

PROYECTO FINAL DE INGENIERÍA

ANÁLISIS ESTRUCTURAL *in silico* DE UN ANTÍGENO DE *Trypanosoma cruzi*

Gallardo, Juan Pablo – LU1090145

Licenciatura en Biotecnología

Tutora:

Dra. Potenza, Mariana

Instituto de Investigaciones en Ingeniería Genética y Biología Molecular “Dr.
Héctor N. Torres” (INGEBI- CONICET)

Co-Tutores:

Dra. Gómez, Karina Andrea

(INGEBI - CONICET)

Dr. Gelman, Diego

Instituto de Biología y Medicina Experimental (IBYME - CONICET)

2022



UNIVERSIDAD ARGENTINA DE LA EMPRESA
FACULTAD DE INGENIERÍA Y CIENCIAS EXACTAS

Agradecimientos

En primer lugar, quiero agradecer a mi familia por haberme dado la oportunidad de estudiar. Desde mamá y papá, cada uno de mis hermanos, cuñados, y hasta mis pequeños sobrinos. Todo el apoyo fue incondicional. Gracias por la confianza y por todo el amor. No hay palabras para describir lo que significan para mí.

Luego, a mis amigos. A Camila, por nuestra convivencia. Con Romerito y Santi (y cuando aparecía Micaela) fueron mi pequeña familia en la gran ciudad. A mis amigas, Sele, Memi, Valen, Lulu, Nai. Desde la escuela (y hasta siempre). A Keili, por todo. A cada uno, y los que también tengo en mente, gracias.

Por último, y no menos importante, gracias al grupo del labo y al INGEBI. Mariana, gracias por ser mi tutora. Junto a Karina, les agradezco por la confianza y haberme dado un lugar en este grupo del cual estoy feliz de formar parte. Junto a Diego, gracias por las correcciones y ser mis guías en este trabajo. Al resto del grupo, gracias por acompañarme en este camino. También, gracias al Dr. Walter Jesús Lapadula (Doctor en Biología) y Dr. Maximiliano Juri Ayub (Doctor en Biología Molecular) del Instituto Multidisciplinario de Investigaciones Biológicas de San Luis (IMIBIO-SL-CONICET) por su ayuda fundamental en el desarrollo del trabajo.

Gracias.

All is full of love

Resumen

La infección por el parásito *Trypanosoma cruzi* es la causante de la Enfermedad de Chagas, que afecta mundialmente alrededor de 7 millones de personas. Esta patología comienza con una etapa aguda, donde los tratamientos quimioterapéuticos actuales son mayoritariamente efectivos. Sin la terapia, se desarrolla la etapa crónica de la enfermedad, donde un 40% de los pacientes presenta síntomas cardíacos o digestivos severos. Dado que en esta fase los niveles de parásitos en sangre son bajos y, por lo tanto, de difícil detección, son necesarias técnicas diagnósticas indirectas en búsqueda de anticuerpos específicos contra *T. cruzi*. En la actualidad, no existen ensayos de referencia o ‘*gold standard*’, por lo que la Organización Mundial de la Salud recomienda al menos dos estudios distintos con resultados positivos para establecer el diagnóstico. En nuestro laboratorio, a partir del ARNm de células B de pacientes con Enfermedad de Chagas crónica (ECC) se aisló un anticuerpo dirigido contra una proteína hipotética de *T. cruzi* denominada Tc323. La misma fue reconocida por plasmas de pacientes con ECC, no así de individuos con leishmaniasis o no-infectados, sugiriendo un potencial uso diagnóstico. En este trabajo, se estudió *in silico* la historia evolutiva y algunas características biológicas de Tc323. Como resultado, se observó que en la familia *Trypanosomatidae* la presencia de proteínas homólogas a Tc323 está acotada a pocas especies de los géneros *Paratrypanosoma* y *Trypanosoma*, estando ausente en *Trypanosoma brucei* y *Leishmania spp.* También se realizó el análisis filogenético de la proteína en base a un dominio de unión al cofactor Pirroloquinolina Quinona denominado PQQ_2 presente en su secuencia. En este estudio se encontró que Tc323 se agrupa en un clado de enzimas alcohol-deshidrogenasas dependientes de PQQ. Modelados *ab initio* mostraron la presencia putativa de estructuras del tipo propulsor beta, plegamiento característico de tales enzimas redox. El análisis bioinformático también predijo que Tc323 se encuentra en una región asociada a membrana, ya sea la plasmática o de alguna vesícula u organela. A partir de estos resultados, este estudio propone que Tc323 podría estar involucrada en la detoxificación del espacio intracelular de compuestos reactivos en *T. cruzi*. Por otro lado, se predijo la presencia de epítopes antigénicos a lo largo de toda la secuencia de Tc323. Su presencia exclusiva en el parásito y su función putativa sustentan la hipótesis que Tc323 es un candidato para el diagnóstico certero de la ECC, como también un nuevo blanco de drogas.

Abstract

Chagas Disease is a neglected tropical disease affecting nearly 7 million people worldwide and is caused by the infection with the parasite *Trypanosoma cruzi*. This ailment starts with an acute phase, in which current chemotherapeutic treatments are mostly effective. If not treated, this disease progresses to a chronic form in which up to 40% of patients develop cardiac or digestive disorders. As this phase exhibits low level of circulating parasites, indirect diagnostic methods are needed to detect *T. cruzi*-specific antibodies. Since no single standard reference test is available, the World Health Organization recommends at least two positive serological tests to establish a precise diagnosis. In our laboratory, from B-cell mRNA of patients with Chronic Chagas Disease (CCD), we isolated an antibody targeted to Tc323, a *T. cruzi* hypothetical protein. Tc323 was recognized by the plasmas of CCD patients, but not by those with leishmaniasis or non-infected donors, suggesting a potential use for diagnosis. In this work, we studied the Tc323 evolutionary history and some of its biologic features *in silico*. Results showed that Tc323 homologues are restricted to few species from *Paratrypanosoma* and *Trypanosoma* genera within *Trypanosomatidae* family, being absent in *Trypanosoma brucei* and *Leishmania spp.* Moreover, phylogenetic studies of the Pirroloquinoline Quinone binding domain (PQQ_2) from Tc323 showed that this protein is positioned in a clade of PQQ-dependent alcohol dehydrogenase enzymes. *Ab initio* modeling predicted the presence of beta propeller domains, a typical folding of PQQ-dependent redox enzymes. Bioinformatic analyses also showed that Tc323 is a membrane-bound protein, located in the plasma membrane, organelle, or vesicle. Overall, we propose that Tc323 might be involved in the intracellular detoxification of reactive species in *T. cruzi*. Furthermore, *in silico* analysis revealed the presence of antigen-specific B cell epitopes along the entire sequence of Tc323. Given that this protein is expressed exclusively in few trypanosome species and has a putative function, we propose Tc323 as a promising candidate for the precise diagnosis of CCD, as well a novel drug target.

Tabla de contenidos

1. Introducción	7
1.1. Generalidades.....	7
1.1.1. Clase <i>Kinetoplastea</i>	7
1.1.2. Enfermedad de Chagas. Epidemiología.....	8
1.1.3. <i>Trypanosoma cruzi</i> . Genética, ciclo de vida y transmisión	9
1.1.4. Manifestaciones clínicas y patología de la enfermedad.....	12
1.1.5. Tratamiento	12
1.1.6. Diagnóstico actual de la enfermedad	13
1.2. Antecedentes	14
1.3. Objetivos.....	19
2. Materiales y Métodos	20
2.1. Filogenia	20
2.2. Predicciones estructurales y localización subcelular	22
2.2.1. Modelados <i>Ab initio</i> de los dominios estructurales en la proteína	22
2.2.2. Análisis de hélices transmembrana y péptido señal	22
2.2.3. Análisis de localización subcelular	23
2.3. Análisis de glicosilación y presencia de epítopes de Tc323	24
2.3.1. Predicción de sitios de <i>N</i> - y <i>O</i> -glicosilación	24
2.3.2. Predicción de epítopes de célula B	24
3. Resultados	25
3.1. Filogenia	25
3.1.1. Filogenia de la proteína Tc323	25
3.1.2. Filogenia del dominio PQQ_2	26
3.2. Modelado estructural y predicción de algunas características biológicas de Tc323	33
3.2.1. Identificación de dominios estructurales en la proteína Tc323	33

3.2.2. Modelado de los dominios estructurales de Tc323	33
3.2.3. Predicción de sitios de <i>N</i> -glicosilación y <i>O</i> -glicosilación	39
3.2.4. Predicción de epítopes de célula B	40
3.3. Localización subcelular de la proteína Tc323	44
3.3.1. Análisis de presencia de regiones transmembrana.....	44
3.3.2. Caracterización de la región N-terminal de Tc323 y análisis del péptido señal	47
3.3.3. Análisis de localización subcelular	52
4. Discusión de resultados	55
5. Bibliografía.....	64

1. Introducción

1.1 Generalidades

Clase *Kinetoplastea*

El nombre de esta clasificación taxonómica se origina por la presencia, en todos los estadios de estos organismos, de la estructura de ADN mitocondrial denominada kinetoplasto (Fig. 1) (Simpson *et al.*, 2006). Cercano a él se encuentra un flagelo, exteriorizado en algunos estadios y ubicado en la región intracelular en otros. Estos organismos se reproducen por división binaria.

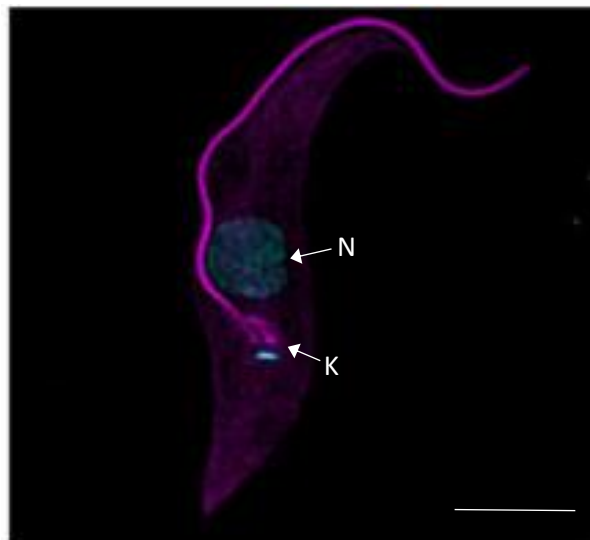


Figura 1: Microscopía de Expansión Ultraestructural (U-ExM, del inglés: Ultrastructural Expansion Microscopy) de formas procíclicas de *Trypanosoma brucei* (clase *Kinetoplastea*, orden *Trypanosomatida*). Las flechas indican la Presencia del kinetoplasto (K) y del ADN nuclear (N). En violeta intenso se evidencia el flagelo asociado al K. La línea blanca de referencia indica 25 micrones. (Adaptado de Kalichava & Ochsenreiter, 2021)

Dentro de esta clase se encuentra el orden *Trypanosomatida* conteniendo a la única familia *Trypanosomatidae* (Fig. 2) (Kaufer *et al.*, 2017). Aquí se hallan los tripanosomátidos, todos ellos parásitos obligados. Algunos cumplen un ciclo de vida monogénico (un solo huésped invertebrado), como los de los géneros *Crithidia*, *Leptomonas*, *Herpetomonas* y *Blastocrithidia*. Otros, en cambio, siguen un ciclo heterogénico (más de un huésped, generalmente siendo un vertebrado/planta alternando con un insecto vector), en particular los de los géneros *Trypanosoma*, *Leishmania* y *Paratrypanosoma*. En este último grupo se encuentran especies de interés médico debido a ser los agentes causantes de

enfermedades parasitarias tropicales (Stevens, 2008). En concreto, en América Latina se distribuye endémicamente *Trypanosoma cruzi*, en África *Trypanosoma brucei*, y mundialmente *Leishmania spp.* (agentes causales de la Enfermedad de Chagas o tripanosomiasis americana, la Enfermedad del sueño o tripanosomiasis africana, y de la Leishmaniasis, respectivamente).

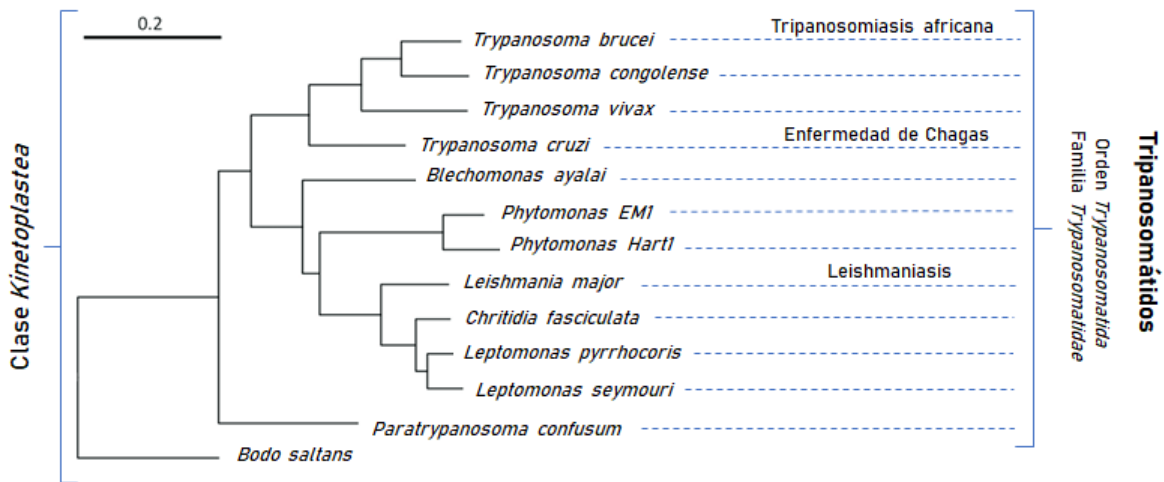


Figura 2: Árbol filogenético de algunas especies dentro de la clase *Kinetoplastea* basado en 50 secuencias nucleotídicas concatenadas de 50 genes de referencia. Se indican a los tripanosomátidos, pertenecientes al orden *Trypanosomatida*, familia *Trypanosomatidae*, y al protista flagelado *Bodo saltans*, parásito de vida libre perteneciente a orden *Bodonia*. La barra horizontal de escala representa 0,2 sustituciones por posición en la secuencia de nucleótidos. (Adaptado de Opperdoes *et al.*, 2016)

Enfermedad de Chagas. Epidemiología

La infección por el parásito *T. cruzi* se transmite naturalmente a humanos por insectos triatomíneos hematófagos (Organización Mundial de la Salud (OMS), 2021). Ésta es una enfermedad tropical endémica en 21 países de América Latina, donde afecta a alrededor de 6-7 millones de personas, y representa la principal patología causada por parásitos en el hemisferio oeste (Stanaway & Roth, 2015). Del mismo modo, es reconocida como una enfermedad desatendida debido a su prevalencia en las zonas más pobres de los países a los que afecta. A causa de las migraciones, la enfermedad está presente en áreas no endémicas de Estados Unidos, Canadá y países de Europa, África, y del Mediterráneo Oriental (Bern *et al.*,

2011; Gascon *et al.*, 2010). En nuestro país, se estima que 7,3 millones de personas estarían expuestas a la infección, con 1,6 millones de infectados y más de 300.000 afectados por cardiopatía chagásica crónica, una de las principales consecuencias de la enfermedad (Crespillo-Andújar *et al.*, 2018).

***Trypanosoma cruzi*. Genética, ciclo de vida y transmisión.**

T. cruzi es un organismo que posee una gran diversidad genética, producto del intercambio génico a través de su evolución clonal (Wang *et al.*, 2021). Los diferentes genotipos de *T. cruzi* presentan una localización geográfica diferente y se agrupan en siete unidades discretas de tipificación o DTU (Discrete Typing Unit, por sus siglas en inglés) (Abrás *et al.*, 2017; Zingales *et al.*, 2012). Para identificar a que DTU pertenece una determinada cepa o aislamiento de *T. cruzi* (proceso denominado tipificación), se analizan una serie de marcadores genéticos, moleculares e inmunológicos (Tabla 1). Si bien es un campo de estudio actual, las implicancias en la patología de la infección en relación con las distintas DTU están parcialmente caracterizadas o comprendidas.

Cepa de <i>T. cruzi</i>	Tipificación	País	Organismo de donde se aisló
Berenice	DTU II	Brasil	<i>Homo sapiens</i>
CL	DTU VI	Brasil	<i>Triatoma infestans</i>
CL Brener	DTU VI	Brasil	<i>Triatoma infestans</i>
Dm28c	DTU I	Venezuela	<i>Didelphis marsupialis</i>
G	DTU I	Brasil	<i>Opossum spp.</i>
Sylvio	DTU I	Brasil	<i>Homo sapiens</i>
Y	DTU II	Brasil	<i>Homo sapiens</i>

Tabla 1: Algunas características de las cepas de *T. cruzi* referidas en este estudio y su correspondiente DTU (Adaptado de Lima *et al.*, 2015; Zingales *et al.*, 2009)

En su ciclo de vida, *T. cruzi* alterna su residencia entre insectos vectores hematófagos como *Triatoma infestans* (vinchuca), y hospedadores mamíferos. A lo largo del ciclo aborda distintos estadios de diferenciación que involucran formas replicativas y no

replicativas (Fig. 3) (Elias, *et al.*, 2007). En la transmisión vectorial, el insecto hematófago se alimenta con la sangre del vertebrado y luego defeca sobre su piel. En el caso de que la materia fecal contenga parásitos (en forma de tripomastigotes metacíclicos), estos ingresan por heridas en la piel o mucosas a raíz de la comezón producida (Fig. 4). En el torrente sanguíneo, los tripomastigotes invaden las células del hospedador por medio de distintos mecanismos (Valente *et al.*, 2019) en los cuales, trans-sialidasas y mucinas juegan un rol fundamental (Barrias *et al.*, 2013; Leiria Campo *et al.*, 2016). Dentro de la célula hospedadora, los parásitos se diferencian en amastigotes no móviles que se dividen activamente por fisión binaria. Los cientos de parásitos por célula infectada se transforman en tripomastigotes sanguíneos móviles, promoviendo la lisis de la célula hospedadora. Estos parásitos son capaces de invadir tejidos cercanos o de diseminarse sistémicamente, iniciando nuevos ciclos de multiplicación intracelular. Cuando un insecto vector se alimenta de la sangre de un mamífero infectado, los parásitos ingresan al sistema digestivo del insecto y se diferencian a epimastigotes replicativos en el intestino medio. Luego, en el intestino grueso, se transforman nuevamente a tripomastigotes metacíclicos que pueden ser eliminados por las heces, completando el ciclo de vida (Teixeira *et al.*, 2012).

