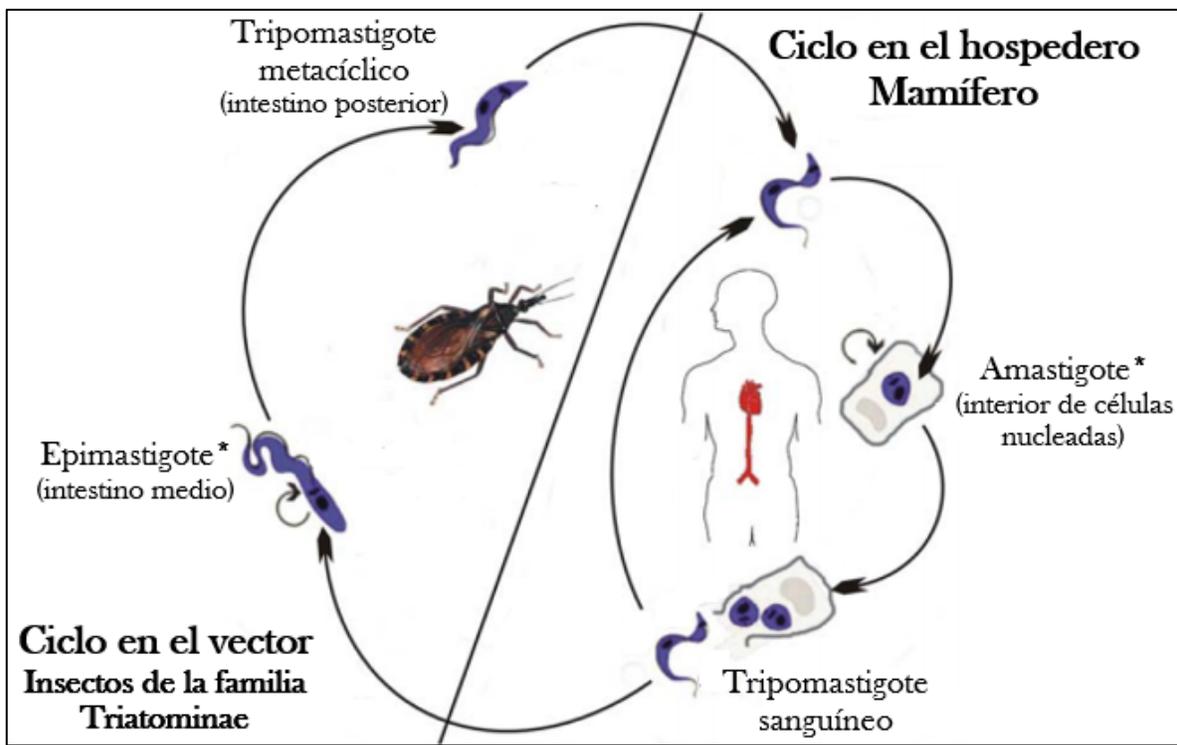


Fig. 1. Ciclo biológico de *Trypanosoma cruzi*. *: estadios replicativos del parásito. Modificado de [27].

Su ciclo comienza cuando el triatominino se alimenta y consume la sangre de un mamífero infectado, ingiriendo formas tripomastigotas sanguíneas del parásito. En el insecto, los tripomastigotes migran al intestino medio en donde se transforman al estadio epimastigote (estadio replicativo en el hospedero invertebrado). Luego de replicarse, los epimastigotes migran a la parte posterior del intestino del vector, en donde se diferencian en tripomastigotes metacíclicos, los cuales son liberados en las heces del insecto cuando éste se alimenta de nuevo, ingresando al mamífero a través de un tejido mucoso o una herida en la piel, en donde el parásito puede invadir cualquier célula nucleada del hospedero. Cuando consigue internalizarse en una

célula, se diferencia en un cuarto estadio llamado amastigote, el cual corresponde a la forma replicativa en el hospedero mamífero, duplicándose cada 12 horas durante un periodo de 4 a 5 días, luego del cual se transforma nuevamente a tripomastigote sanguíneo, siendo liberado producto de la lisis de la célula contenedora, lo que le permite continuar infectando células y aumentando la población de parásitos producto de nuevos ciclos de invasión y replicación, o ser nuevamente consumidos por el vector, iniciando otro ciclo biológico [11,12]. Este complejo ciclo de vida, en el cual se ve sometido a diferentes condiciones de estrés térmico, nutricional y ambiental, le sugiere cambios importantes en su morfología, metabolismo y expresión génica [26].

En *T. cruzi* y en general en los organismos tripanosomátidos, la regulación de la expresión génica está dada principalmente a nivel post-transcripcional, debido a la ausencia de promotores canónicos para la ARN polimerasa II [28,29]. Esta regulación, es mediada por distintos elementos, entre los que se incluyen las proteínas de unión a ARN (PUA) [29,30].

2.4. Proteínas de Unión a ARN (PUA)

Las proteínas de unión a ARN son proteínas que presentan un dominio que les confiere la capacidad de unirse a moléculas de ARN (RRM, KH, dedo de zinc tipo CCCH, entre otros) (**Tabla 1**)[28]. Estas proteínas se encargan de transportar y regular la traducción de los ARN mensajeros (ARNm), mediante su unión, principalmente, a regiones 3' o 5' no traducidas (UTR, del inglés: *Untranslated Region*). De esta manera, la regulación está dada por la capacidad de las PUA de estabilizar o desestabilizar los transcritos ya sintetizados, de manera tal que se favorezca o se disminuya la síntesis de las proteínas correspondientes [29-32]. Dentro de este grupo, se encuentran proteínas pertenecientes a la familia ALBA, las cuales se caracterizan por presentar un dominio específico de unión a ácidos nucleicos (Pfam01918) llamado ALBA [33,34].

Tabla 1. Ejemplos de dominios canónicos de unión a ARN con su respectivo mecanismo.

| Dominios de unión a ARN | Mecanismo de unión ARN-Proteína | Referencias |
|--------------------------------|---|--------------------|
| RRM | Interacción específica con 4 nucleótidos de ARN _{mc} a partir de interacciones electrostáticas entre un residuo Arg/Lys y grupos fosfato, así como interacciones hidrofóbicas entre dos residuos aromáticos con algunas nucleobases. | [35] |
| Dedo de zinc CCHH | Interacciones electrostáticas entre aminoácidos de la sección del dominio que tiene contacto con bucles formados por la molécula de ARN | [36] |
| Dedo de zinc CCCH | Las interacciones de apilamiento entre aminoácidos aromáticos y las nucleobases de regiones ricas en AU, crean una torsión en el ARN que permite el reconocimiento directo de los bordes de Watson-Crick de las bases por parte de la estructura de la proteína. | [37] |
| KH | Reconocimiento de 4 nucleótidos a partir de interacciones hidrofóbicas con bucles de aminoácidos no aromáticos. Adicionalmente se suman otro tipo de interacciones como puentes de hidrógeno, electrostáticas o por complementariedad de forma. | [38] |
| dsRBD | La proteína se une a través de dos surcos menores sucesivos y el surco mayor intermedio en una cara de la hélice ARN _{bc} . Interacción de α -hélice N-terminal con estructuras helicoidales irregulares tales como horquillas o protuberancias por emparejamiento erróneo de bases. | [39] |
| Pumilio | Las interacciones de apilamiento median la unión inespecífica con ARN. La unión específica con esta molécula esta dada por la formación de puentes de hidrógeno entre los bordes de Watson-Crick de las bases con dos aminoácidos ubicados en una hélice alfa. | [40] |
| S1 | La unión a ARN se da a través de interacciones de apilamiento y formación de puentes de hidrógeno entre bases y residuos de la proteína (aromáticos en el caso de las interacciones de apilamiento) | [28] |
| ALBA | En la interacción con ARN _{bc} es importante la formación de un tetrámero en el cual cada dímero de proteína ALBA interactúa con una cadena de ARN. La interacción polar entre aminoácidos específicos y el esqueleto fosfatado del ácido nucleico son necesarios para la formación de un complejo estable ARN-proteína. | [33] |

ARN_{bc}, ARN bicatenario; **CCHH**: sección quelante de iones de Zn compuesta por 2 cisteínas y 2 histidinas; **CCCH**: sección quelante de iones de Zn compuesta por 3 cisteínas y 1 histidina; **dsRBD**, del inglés *double-stranded RNA Binding Domain*; **KH**, del inglés *K-homology*; **RRM**: del inglés *RNA-recognition motif*.

2.5. Proteínas de la familia ALBA

La familia ALBA (del inglés: *Acetylation Lowers Binding Affinity*) es un conjunto de proteínas diméricas de bajo peso molecular que presentan un dominio genérico ALBA (Pfam01918) de unión a ácidos nucleicos [33,34]. Estas proteínas se han reportado en distintas arqueas como *Sulfolobus solfataricus*, *Aeropyrum pernix* y *Methanococcus jannaschii*, agrupadas en un único linaje específico, en donde son altamente conservadas, se unen en su mayoría a moléculas de ADN y están involucradas en procesos de organización y/o empaquetamiento del genoma dependientes de temperatura [33,41-44]. Estas proteínas también han sido descritas en organismos eucariotas como hongos, plantas, nemátodos y protozoos. Sin embargo, en este dominio las proteínas de la familia ALBA se agrupan en dos linajes específicos, Rpp20/Pop7o Rpp25/Mdp2 [33,44], y se caracterizan por ser poco conservadas a nivel de identidad entre secuencias, así como en la función que desempeñan, ya que se ha reportado su participación en distintos procesos desde la adaptación y tolerancia al estrés hídrico en plantas como *Oryza sativa* [45], hasta su intervención en diferenciación y desarrollo celular, regulación postraduccional y/o metabolismo del ARN (regulación transcripcional) de algunos parásitos protozoarios como *Plasmodium falciparum* [46], *Toxoplasma gondii* [47], *Leishmania infantum* [48,49] y *Trypanosoma brucei* [6,50]

2.5.1. Proteínas ALBA en tripanosomátidos

Las proteínas ALBA han sido estudiadas en organismos filogenéticamente cercanos a *T. cruzi* como *T. brucei*, en el cual se han descrito cuatro proteínas con dominio específico ALBA denominadas TbALBA1, TbALBA2, TbALBA3 y TbALBA4, con un tamaño de 125, 117, 190 y 217 aminoácidos, respectivamente, según lo reportado en la bases de datos TritypDB [51] y Pfam 32.0 [34]. En las proteínas TbALBA 1 y 2 se ha descrito la presencia del motivo C-terminal FDXh (presente en las proteínas del linaje Rpp20) y en TbALBA 3 y 4 se ha descrito la presencia de los motivos GYQXP (presente en las proteínas del linaje Rpp25) y RGG C-terminal [50]. Para las tres primeras, se ha sugerido un rol en la regulación de la expresión génica por estabilización de transcritos, puesto que el silenciamiento de los genes codificantes para estas proteínas, mediante ARN de interferencia, conlleva a la disminución en la estabilidad de los ARNm que codifican para una proteína procíclica tipo GPEET [50], la cual es el componente mayoritario de la cubierta del parásito durante su desarrollo en el vector [52]. Adicionalmente, se sugiere que las proteínas TbALBA3 y TbALBA4 son importantes para la viabilidad del *T. brucei*, puesto que su silenciamiento conlleva a la producción de formas aberrantes del parásito,

como la generación de células con un alargamiento excesivo de la sección anterior o sin núcleo. Así mismo, el silenciamiento de estas proteínas bloquea la diferenciación del parásito dentro del vector [6].

Por otra parte, en *L. infantum* se han descrito dos proteínas con dominio específico Pfam01918 denominadas LiALBA1 y LiALBA3 las cuales tienen un tamaño de 121 y 198 aminoácidos, respectivamente, según lo reportado en la bases de datos TritypDB [51] y Pfam 32.0 [34]; además, se ha descrito la presencia del motivo C-terminal FDXh en la proteína LiALBA1, motivo presente en proteínas del linaje Rpp20 [48], y los motivos RGG C-terminal y GYQXP (motivo presente en proteínas del linaje Rpp25) en la proteína LiALBA3 [49]. Esta última proteína parece ser de gran importancia para el parásito puesto que se une a las regiones UTR del ARNm que codifica para la proteína amastina, antígeno importante en el proceso de invasión celular, y su *knockout* conduce a la pérdida de estabilidad de estos ARNm, reduciendo su tiempo de vida media de >150 min a 90 min en el estadio replicativo en el mamífero [49]. Asimismo, el *knockout* de LiALBA3, conlleva a la reducción en la cantidad de ARNm de otras proteínas importantes para su viabilidad, como la proteína HSP70, chaperona que ayuda a proteger proteínas durante condiciones de estrés térmico [53]; y virulencia, como el factor de virulencia A₂ [49,54]. Adicionalmente, LiALBA1 y LiALBA3 interactúan entre sí y con otras moléculas como factores de traducción, proteínas de unión a ARN y subunidades ribosomales [48]. Se ha observado que estas últimas se acumulan en gran cantidad con la proteína denominada LiALBA3 durante la baja traducción de algunos genes generada en condiciones de estrés térmico, lo cual sugiere que estas proteínas pueden estar implicadas en la represión de la traducción y, por ende, en la regulación de la expresión génica durante el desarrollo del protozoo [48], ya que, al igual que *T. cruzi*, este parásito se ve sometido a dicha condición térmica durante su ciclo biológico [55].

2.5.2. Proteínas ALBA en *T. cruzi*

Para *Trypanosoma cruzi* CL Brener Esmeraldo-Like se ha encontrado un reporte que describe la presencia de cuatro genes codificantes de proteínas con dominio ALBA en su genoma, denominadas TcALBA1 (TcCLB.504089.60), TcALBA2 (TcCLB.504089.70), TcALBA30 (TcCLB.510877.30) y TcALBA40 (TcCLB.510877.40). En estas dos últimas se indica la presencia de un dominio ALBA, así como de motivos de repeticiones RGG C-terminales. Sólo para la proteína TcALBA30 se ha descrito una aproximación a su función en donde se evidenció

que su sobreexpresión conlleva a una disminución de la concentración de ARN mensajero de la proteína β -amastina, la cual se sugiere es importante en la adaptación de *T. cruzi* a su vector, debido a que es sobreproducida solo en el estadio epimastigote del parásito [56].

Ensayos de *pull-down* (ARN-Proteínas totales de *T. cruzi*) realizados en el laboratorio de inmunoparasitología molecular con la región 5'UTR del ARNm de la proteína LYT1 de *T. cruzi* [57], permitieron identificar una proteína con dominio ALBA, reportada en la base de datos TritypDB como TcCLB.510877.40 (o TcALBA40 en [56]), que presentó una identidad de 67,89% con la proteína TbALBA3 de *T. brucei*. Sin embargo, aún no es claro la cantidad total de proteínas con dominio ALBA que e *T. cruzi* y cuál es la función de estas proteínas en el parásito. Teniendo en cuenta lo anterior, en el presente proyecto se planea caracterizar molecularmente los genes y las proteínas con dominio específico ALBA, haciendo énfasis en la proteína identificada por el laboratorio de inmunoparasitología molecular (TcCLB.510877.40), como una aproximación inicial al estudio de esta familia proteica en *T. cruzi*.

3. Objetivos

3.1 Objetivo general

Caracterizar los genes codificantes de proteínas con dominio específico ALBA (Pfam01918) en *Trypanosoma cruzi*.

3.2 Objetivos específicos

- a) Identificar y caracterizar secuencias génicas codificantes de las proteínas con dominio ALBA en *T. cruzi* y otros tripanosomátidos.
- b) Caracterizar *in silico* las respectivas proteínas con dominio ALBA identificadas en *T. cruzi*
- c) Analizar comparativamente las características de las proteínas con dominio específico ALBA de *T. cruzi* con otros tripanosomátidos.

4. Metodología

4.1. Identificación de los genes codificantes para proteínas con dominio ALBA en *T. cruzi* y otros tripanosomátidos.

Los genes codificantes de proteínas con dominio ALBA y sus respectivas secuencias proteicas se identificaron en *T. cruzi* y otros tripanosomátidos mediante el uso de la herramienta Blast de proteínas de la base de datos TritypDB (<http://tritypdb.org/tritypdb/showQuestion.do?questionFullName=UniversalQuestions.UnifiedBlast>) y del NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [58], teniendo como referencia las secuencias de aminoácidos de las proteínas de la familia ALBA descritas para *T. brucei* en la base de datos Pfam y TritypDB (TbALBA1, TbALBA2, TbALBA3 y TBALBA4). Se seleccionaron los genes que codifican para proteínas con dominio ALBA presentes en organismos de la familia Tripanosomatidae que obtuvieron un valor de significancia (valor E) menor a 0,01 (se tuvo en cuenta el valor definido por estudio en el que identificaron proteínas ALBA del dominio eucariota) con la herramienta Blast. De las secuencias obtenidas, se eliminaron las secuencias incompletas, repetidas y/o aquellas que no tenían un tamaño semejante con alguna de las 4 TbALBA utilizadas como referencia.