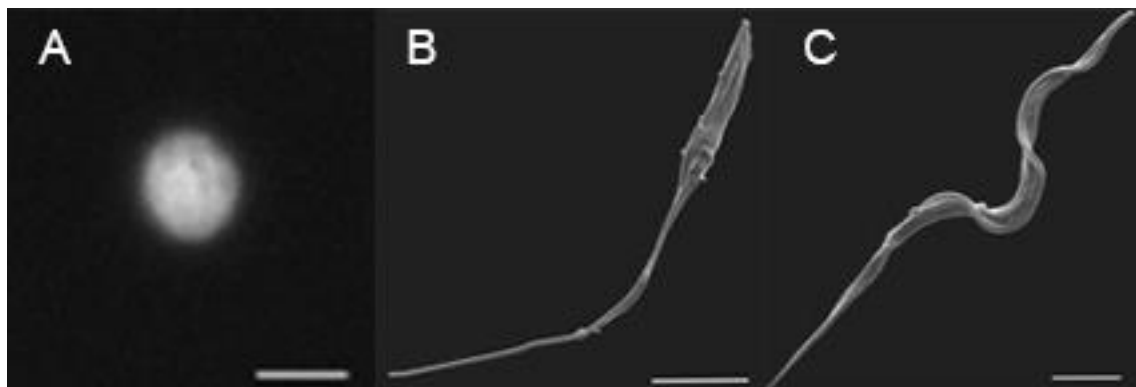


Figura 3: Distintos estadios de *T. cruzi*. (A) Microscopía de barrido electrónico ambiental (ASEM, del inglés: Atmospheric Scanning Electronic Microscopy) de amastigotes, caracterizados por su forma circular (Adaptado de Takagi *et al.*, 2022). (B) Microscopía de barrido electrónico (SEM, del inglés: Scanning Electronic Microscopy) de epimastigotes, caracterizados por su kinetoplasto inmediatamente anterior al núcleo y por un flagelo que emerge de la mitad anterior de la célula. (C) SEM de tripomastigotes, caracterizados por su kinetoplasto en el extremo posterior y por el flagelo asociado a una membrana ondulante a lo largo de casi toda la célula. (Adaptado de Gonçalves *et al.*, 2018). Las barras de escala representan: en (A), 2 micrones; en (B) 5 micrones; en (C), 2,5 micrones.

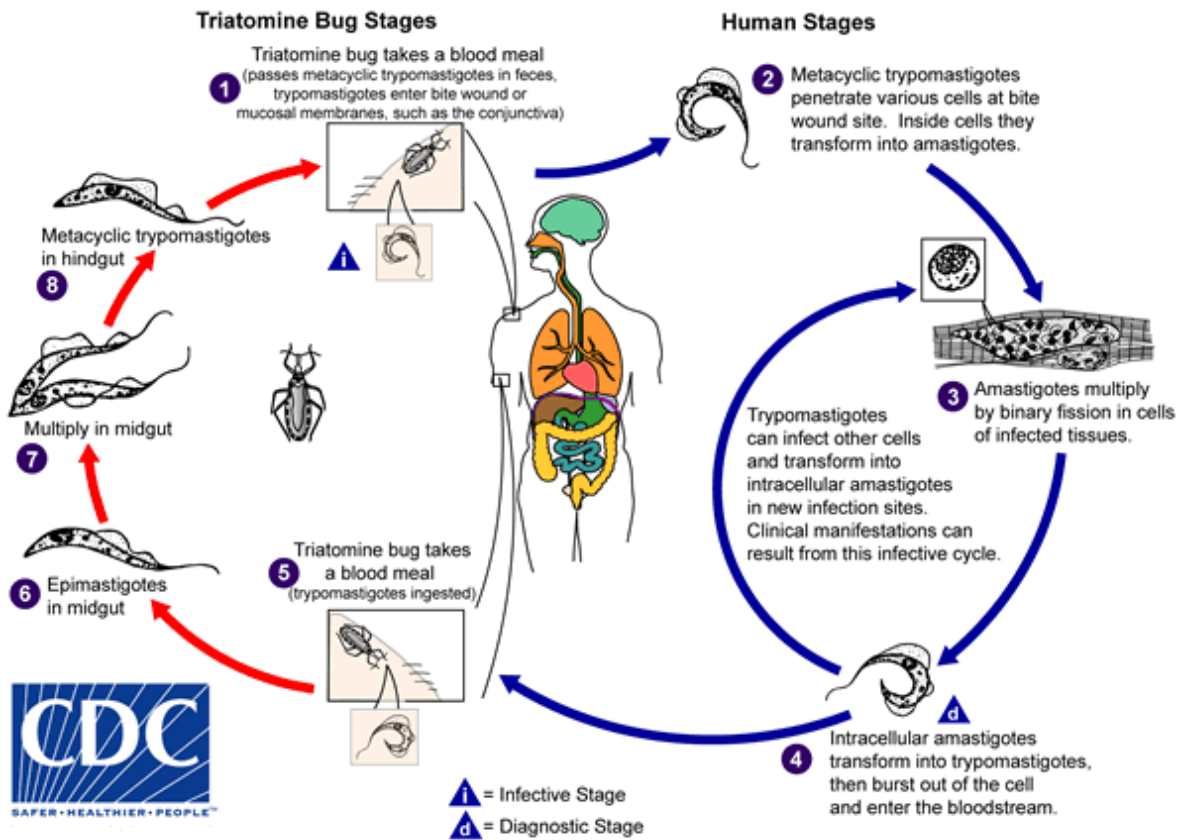


Figura 4: Ciclo de vida del parásito heterogénico *T. cruzi*. (Extraído de CDC: Centros para el Control y la Prevención de Enfermedades, 2019).

La transmisión vectorial de *T. cruzi* es la causa principal de infección en áreas endémicas (Nunes *et al.*, 2013). Estas zonas son aquellas donde coexisten el parásito y el insecto vector, como también los ambientes domiciliarios o cercanos propicios para la transmisión. No menos importante, el parásito también se transmite de una madre infectada a su hijo durante el parto (Cevallos & Hernández, 2014). En América Latina, representa un 22% de las nuevas infecciones por año, siendo un problema importante de salud pública en zonas de bajos recursos (Messenger *et al.*, 2018). En menor medida, la infección puede ocurrir a través del consumo de frutas y/o verduras contaminadas (transmisión oral) y de transfusiones de sangre y órganos provenientes de personas infectadas (Silva-dos-Santos *et al.*, 2017).

Manifestaciones clínicas y patología de la enfermedad

La Enfermedad de Chagas presenta dos fases (Rassi *et al.*, 2012). Comienza con una etapa aguda que puede durar varias semanas luego de la primo-infección, en donde se halla una alta cantidad de parásitos en sangre (parasitemia). Con respecto a las manifestaciones clínicas, algunos pacientes desarrollan síntomas similares a un estado gripal, como son la fiebre, fatiga, dolor de cabeza, náuseas, diarrea y vómitos. En pocas ocasiones se desarrolla un “chagoma”, un nódulo en la piel, o el “signo de Romana”, un edema prolongado en el párpado que no causa dolor, el cual indicaría el sitio de la inoculación del parásito.

En segundo lugar, si el paciente no ha recibido tratamiento en los comienzos de la enfermedad, esta evoluciona hacia la fase crónica (Nunes *et al.*, 2013). En esta etapa hay una pequeña e intermitente, hasta incluso nula, parasitemia. La replicación del parásito se mantiene a niveles bajos por la fuerte y específica respuesta inmune de tipo humoral, y principalmente celular. Aun así, la evasión de esta respuesta y la rápida invasión de las células reducen la eficiencia de la inmunidad (Pérez-Mazliah *et al.*, 2021). En consecuencia, los pacientes quedan con una infección difícil de detectar y tratar (Rassi *et al.*, 2012), durante toda su vida. Generalmente comienza de manera asintomática, mientras que cerca de 20 años después un 40% de las personas infectadas exhiben signos patológicos. Algunos de ellos presentan problemas cardíacos, daño del sistema nervioso periférico o alteraciones en el tracto digestivo como megaesófago y/o megacolon (Ministerio de Salud de la Nación, 2021). Se sabe que un tercio de los pacientes desarrolla la forma cardíaca, la cual es progresiva, irreversible y altamente incapacitante (Juiz *et al.*, 2019).

Tratamiento

Dado el diagnóstico de la infección por *T. cruzi*, el tratamiento actual es la quimioterapia con las drogas Nifurtimox o Benznidazol (Urbina., 2009; Perez-Molina *et al.*, 2020). Aun así, la actividad antiparasitaria se encuentra limitada en la forma crónica de la enfermedad debido principalmente a la residencia intracelular del parásito. Del mismo modo, hay limitaciones debido a un importante reporte de efectos adversos que derivan en la interrupción del tratamiento (Rassi *et al.*, 2010; Apt & Zulantay, 2011).

Diagnóstico actual de la enfermedad

El diagnóstico de la Enfermedad de Chagas está fuertemente condicionado por la fase transcurrida por el paciente (Ferrer *et al.*, 2013). En la fase aguda, la observación de tripomastigotes circulantes en sangre periférica con microscopía convencional es la elección más confiable y sensible (Murcia *et al.*, 2013). Los altos niveles de parasitemia permiten el uso de esta técnica de detección directa en estos pacientes y en los recién nacidos infectados congénitamente. En el caso de la fase crónica, debido a la reducida cantidad de parásitos en sangre, se requieren de métodos de amplificación biológica como el hemocultivo y el xenodiagnóstico (Brenner, 1962), pero estos son costosos, lentos, requiriendo condiciones especiales de bioseguridad en el laboratorio y experiencia de parte del técnico. Más importante aún, esos métodos mostraron resultados negativos en una proporción de pacientes serológicamente positivos, limitando su utilidad en el diagnóstico (Balouz *et al.*, 2017).

Siendo que los pacientes desarrollan seropositividad frente al parásito en los principios de la infección, la detección de anticuerpos anti-*T. cruzi* es una manera indirecta para establecer el diagnóstico de la enfermedad. Las técnicas inmunológicas más utilizadas al momento son la hemaglutinación indirecta (HAI), inmunofluorescencia indirecta (IFI) y el ensayo de inmunoabsorción ligado a enzimas (ELISA) (Murcia *et al.*, 2013). Los ensayos de ELISA de primera generación fueron desarrollados en un principio utilizando lisados totales de parásitos, para luego alternar al uso de fracciones antigénicas del parásito. Entre los que demuestran el mayor potencial diagnóstico se encuentran los antígenos de excreción/secreción de tripomastigote (TESA) y mucinas altamente *O*-glicosiladas de tripomastigote (Balouz *et al.*, 2017). Desafortunadamente, se ha observado reactividad cruzada de los plasmas de pacientes con leishmaniasis y/o enfermedades autoinmunes contra los lisados crudos de *T. cruzi* (Gomes *et al.*, 2009) y con antígenos de *T. cruzi* (Daltro *et al.*, 2019). Ello derivó en el uso de proteínas recombinantes que permitan el diagnóstico específico de la infección del parásito, entre las cuales se encuentran el antígeno repetitivo citoplasmático (CRA) o el antígeno repetitivo flagelar (FRA) (da Silveira *et al.*, 2001).

Sin embargo, ninguno de los ensayos para el diagnóstico muestra un 100% de especificidad ni de sensibilidad (Castro, 2009). Todavía no se cuenta con un ensayo “*gold standard*”, es decir, con un ensayo de referencia de prueba única que permita el diagnóstico

certero de la infección. La OMS recomienda el uso de al menos dos ensayos serológicos basados en distintos principios para llegar a una respuesta concluyente (OMS, 2021). En el caso de que se presente ambigüedad entre los resultados, una tercera técnica debería ser llevada a cabo. Por lo tanto, estas medidas aumentan el costo de diagnosticar a un paciente, y elevan el riesgo de que la persona posiblemente infectada pierda el interés en realizarse más estudios, principalmente en áreas endémicas. Por último, estas limitaciones en el diagnóstico retrasan el inicio de la quimioterapia, lo cual a su vez restringen la eficacia del tratamiento hacia la enfermedad (Junqueira, 2008).

1.2. Antecedentes

Uno de los objetivos de las investigaciones llevadas a cabo por los laboratorios que estudian la enfermedad es establecer un diagnóstico de alta especificidad y sensibilidad. En nuestro grupo de trabajo (Laboratorio de Biología e Inmunología de las Infecciones por Tripanosomátidos dirigido por la Dra. Karina Gómez), se investigó la respuesta inmune humoral desarrollada en pacientes con Enfermedad de Chagas y el repertorio de anticuerpos anti-*T.cruzi*. Para ello, se construyeron bibliotecas de anticuerpos recombinantes a partir de ARN total extraído de células B de médula ósea o sangre periférica de pacientes con cardiopatía chagásica crónica (Dra. Leticia Niborski, Tesis doctoral, 2012). Mediante la técnica de *phage display* (Ledsgaard *et al.*, 2018) se pudo obtener y producir anticuerpos de simple cadena del fragmento variable (scFv, single-chain fragment variable, según sus siglas en inglés) específicos frente a lisado total del parásito. De ellos, debido a su buena expresión y alta reactividad, se seleccionó y expresó el anticuerpo humano recombinante denominado scFv 6B6 (Fig. 5). La proteína reconocida por este anticuerpo presentó un peso molecular de aproximadamente 323 kDa, motivo por el cual se la denominó Tc323 a la proteína diana de scFv 6B6.

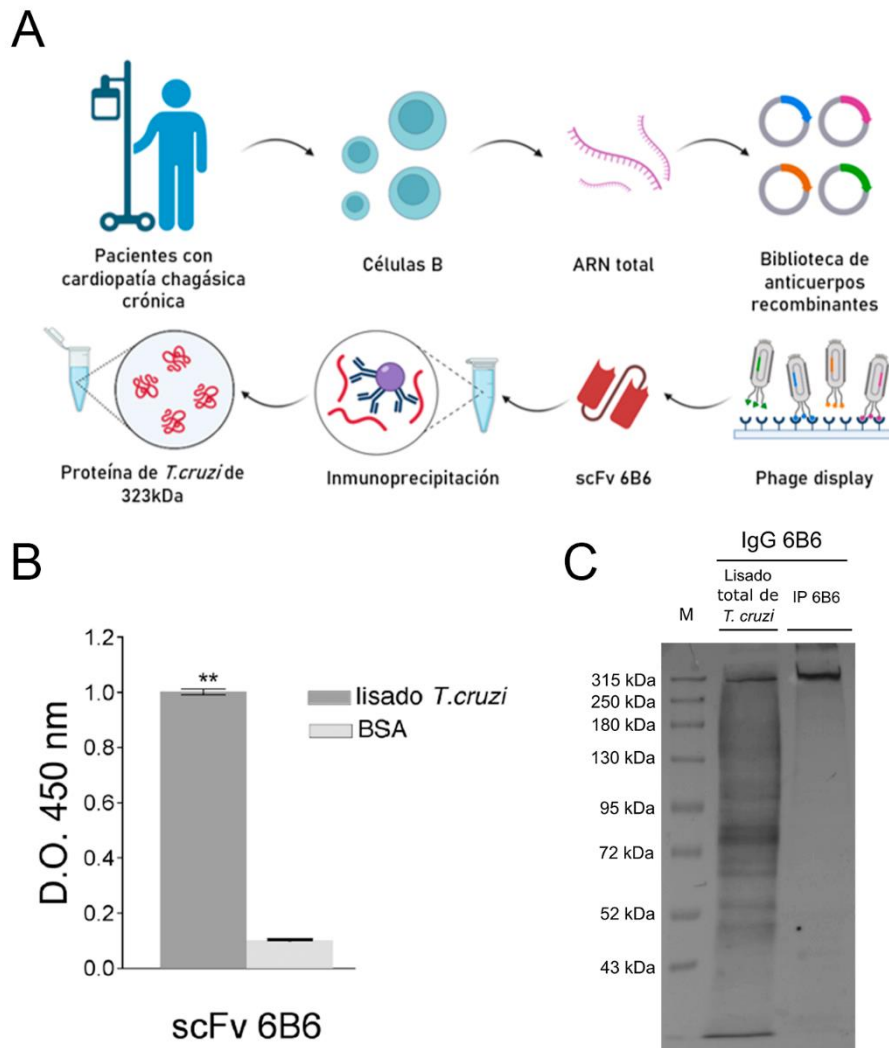


Figura 5: (A) Metodología empleada para la obtención de scFv 6B6 y su proteína diana Tc323. (B) Reactividad de scFv 6B6 utilizando lisado de epimastigotes de *T. cruzi* cepa CL Brener, seroalbúmina bovina (BSA) como control, y scFv 6B6 como anticuerpo primario. (C) Western Blot de Tc323 aislada por inmunoprecipitación, revelado por quimioluminiscencia. En la primera calle, marcador de peso molecular, indicando los pesos correspondientes a las bandas a la izquierda de la figura; en la segunda calle, lisado total de *T. cruzi*; en la tercera calle, proteína obtenida de la inmunoprecipitación del lisado con el anticuerpo quimera 6B6. En las últimas dos calles se observa una banda ubicada aproximadamente a 315 kDa, correspondiente a Tc323.

Posteriormente, mediante ensayos de inmunoprecipitación y Western Blot, se estudió la reactividad de los plasmas de pacientes con Enfermedad de Chagas Crónico (ECC), con Leishmaniasis (con y sin co-infección de *T. cruzi*), y de individuos no infectados, frente a Tc323 (Tabla 2). De las muestras procesadas, todas aquellas provenientes de pacientes infectados con *T. cruzi* fueron reactivas frente a Tc323 (33 de 33), mientras que no se observó reactividad

significativa con los restantes plasmas (1 de 18). Los resultados sugieren a Tc323 como un marcador diferencial de la infección de este parásito.

	Clasificación	Cantidad de plasmas	Positivos (%)
Con infección con <i>T. cruzi</i>	ECC asintomático	12	12 (100%)
	ECC con síntomas cardíacos	14	14 (100%)
	ECC con síntomas digestivos	4	4 (100%)
	Co-infección con <i>Leishmania spp.</i>	3	3 (100%)
Sin infección con <i>T. cruzi</i>	Infección con <i>Leishmania spp.</i>	8	0 (0%)
	No infectado	10	1 (10%)

Tabla 2: Reactividad de plasmas de distintos pacientes frente a Tc323. Se clasifica según la infección transcurrida y las manifestaciones clínicas reportadas. Se muestra el número de plasmas que detectaron a Tc323 en relación con el total.

Por otro lado, en ensayos ELISA utilizando a scFv 6B6 como anticuerpo primario, Tc323 se detectó solamente en lisado de *T. cruzi*. No se halló la presencia de la proteína en otros parásitos de interés médico como *T. brucei* y *L. mexicana*, ni en otros tripanosomátidos como *Chitridia fasciculata* y *Blastochitridia culicis*. Tampoco se reconoció la presencia de la proteína en ensayos de ELISA de extractos proteicos de lisados de *Escherichia coli*, de líneas celulares de linfoma humano *Jurkat* (Fig. 6), o de diferentes órganos de rata y ratón (datos no mostrados). Con estos resultados, se propuso la ausencia de proteínas homólogas a Tc323 en estos organismos.

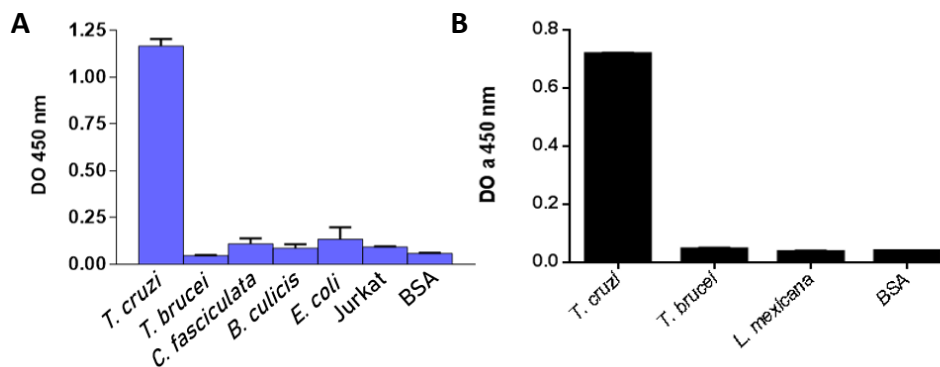


Figura 6 (página anterior): Detección de Tc323 en otros kinetoplástidos, células de mamífero y proteína no relacionada (BSA) en ensayos ELISA utilizando a scFv 6B6 como anticuerpo primario. (A) Reactividad de scFv 6B6 con lisados de *T. cruzi*, *T. brucei*, *C. fasciculata*, *B. culicis*, *E. coli*, células de linfoma humano *Jurkat* y BSA como control (Adaptado de Tesis Doctoral, Dra. Leticia Niborski, 2012). (B) Reactividad de scFv 6B6 con lisados de *T. cruzi*, *T. brucei*, *L. mexicana*, y BSA como control (Adaptado de Lic. Micaela Ossowski, Tesina de Licenciatura, 2018)

Siguiendo con el estudio particular de Tc323, ensayos preliminares utilizando microscopía de inmunofluorescencia indirecta (IFI) mostraron que la proteína se localiza mayoritariamente en todo el espacio intracelular del parásito, dado que fue necesario permeabilizar la membrana plasmática de las células para obtener señales de reconocimiento. Este hallazgo resultó similar para las principales formas evolutivas del parásito (Fig. 7).

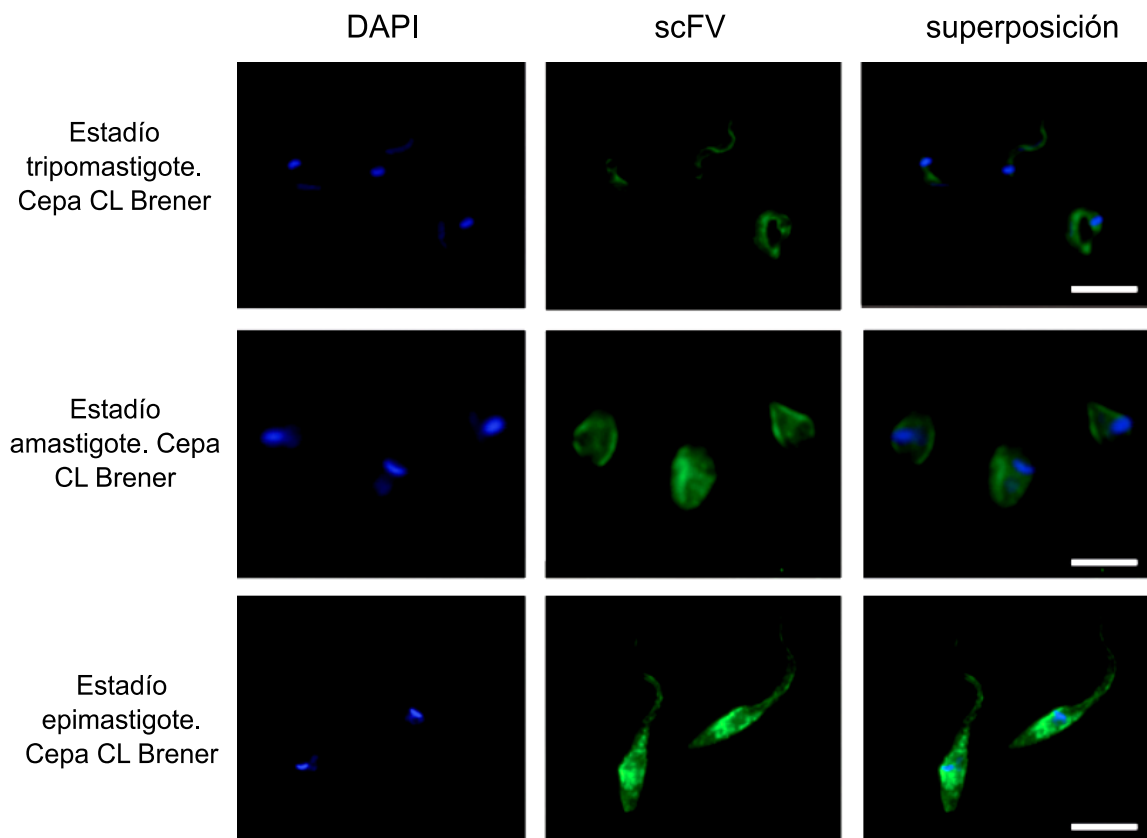


Figura 7: Localización subcelular de Tc323 en *T. cruzi*. IFI sobre tripomastigotes, amastigotes y epimastigotes de la cepa CL Brener permeabilizados con tritón x-100, utilizando a scFv 6B6 como anticuerpo primario. La línea blanca de referencia indica 2,5 micrones para el estadio amastigote y 5 micrones para los restantes estadios. (Adaptado de Tesis de Doctorado, Dra. Leticia Niborski, 2012).

Con el fin de identificar a Tc323, se inmunoprecipitó la proteína a partir de lisados proteicos de *T. cruzi* y luego se analizó por espectrometría de masas en el departamento de Post-Genómica e Interacciones Moleculares, Universidad de Dundee, Escocia, especializado en proteómica de kinetoplastidos. Como resultado, se identificó a Tc323 con el código de acceso TcCLB.504087.20 en la base de datos de kinetoplastidos TriTrypDB (tritrypdb.org), anotada como una proteína hipotética de 2.983 aminoácidos. En dicho sitio se identifica la presencia de dominios funcionales según el servidor InterPro (ebi.ac.uk/interpro), mostrando homología con la superfamilia de las quinoproteínas alcohol-deshidrogenasa (código de identificación, ID: IPR011047).

Las quinoproteínas son enzimas que utilizan como cofactor a una *o*-quinona (Flores-Encarnación *et al.*, 2004), y entre estos compuestos se encuentra la Pirroloquinolina Quinona, o PQQ (Fig. 8A). Las proteínas alcohol-deshidrogenasa dependientes de PQQ (ADH-PQQ), son enzimas redox que transfieren electrones a un citocromo de tipo c soluble, o bien a un citocromo c unido covalentemente al extremo C de la cadena polipeptídica de la proteína (Oubrie *et al.*, 2001). Este grupo de proteínas se encuentran en bacterias metilotróficas o autotróficas, y participan en vías catabólicas que les permiten crecer utilizando distintos alcoholes, como por ejemplo, etanol o metanol, como única fuente de carbono y energía. Estas proteínas aparecen en especies pertenecientes al género *Pseudomonas*, donde se describieron quinoproteínas etanol-deshidrogenasas (EDH) (Görisch, 2003). Por otro lado, en bacterias metilotróficas se hallaron quinoproteínas metanol-deshidrogenasas (MDH) (Anthony & Williams, 2002). Todas estas enzimas presentan en una región de su secuencia, una estructura tridimensional característica denominada propulsor beta, que es el sitio de unión al cofactor (Fig. 8B) (Anthony & Ghosh, 1998). Esta topología puede contar con cuatro a ocho hojas formadas por cuatro láminas beta antiparalelas, dando lugar a una arquitectura de toroide con un hueco en el centro que favorece la coordinación con el cofactor. Esta coordinación es dependiente de la presencia de iones bi o trivalentes tales como el calcio u otros elementos (Anthony, 2004).

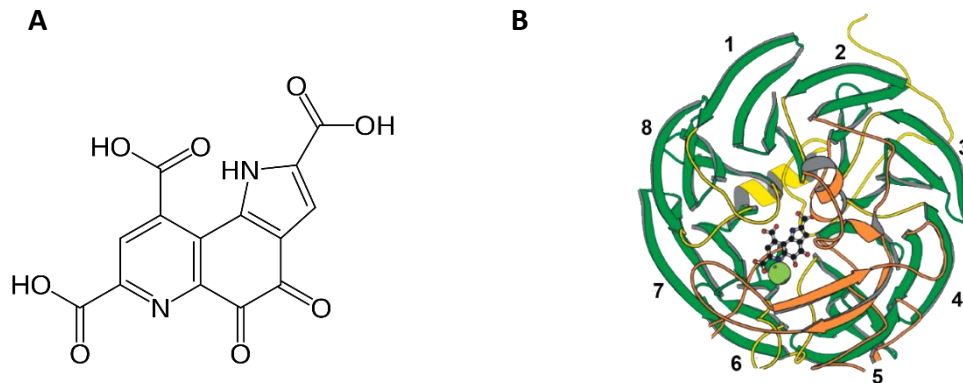


Figura 8. Características de las ADH-PQQ. (A) Estructura química del cofactor PQQ. (B) Estructura terciaria de propulsor beta, característico de las ADH-PQQ. Se observan las ocho hojas del propulsor (1-8), con cuatro láminas beta antiparalelas cada una. Se muestra la coordinación de la quinoproteína (en verde, amarillo y naranja) con su cofactor PQQ (en modelo de barras y esferas) y con el ion calcio (representado como una esfera verde). (Adaptado de Felton & Anthony, 2005; Adaptado de Anthony & Ghosh, 1998)

El conjunto de resultados que muestra la presencia exclusiva de la proteína reconocida por scFv 6B6 en *T. cruzi*, permite proponer el uso de Tc323 para el diagnóstico certero de la Enfermedad de Chagas.

Objetivos

El objetivo general de esta Tesina de Licenciatura es analizar *in silico* la proteína hipotética Tc323 de *T. cruzi* en relación con su inmunogenicidad y potencial uso como antígeno en el diagnóstico de la Enfermedad de Chagas. Para ello, se plantean los siguientes objetivos específicos:

1. Realizar un estudio evolutivo de la proteína Tc323 mediante la construcción de un árbol filogenético.
2. Predecir mediante análisis *in silico*, la estructura tridimensional, presencia de epitopes B, modificaciones postraduccionales y de actividad enzimática de diferentes dominios estructurales de la proteína Tc323 de *T. cruzi*.
3. Predecir la localización subcelular de Tc323 mediante el uso de herramientas bioinformáticas.

2. Materiales y Métodos

2.1. Filogenia

La secuencia de aminoácidos de Tc323 (ID: TcCLB.504087.20) se extrajo de la base de datos TriTrypDB (tritrypdb.org/; Aslett *et al.*, 2010) y se utilizó en todos los análisis siguientes.

El análisis filogenético de la proteína Tc323 dentro de la clase Kinetoplastea se realizó extrayendo las secuencias aminoacídicas de sus homólogas de la base de datos TriTrypDB. Con el conjunto de secuencias aminoacídicas, se eliminó la redundancia del 90% de similitud con el servidor cd-hit (weizhong-lab.ucsd.edu/cdhit-web-server) (Huang *et al.*, 2010; Li & Godzik, 2006), con el fin de eliminar proteínas repetidas. Luego, se realizó un alineamiento múltiple con el servidor on-line MAFFT (mafft.cbrc.jp/alignment/server/) (Kuraku *et al.*, 2013; Katoh *et al.*, 2018). La visualización, recorte y edición de *gaps* de la salida del programa se realizaron con el software MEGA-X (Kumar *et al.*, 2018). En este paso, se eliminan las columnas de un alineamiento múltiple en las posiciones donde se haya forzado un *gap* para la mayoría de las secuencias. Para la obtención del árbol, se calcularon parámetros de filogenia con el programa ProtTest (Darriba *et al.*, 2011), y finalmente se construyó con el servidor PhyML (atgc-montpellier.fr/phyml/) (Guindon & Gascuel, 2003) utilizando la matriz de sustitución WAG según el modelo de máxima verosimilitud. Para el soporte de este resultado se utilizó el método de Bootstrap rápido con 100 réplicas, considerando un soporte robusto si el valor de dicho parámetro para una rama cualquiera era mayor o igual a 60. Para cada una, el programa altera las secuencias originales y arma un nuevo árbol, repitiendo este paso 100 veces. Luego, se registra en cuántos de estos árboles se repite una rama en particular, asignándole el valor de soporte. De esta manera, se pone a prueba si la información contenida en la secuencia de aminoácidos sustenta los resultados del árbol original. Para la visualización y edición del árbol se utilizó el programa FigTree (tree.bio.ed.ac.uk/software/figtree/).

El estudio de la historia evolutiva de Tc323 se realizó buscando regiones conservadas dentro de su secuencia a partir de la predicción de dominios funcionales con el servidor on-line PFAM (pfam.xfam.org/) (Mistry *et al.*, 2021). Se aisló la secuencia del dominio de unión al cofactor PQQ (denominado PQQ_2; PFAM ID: PF13360), correspondiente al posible dominio

funcional de la proteína Tc323 y se utilizó como referencia (*query*) para la búsqueda de otras proteínas en la base de datos de NCBI mediante BLAST (Altschul *et al.*, 1990). En ellas, se corroboró la presencia del dominio PQQ_2 con PFAM, identificando las que contaban con uno o dos dominios de interés. Por un lado, de estas proteínas se aisló la secuencia correspondiente a la región de interés. En caso de hallar dos dominios, se aisló el par PQQ_2 desde el principio de uno de ellos hasta el fin del otro. Por el otro, para cada proteína además se registraron los dominios asociados, la organización de todos los dominios dentro de la secuencia, y la definición dada por NCBI. Esta estrategia se basó en la metodología empleada en el trabajo de Keltjens (2014), que analiza la filogenia de enzimas alcohol/metanol-deshidrogenasa dependientes de PQQ. Las secuencias correspondientes a las proteínas estudiadas en dicho trabajo fueron incluidas en este análisis. Con las secuencias de aminoácidos de los dominios PQQ_2 (individuales o de a pares, de acuerdo con la proteína), se realizó y editó el alineamiento múltiple según lo detallado en el análisis anterior. Luego de calcular los parámetros filogenéticos con ProtTest, se construyó el árbol filogenético con el software RaxML (Stamatakis, 2014) utilizando la matriz de sustitución WAG, según el modelo de máxima verosimilitud. Para el soporte de este se utilizó el método de Bootstrap rápido con 500 réplicas obteniendo los valores finales de TBE (Transfer Bootstrap Expectation) con el servidor on-line BOOSTER (booster.pasteur.fr/) (Lemoine *et al.*, 2018). El valor TBE se asigna a cada rama del árbol indicando la representatividad (entre 0 y 1) dentro de los valores de Bootstrap, donde valores cercanos a 1 indican un buen soporte de la rama. Como en el análisis anterior, se visualizaron los resultados con FigTree. Los distintos clados dentro del árbol fueron analizados por el servidor on-line CATH (cathdb.info/) (Sillitoe *et al.*, 2021), un predictor de clase, arquitectura, topología y homología de secuencias. Lo anterior permite estudiar el tipo de plegamiento que presentan las distintas proteínas en el árbol. El diagrama metodológico empleado en esta sección se muestra en la Figura 9.

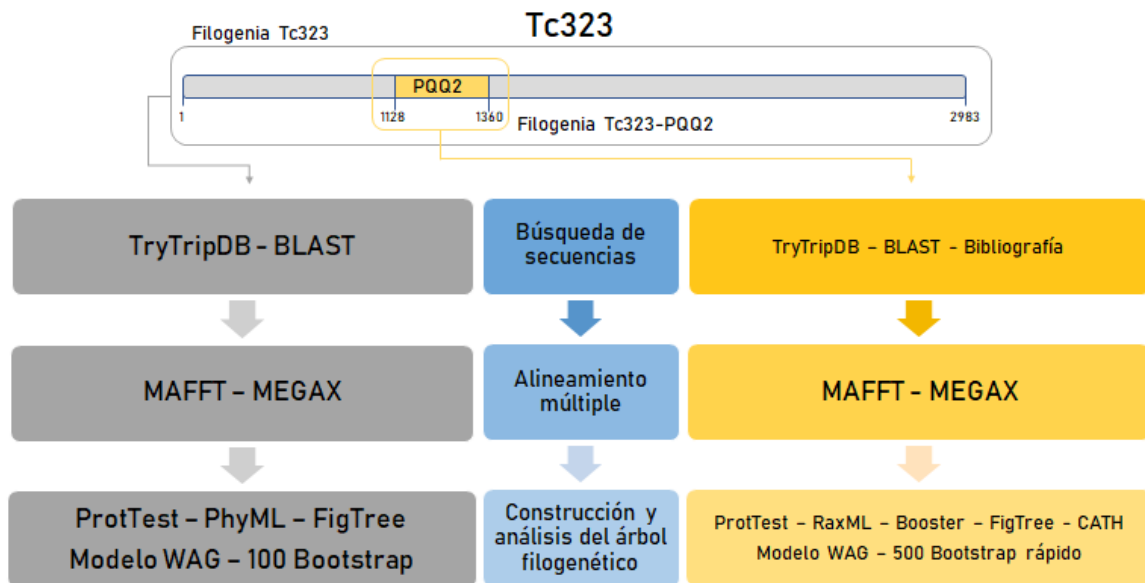


Figura 9: Diagrama de flujo de la metodología empleada para los análisis filogenéticos para la proteína Tc323 y su dominio funcional PQQ_2, predicho por PFAM.

2.2. Predicciones estructurales y localización subcelular

2.2.1. Modelados *Ab initio* de los dominios estructurales en la proteína Tc323

El modelado estructural de los distintos dominios de Tc323 se realizó con el servidor online I-TASSER (zhanglab.ccmb.med.umich.edu/I-TASSER/) (Yang *et al.*, 2015; Roy *et al.*, 2010; Zhang, 2008). Para visualizar los modelados se utilizó el software PyMOL (Schrödinger & DeLano, 2020).