4.2. Caracterización de las proteínas ALBA identificadas en *T. cruzi*

La predicción de las características fisicoquímicas de las proteínas ALBA identificadas en *T. cruzi* se realizó mediante el uso de las herramientas disponibles en el servidor web ProtParam del servidor Expasy (<https://web.expasy.org/protparam/>) [59]. La identificación de motivos y dominios se realizó utilizando la herramienta InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan5/>) [60] o de manera manual, usando como base un alineamiento múltiple de las secuencias proteicas realizado con la herramienta Clustal Omega del EBI (del inglés: *European Bioinformatics Institute*) (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [61].

4.3. Caracterización de la proteína TcCLB.510877.40

4.3.1. Obtención de ADN del parásito

A partir de un cultivo de epimastigotes de *T. cruzi* en medio LIT a 28°C (suplementado con 10% de Suero Fetal Bovino y 1% de Penicilina-Estreptomicina) de los aislados 058-PUJ y Y [62], correspondientes a las DTU TcI y TcII, respectivamente, se realizó la extracción de ADN mediante el estuche comercial High Pure PCR Template Preparation (Roche; Mannheim,

Alemania), siguiendo las instrucciones de la casa comercial. La concentración del ADN extraído se cuantificó por espectrofotometría UV, mediante Nanodrop 2000 [63]. A partir del ADN extraído, se confirmó la DTU de cada cepa mediante el uso de PCR convencional utilizando como blanco una región del gen miniexón [64].

4.3.2. Amplificación del gen codificante de TcCLB.510877.40

Mediante PCR convencional se amplificó el gen de la proteína **TcCLB.510877.40** en los aislados 058PUJ y Y, utilizando los oligonucleótidos TcALBAFw y TcALBARv, diseñados con base en la secuencia génica de la proteína reportada en la base de datos TritypDB [51].

El producto de amplificación obtenido para cada aislado se envió a secuenciar (20µL no purificados a 400µg/mL), obteniéndose así las secuencias genicas correspondientes, para las cuales se confirmó el porcentaje de identidad entre las secuencias obtenidas, la identidad de cada una con el gen TcCLB.510877.40 y con cada uno de los genes codificantes de proteínas ALBA de *T. brucei* utilizando la herramienta Pairwise Sequence Alignment del instituto EBI (<https://www.ebi.ac.uk/Tools/psa/>) [65].

4.3.3. Caracterización *in silico* de la proteína TcCLB.510877.40

Las secuencias de nucleótidos obtenidas del gen codificante de TcCLB.510877.40 fueron traducidas mediante la herramienta Translate de la plataforma ExPaSy (<https://web.expasy.org/translate/>) [59]. Al igual que se describió anteriormente la predicción de las características fisicoquímicas y la identificación de motivos y dominios de cada secuencia proteica se realizó utilizando la plataforma ProtParam del servidor ExPASy [59] y la herramienta InterProScan [60].

El modelamiento molecular de la proteína teórica ALBA se realizó siguiendo la metodología descrita por Nocua y colaboradores [66], utilizando el modo intensivo del servidor Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) el cual utiliza múltiples plantillas de estructuras presentes en el *Protein Data Bank* en conjunto con técnicas *ab initio* para la construcción de la estructura tridimensional teórica [67]. Dicha estructura se visualizó con el programa PyMOL (Molecular Graphics System) versión 2,3, que también fue utilizado para la realización de alineamientos entre el modelo y otras estructuras proteicas. Finalmente, la

validación estereoquímica y energética del modelo se hizo mediante la herramienta Procheck (<https://servicesn.mbi.ucla.edu-/PROCHECK/>) [68].

4.4. Análisis comparativo entre proteínas ALBA de *T. cruzi* y otros tripanosomátidos.

Los análisis de identidad entre las secuencias de las proteínas ALBA de *T. cruzi*, obtenidas *in vitro* a partir de los diferentes aislados del parásito, con las secuencias de proteínas reportados en la base de datos TritypDB [51] y NCBI para otras cepas de *T. cruzi* y de otros tripanosomátidos se realizaron mediante el uso de la herramienta Clustal Omega [60].

Los análisis filogenéticos de las secuencias de aminoácidos identificadas se realizaron mediante el método de Máxima Verosimilitud, y la evaluación estadística de los nodos de los árboles obtenidos se realizó mediante Bootstrap con un mínimo de 1000 réplicas, haciendo uso del programa Mega 7.0 [69], para determinar la relación entre las proteínas ALBA de *T. cruzi* con la de otros tripanosomátidos.

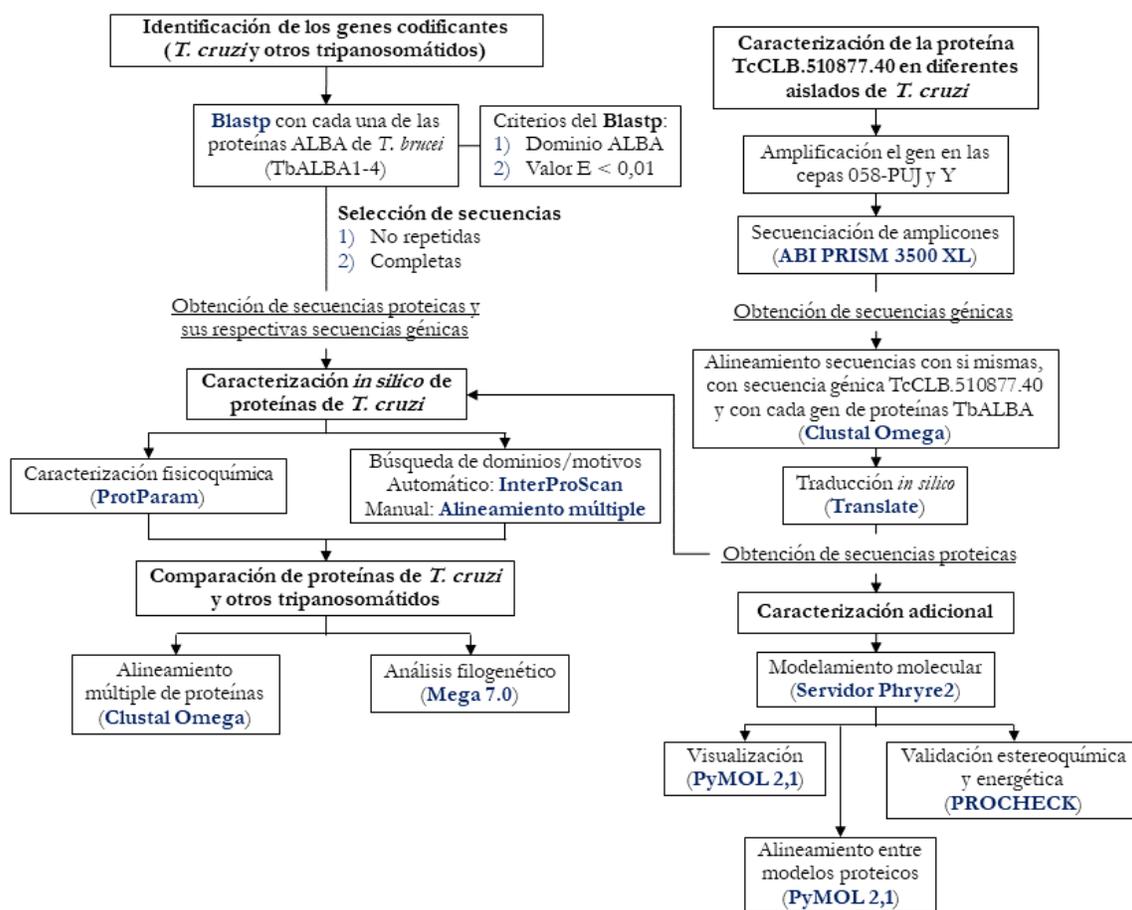


Fig. 2. Resumen de la metodología a desarrollar

5. Resultados

5.1. Secuencias identificadas en *T. cruzi* y otros tripanosomátidos

El análisis Blast utilizando como entrada las proteínas TbALBA permitió la identificación de 94 secuencias proteicas de la familia ALBA reportadas en cepas de distintas especies de tripanosomátidos (**Tabla 2**).

Tabla 2. Secuencias proteicas con dominio ALBA identificadas con Blastp realizado con la secuencia de las proteínas TbALBA en las bases de datos TritypDB y NCBI

| Género | Especie | Cepa | No. sec. | Género | Especie | Cepa | No. sec. |
|---------------------|---------------------|----------------------|-----------------|------------------------|-------------------|-----------------------------|----------|
| <i>Critidia</i> | <i>fasciculata</i> | Cf-Cl | 2 | <i>Leishmania</i> | <i>tropica</i> | L590 | 2 |
| <i>Endotrypanum</i> | <i>monterogeii</i> | LV88 | 2 | <i>Leishmania</i> | <i>turanica</i> | LEM423 | 2 |
| <i>Leishmania</i> | <i>aethiopica</i> | L147 | 2 | <i>Leptomonas</i> | <i>seymouri</i> | ATCC 30220 | 2 |
| | <i>amazonensis</i> | MHOM/BR/71/973/M2269 | 2 | <i>Paratrypanosoma</i> | <i>confusum</i> | CUL13 | 1 |
| | <i>arabica</i> | LEM1108 | 2 | <i>Phytomonas</i> | NR | EM1 | 1 |
| | <i>braziliensis</i> | MHOM/BR/75/M2903 | 2 | <i>Stringomonas</i> | <i>culicis</i> | NR | 1 |
| | | MHOM/BR/75/M2904 | 2 | <i>Trypanosoma</i> | <i>brucei</i> | brucei TREU927 | 4 |
| | <i>donovani</i> | BPK282A1 | 2 | | | gambiense DAL972 | 4 |
| | | CL-SL | 2 | | | Lister 427 | 4 |
| | <i>enriettii</i> | LEM3045 | 2 | | | Lister 427 2018 | 4 |
| | <i>gerbilli</i> | LEM452 | 2 | | <i>cruzi</i> | Dm28c 2018 | 3 |
| | <i>infantum</i> | JPCM5 | 2 | | | CL Brener Esmeraldo-like | 4 |
| | | Friedlin | 2 | | | CL Brener | 2 |
| | | LV39c5 | 2 | | | Dm28c 2017 | 4 |
| | <i>major</i> | SD 75.1 | 2 | | TCC | 4 | |
| | | MHOM/GT/2001/U1103 | 2 | | <i>congolense</i> | IL3000 | 1 |
| | <i>mexicana</i> | MHOM/COL/81/L13 | 2 | <i>evansi</i> | STIB 805 | 4 | |
| | | MHOM/PA/94/PSC-1 | 2 | <i>grayi</i> | ANR4 | 2 | |
| NR | MAR LEM2494 | 2 | <i>theileri</i> | Edinburgh | 3 | | |
| <i>tarentolae</i> | Parrot-TarII | 2 | <i>vivax</i> | Y486 | 2 | | |

N.R.: No reportado; **No. sec.:** Número de secuencias identificadas

Como se observa en la **Tabla 2**, la mayoría de las especies estudiadas presentan homogeneidad en el número de proteínas ALBA identificadas, con un número total de dos de estas proteínas; patrón evidenciado en todas las especies del género *Leishmania* presentes en el estudio. Sin

embargo, en el género *Trypanosoma* no se observó ningún patrón, ya que fueron identificadas secuencias totales de proteínas ALBA que varían de 1 a 4, especialmente en *T. cruzi*, en donde se evidencia una alta heterogeneidad en el número de proteínas identificadas en cada una de las cepas observadas.

5.2. Características génicas de las proteínas ALBA identificadas

Las secuencias de las proteínas ALBA identificadas, tanto génicas como proteicas (**Tablas 3 y 4**, respectivamente), se clasificaron como 1, 2, 3 o 4, de acuerdo con el número de la TbALBA que obtuvo mayor porcentaje de identidad con la secuencia proteica identificada (**Tabla 4**).

Los genes codificantes de proteínas ALBA1, al igual que los genes de proteínas ALBA2, presentan poca variación en el tamaño de su secuencia, exhibiendo tamaños entre 366 – 378 y 363 – 396 nucleótidos, respectivamente; diferente a lo observado en los genes ALBA3, para los que se obtuvieron tamaños entre 573 y 693 nucleótidos; en el caso de los genes de proteínas ALBA4 se evidenció un único tamaño (654 nucleótidos), debido a que solo dos secuencias identificadas fueron clasificadas en este grupo. En relación al porcentaje de Guanina-Citocina (%G-C), se observaron variaciones para los genes de proteínas ALBA1 (45,2 – 55,7 %), ALBA2 (41,0 – 59,2 %) y ALBA3 (59,2 – 71,8 %), diferente de las secuencias génicas de proteínas ALBA4 en donde se obtuvo un único valor (59,9 %) (**Tabla 3A, 3B y 3C**).

Tabla 3A. Características de los genes codificantes de proteínas ALBA identificados en algunos tripanosomátidos

| Género | Especie | Cepa | Gen | Nc | %G-C | Cro. |
|------------------------|--------------------|------------|-------------|-----|------|------|
| <i>Cribsidia</i> | <i>fasciculata</i> | Cf-CI | <i>Cf2</i> | 375 | 59,2 | 22 |
| | | | <i>Cf3</i> | 597 | 68,5 | 29 |
| <i>Endotrypanum</i> | <i>monterogeii</i> | LV88 | <i>Em1</i> | 366 | 55,7 | 13 |
| | | | <i>Em3</i> | 618 | 61,7 | 34 |
| <i>Leptomonas</i> | <i>seymouri</i> | ATCC 30220 | <i>Lps2</i> | 363 | 52,1 | NR |
| | | | <i>Lps3</i> | 615 | 64,2 | NR |
| <i>Paratrypanosoma</i> | <i>confusum</i> | CUL13 | <i>Pc2</i> | 396 | 54,5 | NR |
| <i>Phytomonas</i> | NR | EM1 | <i>Phy3</i> | 618 | 71,8 | NR |
| <i>Stringomonas</i> | <i>culicis</i> | NR | <i>Sc3</i> | 579 | 69,4 | NR |

%G-C: porcentaje Guanina-Citocina; Nc: Número de nucleótidos del gen; Cro.: Cromosoma en el que está ubicado; NR: No reportado.

En el género *Leishmania* se encontraron dos proteínas ALBA en todas las cepas analizadas, estas correspondieron a proteínas ALBA2 y ALBA3 en la mayoría de los casos (*L. major* fue la única especie con ALBA1 en lugar de ALBA2) (**Tabla 4A**). Las secuencias génicas de estas proteínas

se encontraron ubicadas en los cromosomas 13 y 34, respectivamente, en casi todas las cepas del género; ya que en algunos casos el gen de la proteína ALBA3 se evidenció en el cromosoma 20 (Tabla 3B).