2.2.2. Análisis de hélices transmembrana y péptido señal.

La presencia de hélices transmembrana en la proteína Tc323 se estudió a partir de la secuencia completa de aminoácidos con los servidores TMHMM (cbs.dtu.dk/services/TMHMM/) (Käll *et al.*, 2004, 2005) y TopCons2 (topcons.cbr.su.se/) (Tsirigos *et al.*, 2015). Del mismo modo, se utilizaron los siguientes servidores que además predicen presencia de péptido señal: MemBrain (csbio.sjtu.edu.cn/bioinf/MemBrain/) (Feng *et al.*, 2020), Phobius y PolyPhobius (phobius.sbc.su.se/) (Käll *et al.*, 2004, 2007). Asimismo, la

presencia de péptido señal en la proteína Tc323 fue analizada a partir de la región N-terminal dentro de la secuencia de aminoácidos con los servidores:

- SignalP 5.0 (cbs.dtu.dk/services/SignalP/) (Almagro Armenteros *et al.*, 2019).
- Signal-3L 3.0 (csbio.sjtu.edu.cn/bioinf/Signal-3L/) (Zhang *et al.*, 2020).
- DeepSig (deepsig.biocomp.unibo.it/) (Savojardo *et al.*, 2018).
- Predisi (predisi.de/) (Hiller *et al.*, 2004).
- SignalBlast (sigpep.services.came.sbg.ac.at/signalblast.html) (Frank & Sippl, 2008).

2.2.3. Análisis de localización subcelular

La predicción de localización subcelular de la proteína Tc323 se investigó con los servidores que se listan a continuación:

- WolfPSort (wolfpsort.hgc.jp/) (Horton *et al.*, 2007);
- LocTree3 (roslab.org/services/loctree3/) (Goldberg *et al.*, 2014); *DeepLoc* (cbs.dtu.dk/services/DeepLoc-1.0/) (Almagro Armenteros *et al.*, 2017).
- iLoc-Euk (jci-bioinfo.cn/iLoc-Euk) (Chou *et al.*, 2011).
- YLoc (abi-services.informatik.uni-tuebingen.de/yloc) (Briesemeister *et al.*, 2010a, 2010b) en las configuraciones YLoc-Animal-LowRes (4 sitios de predicción), YLoc-Animal-HighRes (9 sitios de predicción) y YLoc+-Animal (9 sitios de predicción con posibilidad de varios sitios a la vez).
- Esl-Pred2 (webs.iitd.edu.in/raghava/eslpred2) (Garg & Raghava, 2008) en sus configuraciones basadas en Composición de aminoácidos, matriz de pesos posicionales, método híbrido.
- MultiLoc2 (abi-services.informatik.uni-tuebingen.de/multiloc2/webloc.cgi) (Blum *et al.*, 2009) en sus configuraciones Animal-HighRes (4 sitios de predicción) y Animal-LowRes (9 sitios de predicción).
- mGOASVM (bioinfo.eie.polyu.edu.hk/mGoaSvmServer/mGOASVM.html) (Wan *et al.*, 2012).
- BUSCA (busca.biocomp.unibo.it/) (Savojardo *et al.*, 2018).

El servidor WolfPSort se empleó también para predecir hélices transmembrana y péptido señal (ver sección 2.2.3).

2.3. Análisis de glicosilación y presencia de epítopes de Tc323.

2.3.1. Predicción de sitios de N- y O-glicosilación.

Para la predicción de sitios de O-glicosilación y N-glicosilación se utilizaron los servidores NetOGlyc 4.0 (cbs.dtu.dk/services/NetOGlyc/) (Steentoft *et al.*, 2013) y NetNGlyc 1.0 (cbs.dtu.dk/services/NetNGlyc/) (Gupta *et al.*, 2004), respectivamente.

2.3.2. Predicción de epítopes de célula B

La predicción de epítopes B se realizó con una combinación de los servidores:

- Bepipred (cbs.dtu.dk/services/BepiPred) (Jespersen *et al.*, 2017), ABCPred (crdd.osdd.net/raghava/abcpred/) (Saha & Raghava, 2006) y SVMTriP (sysbio.unl.edu/SVMTriP/) (Yao *et al.*, 2012), los cuales predicen epítopes lineales con el uso de algoritmos de aprendizaje automático (*machine learning*). Para los dos últimos servidores, se estudiaron secuencias de 10, 12, 14, 16, 18 y 20 aminoácidos en configuraciones independientes.
- BcePred (crdd.osdd.net/raghava/bcepred/) (Saha & Raghava, 2004), el cual predice epítopes lineales según las propiedades fisicoquímicas de los aminoácidos. Los residuos se estudiaron en función de la hidrofiliidad, flexibilidad, accesibilidad, exposición superficial y polaridad en configuraciones independientes.
- DiscoTope (cbs.dtu.dk/services/DiscoTope/) (Kringelum *et al.*, 2012), el cual predice epítopes conformacionales a partir de la estructura tridimensional de la proteína. Se empleó sobre los modelados obtenidos por I-TASSER según la sección 2.2.1.

Para los programas utilizados con más de una configuración, se estableció un consenso de la predicción de aminoácidos que forman parte de epítopes. En los casos en donde se utilizaron seis configuraciones (ABCPred, SVMTriP), si en tres o más se predijo que un mismo residuo está dentro de un epítope, se consideró parte del consenso para dicho programa. Para BcePred, se incluyó en el acuerdo si el residuo aparece en dos de las cinco configuraciones usadas. Luego, se construyó el resultado general entre los seis servidores con las posiciones que aparecieron simultáneamente en tres o más programas (o sus consensos individuales, según se detalló anteriormente). Finalmente, sólo se tomaron en cuenta aquellos epítopes con una longitud mayor o igual 10 aminoácidos, respetando la longitud mínima de regiones antigénicas (Lucchese *et al.*, 2006).

3. Resultados

3.1. Filogenia

3.1.1. Filogenia de la proteína Tc323

En primer lugar, se estudió la relación evolutiva de Tc323 dentro de la familia de los tripanosomátidos. Para esto, se identificaron las proteínas homólogas a Tc323 en la base de datos de kinetoplastidos TriTrypDB, donde se encuentran anotados hasta la fecha consultada (diciembre 2021), 74 genomas de distintas especies y cepas de parásitos pertenecientes a la familia *Trypanosomatidae*. Como resultado, se obtuvieron 18 secuencias proteicas homólogas a Tc323, de las cuales 10 provienen de distintas cepas de *T. cruzi*, 4 de otras especies pertenecientes al género *Trypanosoma* (*Trypanosoma rangeli*, *Trypanosoma theileri*, *Trypanosoma grayi*, *Trypanosoma conorhini*), y 3 de la especie *Paratrypanosoma confusum*. El árbol filogenético obtenido se muestra en la Figura 10. Todos los nodos del árbol muestran valores de soporte del parámetro Bootstrap por sobre el umbral establecido de 60, indicando la robustez de las ramas.

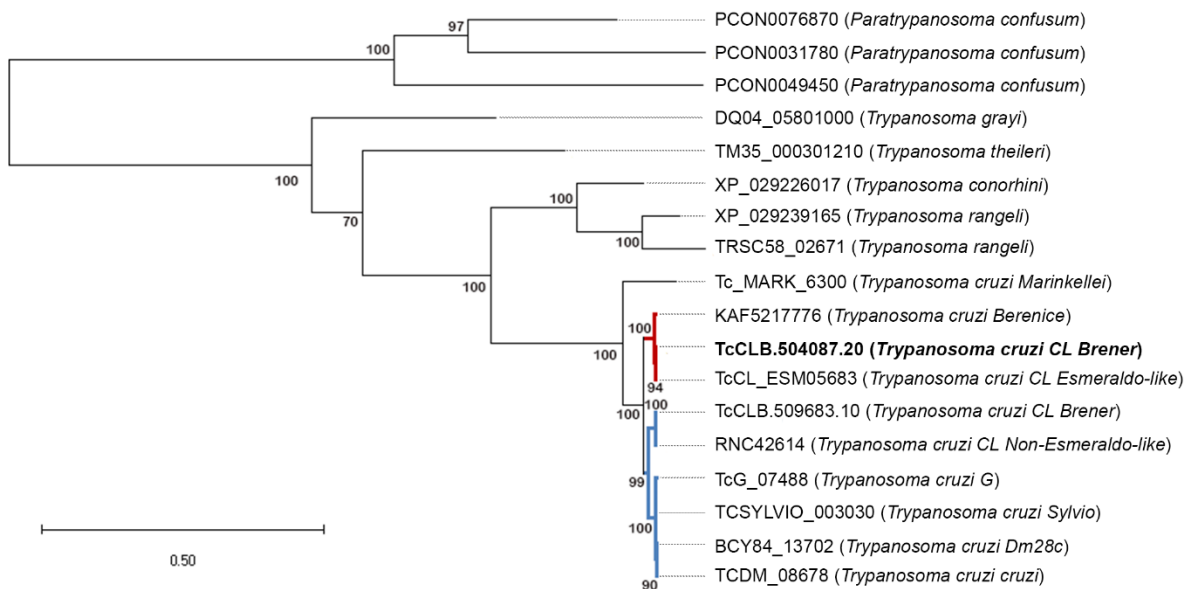


Figura 10: Árbol filogenético de la proteína Tc323 dentro del orden *Trypanosomatida*. Se indica el código de acceso para TriTrypDB o GenBank de las proteínas analizadas y la especie correspondiente. En negrita, Tc323. En azul y rojo se muestran eventos de duplicación independientes de distintas cepas de la especie *T. cruzi*. Se indican los valores de Bootstrap para cada ramificación. La barra horizontal de escala representa 0,50 sustituciones por posición en la secuencia de nucleótidos.

Dado que las proteínas homólogas a Tc323 están presentes tanto en especies del género *Trypanosoma* como en un representante de *Paratrypanosoma* (*P. confusum*), se evidencia que el ancestro común de estas proteínas se originaría en el antecesor que dio lugar a los tripanosomátidos, explicando la aparición en dichos géneros que divergieron en los orígenes de esta familia. Es interesante mencionar que no se encontraron proteínas homólogas de Tc323 en parásitos patógenos de humanos, a excepción de *T. cruzi*. Específicamente, esta proteína está ausente en patógenos como *T. brucei* y especies del género *Leishmania* y presente en sólo 5 especies del género *Trypanosoma*. Por lo tanto, si bien el ancestro de la proteína se encontraría en el antepasado de los géneros mencionados, sólo quedo remanente en ciertas especies del género *Trypanosoma*. Esto implica severos eventos de pérdida en la historia evolutiva de la proteína. Por otro lado, se observaron dos eventos seguidos de duplicación en la especie *P. confusum*, un evento de duplicación en *T. rangeli* y, sobre todo, dos eventos de duplicación independientes dentro de *T. cruzi* (mostrados en la Figura 12 con colores azul y rojo).

3.1.2. Filogenia del dominio PQQ_2

Para estudiar la historia evolutiva de Tc323 en un contexto que abarque organismos más allá de los tripanosomátidos, se realizó una búsqueda de proteínas homólogas utilizando el algoritmo BLASTp en la base de datos de NCBI. Como resultado de esta estrategia, se obtuvieron las proteínas homólogas presentes en tripanosomátidos y una sola secuencia por fuera de esta familia, correspondiente al complejo transportador dependiente de ATP en *Methanophagales spp.*, un Arquea metanogénica, con un 5% de cobertura de alineamiento. En consecuencia, se decidió utilizar regiones conservadas dentro de la secuencia de Tc323 para utilizar como referencia de búsqueda que permita encontrar un mayor número de secuencias con homología a nuestra proteína de estudio. Para esto, se utilizó el dominio PQQ_2 presente en Tc323, que es el único dominio funcional que identifica el programa PFAM, ubicado en las posiciones 1128-1360 (Fig. 11A).

La secuencia aminoacídica de la región PQQ_2 se utilizó como referencia en la búsqueda de proteínas con dominios homólogos en las bases de datos de NCBI y TriTrypDB. Como resultado, se obtuvieron 209 secuencias con uno o más dominios PQQ_2. Sin embargo,

para facilitar el alineamiento múltiple, se decidió trabajar sólo con proteínas con hasta dos dominios de unión a PQQ_2. En caso contrario, las proteínas con tres, cuatro o cinco dominios disturbarían el alineamiento debido a la diferencia en longitud de las secuencias. Como excepción, se incluyó en la selección a la proteína DQ04_05801000. Esta secuencia contiene tres dominios PQQ_2, pero está presente en *T. grayi* e incluida previamente en el árbol filogenético de Tc323, dentro del grupo de los tripanosomátidos (Fig. 10). Como resultado de esta selección, se obtuvieron 198 proteínas, de las cuales 18 provienen del orden *Trypanosomatida* (Reino *Animalia*), 44 provienen de otras especies del Reino *Animalia*, 30 del Reino *Plantae*, 15 del Reino *Fungi*, 60 del Reino *Archaea*, y 31 del Reino *Bacteria* (Fig. 11B). La distribución en función del número de dominios PQQ_2 se esquematiza en la Fig. 11C: 94 proteínas poseen un solo dominio, 103 poseen dos y 1 posee tres (44,44%, 49,76% y 5,79%, respectivamente).

Con respecto a la funcionalidad de estas 198 secuencias, se encontró que 116 tienen una función asignada en NCBI (ya sea por evidencia experimental o predicción putativa), mientras que el resto se encuentran anotadas como proteínas hipotéticas. De las proteínas con función asignada, 53 tienen relación con el cofactor PQQ y 16 están identificadas como enzimas deshidrogenasas (~27% y ~8% del total, respectivamente). Este análisis se muestra en la Tabla 3.

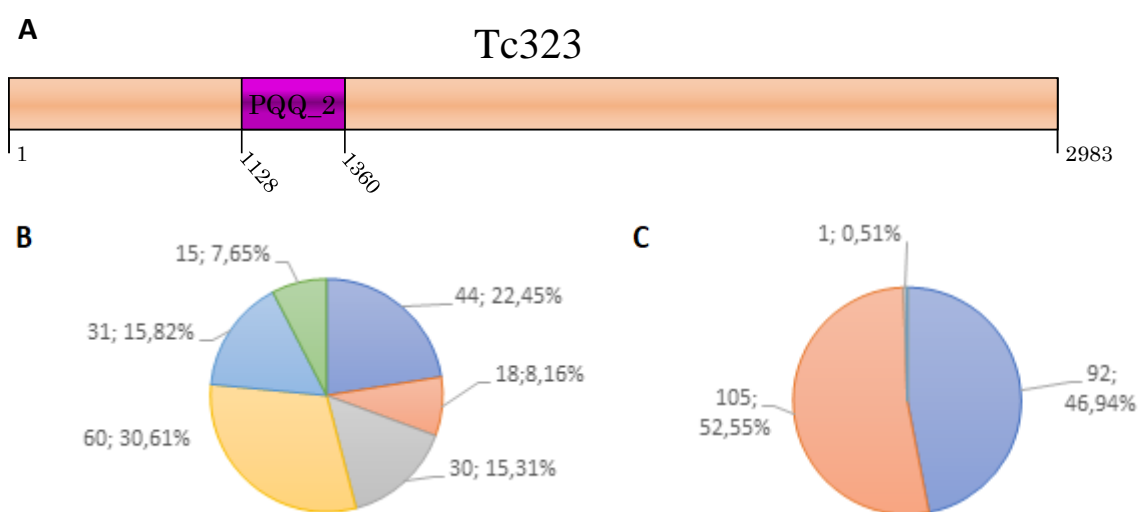


Figura 11 (Página anterior): (A) Posición del dominio funcional PQQ_2 en relación con la secuencia completa de Tc323, indicado por los números que flanquean la región violeta. (B) Representación de los distintos reinos en las proteínas seleccionadas a partir de la búsqueda por BLAST y TriTrypDB con homología al dominio PQQ_2 de Tc323. En azul, proporción de proteínas provenientes de animales distintos a los tripanosomátidos; en marrón, de tripanosomátidos; en gris, de plantas; en amarillo, de arqueas; en celeste, de bacterias; en verde, de hongos. (C) Proporción del número de dominios PQQ_2 en las secuencias obtenidas. El color azul indica un dominio, el rosa dos, el gris tres. Se muestra la frecuencia observada y el porcentaje correspondiente.

<i>Función asignada</i>	<i>Número de secuencias</i>	<i>Porcentaje del total</i>	<i>Porcentaje de las proteínas con función</i>
<i>Metabolismo de acil-CoA</i>	36	18,4%	31,1%
<i>Metabolismo del Cobalto</i>	9	24,6%	7,8%
<i>Unión a PQQ</i>	43	21,9%	37,1%
<i>Deshidrogenasas de PQQ</i>	10	5,1%	8,6%
<i>Deshidrogenasas Otros</i>	6	3,1%	5,2%
<i>Total con función</i>	116	59,2%	
<i>Sin función</i>	82	40,8%	

Tabla 3: Funcionalidad de las proteínas seleccionadas.

Con el fin de incluir un mayor número de secuencias de referencia con función asignada de ADH, se extrajeron los códigos de acceso de las proteínas analizadas en un trabajo publicado por Keltjens, 2014. Este estudio muestra el análisis filogenético de 281 proteínas de bacterias para definir los grupos de alcohol- (etanol, butanol, propilenglicol, alcohol polivinílico) y metanol-deshidrogenasas dependientes de PQQ. Del total de proteínas analizadas en dicho estudio, se pudieron recuperar 229 secuencias que se incluyeron en el análisis de este trabajo. El resto de ellas no contaban con un código de acceso válido por lo que fueron descartadas del estudio. En la Tabla 4 se resume la clasificación de estas proteínas.

	<i>Clasificación</i>	<i>Cantidad</i>	<i>Porcentaje</i>
<i>ADHs</i>	PPQ-ADH tipo 8	6	2,65%
	PQQ-ADH tipo 1	2	0,88%
	PQQ-ADH tipo 2a	6	2,65%
	PQQ-ADH tipo 2b	15	6,64%
	PQQ-ADH tipo 3	8	3,54%
	PQQ-ADH tipo 4 (quinohemoproteínas)	9	3,98%
	PQQ-ADH tipo 5	2	0,88%
	PQQ-ADH tipo 6a	5	2,21%
	PQQ-ADH tipo 6b	4	1,77%
	PQQ-ADH tipo 7	5	2,21%
	PQQ-ADH tipo 9	18	7,96%
<i>Metanol-deshidrogenasas</i>	MxaF	32	14,16%
	XoxF1	6	2,65%
	XoxF2	5	2,21%
	XoxF3	13	5,75%
	XoxF4	12	5,31%
	XoxF5	78	34,51%

Tabla 4: Clasificación de las proteínas obtenidas del trabajo publicado por Keltjens (2014). Se detalla la cantidad y el porcentaje de cada grupo respecto el total de proteínas.

En su conjunto, para el análisis filogenético de Tc323 se incluyeron un total de 421 secuencias conteniendo uno o dos dominios PQQ_2, que se analizaron con los programas ProtTest y RaxML. El árbol resultante muestra la historia evolutiva de PQQ_2, único dominio funcional de la proteína Tc323 (Fig. 12). En este análisis, se identificaron manualmente clados de proteínas de acuerdo con las definiciones en común dadas por NCBI, ya sea por poseer iguales dominios distintos a PQQ_2, o por tener una misma función asociada. Es interesante destacar que las enzimas alcohol- y metanol-deshidrogenasas (ramas de color rojo o celeste, respectivamente) de bacterias y arqueas extraídas del trabajo de Keltjens (2014) forman clados independientes, y son de utilidad como referencia de proteínas con función asignada

experimentalmente. Del mismo modo, las proteínas de bacterias, hongos y plantas con función ADH extraídas por búsqueda en bases de datos de NCBI, se agrupan en un mismo clado (ramas de color verde). El clado de proteínas provenientes de organismos del género *Trypanosoma* (rama violeta), se ubica cercano a estos dos grupos de enzimas ADH. Esto significa que los clados anteriormente mencionados se agrupan de manera monofilética, implicando que comparten un mismo ancestro en común. El agrupamiento monofilético de Tc323 y sus homólogos con los clados de ADH sugiere una función similar para este grupo de proteínas. Como excepción, las secuencias de *T. conorhini* y *P. confusum* agrupan por fuera de los clados mencionados de ADH. Los parámetros que analizan la robustez de los nodos involucrados presentan valores que soportan al árbol como evidencia de la historia evolutiva presentada (valores TBE cercanos a 1).

Luego, se analizaron las secuencias incluidas en la construcción del árbol filogenético con el programa CATH. Este programa clasifica a las distintas proteínas de los diferentes grupos según su clase (acorde al contenido de estructuras secundarias), arquitectura (similitud al nivel tridimensional sin considerar homología), topología/plegamiento (ordenamiento e interacción de las estructuras secundarias) y homología (acorde a relaciones evolutivas). Los resultados del servidor se indican en la Figura 12. Con este análisis se encontraron dos grupos de deshidrogenasas. Por un lado, un grupo de proteínas con un plegamiento semejante a la superfamilia YVTN, enzimas con repeticiones de los aminoácidos Tirosina-Valina-Treonina-Asparagina, que forman una estructura de propulsor beta. La función de estas enzimas es ajena a la reducción de alcoholes, ya que está asociada a quinoproteínas amino-deshidrogenasas (CATH Superfamily 2.130.10.10) (Jing *et al.*, 2002). Por otro lado, se encontró otro grupo, que incluye a las proteínas pertenecientes a los clados de enzimas alcohol- y metanol-deshidrogenasas (clados rojo/rosa y celeste, respectivamente, en el árbol filogenético de la Figura 12), que cuentan con un plegamiento característico de la superfamilia de las proteínas similares a las enzimas ADH-PQQ (CATH Superfamily 2.140.10.10). Si bien para el clado de secuencias de tripanosomátidos no se obtuvo resultados en CATH, este agrupa de manera monofilética con el último grupo mencionado. Este grupo se muestra representado de manera ampliada en la Figura 13. Esto permite hipotetizar que Tc323 y sus homólogas presentarían este tipo de estructura, aportando evidencia de una posible función a partir del plegamiento asociado.

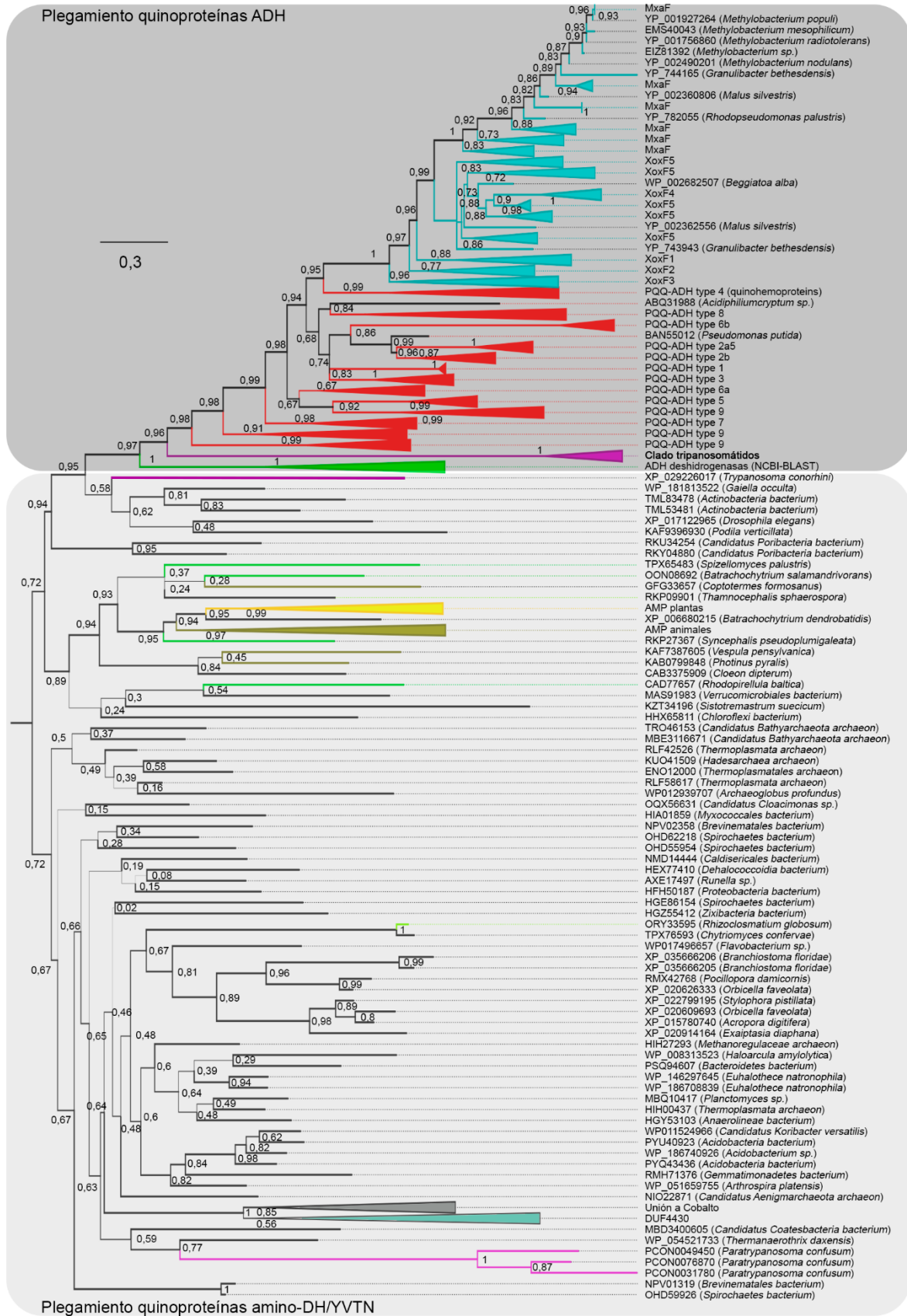


Figura 12 (Página anterior): Historia evolutiva de Tc323. Se muestra el árbol filogenético del dominio PQQ_2. Los diferentes colores en las ramas indican de donde se extrajeron las secuencias correspondientes: en violeta, proteínas de tripanosomátidos; en rojo o celeste, alcohol- o metanol-deshidrogenasas, respectivamente, obtenidas de Keltjens (2014); en verde, ADH recuperadas de NCBI por búsqueda con BLAST; en marrón o amarillo, proteínas de animales o plantas que unen AMP, respectivamente; en gris, proteínas que unen Cobalto; en turquesa: proteínas con dominio DUF4430 (dominio de función desconocida presente en proteínas hipotéticas). En negro, proteínas sin función asignada. En cada rama se muestra el valor de soporte TBE correspondiente, que representa la proporción de taxones que agruparon juntos en las distintas réplicas en el test de *Bootstrap*. Se indican los valores de TBE para cada ramificación. La barra horizontal de escala representa 0,3 sustituciones por posición en la secuencia de nucleótidos. Con fondos de distinto color se recuadran a las proteínas que presentan predicción de plegamiento característico de enzimas con repeticiones de aminoácidos YVTN/ amino-deshidrogenasa en gris claro, o a las quinoproteínas alcohol-deshidrogenasa en gris oscuro, según el servidor CATH.

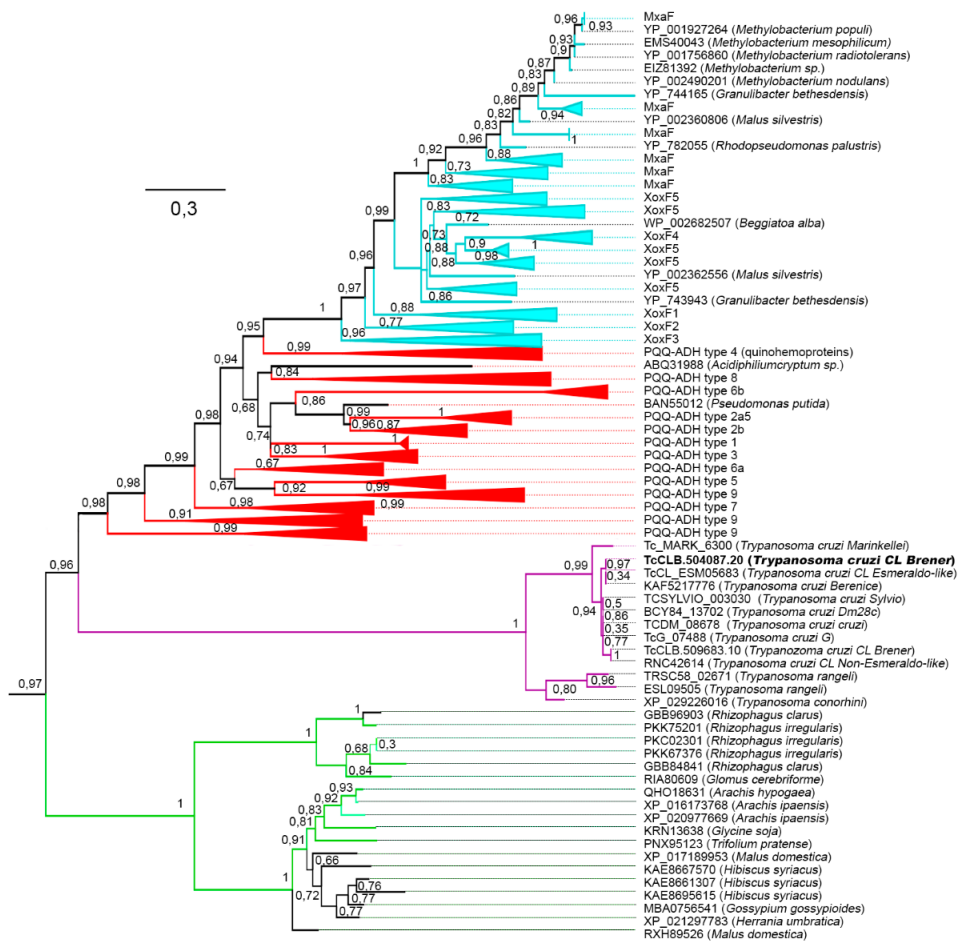


Figura 13: Filogenia del dominio PQQ_2. Región ampliada del árbol esquematizado en la Fig. 12, mostrando las proteínas con plegamiento predicho de quinoproteínas ADH.

3.2. Modelado estructural y predicción de algunas características biológicas de Tc323

3.2.1 Identificación de dominios estructurales en la proteína Tc323

En trabajos anteriores del laboratorio realizados en colaboración con el Dr. Javier Santos del Departamento de Fisiología, Biología Molecular y Celular de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, se pudo predecir la estructura secundaria y terciaria de la proteína Tc323 a partir de servidor RaptorX (raptorx.uchicago.edu). Como resultado, este servidor definió 6 dominios estructurales, los cuales fueron nombrados del 1-6. En la Figura 14 se observa la posición de cada uno de ellos en la secuencia de Tc323.

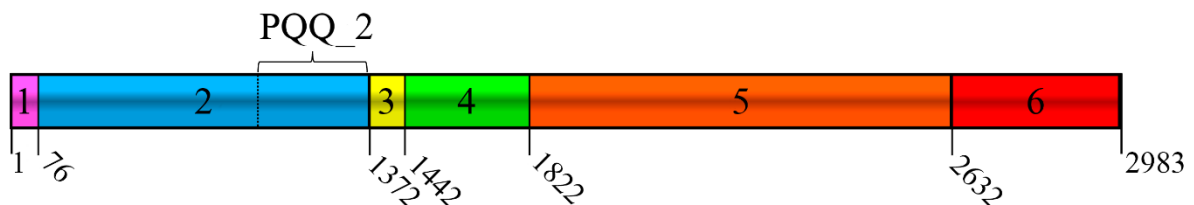


Figura 14: Ubicación de los seis dominios estructurales en la proteína Tc323 de *T. cruzi*, indicando la posición de inicio de cada uno dentro de la secuencia completa. Se indica la ubicación del dominio funcional PQQ_2 predicho por PFAM.

3.2.2. Modelado de los dominios estructurales de Tc323

Los dominios definidos en el apartado anterior se analizaron con el servidor I-TASSER para predecir su estructura tridimensional. Este programa realiza alineamientos según un reconocimiento de plegamiento en ausencia de homología con otras proteínas para predecir una estructura (denominado modelado *ab initio*). Los resultados obtenidos para los seis dominios estructurales de Tc323 se muestran en la Figura 15 y los valores asociados a esta predicción, en la Tabla 5.

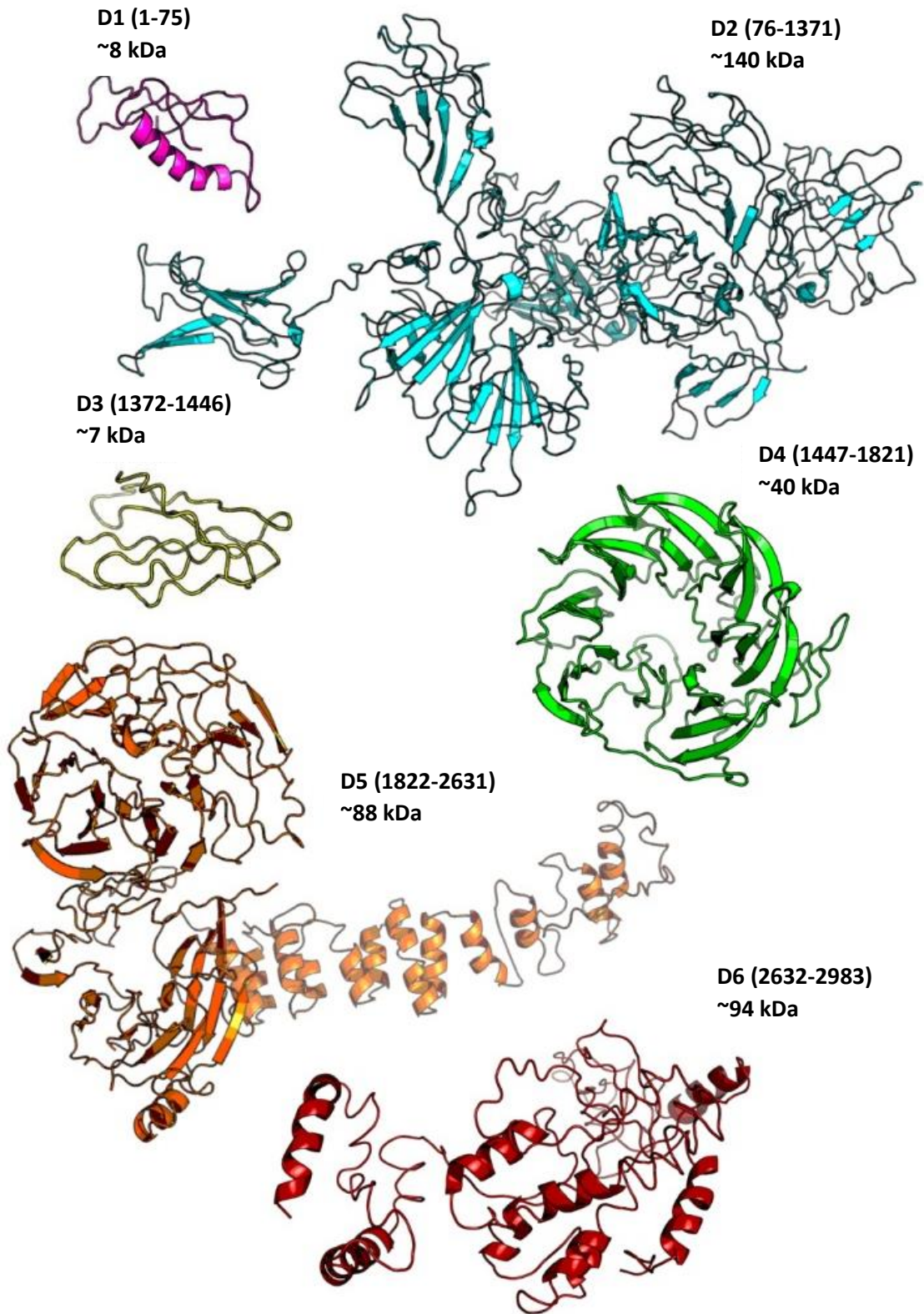


Figura 15: Dominios estructurales de Tc323, modelados por I-TASSER. Adyacente a cada modelado se indica el número de dominio, su respectiva posición dentro de la proteína y peso molecular estimado calculado en kDa en función de la composición aminoacídica.

<i>Dominio y posición en Tc323</i>	<i>Primera referencia utilizada para el modelado de Tc323</i>	<i>C-score</i>	<i>TM-Score</i>
1 (1-75)	3ekmD: diaminopimelato epimerasa (Pillai <i>et al.</i> , 2009)	3,48	0,33±11
2 (76-1371)	5ifeA: factor de <i>splicing</i> 3b subunidad 5 (Cretu <i>et al.</i> , 2016)	-1,87	0,49±0,15
3 (1372-1446)	5h5vA: proteína asociada al flagelo (Song <i>et al.</i> , 2017)	-3,37	0,34±0,11
4 (1447-1821)	3p11A: lipoproteína BamB (Kim & Paetzel, 2011)	-2,37	0,44±0,14
5 (1822-2631)	4a21A: sensor de histidina kinasa/ reguladora (Lowe <i>et al.</i> , 2012)	-2,14	0,46±0,15
6 (2632-2983)	5zaz: integrina beta-2 (Guo <i>et al.</i> , 2018)	-4,08	0,28±0,09

Tabla 5: Parámetros del modelado por I-TASSER de los dominios estructurales de Tc323. Se muestra la posición dentro de la secuencia, el molde que se consideró en primer lugar para la construcción del modelo indicando código de acceso para Protein Data Bank y los valores de los parámetros *C-score* y *TM-score* para la predicción final.