Tabla 3B. Características de los genes codificantes de proteínas ALBA identificados en *Leishmania* spp.

| Género | Especie | Cepa | Gen | Nc | %G-C | Cro. | Género | Especie | Cepa | Gen | Nc | %G-C | Cro. |
|-------------------|---------------------|---------------------|-------------|------|------|------------|-------------------|-------------------|--------------------|-------------------|--------------|------------|------|
| <i>Leishmania</i> | <i>aethiopica</i> | L147 | <i>La2</i> | 366 | 53,3 | 13 | <i>Leishmania</i> | <i>major</i> | Friedlin | <i>Lm1</i> | 366 | 53,6 | 13 |
| | | | <i>La3</i> | 618 | 64,6 | 34 | | | | <i>Lm3</i> | 618 | 64,4 | 34 |
| | <i>amazonensis</i> | MHOM/BR/71973/M2269 | <i>Lam2</i> | 366 | 53,6 | NR | | | LV39c5 | <i>Lm1</i> | 366 | 54,1 | 13 |
| | | | <i>Lam3</i> | 618 | 64,7 | NR | | | | <i>Lm3</i> | 618 | 64,4 | 34 |
| | <i>arabica</i> | LEM1108 | <i>Lar2</i> | 366 | 53,0 | 13 | | | SD 75.1 | <i>Lm1</i> | 366 | 53,8 | 13 |
| | | | <i>Lar3</i> | 618 | 64,7 | 34 | | | | <i>Lm3</i> | 618 | 64,4 | 34 |
| | <i>braziliensis</i> | MHOM/BR/75/M2903 | <i>Lb2</i> | 366 | 54,4 | 13 | | <i>mexicana</i> | MHOM/GT/2001/U1103 | <i>Lmx2</i> | 366 | 53,8 | 13 |
| | | | <i>Lb3</i> | 597 | 62,8 | 20 | | | | <i>Lmx3</i> | 618 | 64,9 | 34 |
| | | MHOM/BR/75/M2904 | <i>Lb2</i> | 366 | 54,4 | 13 | | <i>panamensis</i> | MHOM/CO L/81/L13 | <i>Lp2</i> | 366 | 53,8 | 13 |
| | | | <i>Lb3</i> | 597 | 62,8 | 20 | | | | <i>Lp3</i> | 597 | 62,5 | 20 |
| | <i>donovani</i> | BPK282A1 | <i>Ld2</i> | 366 | 53,3 | 13 | | NR | MAR LEM2494 | <i>Lsp2</i> | 366 | 54,4 | 13 |
| | | | <i>Ld3</i> | 618 | 63,9 | 34 | | | | <i>Lsp3</i> | 618 | 66,3 | 34 |
| | | CL-SL | <i>Ld2</i> | 366 | 53,3 | 13 | | | | <i>tarentolae</i> | Parrot-TarII | <i>Lt2</i> | 366 |
| | <i>Ld3</i> | | 618 | 63,9 | 34 | <i>Lt3</i> | | 624 | 62,8 | | | 34 | |
| | <i>enriettii</i> | LEM3045 | <i>Le2</i> | 366 | 54,1 | 13 | | <i>tropica</i> | L590 | <i>Ltr2</i> | 366 | 53,6 | 13 |
| | | | <i>Le3</i> | 633 | 65,6 | 34 | | | | <i>Ltr3</i> | 618 | 64,9 | 34 |
| | <i>gerbilli</i> | LEM452 | <i>Lg2</i> | 366 | 52,7 | 13 | | <i>turanica</i> | LEM423 | <i>Ltu2</i> | 366 | 54,1 | 13 |
| | | | <i>Lg3</i> | 618 | 64,9 | 34 | | | | <i>Ltu3</i> | 618 | 64,9 | 34 |
| <i>infantum</i> | JPCM5 | <i>Li2</i> | 366 | 53,0 | 13 | | | | | | | | |
| | | <i>Li3</i> | 597 | 63,7 | 34 | | | | | | | | |

%G-C: porcentaje Guanina-Citocina; **Nc:** Número de nucleótidos del gen; **Cro.:** Cromosoma en el que está ubicado; **NR:** No reportado.

Diferente a lo observado en el género *Leishmania*, en las especies de *Trypanosoma* encontradas en el estudio se hallaron números variables de proteínas ALBA; desde la presencia de las cuatro proteínas (ALBA1 – 4) en las cepas de *T. brucei* (Tb) y *T. evansi* (Te) hasta casos en los que se identificó una única proteína ALBA como ocurrió con *T. congolense* (Tco) (Tabla 4B). Adicionalmente se observó que la mayoría de las especies comparten el cromosoma en donde se ubican los genes de proteínas ALBA; los genes de las proteínas ALBA1 y 2 se encuentran juntos en el cromosoma 11, al igual que los genes de ALBA3 y 4 presentes en el cromosoma 4 (Tabla 3C).

Tabla 3C. Características de los genes codificantes de proteínas ALBA identificados en *Trypanosoma* spp.

| Género | Especie | Cepa | Gen | Nc | %G-C | Cro. | Género | Especie | Cepa | Gen | Nc | %G-C | Cro. | |
|--------------------|--------------------|---------------------|------------|------------|------|------------|--------------------|--------------|-------------|---------------|-------------|----------|------------|-----|
| <i>Trypanosoma</i> | <i>brucei</i> | TREU927 | <i>Tb1</i> | 378 | 49,5 | 11 | <i>Trypanosoma</i> | <i>cruzi</i> | CL Brener | <i>Tc1</i> | 369 | 47,2 | 14 | |
| | | | <i>Tb2</i> | 354 | 45,5 | 11 | | | | <i>Tc2</i> | 354 | 43,2 | 14 | |
| | | | <i>Tb3</i> | 573 | 59,2 | 4 | | | | Dm28c 2017 | <i>Tc1</i> | 369 | 47,2 | NR |
| | | | <i>Tb4</i> | 654 | 59,9 | 4 | | | | | <i>Tc2</i> | 354 | 42,9 | NR |
| | | gambiense DAL972 | <i>Tb1</i> | 378 | 49,5 | 11 | | | <i>Tc3a</i> | | 693 | 61,9 | NR | |
| | | | <i>Tb2</i> | 354 | 45,5 | 11 | | | <i>Tc3b</i> | | 681 | 61,8 | NR | |
| | | | <i>Tb3</i> | 573 | 59,2 | 4 | | | TCC | <i>Tc1</i> | 369 | 47,2 | NR | |
| | | | <i>Tb4</i> | 654 | 59,9 | 4 | | | | <i>Tc2</i> | 354 | 43,8 | NR | |
| | | Lister 427 | <i>Tb1</i> | 378 | 49,5 | 11 | | | | <i>Tc3a</i> | 693 | 61,9 | NR | |
| | | | <i>Tb2</i> | 354 | 45,5 | 11 | | | <i>Tc3b</i> | 681 | 61,7 | NR | | |
| | | | <i>Tb3</i> | 573 | 59,2 | 4 | | | congolense | IL3000 | <i>Tco3</i> | 561 | 58,8 | 4 |
| | | | <i>Tb4</i> | 654 | 59,9 | 4 | | | | | evansi | STIB 805 | <i>Tc1</i> | 378 |
| | Lister 427 2018 | <i>Tb1</i> | 378 | 49,5 | 11 | <i>Tc2</i> | | 354 | 45,5 | 11 | | | | |
| | | <i>Tb2</i> | 354 | 45,5 | 11 | <i>Tc3</i> | | 573 | 59,2 | 4 | | | | |
| | | <i>Tb3</i> | 573 | 59,2 | 4 | <i>Tc4</i> | | 654 | 59,9 | 4 | | | | |
| | | <i>Tb4</i> | 654 | 59,9 | 4 | grayi | | ANR4 | <i>Tg1</i> | 378 | 48,7 | NR | | |
| | Dm28c 2018 | <i>Tc1</i> | 369 | 47,2 | NR | | | | <i>Tg2</i> | 354 | 42,9 | NR | | |
| | | <i>Tc2</i> | 354 | 42,9 | NR | theileri | | Edinburgh | <i>Tt1</i> | 378 | 45,5 | NR | | |
| | | <i>Tc3b</i> | 681 | 61,7 | NR | | | | <i>Tt2</i> | 357 | 42,6 | NR | | |
| | | <i>cruzi</i> | CL Brener | <i>Tc1</i> | 369 | 47,2 | | 14 | <i>Tt3</i> | 639 | 63,2 | NR | | |
| <i>Tc2</i> | 354 | | | 43,8 | 14 | vivax | Y486 | <i>Tv1</i> | 378 | 45,2 | 11 | | | |
| Esmeraldo-like | <i>Tc3a</i> | | 693 | 61,8 | 34 | | | <i>Tv2</i> | 354 | 41,0 | 11 | | | |
| | <i>Tc3b</i> | | 681 | 61,5 | 34 | | | | | | | | | |

%G-C: porcentaje Guanina-Citocina; **Nc:** Número de nucleótidos del gen; **Cro.:** Cromosoma en el que está ubicado; **NR:** No Reportado.

En el caso de *T. cruzi*, se identificaron 4 secuencias correspondientes a proteínas ALBA; ALBA1 (*Tc1*), ALBA2 (*Tc2*) y otras dos que corresponden a un mismo tipo (ALBA3), pero presentan tamaños diferentes, así como distintas posiciones en el genoma, por lo que fueron denominadas como dos proteínas diferentes: *Tc3a* y *Tc3b* (**Tabla 4B**). Los genes de las proteínas *Tc1* y *Tc2*, y los genes de *Tc3a* y *Tc3b*, se encuentran organizados en tándem en los cromosomas 14 y 34, respectivamente (**Tabla 3C; Figura 3**); dicha organización es compartida por los genes de las proteínas *Tb1* y *Tb2* (ALBA1 y 2 de *T. brucei*), al igual que por los genes de *Tb3* y *Tb4* (ALBA3 y 4 de *T. brucei*). Asimismo se evidenció que los genes de las proteínas ALBA1 y 2 en *T. cruzi* y *T. brucei* comparten varias secuencias génicas cercanas entre si y, en un menor número, con el gen de la proteína ALBA2 de *L. infantum*. En el caso de los genes de las proteínas *Tc3a* y *Tc3b* estos se encuentran separados por una región de 3000 nucleótidos aproximadamente; sin embargo, comparten los mismos genes adyacentes con las secuencias génicas de *Tb3* y *Tb4*, así como del gen de la proteína Li3 (ALBA3 de *L. infantum*). Lo anterior sugeriría que los genes codificantes de proteínas ALBA en *T. cruzi*, *T. brucei* y *L. infantum* son sinténicos. (**Figura 3**).

Tabla 4A. Características de proteínas ALBA identificadas en *Leishmania* spp. y otros tripanosomátidos

| Género | Especie | Cepa | Prot | aa | Género | Especie | Cepa | Prot | aa |
|------------------------|---------------------|---------------------|------|-----|-------------------|-------------------|--------------------|------|-----|
| <i>Crithidia</i> | <i>fasciculata</i> | Cf-Cl | Cf2 | 124 | <i>Leishmania</i> | <i>enriettii</i> | LEM3045 | Le2 | 121 |
| | | | Cf3 | 198 | | | | Le3 | 210 |
| <i>Endotrypanum</i> | <i>monterogeii</i> | LV88 | Em1 | 121 | | <i>gerbilli</i> | LEM452 | Lg2 | 121 |
| | | | Em3 | 205 | | | | Lg3 | 205 |
| <i>Leptomonas</i> | <i>seymouri</i> | ATCC 30220 | Lps2 | 120 | | <i>infantum</i> | JPCM5 | Li3 | 198 |
| | | | Lps3 | 204 | | | | Li2 | 121 |
| <i>Paratrypanosoma</i> | <i>confusum</i> | CUL13 | Pc2 | 131 | | <i>major</i> | Friedlin | Lm1 | 121 |
| | | | | | | | | Lm3 | 205 |
| <i>Phytomonas</i> | NR | EM1 | Phy3 | 203 | | | LV39c5 | Lm1 | 121 |
| | | | | | | | | Lm3 | 205 |
| <i>Stringomonas</i> | <i>culicis</i> | NR | Sc3 | 193 | | | SD 75.1 | Lm1 | 121 |
| | | | | | | | | Lm3 | 205 |
| <i>Leishmania</i> | <i>aethiopica</i> | L147 | La2 | 121 | | <i>mexicana</i> | MHOM/GT/2001/U1103 | Lmx2 | 121 |
| | | | La3 | 205 | | | | Lmx3 | 205 |
| | <i>amazonensis</i> | MHOM/BR/71973/M2269 | Lam2 | 121 | | <i>panamensis</i> | MHOM/CO L/81/L13 | Lp2 | 121 |
| | | | Lam3 | 205 | | | | Lp3 | 198 |
| | <i>arabica</i> | LEM1108 | Lar2 | 121 | MHOM/PA/94/PSC-1 | | Lp2 | 121 | |
| | | | Lar3 | 205 | | | Lp3 | 198 | |
| | <i>braziliensis</i> | MHOM/BR/75/M2903 | Lb2 | 121 | NR | MAR LEM2494 | Lsp2 | 121 | |
| | | | Lb3 | 198 | | | Lsp3 | 205 | |
| | | MHOM/BR/75/M2904 | Lb2 | 121 | <i>tarentolae</i> | Parrot-TarII | Lt2 | 121 | |
| | | | Lb3 | 198 | | | Lt3 | 207 | |
| | <i>donovani</i> | BPK282A1 | Ld2 | 121 | <i>tropica</i> | L590 | Ltr2 | 121 | |
| | | | Ld3 | 205 | | | Ltr3 | 205 | |
| | | CL-SL | Ld2 | 121 | <i>turanica</i> | LEM423 | Ltu2 | 121 | |
| | | | Ld3 | 205 | | | Ltu3 | 205 | |

aa: Número de aminoácidos de la proteína; **N.R.:** No Reportado; **% Id:** Porcentaje de identidad con TbALBA con mayor homología; **Prot:** Proteína.

Para los análisis posteriores se planteó utilizar las secuencias proteicas ALBA de la cepa de referencia correspondiente, en el caso en que haya más de una cepa identificada por especie. Esto fue realizado, ya que se observó que las secuencias de cada tipo de ALBA entre cepas de una misma especie presentan un porcentaje de similitud del 100%, salvo para las proteínas de *T. cruzi* y la TbALBA2 de *T. brucei* en las que se evidencian pequeñas variaciones (% identidad > 97,3 %) (**Tabla 5**). En las secuencias seleccionadas, se evidenció la presencia de proteínas ALBA descritas en la literatura tales como LiALBA1 (Li2) y LiALBA3 (Li3) de *L. infantum* JPCM5 y las proteínas TcALBA1 (Tc1), TcALBA2 (Tc2), TcALBA30 (Tc3a) y TcALBA40/TcCLB.510877.40 (Tc3b) de *T. cruzi* CL Brener Esmeraldo-like (**Tabla 5**).

Si bien las proteínas ALBA identificadas en *T. cruzi* corresponden a las proteínas descritas en la literatura, se propuso realizar la caracterización de la secuencia *in vitro* del gen codificante de la proteína TcCLB.510877.40 o TcALBA40 como parte del estudio detallado de esta proteína en diferentes DTU o grupos génicos del parásito.

Tabla 4B. Características de proteínas ALBA identificadas en *Trypanosoma* spp.

| Género | Especie | Cepa | Prot | aa | Género | Especie | Cepa | Prot | aa |
|--------------------|---------------|---------------------------------|------|--------------|--------------------|-------------------|---------------|------|-----|
| <i>Trypanosoma</i> | <i>brucei</i> | brucei TREU927 | Tb1 | 125 | <i>Trypanosoma</i> | <i>cruzi</i> | CL Brener | Tc1 | 122 |
| | | | Tb2 | 117 | | | | Tc2 | 117 |
| | | | Tb3 | 190 | | | Dm28c 2017 | Tc1 | 122 |
| | | | Tb4 | 217 | | | | Tc2 | 117 |
| | | gambiense DAL972 | Tb1 | 125 | | | | Tc3a | 230 |
| | | | Tb2 | 117 | | | Tc3b | 226 | |
| | | | Tb3 | 190 | | | TCC | Tc1 | 122 |
| | | Tb4 | 217 | Tc2 | | | | 117 | |
| | | Lister 427 | Tb1 | 125 | | | | Tc3a | 230 |
| | | | Tb2 | 117 | | | Tc3b | 226 | |
| | | | Tb3 | 190 | | <i>congolense</i> | IL3000 | Tco3 | 186 |
| | | | Tb4 | 217 | | | | | |
| | | Lister 427 2018 | Tb1 | 125 | | <i>evansi</i> | STIB 805 | Te1 | 125 |
| | | | Tb2 | 117 | | | | Te2 | 117 |
| | | | Tb3 | 190 | | | | Te3 | 190 |
| | | | Tb4 | 217 | | | | Te4 | 217 |
| | <i>cruzi</i> | Dm28c 2018 | Tc1 | 122 | <i>grayi</i> | ANR4 | Tg1 | 125 | |
| | | | Tc2 | 117 | | | Tg2 | 117 | |
| | | | Tc3 | 226 | <i>theileri</i> | Edinburgh | Tt1 | 125 | |
| | | CL Brener Esmeraldo- like | Tc1 | 122 | | | Tt2 | 118 | |
| Tc2 | | | 117 | Tt3 | | | 212 | | |
| Tc3a | | | 230 | <i>vivax</i> | Y486 | Tv1 | 125 | | |
| Tc3b | | | 226 | | | Tv2 | 117 | | |

aa: Número de aminoácidos de la proteína; **NR:** No Reportado; **% Id:** Porcentaje de identidad con TbALBA con mayor homología; **Prot:** Proteína

La confirmación de la DTU de las cepas de *T. cruzi* en las cuales se planeó realizar la amplificación del gen codificante de TcCLB.510877.40, mediante el uso de la PCR miniexón, permitió verificar que la cepa 058PUJ y la cepa Y corresponden a la DTU I y II, respectivamente, ya que en ambos casos se observó que el tamaño del amplicón obtenido, corresponde al tamaño del producto para su respectivo grupo génico (DTU I: 350 pb; DTU II: 300 pb) (**Figura 4**).

Tabla 5. Variaciones en la secuencia de proteínas ALBA de la misma clase entre cepas de una misma especie.

| Género | Especie | Prot. | aa | % Id EP | Género | Especie | Prot. | aa | % Id EP |
|-------------------|---------------------|-------|-----|---------|--------------------|---------------|-------|------------|-------------|
| <i>Leishmania</i> | <i>braziliensis</i> | Lb2 | 121 | 100 | <i>Trypanosoma</i> | <i>brucei</i> | Tb1 | 125 | 100 |
| | | Lb3 | 198 | 100 | | | Tb2 | 117 | 97,4 - 98,3 |
| | <i>donovani</i> | Ld2 | 121 | 100 | | | Tb3 | 190 | 100 |
| | | Ld3 | 205 | 100 | | | Tb4 | 217 | 100 |
| | <i>major</i> | Lm1 | 121 | 100 | | <i>cruzi</i> | Tc1 | 122 | 98,4 - 100 |
| | | Lm3 | 205 | 100 | | | Tc2 | 117 | 99,4 - 100 |
| | <i>panamensis</i> | Lp2 | 121 | 100 | | | Tc3a | 230 | 97,4 - 98,7 |
| | | Lp3 | 198 | 100 | | Tc3b | 226 | 97,3 - 100 | |

aa: Número de aminoácidos de la proteína;; % Id EP: Porcentaje de identidad entre proteínas ALBA de cepas de una misma especie; Prot.: Proteína.

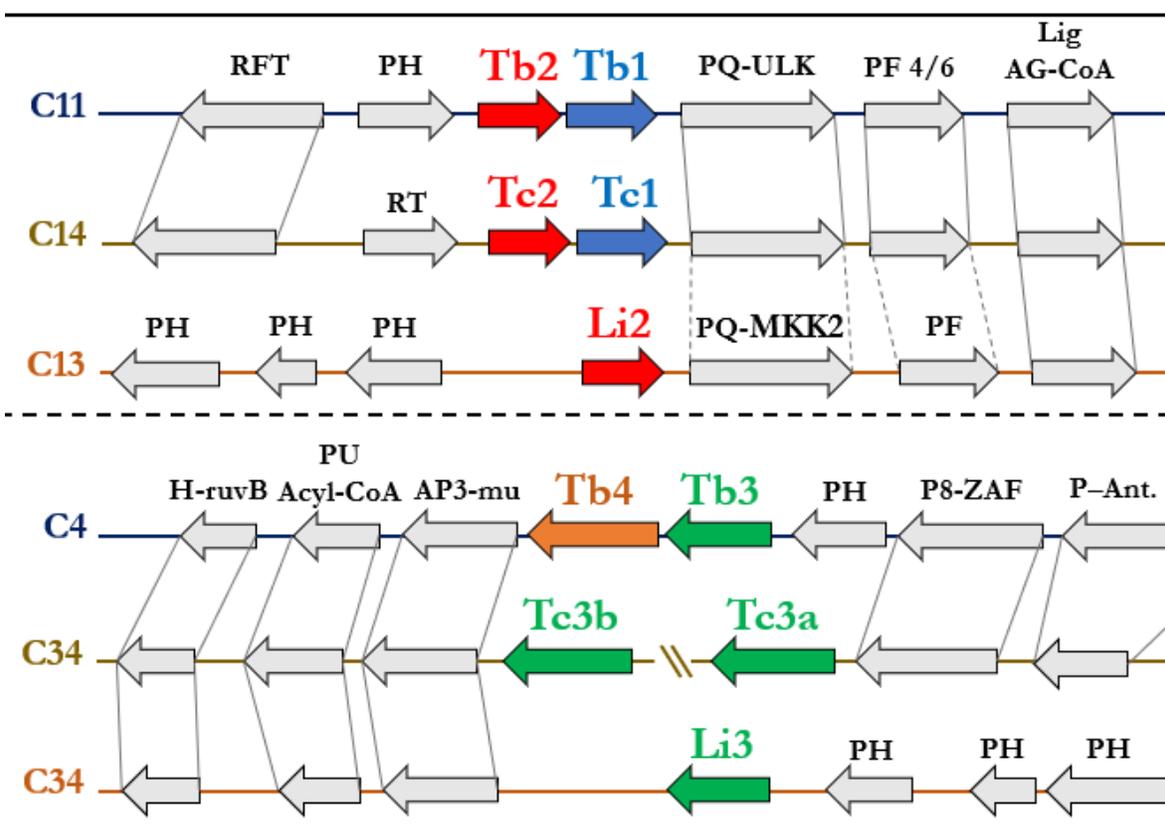


Fig. 3. Organización de los genes codificantes de proteínas ALBA de *T. brucei*, *T. cruzi* y *L. infantum*, descrita en la base de datos TritypDB. C 4, 11, 13, 14 o 34: Cromosoma 4, 11, 14 o 34; **Lineas continuas:** Genes codificantes de proteínas iguales; **Lineas discontinuas:** Genes codificantes de proteínas con funciones semejantes; **RFT:** Reductasa Férrica Transmembranal; **PH:** Proteína Hipotética; **PQ-ULK o MKK2:** Quinasa de la familia ULK o activada por mitógenos; **PF:** Proteína flagelar; **Lig Ac-CoA:** Ligasa de CoA y ácidos grasos de cadena larga; **H-ruvB:** Helicasa de ADN similar a ruvB; **PU Acyl-CoA:** Proteína de Unión a Acyl-CoA; **AP3-mu:** Subunidad mu del complejo proteico AP3 ; **P8-ZAF:** Proteína 8 – Zona de anclaje de flagelo ; **P-Ant.:** Proteína Antigénica; **RT:** Retrotransposón.

Luego de confirmar el grupo génico de las cepas del parásito se prosiguió a diseñar primers para la amplificación del gen codificante para la proteína TcCLB.510877.40 a partir de la secuencia de este gen reportada en la plataforma TritypDB la cual presenta un tamaño de 681 pb (**Figura 5**). Como producto de la amplificación del gen de interés en cada cepa evaluada, se obtuvo un amplicon con un tamaño aproximado a los 681 pb esperados, evidenciado en la corrida electroforética de cada producto, sin la presencia de amplificaciones inespecíficas (**Figura 6**).

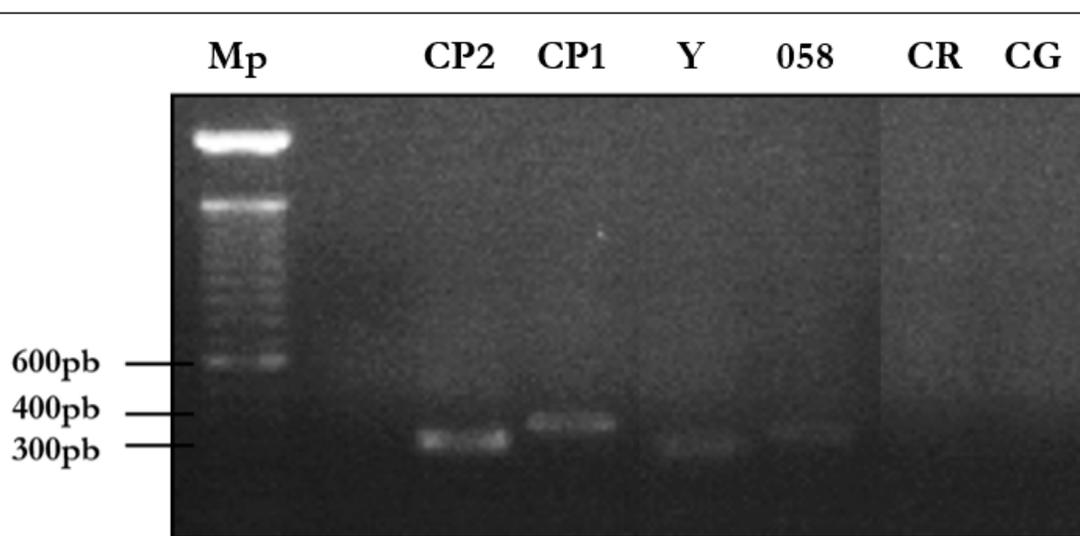


Fig. 4. Electroforesis de PCR miniexón confirmatoria de la DTU de las cepas Y (**Y**) y 058PUJ (**058**) de *T. cruzi*. **CP1**, control positivo para DTU I (350pb), **CP2**, control positivo para DTU II (300pb) **CR**, control de reacción; **CG**, Control gris.

Las secuencias consenso obtenidas de la secuenciación del producto de amplificación del gen TcCLB.510877.40 de las cepas Y (TcY) y 058PUJ (Tc058) (**Figura 7A**), presentaron un mismo tamaño molecular de 656 nucleótidos, y como se esperaba, se obtuvo un alto porcentaje de identidad del 98,8% entre la secuencia del gen de la proteína TcCLB.510877.40, reportada en la plataforma TritypDB, y la secuencia obtenida de la cepa Y y del 98,6% entre la secuencia del gen de la proteína TcCLB.510877.40 y la secuencia de la cepa 058PUJ. Adicionalmente se realizó un alineamiento entre las secuencias obtenidas de los aislados evaluados obteniendo un porcentaje de identidad del 98,6% (**Figura 7B**). Estas secuencias también fueron comparadas con los genes de las cuatro proteínas TbALBA, lo que reveló que ambas secuencias consenso presentaron mayor porcentaje de identidad con el gen de la proteína TbALBA3 (secuencia de Y: 71,1%; secuencia de 058PUJ: 71,6%), indicando que corresponden a proteínas ALBA3 de *T. cruzi*. Por otra parte, se realizó la traducción de las secuencias consenso obtenidas para la cepa Y

(TcY) y 058PUJ (Tc058), y se determinó el porcentaje de identidad entre ambas el cual fue de 97,2 % (Figura 7B).

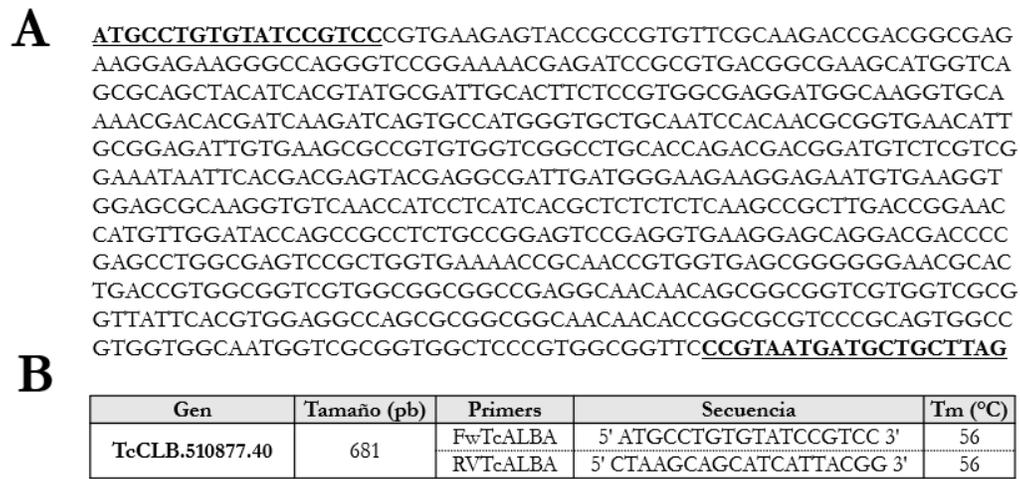


Fig. 5. Información utilizada como base para la amplificación del gen codificante de proteína TcCLB.510877.40 **A.** Secuencia de gen de referencia (TcCLB.510877.40), **regiones marcadas:** regiones base para el diseño de los primers. **B.** Gen de referencia a amplificar y set de primers elaborados; **Tm:** Temperatura melting; **FwTcALBA:** Cebador delantero del gen; **RvTcALBA:** Cebador reverso del gen.

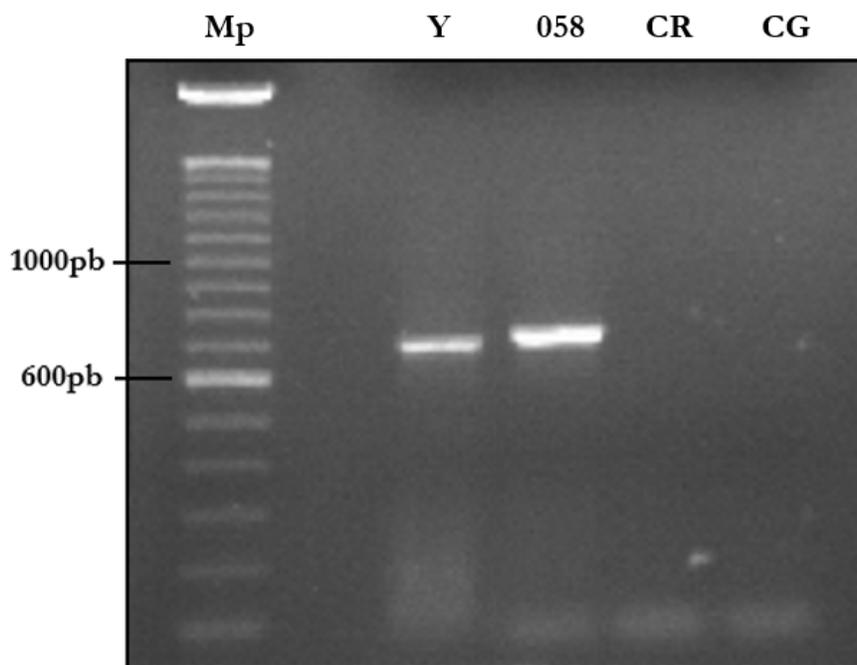


Fig. 6. Electroforesis del producto de amplificación del gen codificante de TcCLB.510877.40 en *T. cruzi*. Resultados para las cepas Y (**Y**) y 058PUJ (**058**) de *T. cruzi*. **CR**, control de reacción; **CG**, Control gris

A

```

-----
Tc058 1 CGTCCCCTGAAGAGTACCCCGCTGTTCGCAAGACCGACGGCGAGAAGGAGAAGGGCCAGGGTCCGGAAAAACGAGATCCGGGTGACGGCGAAGCATGGTCA 100
      |
      |
      |
TcY 1 CGTCCCCTGAAGAGTACCCCTCGTGTTCGCAAGACCGACGGCGAGAAGGAGAAGGGCCAGGGTCCGGAAAAACGAGATCCGGGTGACGGCGAAGCATGGTCA 100
-----
Tc058 101 GCGCAGCTACATCAGGTATCGGATTGCACTCCTCCGTGGCGAGGATGGCAAGGTGCAAAAACGACACGATCAAGATCAGTGCCATGGGTGCTGCAATCCAC 200
      |
      |
      |
TcY 101 GCGCAGCTACATCAGGTATCGGATTGCACTCCTCCGTGGCGAGGATGGCAAGGTGCAAAAACGACACGATCAAGATCAGTGCCATGGGTGCTGCAATCCAC 200
-----
Tc058 201 AACCGGTGAACATTGCGGAGATTGTGAAGCGCCGTGTGGTCCGGCTGCACCAGACGACGGATGTCTCGTCCGAGATAATTCACGACGAGTACGAGGGCA 300
      |
      |
      |
TcY 201 AACCGGTGAACATTGCGGAGATTGTGAAGCGCCGTGTGGTCCGGCTGCACCAGACGACGGATGTCTCGTCCGAGATAATTCACGACGAGTACGAGGGCA 300
-----
Tc058 301 TTGATGGGAAGAAGGAGAATATGAAGGTGGAGCGCAAGGTGTCAACCATCCTCATCACGCTCTCTCTCAAGCCGCTTGACCGGAACCATGTTGGATACCA 400
      |
      |
      |
TcY 301 TTGATGGGAAGAAGGAGAATATGAAGGTGGAGCGCAAGGTGTCAACCATCCTCATCACGCTCTCTCTCAAGCCGCTTGACCGGAACCATGTTGGATACCA 400
-----
Tc058 401 GCCGCCTCTGCCGGAGTCCGAGGTGAAGGAGCAGGATGACCCCGAGCCTGGCGAGTCCGCTGGTGA AAAACCGCAACCGTGGT GAGCGGGGGAAACGCACT 500
      |
      |
      |
TcY 401 GCCGCCTCTGCCGGAGTCCGAGGTGAAGGAGCAGGATGACCCCGAGCCTGGCGAGTCCGCTGGTGA AAAACCGCAACCGTGGT GAGCGGGGGAAACGCACT 500
-----
Tc058 501 GACCGTGGCGGTCTGTGGCGCGCCGAGGCAACAACAGCGCCGCTCGTGGTCCGGGTTATTACAGTGGGGGCCAACCGCGCGCAGCAACACCGGCCGCT 600
      |
      |
      |
TcY 501 GACCGTGGCGGTCTGTGGCGCGCCGAGGCAACAACAGCGCCGCTCGTGGTCCGGGTTATTACAGTGGAGGCCAGCGCGCGCAACAACACCGGCCGCT 600
-----
Tc058 601 CCCGCAGTGGCCGTGGAGGCAATGGTCCGCGTGGCTCCCGTGGCGGTTCCCGTAAT 656
      |
      |
      |
TcY 601 CCCGCAGTGGCCGTGGAGGCAATGATCGCGGTGGCTCCCGTGGCGGTTCCCGTAAT 656
-----

```

B

| Secuencia | Nc | % Id. entre genes | % Id. con gen de TcCLB.510877.40 | % Id. con gen de TbALBA3 | aa | % Id. entre proteínas |
|-----------|-----|-------------------|----------------------------------|--------------------------|-----|-----------------------|
| TcY | 656 | 98,6 | 98,8 | 71,1 | 218 | 97,2 |
| Tc058 | 656 | | 98,6 | 71,6 | 218 | |

Fig. 7. Características de las secuencias del gen codificante de TcCLB.510877.40 de la cepa Y y 058PUJ de *T. cruzi*. **A.** Secuencias consenso obtenidas para gen de estudio codificante de proteína ALBA en la cepa Y (TcY) y 058-PUJ (Tc058) de *T. cruzi*; Imagen base obtenida de alineamiento con *Pairwise Sequence Alignment* del EBI). **B.** Características del producto de amplificación obtenido en cada una de las cepas evaluadas. %Id: Porcentaje de identidad.