Con respecto a los parámetros mostrados en la Tabla 5, mientras el *C-score* (*confidence score*) juzga la confianza de la simulación, el *TM-score* (*template modeling score*) mide la calidad absoluta del modelo final en función de la similitud estructural independiente del tamaño de las proteínas (Zhang, 2008). El *C-score* se calcula basándose en la significancia del alineamiento con el templado y los parámetros de convergencia en las simulaciones. Este se encuentra en un rango de [-5, 2], teniendo una alta confianza de predicción a mayores valores. El *TM-score* varía entre [0, 1], en donde valores más elevados indican mejores modelados. Estadísticamente, un valor de este último menor o igual a 0,17 corresponde a una similitud aleatoria, y uno mayor a 0,5 indica un modelado con correcta topología (Zhang & Skolnick, 2004). En general, los modelos obtenidos presentan valores aceptables de *C-score* y *TM-score* dentro de los rangos establecidos para cada parámetro. Para todas las simulaciones, este último parámetro indica una topología que se cree correcta, con bajas probabilidades de que sean resultados aleatorios. Para el Dominio 1, presente en la región N-terminal de la proteína

completa, se muestra una estructura corta de alfa hélice compatible con la presencia de un péptido señal (von Heijne, 1990).

Es importante destacar que el servidor predice que tanto el Dominio 4 como parte del Dominio 5 poseen una conformación típica de propulsor beta, de siete hojas con tres láminas beta cada una. En el caso del Dominio 4, la estructura de propulsor beta abarca la totalidad de la secuencia, en cambio en el Dominio 5, sólo una parte de éste presenta esta conformación (Fig. 16). Como se mencionó anteriormente, esta estructura está presente en las enzimas ADH-PQQ, formando una especie de *hueco* en su interior que es el sitio de unión al cofactor (Takeda *et al.*, 2019).

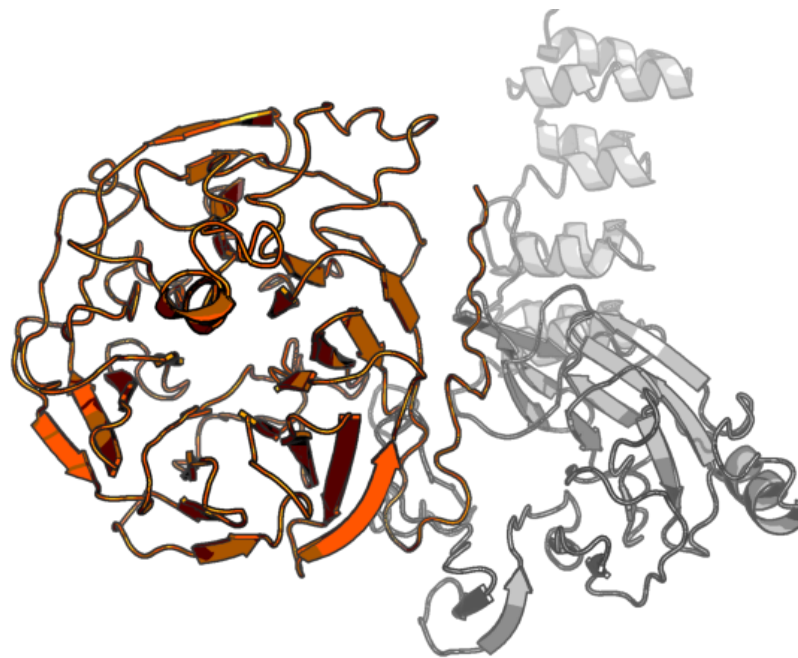


Figura 16: Dominio 5, modelado por I-TASSER. Se muestra en naranja la estructura de propulsor beta, abarcando los aminoácidos 1822-2182 de Tc323.

Dado que en el mismo servidor se puede predecir qué tipo de molécula, de existir, puede unirse a las estructuras simuladas (a través del algoritmo COFACTOR), se decidió realizar este análisis sobre el Dominio 4 y sobre la región que comprende el propulsor beta del Dominio 5 (aminoácidos 1822-2182). Como resultado en el caso del Dominio 4, el algoritmo predice que el cofactor PQQ se une al sitio interno de la estructura de propulsor beta, con un alto valor de confianza del parámetro *C-score* (0,19). I-TASSER otorga una serie de

modelos de interacción con el cofactor, tanto de moléculas como de la orientación que tome cada una de ellas. Para el Dominio 4, el cofactor PQQ mostró tener el mayor valor de *C-score*. En la Figura 17 se esquematiza la interacción más representativa del cofactor con los aminoácidos del Dominio 4 en el centro de su estructura tridimensional. Debido a limitaciones del servidor online, no es posible identificar con exactitud los aminoácidos que interaccionan con dicha molécula. De acuerdo con la orientación del cofactor dentro del sitio de unión, son distintos los residuos con los que establece uniones del tipo puente de hidrógeno.

También, este servidor predice una posible unión del Dominio 4 al ion calcio. Cabe recordar que los iones bi o trivalente en las ADH-PQQ participan en la coordinación de las fuerzas no covalentes para la unión con su cofactor. El servidor I-TASSER no fue útil para evaluar la interacción conjunta de PQQ y el calcio dentro de la estructura del Dominio 4. Esto se debe a que, para cualquier proteína en estudio, la plataforma sólo puede analizar o la coordinación con algún ion o la interacción con el cofactor, pero no ambas simultáneamente.

Por el contrario, el programa no predice unión a cofactor para el Dominio 5, a pesar de su similitud estructural.

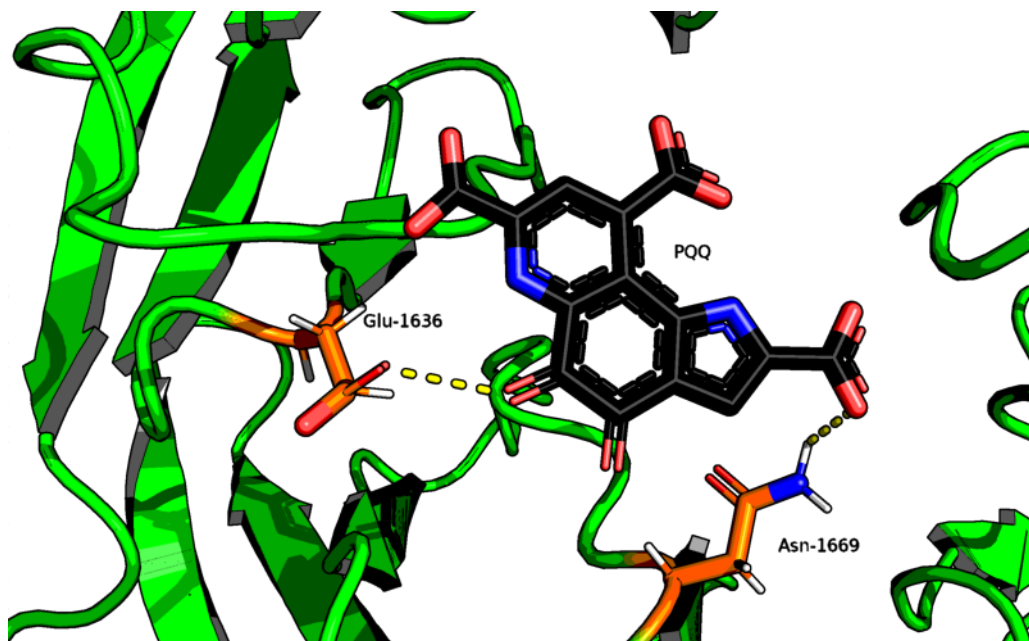


Figura 17: Posible interacción del cofactor PQQ (color negro) con los aminoácidos Glu-1636 y Asn-1669 (en naranja) del Dominio 4 de Tc323 (en verde) (I-TASSER; PyMOL).

Llamativamente, el programa I-TASSER predice dos estructuras con similitud estructural a propulsor beta presentes en los Dominios 4 y 5, pero no identifica otro dominio

funcional. Esto difiere de la predicción del programa PFAM, que identifica el dominio PQQ_2 en la región C-terminal del Dominio 2 de Tc323, posicionado en los residuos 1128-1360. Por este motivo, se analizó la estructura tridimensional de esta región de aminoácidos del Dominio 2 con I-TASSER. Como resultado, se encontró que el modelado muestra semejanzas con la estructura de propulsor beta presentes en ADH-PQQ (Fig. 18). Sin embargo, también se observa que es una estructura parcial, ya que la misma posee tres de las siete hojas con tres laminas beta cada una, características del propulsor beta.



Figura 18: Modelado estructural del dominio funcional PQQ_2

La discrepancia entre el modelo obtenido y la estructura completa de propulsor beta hallado en enzimas ADH-PQQ podría deberse a que, el dominio funcional de unión a este cofactor abarque una región más extensa que la predicha por PFAM. Por otro lado, el servidor no predijo una posible unión al cofactor PQQ, lo que contrasta con la función hipotética predicha por PFAM. Aun así, I-TASSER sí predice la unión al ion calcio, otra de las características de estos sitios activos dependientes de PQQ. Además, y aunque con bajo score, esta estructura se modela a partir del tercer lugar con deshidrogenasas dependientes de quinonas.

Luego, se realizó el alineamiento y superposición de los modelados correspondientes a la región del dominio PQQ_2 y la estructura tridimensional del Dominio 2 (Fig. 19). En la superposición, si bien ambas estructuras fueron predichas por I-TASSER, éstas difieren groseramente. Estos resultados indican que el modelado del posible dominio funcional no pudo ser detectado en el contexto de los dominios estructurales, evidenciando las limitaciones de los servicios de predicción. La misma secuencia es modelada de distinta manera en el caso de que se encuentre aislada o forme parte de una estructura de mayor longitud. Este puede ser el motivo por el cual el programa no predice la unión al cofactor PQQ.

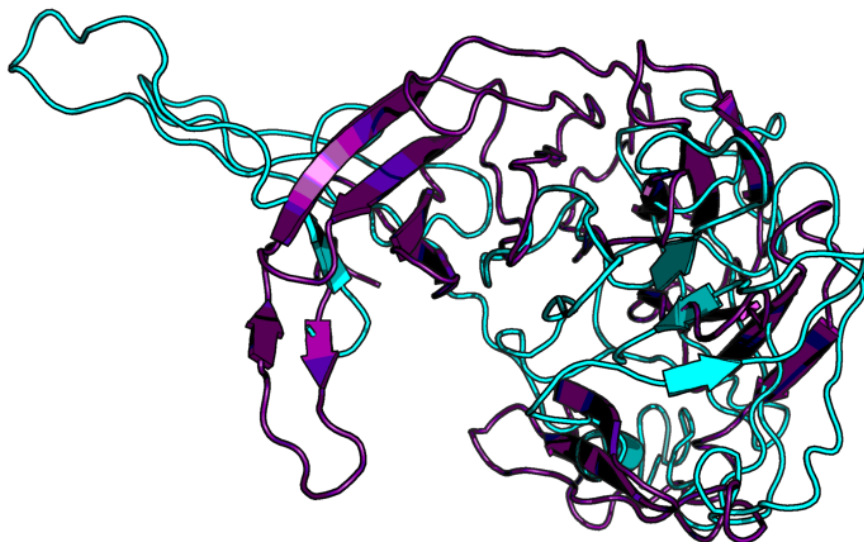


Figura 19: Superposición del modelado del Dominio 2 de Tc323 (en celeste) con el dominio funcional PQQ2 (en violeta), ambos predichos por I-TASSER

Finalmente, para el Dominio 3 se predice la unión con el ion fosfato. Para el resto de los dominios no se detectó ningún posible cofactor.

3.2.3. Predicción de sitios de *N*-glicosilación y *O*-glicosilación

Dado que las distintas modificaciones postraduccionales de una proteína tienen potencial implicancia en sus propiedades antigénicas e inmunogénicas (Lisowska, 2002;

Jefferis, 2016), se realizó un análisis bioinformático de los distintos dominios con el fin de predecir la presencia de residuos glicosilados. Empleando los servidores descritos en la sección 2.3.1., se predijeron distintos sitios de *O*-glicosilación y *N*-glicosilación a lo largo de Tc323 (Fig. 20).

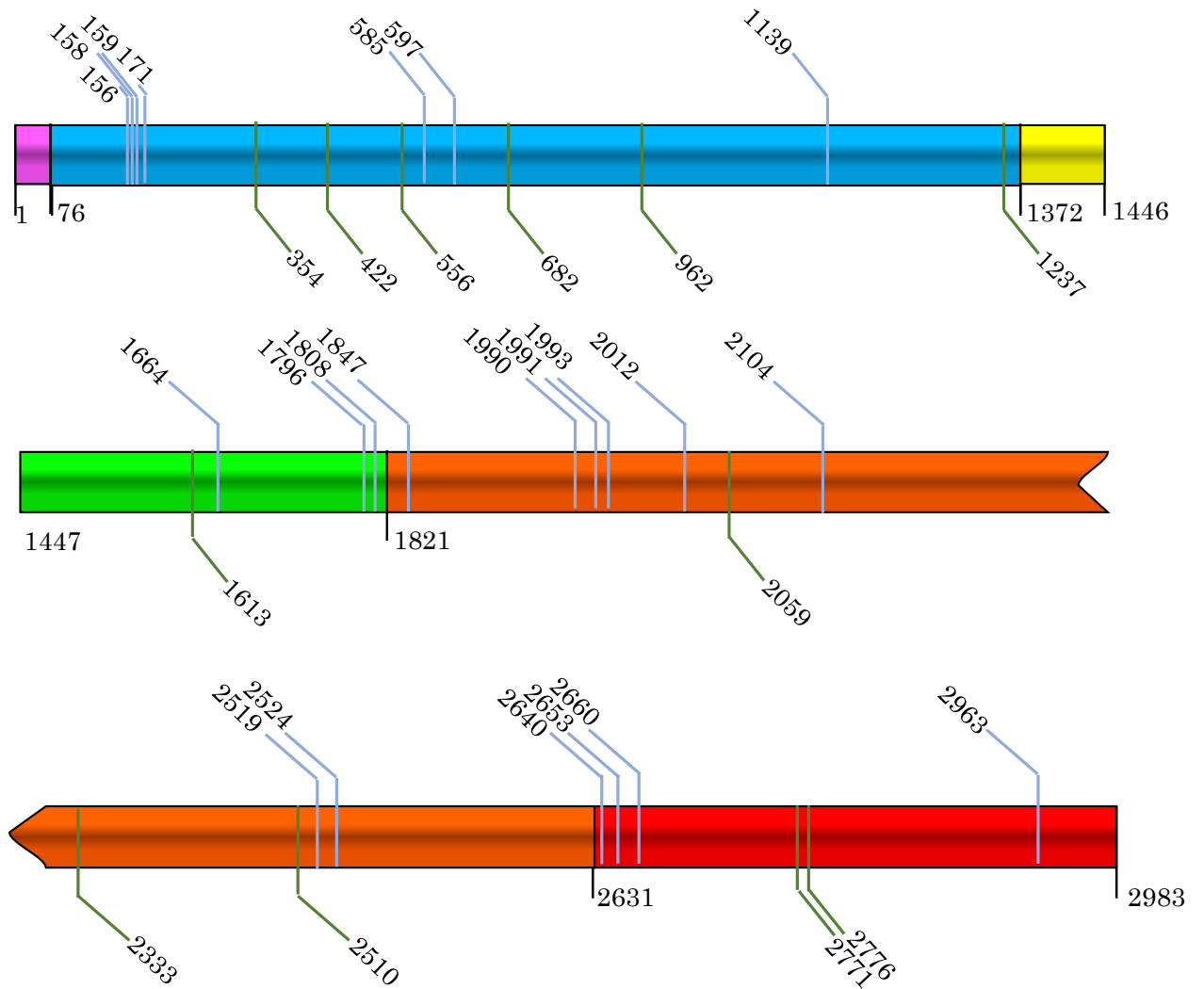


Figura 20: Sitios de *O*-glicosilación (línea celeste) y *N*-glicosilación (línea verde) predichos en los dominios de Tc323 (NetOGlyc 4.0 y NetNGlyc 1.0, respectivamente). En rosa, Dominio 1; en celeste, Dominio 2; en amarillo, Dominio 3; en verde, Dominio 4; en naranja, Dominio 5; en rojo, Dominio 6

3.2.4. Predicción de epítopes de célula B

Dado que se propone a Tc323 o alguno de sus dominios como antígeno de diagnóstico para la Enfermedad de Chagas, es sumamente relevante caracterizar, tanto *in silico* como experimentalmente, la presencia de epítopes de células B en la molécula de estudio. La presencia de epítopes B se estudió con los servidores ABCPred, BCEPred, Bepipred, Discotope, y SVMTriP. Si bien estos servidores utilizan diferentes algoritmos de predicción, se encontró que algunos epítopes fueron identificados por dos o más programas, por lo que fueron los únicos que se tomaron en cuenta para el consenso general. De esta manera, se identificó la presencia de posibles determinantes antigénicos lineales en todos los dominios (Tabla 6), representados en los respectivos modelados en color azul (Fig. 21). La predicción de epítopes conformacionales se analizó con el servidor Discotope, y se incluyó en el armado del consenso. Cabe destacar que las predicciones de los algoritmos que resultaron en epítopes de menos de cinco aminoácidos fueron descartadas por su baja probabilidad de existir como tales.

<i>Dominio y posición en Tc323</i>	<i>Número de regiones antigénicas</i>	<i>Porcentaje de aminoácidos dentro de epítopes (residuos antigénicos sobre residuos totales)</i>
1 (1-75)	2	15/75: 20,00%
2 (76-1371)	17	259/1295: 20,00%
3 (1372-1446)	2	27/74: 36,40%
4 (1447-1821)	6	109/374: 29,14%
5 (1822-2631)	7	178/809: 22,00%
6 (2632-2983)	7	212/352: 60,40%

Tabla 6: Regiones antigénicas de Tc323. Para cada dominio estructural de Tc323 se muestra el número de regiones antigénicas predichas, y la cantidad de posiciones dentro de epítopes respecto el número de residuos totales por dominio.