5.3. Características de las proteínas ALBA en *T. cruzi*

5.3.1. Propiedades fisicoquímicas

Las secuencias de *T. cruzi* obtenidas *in vitro*, junto con las secuencias de las cuatro (4) proteínas ALBA identificadas en la cepa de referencia de *T. cruzi* (Cepa CL Brener – Esmeraldo like) con la herramienta Blast, fueron objeto del análisis fisicoquímico realizado con la herramienta ProtParam del servidor ExPASy. Como se observa en la **Tabla 6**, las proteínas ALBA3 de *T. cruzi*, tanto obtenidas del análisis Blast como obtenidas *in vitro*, y la proteína ALBA1 presentan un carácter básico, debido a que presentan un elevado punto isoelectrico (PI < 9,0), y una carga positiva representada por un mayor porcentaje de residuos positivos que negativos [70], a diferencia de la proteína ALBA2 cuya carga es menos positiva debido a que su punto isoelectrico es cercano a la neutralidad y su porcentaje de residuos positivos es poco mayor al porcentaje de residuos negativos. La mayoría de proteínas obtuvieron un valor de índice de inestabilidad menor a 40, por lo cual se presume que son estables en condiciones *in vitro*; así las proteínas Tc2 y Tc3a

son consideradas poco estables al superar dicho valor [71]. Sin embargo, la proteína Tc2 junto con la Tc1, obtuvieron los índices alifáticos de mayor valor, asegurando que presentan mayor termoestabilidad [72], en comparación con las demás proteínas ALBA. Cabe resaltar que todas las proteínas del estudio son hidrofílicas debido a que en todos los casos el valor GRAVY fue negativo [73], sin embargo, las proteínas ALBA3 fueron las que presentaron una mayor propiedad hidrofílica representada por la obtención de los menores valores entre el grupo estudiado.

Tabla 6. Características generales y propiedades fisicoquímicas de las proteínas ALBA en *T. cruzi*.

| Genes | Tamaño (pb) | Peso molecular (KDa) | PI | II | IA | GRAVY | %R+ | %R- |
|-------|-------------|----------------------|-----|------|------|--------|------|------|
| Tc1 | 122 | 13,5 | 9,1 | 20,8 | 80,7 | -0,619 | 17,2 | 14,8 |
| Tc2 | 117 | 13,0 | 7,8 | 43,7 | 89,8 | -0,452 | 16,2 | 15,4 |
| Tc3a | 230 | 25,1 | 9,8 | 44,5 | 58,9 | -1,109 | 18,7 | 14,8 |
| Tc3b | 226 | 24,4 | 9,7 | 39,0 | 60,0 | -1,133 | 17,7 | 14,2 |
| Tc058 | 218 | 23,6 | 9,8 | 37,8 | 58,1 | -1,196 | 18,3 | 14,2 |
| TcY | 218 | 23,7 | 9,7 | 36,7 | 62,2 | -1,175 | 18,3 | 14,7 |

PI: Punto Isoeléctrico; **II:** Índice de Inestabilidad; **GRAVY:** del inglés, *grand average of hydropathy* (promedio general de hidropatía); **%R+:** Porcentaje de residuos con carga positiva; **%R-:** Porcentaje de residuos con carga negativa.

5.3.2. Dominios y/o motivos identificados

El análisis de dominios y motivos utilizando el programa InterProScan, permitió identificar la presencia del dominio ALBA en las seis secuencias de proteínas ALBA de *T. cruzi*, reportado en la base de datos de las superfamilias como IPRO36882, el cual corresponde a la región que contiene las estructuras secundarias que conforman el dominio y que son compartidas por todas las proteínas de la familia, además de contener la secuencia del dominio Pfam01918 (**Figura 8**). Como se observa en la **Figura 8A**, el dominio se encuentra comprendido entre los aminoácidos 9 y 117 en Tc1, 11 y 117 en Tc2, 24 y 143 en Tc3a y Tc3b y, 19 y 138 en TcY y Tc058, con un tamaño entre 106 a 119 aminoácidos, el cual es igual en todas las proteínas ALBA3 de *T. cruzi* analizadas.

Los resultados de la búsqueda manual de dominios y motivos usando como base el alineamiento múltiple en Clustal Omega de las proteínas ALBA de *T. cruzi*, evidenciaron para las proteínas ALBA3 la presencia de motivos RGG, correspondiente a repeticiones de aminoácidos ricas en

arginina y glicina (RGG: 1 arginina y 2 glicinas) presentes en secuencias dobles o triples de RGG en la región C-terminal [74], así como la presencia de un motivo GYQXP (denominado por su composición de aminoácidos) contenido en el dominio ALBA. Por su parte, para las proteínas ALBA2, sólo se encontró el motivo FDXh dentro de la región del dominio ALBA, sin presencia de otros motivos. Sin embargo en la proteína Tc1 no se hayaron otros dominios o motivos acompañantes del dominio ALBA (**Figura 8**).

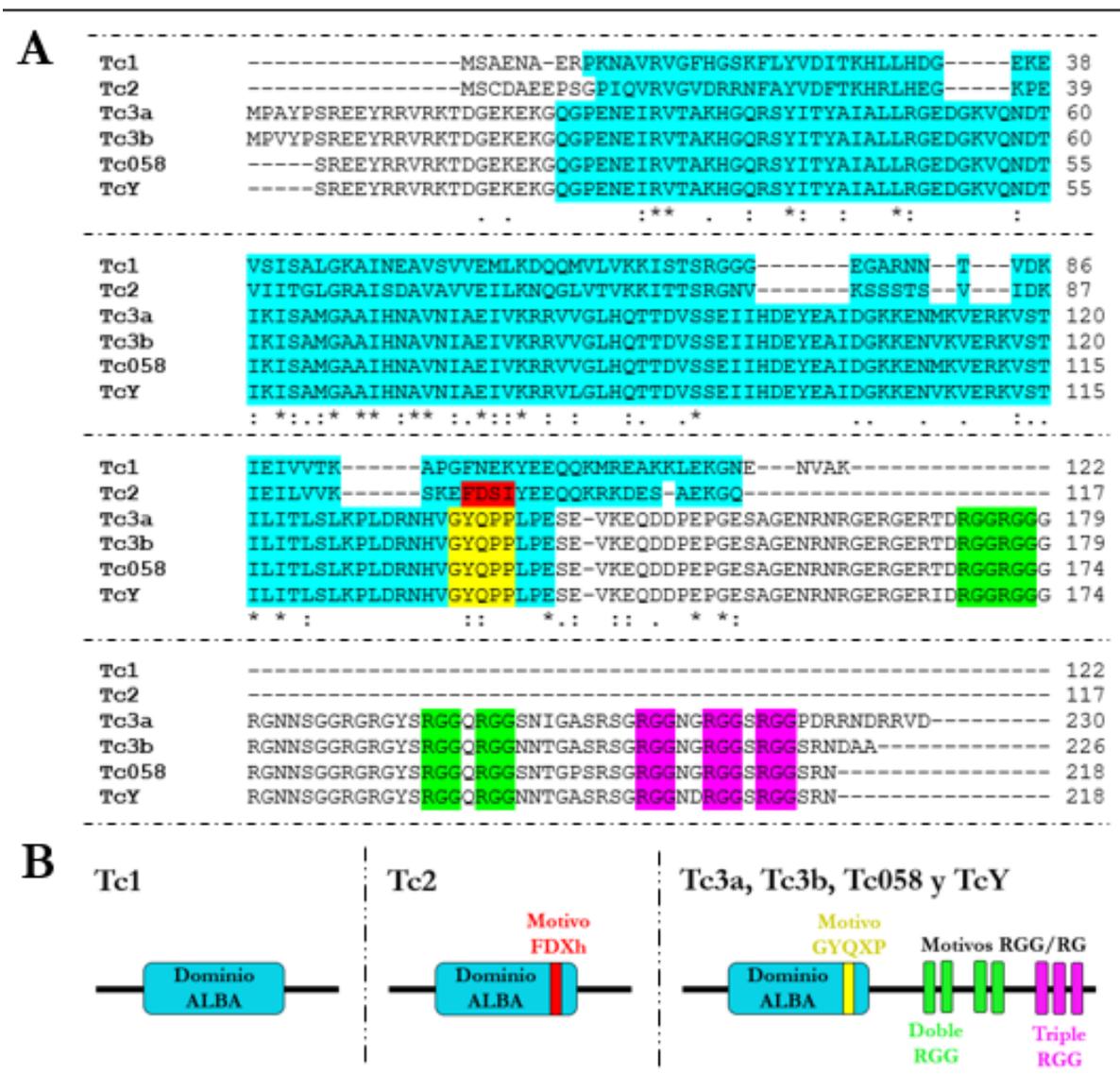


Fig. 8. Organización estructural de los dominios y motivos identificados en las proteínas ALBA de *T. cruzi*. **A.** Alineamiento de las secuencias proteicas de *T. cruzi*, base del esquema, realizado con la herramienta Clustal Omega del instituto EBI. Residuo con alto (*), mediano (:) o bajo (.) grado de conservación. **B.** Esquemas propuestos para describir la organización de dominios y motivos en las proteínas TcALBA

5.4. Obtención del modelo tridimensional de la proteína TcY (TcALBA3)

Teniendo en cuenta los antecedentes descritos para la proteína TcCLB.510877.40, se decidió realizar un análisis adicional correspondiente al modelamiento de la estructura tridimensional de esta proteína, para evidenciar si a partir de la secuencia proteica obtenida *in vitro* se consigue hacer un acercamiento a la conformación real que adquiere la proteína y con ello proponer un modelo para esta proteína. Este modelamiento fue realizado a partir de la secuencia obtenida de la cepa Y puesto que se ha observado a lo largo de los resultados que las secuencias de este gen en las diferentes cepas analizadas, son bastante similares entre si.

La **Figura 9A** presenta la estructura tridimensional de la proteína TcY obtenida con el servidor Phyre2, el cual utilizó como base la subunidad P25 de la proteína ribonucleasa P humana identificada en el Protein Data Bank como 6CWX. El 57% del modelo de TcY (124 aminoácidos), fue construido con un porcentaje de confianza mayor al 90 %, esta región está situada entre el aminoácido 14 y 138, en donde fue determinada la región del dominio ALBA. Se encontró que este dominio presenta una topología de sucesiones de pliege Alfa/Beta compuesta por dos (2) hélices alfa (α) y cuatro cadenas Beta (β) interconectadas por medio de regiones desordenadas ($\text{N}-\beta_1-\alpha_1-\beta_2-\alpha_2-\beta_3-\beta_4-\text{C}$), y se evidencia la formación de dos hojas Beta, una paralela ($\beta_1-\beta_2$) y otra antiparalela ($\beta_3-\beta_4$). En el caso de los motivos identificados, estos se localizan en regiones desordenadas (**Figura 9A**).

La validación estereoquímica del modelo se presentó en un diagrama de Ramachandran (**Figura 9B**), el cual indicó que un 96% de los aminoácidos del modelo generado se encontraban en las regiones más favorecidas o adicionales permitidas (regiones en las que los ángulos Phi y Psi de los aminoácidos son posibles en la estructura obtenida) y el otro 4% faltante se encontraban en regiones generosamente permitidas o no permitidas (regiones en las que los ángulos Phi y Psi de los aminoácidos son improbables para la estructura obtenida), lo que sugiere que el modelo puede ser una buena aproximación a la conformación nativa de la proteína.

Como parte de la verificación del modelo obtenido se realizó un alineamiento entre las estructuras del dominio ALBA de la proteína TcY (al ser la única región con estructuras secundarias y alto grado de conservación) con el dominio de la subunidad P25 de la ribonucleasa humana (Rpp25) (**Figura 10**). De este alineamiento se obtuvo un valor de RMSD (del inglés, *Root-mean-square deviation* o desviación de la raíz cuadrada media); cuanto más bajo sea el valor de

RMSD obtenido, mayor será la similitud entre las estructuras comparadas. El valor de 0,443 obtenido en este caso indica una alta similitud entre la estructura del dominio en ambas proteínas, resultado que apoya nuestra hipótesis en la cual se sugiere que el modelo obtenido es un buen acercamiento a la estructura real de la proteína.

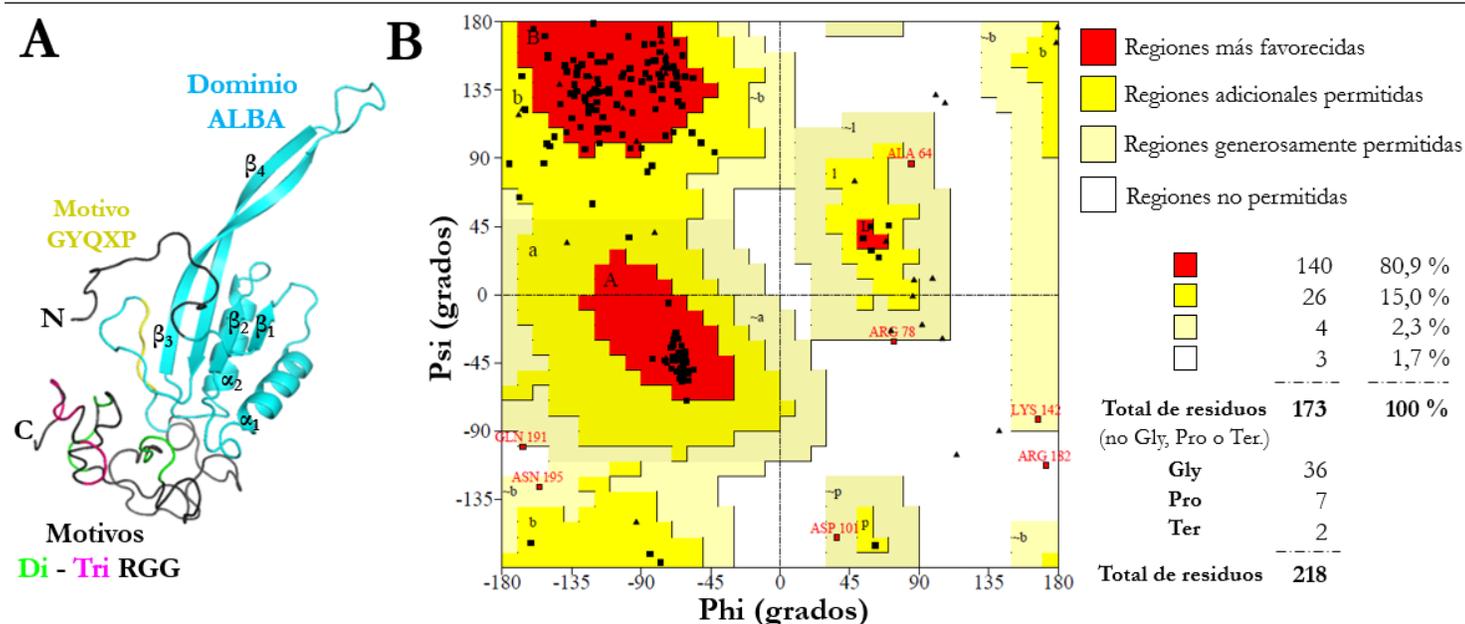


Fig. 9. Predicción de estructura 3D de la proteína TcY de *T. cruzi*. **A.** Modelo tridimensional de la proteína TcY obtenido en servidor Phyre2 y visualizado con Pymol 2,1. **B.** Validación estereoquímica del modelo, realizada con la herramienta PROCHECK y representada en diagrama de Ramachandran. **Gly:** Glicina; **Pro:** Prolina; **Ter.:** Residuos terminales.

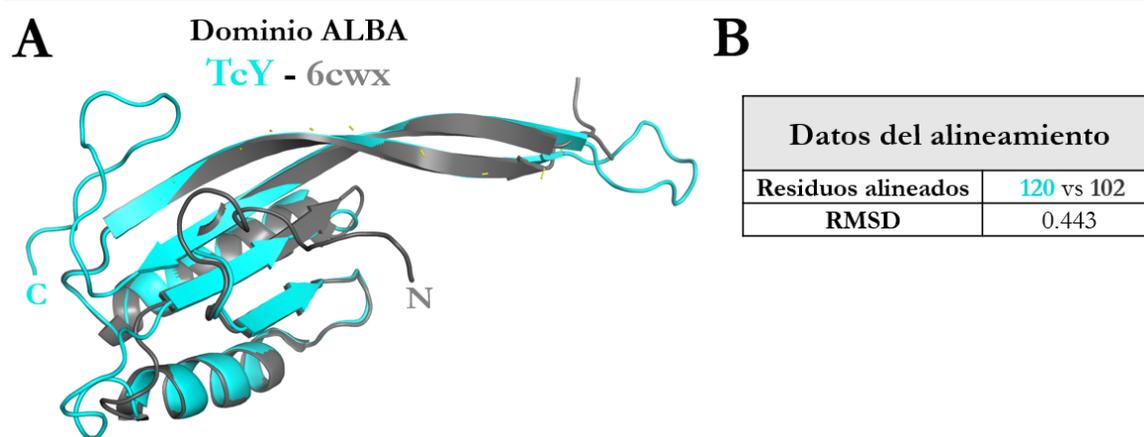


Fig. 10. Alineamiento de las estructuras del dominio ALBA entre el modelo de la proteína TcY y el de la proteína base para su construcción (Rpp25). **A.** Superposición de los modelos de dominio ALBA obtenida del alineamiento realizado en Pymol 2,1 entre TcY y Rpp25. **B.** Datos informativos acerca del alineamiento, obtenidos en el mismo programa.

5.5. Análisis comparativo de las proteínas ALBA de *T. cruzi* y otros tripanosomátidos

Para el análisis comparativo de las proteínas ALBA identificadas, se utilizaron las secuencias de las cepas de referencia de cada una de las especies de tripanosomátidos analizadas.

En la **Figura 11**, se observa el árbol filogenético consenso de estas secuencias, en el cual hay presencia de dos grandes clados soportados por valores bootstrap de 100, los cuales corresponden a los dos linajes eucariotas de la familia ALBA, ya que en cada uno se agrupan proteínas que han sido previamente clasificadas en cada uno de dichos grupos monofiléticos, como LiALBA1 (Li2) y TbALBA3 (Tb3) las cuales pertenecen al linaje Rpp20 y Rpp25, respectivamente. Además, se observa que ambos linajes se dividen en dos clados inferiores en los que se evidencia una separación entre las proteínas ALBA de los géneros *Trypanosoma* y *Leishmania*. En el género *Trypanosoma*, las proteínas ALBA1 y 2 presentan una clara diferenciación al agruparse en dos clados separados, diferente a lo observado en *Leishmania* spp., en donde la única ALBA1 identificada (Lm1) se encuentra inmersa en el clado de las proteínas ALBA2 del mismo género, lo que indicaría que todas corresponden a un mismo tipo de proteína ALBA. Adicionalmente, en el género *Leishmania*, se observa que las proteínas ALBA2 y ALBA3 son conservadas en todas las especies identificadas, lo que sugiere que son conservadas en todo el género.

Teniendo en cuenta que la proteína TcY analizada corresponde a una proteína ALBA3, se decidió realizar un análisis filogenético de estas proteínas, mediante la elaboración de un árbol filogenético, el cual es evidenciado en la **Figura 12**. En este caso se observa nuevamente la separación de las proteínas ALBA de *Trypanosoma* spp. y *Leishmania* spp. en diferentes clados, y en el caso de la proteína TcY, esta se agrupó en un mismo clado con la proteína Tc3b, resultado esperado, debido a que la proteína Tc3b corresponde a la secuencia de TcCLB.510877.40 cuyo gen fue el blanco utilizado para la obtención de la secuencia génica, y posteriormente proteica, de TcY. Por otra parte, se observó que las proteínas ALBA3 de *T. cruzi* se agruparon con la proteína ALBA3 de *T. theileri* separadas de las proteínas ALBA3 y 4 de *T. brucei* que se encontraban agrupadas con las proteínas ALBA3 de *T. congolense* y ALBA3 y 4 de *T. evansi* en el grupo de proteínas de *Trypanosoma* spp., esta separación entre *T. cruzi* y *T. brucei* también se evidencia en los clados de para las proteínas ALBA1 y ALBA2 para este género.

Sumado a lo anterior, los resultados de los análisis de identidad de estas secuencias mediante la herramienta Clustal Omega (**Tabla 7**), permitieron evidenciar mayores porcentajes de identidad entre las proteínas ALBA 2 del género *Leishmania* (82,6 a 100,0 %) en comparación con lo obtenido para *Trypanosoma* spp. (51,3 – 100,0 %), este mismo resultado fue obtenido en el caso de las proteínas ALBA3, lo cual sugiere un mayor grado de conservación a nivel de secuencia para las proteínas ALBA2 y 3, en el género *Leishmania* en comparación con el género *Trypanosoma*.

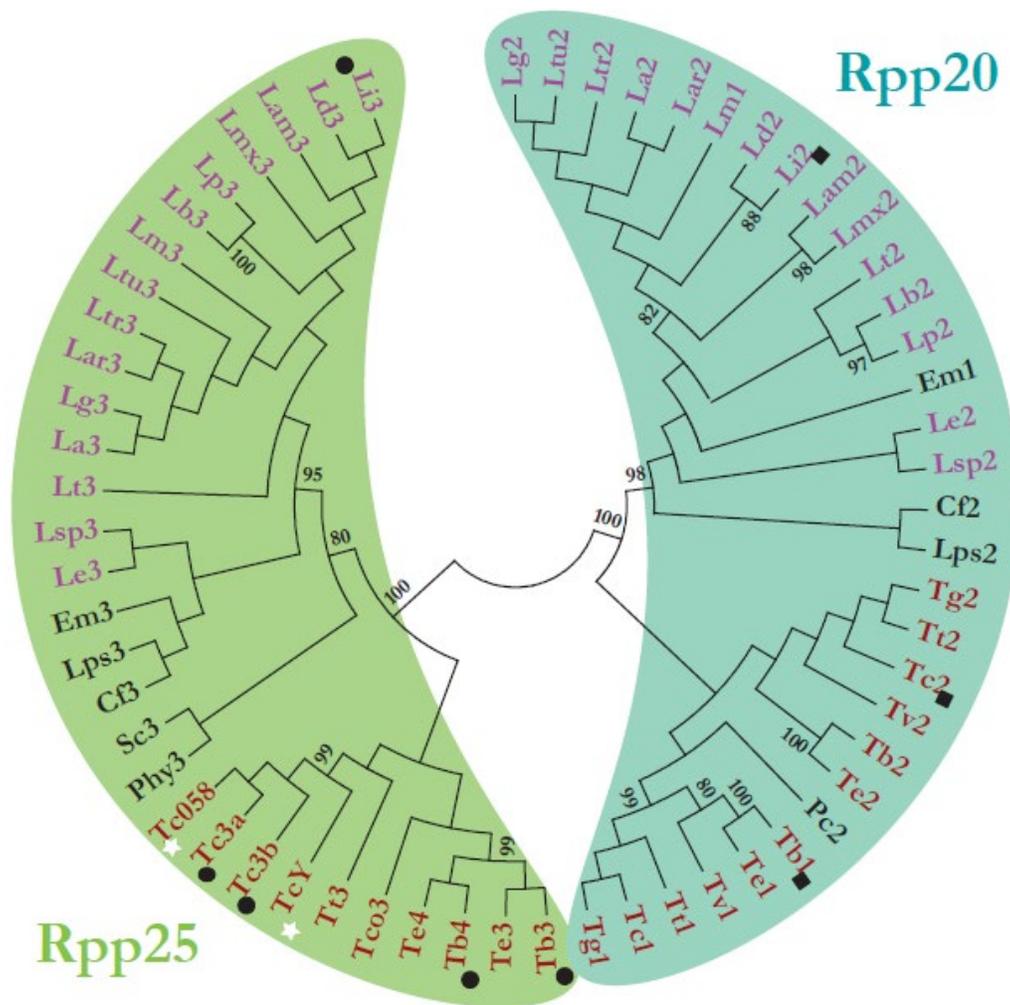


Fig. 11. Árbol filogenético consenso de las proteínas ALBA identificadas, elaborado por el método de Máxima verosimilitud en MEGA 7,0 con Bootstrap de 1000 réplicas. Se muestran los dos linajes eucariotas de la familia ALBA (**Forma azul:** proteínas similares a la proteína **Rpp20**; **Forma verde:** proteínas similares a la proteína **Rpp25**), las Proteínas ALBA de *Trypanosoma* spp. (letra naranja) y de *Leishmania* spp. (letra morada), y los valores de soporte Bootstrap mayores a 80%. ☆: proteínas ALBA de *T. cruzi* obtenidas *in vitro*, ●: Proteínas ALBA reportadas en el linaje Rpp25; ■ Proteínas ALBA reportadas en el linaje Rpp20.

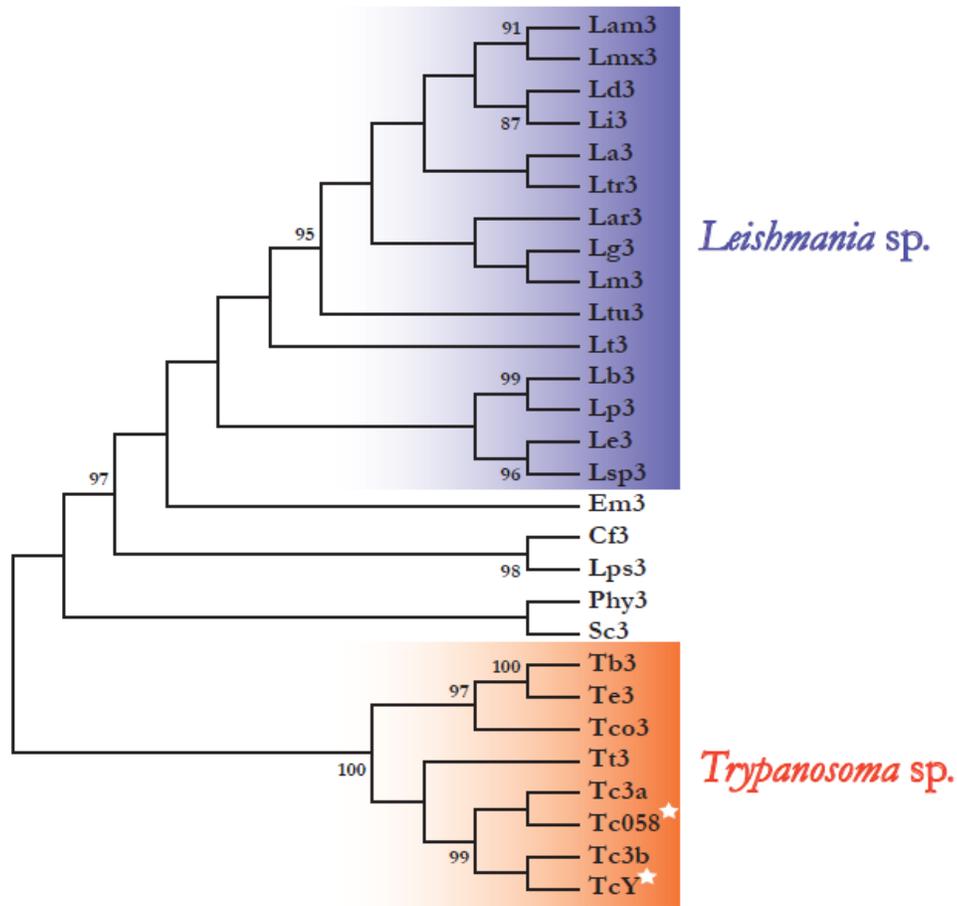


Fig. 12. Árbol filogenético consenso de las proteínas ALBA 3 identificadas, elaborado por el método de Máxima verosimilitud en MEGA 7,0 con Bootstrap de 1000 réplicas. Se muestran los Bootstrap mayores a 80% y se destacan los clados de las Proteínas ALBA3 de *Trypanosoma* spp. (forma naranja) de *Leishmania* spp. (forma morada). ☆: proteínas ALBA de *T. cruzi* obtenidas *in vitro*, ■: Proteínas ALBA reportadas en la literatura (Li3: LiALBA3, Tb3: TbALBA3, Tc3a: TcALBA30, Tc3b: TcALBA40).

Tabla 7. Comparación entre las clases definidas de proteínas ALBA

| Grupos | Porcentaje de identidad entre proteínas | | | |
|---|---|--------------|--------------|-------|
| | ALBA1 | ALBA2 | ALBA3 | ALBA4 |
| <i>Trypanosoma</i> spp. | 72,8 - 100,0 | 51,3 - 100,0 | 65,2 - 100,0 | 100 |
| <i>Leishmania</i> spp. | NP | 82,6 - 100,0 | 82,8 - 100,0 | NP |
| Todos los tripanosomátidos presenciados | 45,4 - 100,0 | 36,8 - 100,0 | 54,2 - 100,0 | 100 |

NP: El grupo no presenta al menos dos proteína ALBA de ese tipo.

6. Discusión

En este estudio se planteó identificar los genes codificantes de la familia ALBA en *T. cruzi*. Hasta el momento para este parásito se han descrito cuatro proteínas ALBA (TcALBA1, TcALBA2, TcALBA30 y TcALBA40) cuyos genes se encuentran ordenados en parejas y organizadas en tándem entre la ALBA1 y 2, y la ALBA30 y 40. En nuestro caso también se reporta la presencia de cuatro secuencias codificantes de proteínas ALBA con el mismo patrón de organización. Sin embargo, se dio a conocer que la organización en tándem es compartida entre los genes de proteínas ALBA de *T. cruzi* y de *T. brucei*, además de compartir de 4 a 5 genes adyacentes entre sí, y 3 de estos con los genes codificantes de las proteínas ALBA2 y ALBA3 de *L. infantum*, así como el orden en el que estos se encuentran organizados; dicho fenómeno corresponde a un evento de sintenia compartida. Este evento se ha reportado para otros genes del género *Trypanosoma* como es el caso del gen PRAC (gen codificante de la Prolina-racemasa) cuyas regiones adyacentes exhibieron elevada sintenia en *T. cruzi*, *T. c. marinkellei*, *T. erneyi*, *T. dionisii*, *T. vivax* y *T. grayi* con al menos 8 genes organizados en el mismo orden en todas las especies mencionadas [75]. Asimismo, se ha observado que entre *T. brucei*, *T. cruzi* y *L. major* hay alrededor de 6158 genes conservados ubicados en clusters con sintenia compartida entre estas especies [76].