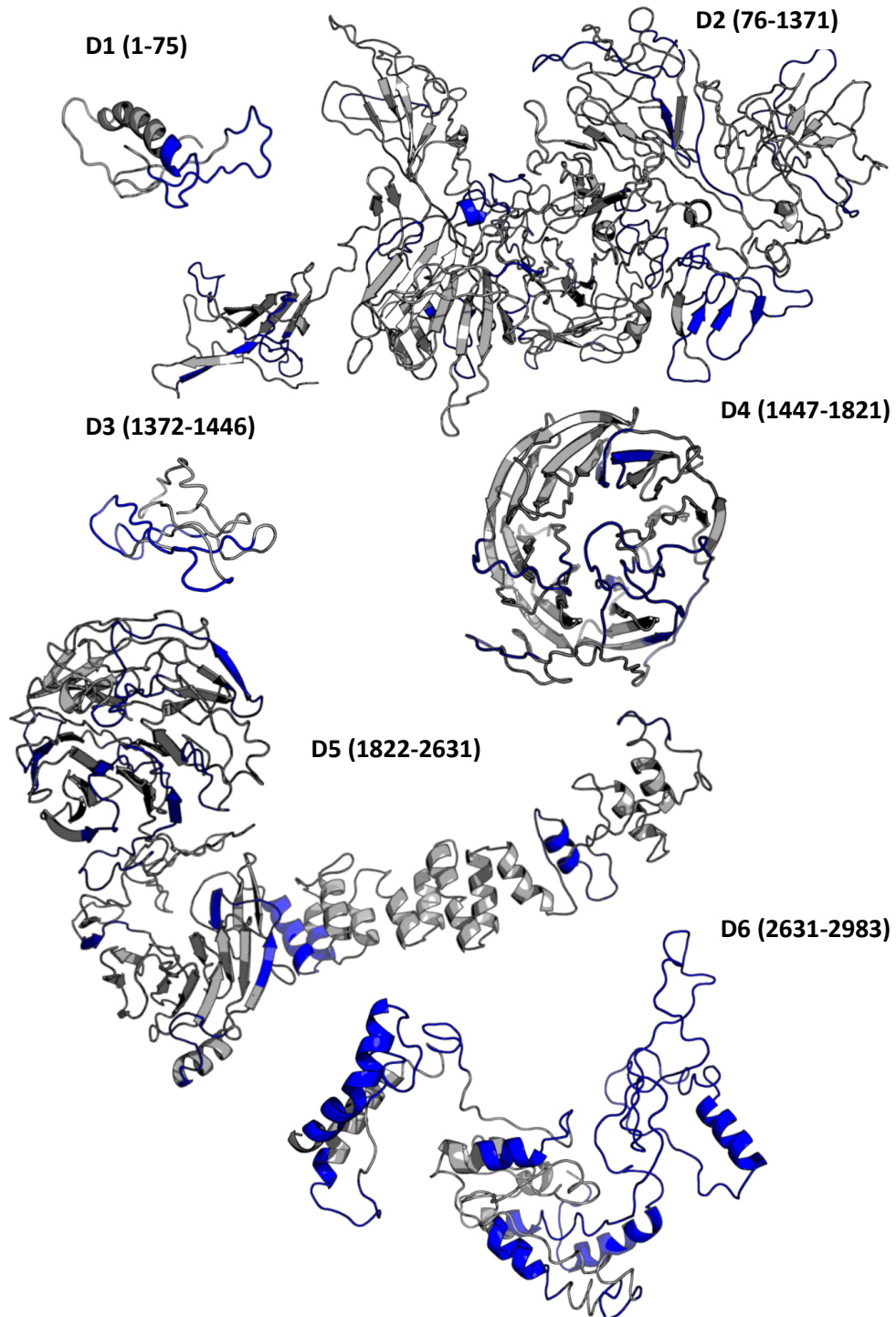


Figura 21: Regiones antigénicas consenso (indicadas en azul) en el contexto de las estructuras predichas de los seis dominios estructurales de Tc323.

Con respecto a los epítopes predichos para el Dominio 4, se observa que gran parte de las regiones antigénicas (4 de 6) se encuentran fuera de las láminas plegadas beta que conforman la estructura del propulsor, mientras que una baja proporción de las regiones se encuentran en el centro de la proteína, posible sitio de unión al cofactor. Específicamente, varios de los epítopes que están externos al “hueco” del propulsor se ubicarían en zonas sin estructura secundaria, similar a lo que se observa en el modelado del Dominio 6. En este último, se predice una región antigénica de 135 residuos en el extremo C-terminal que corresponde a un 38,35% de este dominio estructural. Es importante tener en cuenta que estas regiones de la cadena peptídica sin plegamiento pueden existir en la estructura nativa de la proteína o pueden ser el resultado de que el servidor no haya encontrado una estructura tridimensional de referencia para realizar la simulación. Por último, hay que considerar que en la conformación completa de Tc323 se encontrarían epítopes en regiones internas de la proteína que serían imposibles de ser reconocidas por los actores del sistema inmune humoral, por lo que estos deberían ser desestimados.

3.3. Localización subcelular de la proteína Tc323

3.3.1. Análisis de presencia de regiones transmembrana

Con el objetivo de estudiar con mayor profundidad la presencia de estructuras secundarias que podrían estar relacionadas con la función de Tc323, se decidió analizar la secuencia completa de aminoácidos de la proteína en búsqueda de hélices transmembrana (HTM). Para esto se utilizaron los servidores TMHMM, MemBrain, Phobius/PolyPhobius, TopCons2 y WolfPSort. Con el programa TMHMM, se predicen dos HTM cercanas a las regiones N- y C-terminal de la proteína, respectivamente (Fig. 22). Este servidor predice que, sumadas las dos hélices, un total de 49 aminoácidos de Tc323 constituyen HTM, indicando que si este número es mayor a 18 es probable que la proteína estudiada esté anclada a membrana o que contenga un péptido señal (PS). Además, 22 de dichos residuos se encuentran entre los primeros 60 aminoácidos de la proteína, por lo que sería indicio de un posible PS. Por último, el servidor señala que es muy probable que la región N-terminal se encuentre orientada al citoplasma. En resumen, este servidor predice que Tc323 está anclada a membrana, que presenta con un PS y que una gran proporción de la proteína se encuentra en una región no citoplasmática.

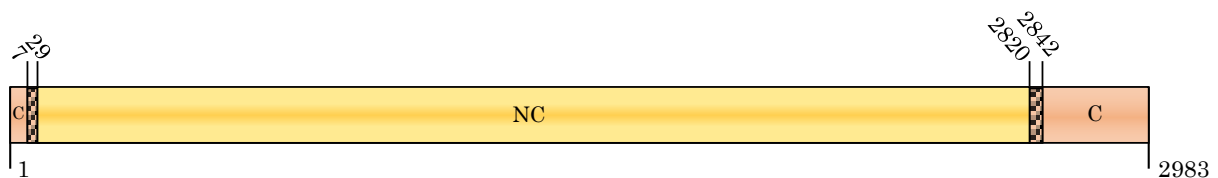


Figura 22: Predicción de hélices transmembrana por TMHMM. Las HTM se señalan con franjas diagonales, detallando las posiciones iniciales y finales para dichas regiones. C y NC: regiones citoplasmática y no citoplasmática, respectivamente.

El servidor MemBrain, a diferencia de la mayoría de las plataformas existentes para el estudio de HTM, permite predecir dichas estructuras de longitud irregular (menores a 15 aminoácidos o mayores a 35 aminoácidos). Por lo tanto, detectaría HTM de menor o mayor tamaño a lo regular a diferencia del resto de los servidores utilizados. En cuanto al estudio de Tc323, este programa predice tres HTM, pero no informa la presencia de PS (Fig. 23). En contraste con el servidor TMHMM, MemBrain predice que una mayor proporción de la proteína se encontraría en una región del citoplasma.

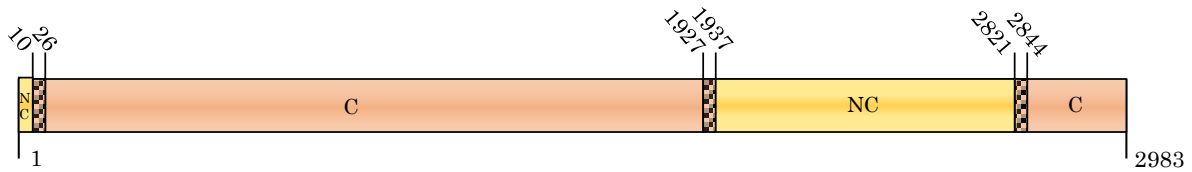


Figura 23: Predicción de hélices transmembrana por MemBrain. Las HTM y las regiones celulares se indican como en la Figura 22.

Luego, se utilizaron servidores que trabajan con modelos ocultos de Markov, como Phobius y PolyPhobius. Este último, además, recurre al estudio de la homología de la proteína como soporte de los resultados. Ambos programas predicen la presencia de un PS (con sitio de clivaje entre las posiciones 38 y 39) y al menos una HTM en la región C-terminal (Figura 24). Se predice que una mayor proporción de la proteína se encuentra en una región no citoplasmática.

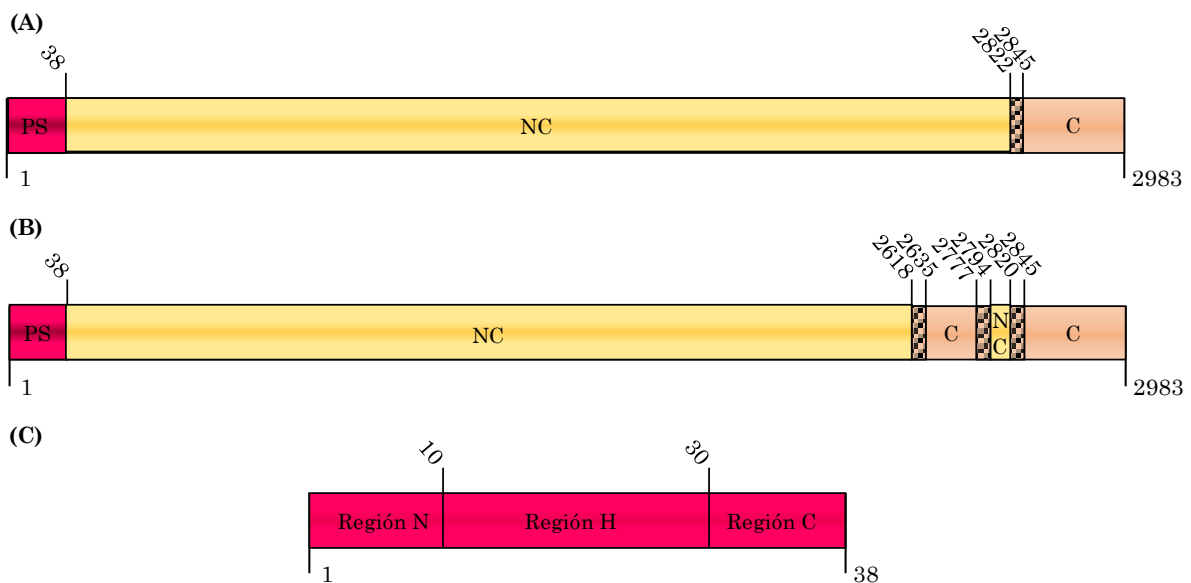


Figura 24: Predicción de hélices transmembrana y PS para la proteína Tc323 mediante Phobius (A) y Polyphobius (B). Las HTM y las regiones celulares se indican como en la Figura 22. PS: Péptido señal (C). Análisis del PS según ambos servidores, indicando regiones N, H y C con sus respectivas posiciones de inicio y fin.

Por otro lado, el servidor TopCons2 también indica la presencia de PS y de tres HTM cercanas a la región C-terminal (Fig. 25). Además, se predice que la mayor parte de la

proteína presenta ubicación no citoplasmática. Estos resultados se obtienen luego de realizar el consenso de predicciones entre los servidores OCTOPUS, Philius, PolyPhobius, SCAMPI, SPOCTOPUS (Fig. 26).



Figura 25: Predicción de hélices TM y PS por TopCons2. Las HTM, PS y regiones celulares se indican como en la Figura 22.

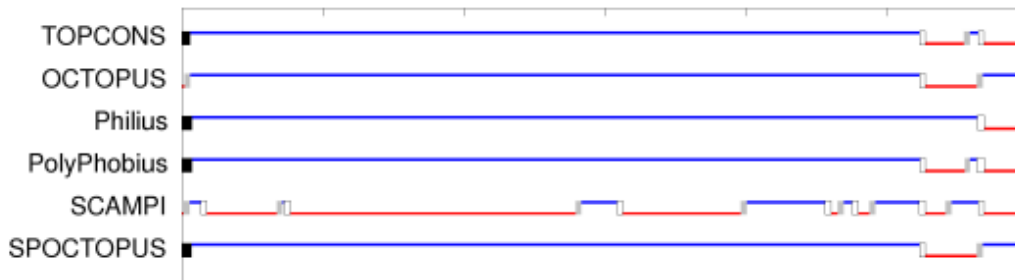


Figura 26: Consenso entre servidores dentro de TopCons2. En negro, PS; en azul, región de la proteína fuera del citoplasma; en rojo, región de la proteína dentro del citoplasma; en gris, HTM orientada desde adentro hacia afuera del citoplasma; en blanco, HTM orientada desde afuera hacia adentro del citoplasma.

Finalmente, si bien WolfPSort es un predictor de localización subcelular, sus resultados se presentan en función de la presencia de señales de direccionamiento dentro de la secuencia de la proteína como es el PS. Por lo tanto, entre sus resultados detalla el análisis de topología y PS. Para Tc323, se predice la existencia de tanto PS como de 7 HTM (Fig. 27).

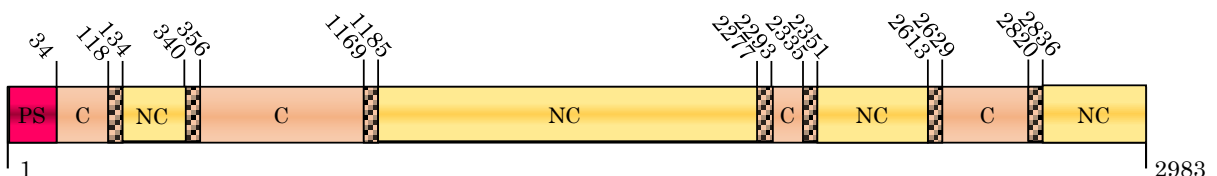


Figura 27: Predicción de WolfPSort para hélices TM y PS. Las HTM, PS y regiones celulares se indican como en la Figura 22.

En la Figura 28 se muestra un resumen de los resultados de predicción de HTM.

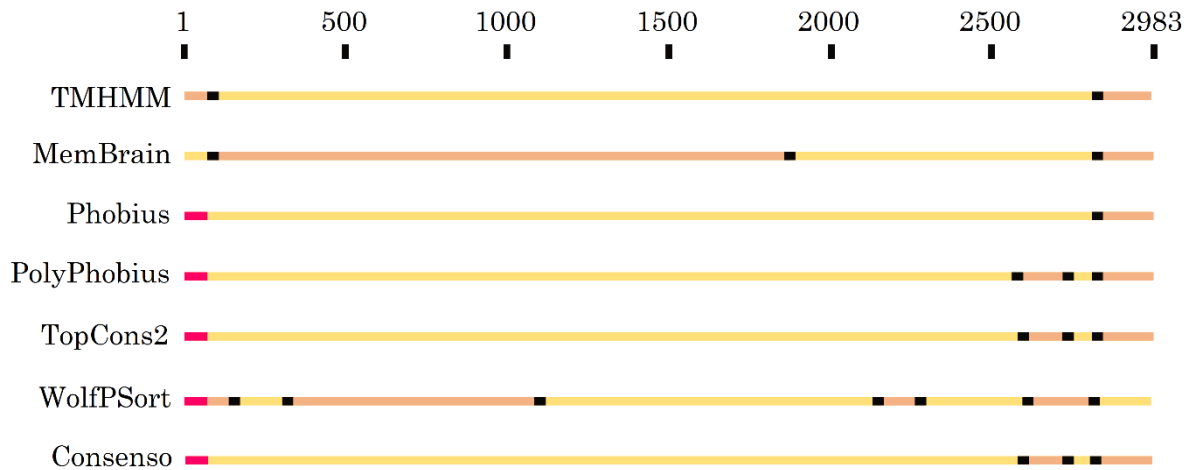


Figura 28: Resumen de los resultados obtenidos en los distintos servidores de predicción de HTM en Tc323. Se compara la presencia y posiciones de dichas regiones. La línea horizontal representa a la proteína Tc323 de *T. cruzi* de 2983 aminoácidos de longitud. En rosa, PS; en negro, hélice TM; en amarillo, región no citoplasmática; en marrón, región citoplasmática.

Estableciendo un consenso entre las estimaciones de HTM, todas las predicciones coinciden en que entre los residuos 2820 y 2845 de Tc323 hay una región transmembrana. Además, la mitad de los servidores indican presencia de dos HTM cercanos a las posiciones 2615-2630 y 2775-2790. En cuanto a las HTM de la región N-terminal, más de la mitad de los servidores predicen la presencia de un PS. En conjunto, los datos sugieren que la proteína podría estar anclada por su región C-terminal a la membrana de alguna organela o a la membrana plasmática, asociada mayoritariamente a una región no citoplasmática.

3.3.2 Caracterización de la región N-terminal de Tc323 y análisis del péptido señal

Teniendo en cuenta que los servidores de predicción de PS estudian la región N-terminal en las proteínas, se realizó un análisis de los primeros 70 aminoácidos de la secuencia de Tc323, correspondiente al Dominio 1. Como se observó anteriormente en el modelado por I-TASSER, en este dominio se predice la presencia de una estructura secundaria de alfa hélice de 16 aminoácidos en las posiciones 12-28 (Fig. 29). Para discernir si este plegamiento

correspondía a un componente de un PS o a una HTM, se realizó un estudio comparativo utilizando distintos servidores.

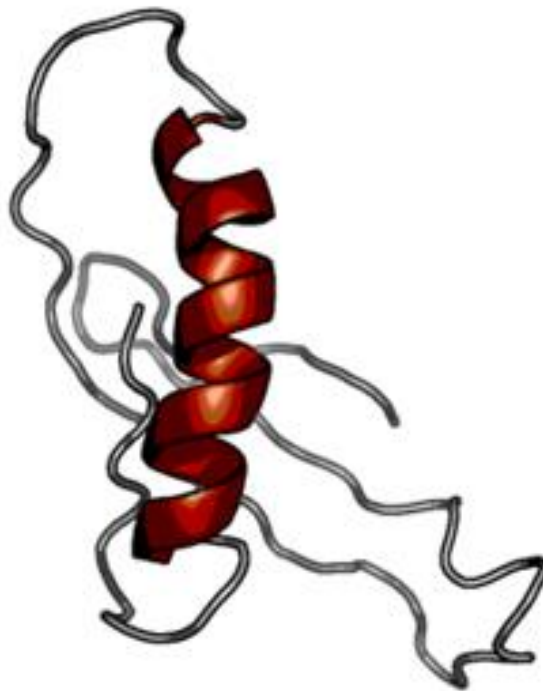


Figura 29: Modelado estructural del Dominio 1, aminoácidos 1-75 dentro de Tc323. En color se indica la estructura de alfa hélice (posiciones 12-28).

En primer lugar, se analizó la composición aminoacídica de los primeros 40 aminoácidos en la secuencia de Tc323, ya que, en caso de existir un PS, éste se encontraría en este rango de residuos, considerando una longitud media de 15-35 aminoácidos (Owji *et al.*, 2018) (Fig. 30). Como resultado, se encontró una alta proporción de residuos hidrofóbicos en la región N-terminal de la proteína, característica común en los PS (Zalucki & Jennings, 2017). Está descrito que las regiones que contienen PS comienzan con una alta proporción de aminoácidos con carga positiva (Owji *et al.*, 2018). Para el caso de la región analizada aquí, se observó que, de los diez primeros aminoácidos presentes, sólo uno presenta carga positiva mientras que el resto tiene carga neutra. Esto tiene un impacto potencial en la predicción de la presencia de PS en Tc323.