La evidencia de sintenia entre genes de proteínas ALBA en *T. cruzi*, *T. brucei* y *L. infantum* soporta su ortología [75], sugiriendo que las proteínas ALBA en estos parásitos podrían tener características estructurales y funcionales similares, como se evidencia en el caso de las proteínas TcALBA30 (correspondiente a la proteína aquí denominada Tc3a), LiALBA3 y TbALBA3, las cuales presentan motivos compartidos como repeticiones de RRG C-terminales, tienen capacidad de unión a ARN mensajero e intervienen en procesos relacionados con la estabilización/desestabilización de ARNm de proteínas en su respectivo parásito [49,50,56]

De acuerdo a las propiedades fisicoquímicas analizadas, se determinó en este estudio, que la mayoría de las proteínas ALBA de *T. cruzi* son básicas, ya que estas presentan valores de punto isoeléctrico (PI) entre 9,1 y 9,8. Esto es congruente con lo descrito en la literatura para proteínas de esta familia, en donde se atribuye dicho carácter básico por presentar altos valores de PI (10,34 en el caso de la proteína TcALBA30) y la importancia del mismo en la unión con ácidos nucleicos [33,56,77]. Sin embargo, la proteína Tc2, obtuvo un punto isoeléctrico cercano a la

neutralidad (7,8) y, a su vez, porcentajes de residuos negativos y positivos cercanos, indicando que probablemente la función que desempeña en el parásito no es dependiendo de su capacidad de unión a estas moléculas, como en el caso de la proteína ALBA MthALBA de *Methanobacterium thermoautotrophicum* (MthAlba) la cual carece de carácter básico y presenta varias sustituciones de aminoácidos básicos por aminoácidos neutros, siendo incapaz de unirse a ácidos nucleicos (tanto ADN como ARN) en condiciones *in vitro* [78].

En cuanto a la estructura observada para las proteínas ALBA3 de *T. cruzi*, tanto las obtenidas *in vitro* como en el análisis Blast, se describió la presencia de motivos de repeticiones RGG C-terminal. Este motivo es formado por repeticiones cercanas de dobles o triples RGG y están implicados en las interacciones proteína-ARN, siendo un motivo de unión a ARN frecuente en proteínas del dominio eucariota [79,80]. Si bien no es claro el mecanismo de unión se sugiere que las interacciones electrostáticas entre el esqueleto fosfatado del ácido nucleico y las argininas del motivo cumplen un papel fundamental en el establecimiento de la unión [79,81], por lo que el punto isoeléctrico de las argininas (pH =10) [82] es importante, ya que en condiciones de pH inferiores a este punto, el aminoácido se encuentra cargado positivamente y podrá efectuar la interacción electrostática. En los estadios epimastigote, amastigote y tripomastigote de *T. cruzi*, se reporta un valor de pH intracelular aproximado de 7,5 [83,84], lo que sugiere que en los tres estadios del parásito, los motivos se encuentran cargados permitiendo la interacción electrostática entre la proteína con la molécula de ARN y el desarrollo de su posible función asociada a esta molécula.

Este motivo se ha descrito presente en estructuras desordenadas de proteínas de unión a ARN [79,85], tal y como se evidenció en el modelo obtenido en Phyre2 de la proteína TcY, en el que las repeticiones RGG se encontraron en amplias regiones desordenadas en el extremo C-terminal de la proteína. En este mismo modelo teórico el dominio ALBA de *T. cruzi* obtuvo una alta similitud con el dominio ALBA obtenido por cristalografía de la subunidad P25 de la ribonucleasa humana representado por un valor de RMSD cercano a cero. Lo anterior indica que el modelo propuesto para la proteína TcY parece ser un buen acercamiento a su estructura en la naturaleza, ya que presenta estructuras observadas en modelos proteicos existentes (dominio ALBA de Rpp25 y la estructura desordenada descrita para el motivo RGG), así como una elevada cantidad de aminoácidos con una conformación estereoquímica adecuada para la estructura obtenida, observado en el diagrama de Ramachandran, que aunque no mostró un total

de 90% de aminoácidos ubicados en regiones más favorables (sugerido por el programa para definirlo como un buen modelo [86]) el porcentaje obtenido fue cercano a este valor (81%).

Por último, en el análisis comparativo descrito por los análisis filogenéticos realizados, se observó una agrupación entre las proteínas de *Leishmania* spp. y *Trypanosoma* spp.; aunque ambos presentan un número diferente de secuencias ALBA (4 en *Trypanosoma* y 2 en *Leishmania*) estos se agruparon de manera que en cada uno de los grandes linajes evidenciados estaba presente una proteína ALBA de *Leishmania* spp. y dos de *Trypanosoma* spp., encontrándose agrupadas proteínas de cada género en clados diferentes como ha sido previamente reportado [87]. Esta misma distribución ha sido reportada en otro análisis filogenético de proteínas ALBA en estos dos organismos, en donde se ha sugerido que esta diferencia, en cuanto al número de genes, se debe a que los genes de las proteínas ALBA en el género *Trypanosoma* sufrieron un proceso de duplicación después de la especiación con el género *Leishmania* [48]. Dicho proceso de duplicación génica post-especiación se ha sugerido en otros genes en estos dos géneros, como aquellos codificantes de la superfamilia de proteínas ABC (proteínas transportadoras dependientes de ATP), en donde se ha evidenciado que el género *Leishmania* presenta 42 genes *ABC*, a diferencia de *T. cruzi* y *T. brucei* que presentan tan solo 22 y 28, respectivamente, y se ha descrito que esta diferencia se debe principalmente al aumento de los genes codificantes de subfamilias de proteínas ABC en *Leishmania* spp., el cual, se sugiere, ocurrió debido a eventos de duplicación génica en estos parásitos después de su división del linaje compartido con *Trypanosoma* spp., ya que los múltiples genes de estas subfamilias se encuentran agrupados en forma de tándem [88, 89], lo que es compatible con un evento de duplicación de genes; organización que también comparten las parejas de genes que codifican para las proteínas que se agrupan en un mismo linaje de proteínas ALBA (Tb1 y Tb2; Tb3 y Tb4)-

Adicionalmente se observa una elevada conservación en el número de proteínas ALBA para el género *Leishmania*, la cual fue descrita en un análisis realizado con un menor número de especies del género [87], por lo que este estudio soporta y complementa dicha hipótesis, al proponer que las proteínas ALBA también son conservadas a nivel de secuencia proteica al evidenciarse porcentajes de identidad de 82 y 100% entre proteínas del mismo tipo de ALBA en diferentes especies. Sin embargo, en *Trypanosoma* spp., no se evidenció dicha conservación en el número o en la secuencias de proteínas ALBA. Por otra parte, en el caso de *T. cruzi*, se observó que sus proteínas se encuentran distribuidas en cada uno de los dos grandes grupos de proteínas ALBA,

así como también dichas proteínas se organizan en los clados de proteínas de género *Trypanosoma* spp en cada grupo. Sin embargo, se evidenció que dentro de estos clados se forman agrupaciones que separa proteínas de *T. cruzi* de las proteínas de *T. brucei*, y en algunos casos no se evidencia la separación por clases de ALBA como ocurrió con las clases ALBA3 y 4, en donde las proteínas ALBA 3 de *T. brucei* y *T. evansi*, así como sus proteínas ALBA4, se agruparon, pero todas estas formando un grupo dentro del clado que contiene las otras proteínas ALBA3 de *Trypanosoma* spp. Esto puede deberse a que algunas especies presentan mayor cercanía filogenética derivada de su origen compartido, ya que las proteínas de *T. brucei* se agruparon con proteínas de otras especies de *Trypanosoma* spp., que al igual que este, son originarias del continente africano, diferente a las proteínas de *T. cruzi* que se agruparon con proteínas de especies, que al igual que *T. cruzi*, son americanas [90]. Esto se evidencia con mayor claridad en el árbol filogenético de proteínas ALBA3 (**Figura 12**), en el que las proteínas de las especies africanas *T. congolense*, *T. evansi* y *T. brucei* y las proteínas ALBA3 de las cepas del continente americano *T. cruzi* y *T. theileri* se agrupan en clados separados.

7. Conclusiones

Las proteínas ALBA son más conservadas a nivel de número y secuencia en el género *Leishmania* sp. en comparación con las proteínas del género *Trypanosoma* spp.

En *T. cruzi*, se evidenció la presencia de dos proteínas ALBA1 y 2, así como dos proteínas correspondientes a la clase ALBA3, patrón de número y clases de proteínas ALBA que no fue evidenciado en otra especie del género.

Los genes de las proteínas ALBA en *T. cruzi*, *T. brucei* y *L. infantum* presentan una organización compartida o sintenia, que es más notoria entre especies de un mismo género.

Las características estructurales y fisicoquímicas de las proteínas ALBA de *T. cruzi* corresponden con características de la familia de proteínas ALBA descritas en la literatura.

El modelo predicho para la proteína TcY presenta una región con estructuras secundarias ordenadas que comprenden el dominio ALBA y son compatibles con lo que reporta la literatura

T. cruzi presenta proteínas pertenecientes a cada uno de los dos grandes grupos de proteínas ALBA, así como representantes de cada uno de los clados formados de proteínas de *Trypanosoma* spp. y junto a las proteínas de especies americanas se agrupan en clados diferentes a los formados por especies africanas.

8. Referencias

1. World Health Organization [WHO]. Fourth WHO report on neglected tropical diseases. Integrating neglected tropical diseases into global health and development. Geneva (Switzerland). 155-62, 2017.
Disponible en: https://www.who.int/neglected_diseases/resources/9789241565448/en/
2. Jannin J, Villa L. An overview of CD treatment, *Memórias do Instituto Oswaldo Cruz*, 102(Suppl D): 95–97, 2007.
doi:[10.1590/S0074-02762007005000106](https://doi.org/10.1590/S0074-02762007005000106)
3. Urbina JA. Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches, *Acta Tropica*, 115(1-2): 55-68, 2010.
doi:[10.1016/j.actatropica.2009.10.023](https://doi.org/10.1016/j.actatropica.2009.10.023)
4. Fernandes CD, Tiecher FM, Balbinot MM, Liarte DB, Scholl D, Steindel M, Romanha A. Efficacy of benznidazol treatment for asymptomatic Chagasic patients from state of Rio Grande do Sul evaluated during a three years follow-up, *Memórias do Instituto Oswaldo Cruz*, 104(1): 27–32, 2009.
doi:[10.1590/S0074-02762009000100004](https://doi.org/10.1590/S0074-02762009000100004)
5. Garcia S, Ramos CO, Senra JF, Vilas-Boas F, Rodrigues MM, Campos-de-Carvalho AC, Ribeiro-Dos-Santos R, Soares MB. Treatment with benznidazole during the chronic phase of experimental Chagas' disease decreases cardiac alterations, *Antimicrobial Agents and Chemotherapy*, 49(4): 1521–1528, 2005.
doi:[10.1128/AAC.49.4.1521-1528.2005](https://doi.org/10.1128/AAC.49.4.1521-1528.2005)
6. Subota I, Rotureau B, Blisnick T, Ngwabyt S, Durand-Dubief M, Engstler M, Bastin P. ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation, *Molecular Biology of the Cell*, 22(22): 4205-19, 2011.
doi:[10.1091/mbc.E11-06-0511](https://doi.org/10.1091/mbc.E11-06-0511)

7. Rassi A Jr, Rassi A, Marcondes de Rezende J. American trypanosomiasis (Chagas disease), *Infectious Disease Clinics of North America*, 26(2): 275-91, 2012.
doi:10.1016/j.idc.2012.03.002
8. Laranja FS, Dias E, Nobrega G, Miranda A. Chagas' disease. A clinical, epidemiologic, and pathologic study, *Circulation*, 14(6): 1035–60, 1956 .
PMID:13383798.
9. Pinto AY, Valente SA, Valente VDC, Ferreira Junior AG, Coura JR. Acute phase of Chagas disease in the Brazilian Amazon region: study of 233 cases from Pará, Amapá and Maranhão observed between 1988 and 2005, *Revista da Sociedade Brasileira de Medicina Tropical*, 41(6): 602–14, 2008.
doi:[10.1590/S0037-86822008000600011](https://doi.org/10.1590/S0037-86822008000600011)
10. Dias JC. The indeterminate form of human chronic Chagas' disease. A clinical epidemiological review, *Revista da Sociedade Brasileira de Medicina Tropical*, 22(3): 147–56, 1989.
PMID: 2486527
11. Pérez-Molina JA, Molina I. Chagas disease, *Lancet*, 391(10115): 82-94, 2017.
doi:[10.1016/s0140-6736\(17\)31612-4](https://doi.org/10.1016/s0140-6736(17)31612-4)
12. Bern C. Chagas' disease, *The New England Journal of Medicine*, 373(5): 456-66, 2015.
doi:10.1056/NEJMra1410150
13. Organización Panamericana de la Salud [OPS]. Estimación cuantitativa de la enfermedad de Chagas en las Américas. Montevideo (Uruguay). 2006.
Disponible en: <http://ops-uruguay.bvsalud.org/pdf/chagas19.pdf>
14. Pan American Health Organization [PAHO]. Neglected, Tropical and Vector Borne Diseases. Chagas disease. Washington, D.C. [United States of America]. 2018.
Disponible en: <https://bit.ly/2QxD7ID>
15. Lee BY, Bacon KM, Bottazzi ME, Hotez PJ. Global economic burden of Chagas disease: a computational simulation model, *The Lancet Infectious Diseases*, 13(4): 342-8, 2013.
doi:[10.1016/S1473-3099\(13\)70002-1](https://doi.org/10.1016/S1473-3099(13)70002-1)
16. Castillo-Riquelme M, Guhl F, Turriago B, Pinto N, Rosas F, Flórez-Martínez M, Fox-Rushby J, Davies C, Campbell-Lendrum D. The costs of preventing and treating Chagas disease in Colombia, *PLOS Neglected Tropical Diseases*, 2(11): e336, 2008.
doi:[10.1371/journal.pntd.0000336](https://doi.org/10.1371/journal.pntd.0000336)

17. Rodríguez-Coura J, De Castro SL. A critical review on Chagas disease chemotherapy, *Memórias do Instituto Oswaldo Cruz*, 97(1): 3-24, 2002.
doi:[10.1590/S0074-02762002000100001](https://doi.org/10.1590/S0074-02762002000100001)
18. Castro JA, De Mecca MM, Bartel LC. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis), *Human and Experimental Toxicology*, 25(8): 471-9, 2006.
doi:[10.1191/0960327106het653oa](https://doi.org/10.1191/0960327106het653oa)
19. Urbina JA, Docampo R. Specific chemotherapy of Chagas disease: controversies and advances, *Trends in Parasitology*, 19(11): 495–501, 2003.
doi:[10.1016/j.pt.2003.09.001](https://doi.org/10.1016/j.pt.2003.09.001)
20. Zingales B. *Trypanosoma cruzi* genetic diversity: Something new for something know about Chagas disease manifestations, serodiagnosis and drug sensitivity. *Acta Tropica*, 184: 38-52, 2018.
doi:[10.1016/j.actatropica.2017.09.017](https://doi.org/10.1016/j.actatropica.2017.09.017)
21. Brenière SF, Waleckx E, Barnabé C. Over Six Thousand *Trypanosoma cruzi* Strains Classified into Discrete Typing Units (DTUs): Attempt at an Inventory. *PLOS Neglected Tropical Diseases*, 10(8): e0004792, 2016.
doi:[10.1371/journal.pntd.0004792](https://doi.org/10.1371/journal.pntd.0004792)
22. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infection, genetics and evolution*, 12(2): 240-253, 2012.
doi:[10.1016/j.meegid.2011.12.009](https://doi.org/10.1016/j.meegid.2011.12.009)
23. Hernández C, Cucunubá Z, Flórez C, Olivera M, Valencia C, Zambrano P, et al. Molecular Diagnosis of Chagas Disease in Colombia: Parasitic Loads and Discrete Typing Units in Patients from Acute and Chronic Phases, *PLOS Neglected Tropical Diseases*, 10(9): e0004997, 2016.
doi:[10.1371/journal.pntd.0004997](https://doi.org/10.1371/journal.pntd.0004997)
24. Guhl F, Ramírez JD. Retrospective molecular integrated epidemiology of Chagas disease in Colombia, *Infection, Genetics and Evolution*, 20: 148-154, 2013.
doi:[10.1016/j.meegid.2013.08.028](https://doi.org/10.1016/j.meegid.2013.08.028)
25. Izeta-Alberdi A, Ibarra-Cerdeña CN, Moo-Llanes DA, Ramsey JM. Geographical, landscape and host associations of *Trypanosoma cruzi* DTUs and lineages, *Parasites & Vectors*, 9: 631,

2016.
doi:[10.1186/s13071-016-1918-2](https://doi.org/10.1186/s13071-016-1918-2)
26. Hendriks EF, Matthews KR. In *Trypanosomes: After the Genome*, Norfolk: Horizon Bioscience, United Kingdom; 2007.
 27. Vazquez MP. The Genetics and Genomics of *Trypanosoma cruzi*, *International Journal of Biomedical and Pharmaceutical Sciences*, 1(1): 1-11, 2007.
 28. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function, *Nature Reviews Molecular Cell Biology*, 8(6): 479-90, 2007.
doi:[10.1038/nrm2178](https://doi.org/10.1038/nrm2178)
 29. De Gaudenzi JG, Noé G, Campo VA, Frasch AC, Cassola A. Gene expresión regulation in trypanosomatids, *Essays In Biochemistry*, 51: 31-46, 2011.
doi:[10.1042/bse0510031](https://doi.org/10.1042/bse0510031)
 30. Haile S, Papadopoulou B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa, *Current Opinion in Microbiology*, 10(6): 569-77, 2007.
doi:[10.1016/j.mib.2007.10.001](https://doi.org/10.1016/j.mib.2007.10.001).
 31. Da Silva MS, Cano MIN. *Molecular and Cellular Biology of Pathogenic Trypanosomatids*. 1ª ed. São Paulo: Frontiers in Parasitology, Brazil; 2017.
 32. Oliveira C, Faoro H, Alves LR, Goldenberg S. RNA-binding proteins and their role in the regulation of gene expression in *Trypanosoma cruzi* and *Saccharomyces cerevisiae*, *Genetics and Molecular Biology*, 40(1): 22-30, 2017.
doi:[10.1590/1678-4685-GMB-2016-0258](https://doi.org/10.1590/1678-4685-GMB-2016-0258)
 33. Goyal M, Banerjee C, Nag S, Bandyopadhyay U. The Alba protein family: Structure and function, *Biochimica et Biophysica Acta*, 1864(5): 570-83, 2016.
doi:[10.1016/j.bbapap.2016.02.015](https://doi.org/10.1016/j.bbapap.2016.02.015)
 34. Finn RD, Bateman A, Clements J et al. Pfam: the protein families database, *Nucleic Acids Research*, 42(Database issue): D222-D230, 2014.
doi:[10.1093/nar/gkt1223](https://doi.org/10.1093/nar/gkt1223)
 35. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *The FEBS Journal.*, 272: 2118–31, 2005.
doi:[10.1111/j.1742-4658.2005.04653.x](https://doi.org/10.1111/j.1742-4658.2005.04653.x)

36. Lu D, Searles MA, Klug A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition, *Nature*, 426: 96–100, 2003.
doi:[10.1038/nature02088](https://doi.org/10.1038/nature02088)
37. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d, *Nature Structural & Molecular Biology*, 11: 257–264, 2004.
doi:[10.1038/nsmb738](https://doi.org/10.1038/nsmb738)
38. Grishin NV. KH domain: one motif, two folds. *Nucleic Acids Research*, 29: 638–43, 2001.
doi:[10.1093/nar/29.3.638](https://doi.org/10.1093/nar/29.3.638)
39. Ryter JM, Schultz SC. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA, *The EMBO Journal*, 17: 7505–13, 1998.
doi:[10.1093/emboj/17.24.7505](https://doi.org/10.1093/emboj/17.24.7505)
40. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain, *Cell*, 10(4): 501-12, 2002.
doi:[10.1016/S0092-8674\(02\)00873-5](https://doi.org/10.1016/S0092-8674(02)00873-5)
41. Wang G, Guo R, Bartlam M, Yang H, Xue H, Liu Y, et al. Crystal structure of a DNA binding protein from the hyperthermophilic euryarchaeon *Methanococcus jannaschii*, *Protein Science*, 12(12): 2815-22, 2003.
doi:[10.1110/ps.03325103](https://doi.org/10.1110/ps.03325103)
42. Brent MM, Iwata A, Carten J, Zhao K, Marmorstein R. Structure and biochemical characterization of protein acetyltransferase from *Sulfolobus solfataricus*, *Journal of Biological Chemistry*, 284(29): 19412-9, 2009.
doi:[10.1074/jbc.M109.014951](https://doi.org/10.1074/jbc.M109.014951)
43. Tanaka T, Padavattan S, Kumarevel T. Crystal structure of archaeal chromatin protein Alba2-double-stranded DNA complex from *Aeropyrum pernix* K1, *Journal of Biological Chemistry*, 287(13): 10394-402, 2012.
doi:[10.1074/jbc.M112.343210](https://doi.org/10.1074/jbc.M112.343210)
44. Aravind L, Iyer LM, Anantharaman V. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism, *Genome Biology*, 4(10): R64, 2003.
doi:[10.1186/gb-2003-4-10-r64](https://doi.org/10.1186/gb-2003-4-10-r64)

45. Verma JK, Gayali S, Dass S, Kumar A, Parveen S, Chakraborty S, Chakraborty N. OsAlba1, a dehydration-responsive nuclear protein of rice (*Oryza sativa* L. ssp. indica), participates in stress adaptation, *Phytochemistry* 100: 16-25, 2014.
doi:[10.1016/j.phytochem.2014.01.015](https://doi.org/10.1016/j.phytochem.2014.01.015)
46. Chêne A, Vembar SS, Rivière L, Lopez-Rubio JJ, Claes A, Siegel TN. PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites, *Nucleic Acids Research*, 40(7): 3066-77, 2011.
doi:[10.1093/nar/gkr1215](https://doi.org/10.1093/nar/gkr1215)
47. Gissot M, Walker R, Delhaye S, Alayi TD, Huot L, Hot D, et al. *Toxoplasma gondii* Alba proteins are involved in translational control of gene expression, *Journal of Molecular Biology* . 425(8): 1287-301, 2013.
doi:[10.1016/j.jmb.2013.01.039](https://doi.org/10.1016/j.jmb.2013.01.039)
48. Dupé A, Dumas C, Papadopoulou B. Differential Subcellular Localization of *Leishmania* ALBA-Domain Proteins throughout the Parasite Development, *PLOS One*, 10(9): e0137243, 2015.
doi:[10.1371/journal.pone.0137243](https://doi.org/10.1371/journal.pone.0137243)
49. Dupé A, Dumas C, Papadopoulou B. An Alba-domain protein contributes to the stage-regulated stability of amastin transcripts in *Leishmania*, *Molecular Microbiology*, 91(3): 548-61, 2013.
doi:[10.1111/mmi.12478](https://doi.org/10.1111/mmi.12478)
50. Mani J, Güttinger A, Schimanski B, Heller M, Acosta-Serrano A, Pescher P, et al. Alba-Domain Proteins of *Trypanosoma brucei* Are Cytoplasmic RNA-Binding Proteins That Interact with the Translation Machinery, *PLOS One*, 6(7): e22463, 2011.
doi:[10.1371/journal.pone.0022463](https://doi.org/10.1371/journal.pone.0022463)
51. Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae, *Nucleic Acids Research* 38(Database issue): D457-D462, 2010.
doi:[10.1093/nar/gkp851](https://doi.org/10.1093/nar/gkp851)
52. Bütikofer P, Ruepp S, Boschung M, Roditi I. 'GPEET' procyclin is the major surface protein of procyclic culture forms of *Trypanosoma brucei* strain 427, *Biochemical Journal*, 326(Pt2)(Pt2): 415-23, 1997.
doi:[10.1042/bj3260415](https://doi.org/10.1042/bj3260415)

53. Gupta RS, Golding GB. Evolution of HSP70 gene and its implications regarding relationships between archaeobacteria, eubacteria, and eukaryotes, *Journal of Molecular Evolution*, 37(6): 573-82, 1993.
doi:[10.1007/BF00182743](https://doi.org/10.1007/BF00182743)
54. McCall LI, Matlashewski G. Localization and induction of the A2 virulence factor in *Leishmania*: evidence that A2 is a stress response protein, *Molecular Microbiology*, 77(2): 518-30, 2010.
doi:[10.1111/j.1365-2958.2010.07229.x](https://doi.org/10.1111/j.1365-2958.2010.07229.x)
55. Clos J, Krobitch S. Heat Shock as a Regular feature of the Life Cycle of *Leishmania* Parasites, *American Zoologist*, 39(6): 848-856, 1999.
doi:[10.1093/icb/39.6.848](https://doi.org/10.1093/icb/39.6.848)
56. Pérez-Díaz L, Silva TC, Teixeira SM. Involvement of an RNA binding protein containing Alba domain in the stage-specific regulation of beta-amastin expression in *Trypanosoma cruzi*, *Molecular and biochemical parasitology*, 211: 1-8, 2017.
doi:[10.1016/j.molbiopara.2016.12.005](https://doi.org/10.1016/j.molbiopara.2016.12.005)
57. Grupo de Enfermedades Infecciosas - Laboratorio de Inmunoparasitología Molecular. Comunicación personal: Estudios de proteínas con capacidad de unión a la región 5' UTR de la proteína LYT1 de *Trypanosoma cruzi*. **08/06/2018**.
58. Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis, *Nucleic Acids Research*, 34(Web Server issue): W6-9, 2006.
doi:[10.1093/nar/gkl164](https://doi.org/10.1093/nar/gkl164)
59. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Research*, 31(13): 3784-8, 2003.
doi:[10.1093/nar/gkg563](https://doi.org/10.1093/nar/gkg563)
60. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R.. InterProScan: protein domains identifier. *Nucleic Acids Research*, 33(Web Server issue): W116-20, 2005.
doi:[10.1093/nar/gki442](https://doi.org/10.1093/nar/gki442)
61. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Research*, 31(13): 3497-

- 500, 2003.
doi:[10.1093/nar/gkg500](https://doi.org/10.1093/nar/gkg500)
62. Pavia PX, Thomas MC, López MC, Puerta CJ. Molecular characterization of the short interspersed repetitive element SIRE in the six discrete typing units (DTUs) of *Trypanosoma cruzi*, *Experimental Parasitology*, 132(2): 144-50, 2012.
doi:[10.1016/j.exppara.2012.06.007](https://doi.org/10.1016/j.exppara.2012.06.007)
63. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids, *Journal of visualized experiments*, 45: e2565, 2010.
doi:[10.3791/2565](https://doi.org/10.3791/2565)
64. Fernandes O, Souto RP, Castro JA, Pereira JB, Fernandes NC, Junqueira AC, et al. Brazilian isolates of *T. cruzi* from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences, *American Journal of Tropical Medicine and Hygiene*, 58(6): 807-11, 1998.
PMID: 9660469
65. McWilliam H, Li W, Uludag M, et al. Analysis Tool Web Services from the EMBL-EBI, *Nucleic Acids Research*, 41(Web Server issue): W597–W600, 2013.
doi:[10.1093/nar/gkt376](https://doi.org/10.1093/nar/gkt376)
66. Nocua PA, Ramirez CA, Requena JM, Puerta CJ. *Leishmania braziliensis* SCD6 and RBP42 proteins, two factors with RNA binding capacity, *Parasites & Vectors*, 10(1): 610, 2017.
doi:[10.1186/s13071-017-2557-y](https://doi.org/10.1186/s13071-017-2557-y)
67. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis, *Nature Protocols*, 10(6): 845-58, 2015.
doi:[10.1038/nprot.2015.053](https://doi.org/10.1038/nprot.2015.053)
68. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography*, 26: 283–291, 1993.
doi:[10.1107/S0021889892009944](https://doi.org/10.1107/S0021889892009944)
69. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, *Molecular Biology and Evolution*, 33(7): 1870-4, 2016.
doi:[10.1093/molbev/msw054](https://doi.org/10.1093/molbev/msw054)

70. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa, *Protein science: a publication of the Protein Society*, 10(6): 1206–1215, 2001. doi:10.1110/ps.440101
71. Guruprasad K1, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence, *Protein Engineering, Design and Selection*, 4(2): 155-61, 1990. doi:10.1093/protein/4.2.155
72. Ikai A. Thermostability and aliphatic index of globular proteins, *The Journal of Biochemistry*, 88(6): 1895-8, 1980.
73. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein, *Journal of molecular biology*, 157: 105-132, 1982. doi:10.1016/0022-2836(82)90515-0
74. Rajyaguru P, Parker R. RGG motif proteins: modulators of mRNA functional states, *Cell Cycle*, 11(14): 2594–2599, 2012. doi:10.4161/cc.20716
75. Caballero ZC, Costa-Martins AG, Ferreira RC, et al. Phylogenetic and syntenic data support a single horizontal transference to a *Trypanosoma* ancestor of a prokaryotic proline racemase implicated in parasite evasion from host defences. *Parasites & vectors*, 8: 222, 2015. doi:10.1186/s13071-015-0829-y
76. El-Sayed NM. Comparative Genomics of Trypanosomatid Parasitic Protozoa. *Science*, 309(5733): 404–409, 2005. doi:10.1126/science.1112181
77. Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. *Nature reviews. Molecular cell biology*, 15(11): 749–760, 2014. doi:10.1038/nrm3884
78. Liu YF, Zhang N, Yao HW, Pan XM, Ge M. Mth10b, a unique member of the Sac10b family, does not bind nucleic acid. *PLoS One*, 6(5): e19977. doi:10.1371/journal.pone.0019977
79. Ozdilek BA, Thompson VF, Ahmed NS, White CI, Batey RT, Schwartz JC. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.*, 45(13): 7984–7996, 2017. doi:10.1093/nar/gkx460

80. Thandapani P, O'Connor TR, Bailey TL, Richard S. Defining the RGG/RG motif, *Molecular cell*, 50(5): 613-23, 2013.
doi:10.1016/j.molcel.2013.05.021
81. Blackwell E, Zhang X, Ceman S. Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA, *Human molecular genetics*, 19(7): 1314–1323, 2010.
doi:10.1093/hmg/ddq007
82. Pu W, Zhao H, Huang C, Wu L, Xu D. Visual detection of arginine based on the unique guanidino group-induced aggregation of gold nanoparticles, *Analytica chimica acta*, 764: 78-83, 2013.
doi:10.1016/j.aca.2012.12.026
83. Gil JR, Soler A, Azzouz S, Osuna A. Ion regulation in the different life stages of *Trypanosoma cruzi*, *Parasitology research*, 90(4): 268-72, 2003.
doi:10.1007/s00436-003-0847-0
84. Van Der Heyden N, Docampo R. Intracellular pH in mammalian stages of *Trypanosoma cruzi* is K⁺-dependent and regulated by H⁺-ATPases, *Molecular and biochemical parasitology*, 105(2): 237-51, 2000.
doi:10.1016/S0166-6851(99)00184-X
85. Järvelin AI, Noerenberg M, Davis I, Castello A. The new (dis)order in RNA regulation, *Cell Communication and Signaling*, 14: 9, 2016.
doi:10.1186/s12964-016-0132-3
86. Balaji S, Kalpana R, Shapshak P. Paradigm development: comparative and predictive 3D modeling of HIV-1 Virion Infectivity Factor (Vif), *Bioinformatics*, 1(8): 290–309, 2006.
doi:10.6026/97320630001290.
87. da Costa KS, Galúcio JMP, Leonardo ES, Cardoso G, Leal É, Conde G, Lameira J. Structural and evolutionary analysis of *Leishmania Alba* proteins, *Molecular and biochemical parasitology*, 217: 23-31, 2017.
doi:10.1016/j.molbiopara.2017.08.006
88. Leprohon P, Légaré D, Girard I, Papadopoulou B, Ouellette M. Modulation of *Leishmania* ABC protein gene expression through life stages and among drug-resistant parasites, *Eukaryotic cell*, 5(10): 1713–1725, 2006.
doi:10.1128/EC.00152-06

89. Niels, S. ABC Transporters in Microorganisms, Research, Innovation and Value as Targets against Resistance, *International Journal of Food Microbiology*, 136: 386-386, 2009. doi:10.1016/j.ijfoodmicro.2009.10.027
90. Hughes AL, Piontkivska H. Molecular phylogenetics of Trypanosomatidae: contrasting results from 18S rRNA and protein phylogenies. *Kinetoplastid biology and disease*, 2(1): 15, 2003. doi:10.1186/1475-9292-2-15