MSVIVRNSTALLCGAVVHIFFFIFILLFGSGTAVHVAEV
 |1 |10 |20 |30

Figura 30: Composición aminoacídica de los primeros 40 aminoácidos de Tc323. En rojo, residuos con carga positiva; en verde, carga negativa; en azul, residuos no polares (carga neutra); en negro, residuos hidrofóbicos. Se subrayan los residuos que componen la estructura de alfa hélice predicha por I-TASSER para el Dominio 1 de la proteína.

Paralelamente, se buscó la presencia de un posible PS en la región N-terminal de la proteína Tc323 utilizando distintos servidores. El programa SignalP 5.0 predijo que Tc323 no presenta una región compatible con un PS, asignando una mayor probabilidad a que la secuencia de la región N-terminal cumpla otra función (Fig. 31). Del mismo modo, los programas Signal-3L 3.0 y DeepSig tampoco identificaron la presencia de un PS en esa región de la proteína. El programa DeepSig indica que la estructura de alfa hélice hallada en la región N-terminal corresponde a una HTM.

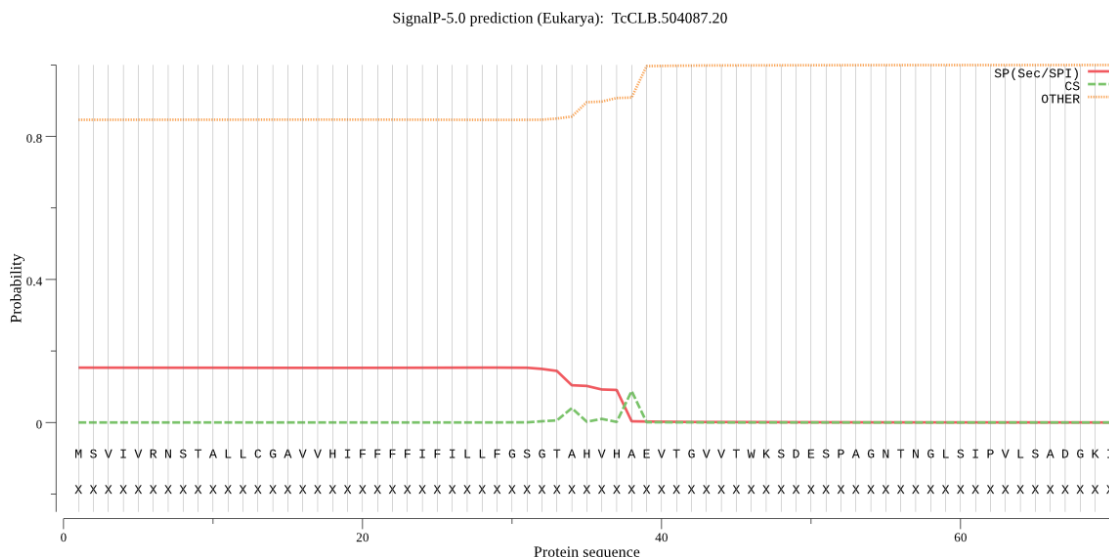


Figura 31: Análisis de Tc323 por SignalP 5.0. El gráfico muestra las curvas de probabilidad para cada posición dentro de los primeros 70 aminoácidos de la secuencia de Tc323. En verde, probabilidad de presencia de un sitio de clivaje; en rojo, PS; en naranja, otra función putativa para esa región. Se observa que la probabilidad de que la región N-terminal tenga un PS es menor a la que indica una función distinta.

En contraste, como resultado del análisis realizado utilizando servidor Predisi, se pudo identificar la presencia de un PS putativo, así como también el sitio de clivaje entre las posiciones 34 y 35 (Fig. 32).

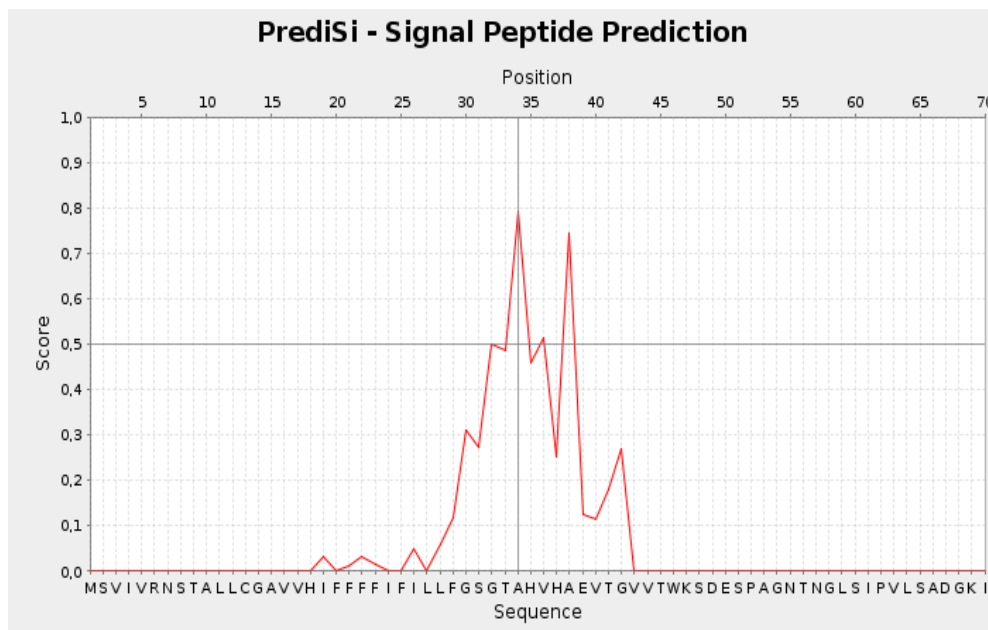


Figura 32: Predicción de sitio de clivaje (posiciones 34-35) para el PS de Tc323 a partir del análisis por *Predisi*.

Finalmente, se utilizó el programa SignalBlast, el cual realiza comparaciones por homología para predecir la presencia de un PS. Como resultado, se predice la presencia de un PS en la secuencia de Tc323 a partir de su similitud con la proteína secretoria *Tripeptidil peptidasa* de hongos (Nro. de acceso: SED1_ARTOC). Este programa también predice que el sitio de clivaje ocurre en el aminoácido 34 de Tc323 (Fig. 33). En el alineamiento realizado por el programa se compararon los primeros 34 residuos de ambas proteínas, donde la identidad fue del 41% (14 aminoácidos idénticos) con una similitud del 61% (21 aminoácidos de similares características), mientras que un 14% de las posiciones corresponden a inserciones o deleciones (5 de los 34 corresponden a *gaps*).

```

Most Significant SValue:
>SED1_ARTOC (Cleavage Site after AA 20)
    Length = 60

    Score = 28.1 bits (61), Expect = 0.94
    Identities = 14/34 (41%), Positives = 21/34 (61%), Gaps = 5/34 (14%)

Query: 1  MSVIVRNSTALLCGAVVHIFFFIFILLFGSGTAHVHAEVTGVVTTWKSDESPAGNTNGLS
60
          IF +FI+I+L+ SG A      HVH + +  V W+
Sbjct:   5  IFMYFIYIVLYASGIAANLSYHVHEKRSIPVWWQ 38
          /
Result: Signal Peptide, putative Cleavage Site after AA 34 (by
similarity to SED1_ARTOC)
  
```

Figura 33: Predicción de PS de Tc323 por SignalBlast. Se muestra el alineamiento realizado por el programa, correspondientes a las primeras 60 posiciones de Tc323 con los aminoácidos 5-38 de la proteína SED1_ARTOC, así como también la posición del sitio de clivaje.

En la Tabla 7 se muestra un resumen de los resultados obtenidos con los programas de predicción, en la cual se observa que, de los diez servidores utilizados, seis predicen la presencia de PS en la secuencia de Tc323.

<i>Servidor</i>	<i>Presencia de PS</i>	<i>Longitud predicha de PS (número de aminoácidos)</i>
MemBrain	No	-
Phobius	Si	38
PolyPhobius	Si	38
TopCons2	Si	30
WolfPSort	Si	34
SignalP 5.0	No	-
Signal-3L 3.0	No	-
DeepSig	No	-
Predisi	Si	34
SignalBlast	Si	34

Tabla 7: Resumen de los resultados obtenidos en los distintos servidores de predicción de PS. Se compara la presencia de dicha secuencia, indicando la longitud predicha en caso de un resultado afirmativo.

3.3.3. Análisis de localización subcelular

La predicción de localización se realizó con los servidores WolfPSort, LocTree3, DeepLoc y iLoc-Euk, basados en Deep Learning; Yloc, Esl-Pred2, MultiLoc2, servidores híbridos; mGOASVM, basado en el estudio de anotaciones; *BUSCA*, basado en el consenso entre servidores. Se informan los resultados de cada uno.

En un primer lugar, WolfPSort predice que Tc323 se localiza en la membrana plasmática. Como se mostró en la sección 3.3.1, su predicción se basa en las señales de direccionamiento dentro de la proteína. En este caso, la presencia de un PS resulta en un posible direccionamiento a vacuola. Además, este programa detecta motivos dentro de la proteína. En este caso, se detecta un motivo de unión a ADN en la posición 670 de 22 aminoácidos (LGGTDACLFEVNDSLVARVTSL), mientras que también predice un posible motivo de direccionamiento a vacuola en la posición 2511 de 4 residuos (TLPK).

DeepLoc, por su lado, predice como localización la membrana de algún compartimento como vacuola o lisosoma. En contraste, el servidor LocTree3 predice como destino de Tc323 el citoplasma, mientras que iLoc-Euk predice una localización nuclear.

De los servidores basados en métodos híbridos, las distintas configuraciones para YLoc predicen que Tc323 participa de la vía secretoria, teniendo como destino la membrana plasmática. Si bien para la predicción de “vía secretoria” la confianza es baja, para “membrana plasmática” la confianza es alta. Este último resultado se obtuvo con las versiones de YLoc que tienen mayor resolución de predicción.

YLoc es un predictor interpretado, por lo que los resultados arrojados incluyen una justificación por parte del servidor. En los casos donde se predijo como localización la membrana plasmática, YLoc se basó en la larga extensión de la región más hidrofóbica de la proteína. El programa argumenta que este atributo es compartido similarmente por el 67% de proteínas de la membrana plasmática, mientras que esto sólo lo comparte el 2% de proteínas del citoplasma, espacio extracelular, mitocondria y núcleo.

El siguiente servidor utilizado fue Esl-Pred2. Este cuenta con distintas configuraciones, basadas en composición de aminoácidos, matriz de pesos posicionales, o método híbrido. Para todas estas modalidades, se predijo a la región extracelular como destino de la proteína. Por lo tanto, este resultado indica que Tc323 participa de la vía de secreción.

Finalmente, MultiLoc2 fue utilizado en sus configuraciones de baja o alta resolución (4 o 9 sitios de predicción, respectivamente). Para el primero, se predice que Tc323 participa de la vía secretoria, mientras que el segundo indica que el destino de la proteína es la región extracelular.

Por otra parte, se realizó el estudio la localización con el servidor mGOASVM, siendo este un predictor basado en el análisis de las anotaciones de la proteína en cuestión. Para Tc323, se predice que se localiza en el citoplasma.

Por último, BUSCA es un predictor que basa sus resultados en el consenso de otros servidores. Para Tc323, el resultado indicaría que esta proteína se localiza en la membrana plasmática o en la membrana de alguna organela.

En la Tabla 8 se muestra un resumen de los resultados obtenidos para esta sección.

	<i>Predicción</i>	<i>Consenso del servidor</i>
WolfPSort	Membrana plasmática	
DeepLoc	Vacuola/Lisosoma	
LocTree3	Citoplasma	
iLoc-Euk	Núcleo	
YLoc	Animal-LowRes: Via secretoria (Baja confianza)	
	Animal-HighRes: Membrana plasmática (Alta confianza)	Membrana plasmática
	+Animal: Membrana plasmática (Alta confianza)	
	Composición de aminoácidos: Extracelular	
Esl-Pred2	Matriz de pesos posicionales: Extracelular	Extracelular
	Método híbrido: Extracelular	
mGOASVM	Citoplasma	

BUSCA	Membrana plasmática
MultiLoc2	Baja resolución: Vía secretoria
	Alta resolución: Extracelular

Tabla 8: Resumen de los resultados para localización subcelular. Para YLoc, se indica la confianza de la predicción para cada configuración. En el caso de que el servidor utilizado presente distintas configuraciones, se indica el resultado de cada una y el consenso para ese servidor.

4. Discusión de Resultados

En este trabajo de Tesina se realizó la caracterización *in silico* de la proteína Tc323 de *T. cruzi*. Se diseñó una estrategia de búsqueda de secuencias por medio del estudio del dominio conservado PQQ_2, única región funcional predicha por el servidor PFAM en la secuencia de la proteína. Esto permitió incluir cerca de 200 proteínas para la construcción de un árbol filogenético. Como resultado, se pudo analizar su historia evolutiva, que ubica a esta proteína en el clado de las enzimas con función de ADH-PQQ. Además, se determinó que su secuencia de aminoácidos presenta al menos tres regiones de plegamiento similar a las estructuras de propulsor beta presentes en enzimas ADH, y que al menos dos de estos dominios podrían asociarse al compuesto PQQ como cofactor.

El análisis evolutivo de Tc323 realizado en este trabajo sugiere que el ancestro del dominio PQQ_2 dio lugar en un principio a las enzimas ADH-PQQ, encontrándolas en bacterias, arqueas, hongos, plantas y tripanosomátidos, no así en otros metazoos o animales. Siguiendo la historia evolutiva, a partir de este tipo de enzimas surgieron las proteínas metanol-deshidrogenasas, de la misma manera que se infiere en el trabajo de Keltjens (2014). Por otro lado, el árbol permite confirmar que el ancestro del dominio PQQ_2 fue asociándose con otros dominios funcionales para generar proteínas con diversas funciones. Como ejemplo, PQQ_2 se asoció con el dominio de unión a AMPc para formar enzimas participes del metabolismo del acil-CoA (Akagawa *et al.*, 2016). Particularmente con las alcohol- y metanol-deshidrogenasas, en la mayoría de las proteínas este dominio quedó aislado, mientras que en otras se asoció a dominios PQQ y/o PQQ_3. Estos últimos son dominios funcionales relacionados que unen al cofactor PQQ y forman estructura de propulsor beta, diferenciándose uno con el otro en las secuencias consenso que los componen. Además, es interesante notar que a lo largo de la historia evolutiva el dominio también sufrió eventos de duplicación, dado que cerca del 50% de las secuencias analizadas poseen dos dominios PQQ_2. Estos eventos de duplicación se correlacionan con la existencia de sustratos de las diferentes ADH-PQQ, ya que este es uno de los mecanismos de la diversificación funcional de las proteínas a través de la evolución (Aluru & Singh, 2021).

Una pregunta que surge de estos conocimientos y lo encontrado en este trabajo (esto es, la presencia de tres motivos putativos de dominios de unión a PQQ y la agrupación en

los clados de las PQQ-ADH): ¿es posible que eventos de duplicación hayan conferido nuevas características enzimáticas a Tc323? Tomando en cuenta la posible función de Tc323 y su presunta capacidad de unión al cofactor, las simulaciones estructurales obtenidas por I-TASSER sugieren la existencia de al menos tres regiones que adquieren la conformación de propulsor beta, distribuidas a lo largo de la proteína (ubicadas en los Dominios 2, 4 y 5). Si bien la unión a PQQ se predijo para los propulsores de los Dominios 4 y 5, hay que considerar que el propulsor presente en el Dominio 2 puede tener una secuencia de aminoácidos no conservada y que los programas de predicción no encuentren una molécula de unión para este dominio. Por otro lado, en ADH-PQQ caracterizadas estructuralmente se reportaron proteínas oligoméricas y monoméricas (Sarmiento-Pavía & Sosa-Torres, 2021; Toyama *et al.*, 2004). Como ejemplo de las oligoméricas, la quinoproteína MDH de *Methylobacterium extorquens* (PDB: 1W6S) presenta cuatro subunidades; dos de las cuales (denominadas alfa), muestran estructura de propulsor beta. Por el contrario, la quinoproteína monomérica ADH-IIB de *Pseudomonas putida* (PDB: 1KV9), presenta un único propulsor donde une dos moléculas de cofactor. Esto es interesante ya que muestra que en un propulsor beta pueden haber uno o más dominios de unión a PQQ. La mayoría de las ADH-PQQ muestran estos patrones en su estructura. Si bien Tc323 presentaría tres propulsores beta saliendo de esta clasificación propuesta, es necesario recordar que la mayoría de estas ADH-PQQ están descritas en bacterias y arqueas, como también que presentan tamaños relativamente menores a los 2983 aminoácidos de esta proteína de *T. cruzi*. Paralelamente, si bien no hubo predicción de sitio activo en Tc323, en estudios cristalográficos de otras ADH-PQQ se reportó que el sitio activo para la oxidación de los sustratos se ubica cercano al sitio de unión al cofactor (Rozeboom *et al.*, 2015; White *et al.*, 2002; Anthony, 1996).

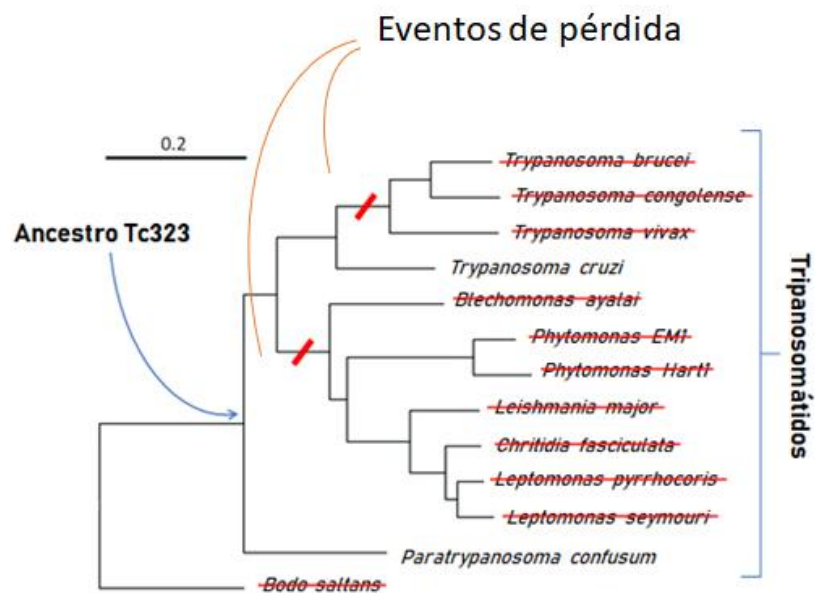
Dicho lo anterior, es necesario contar con evidencia experimental para determinar la función enzimática de Tc323 y si es dependiente del cofactor PQQ. En este sentido, se podrían realizar cromatografías de afinidad utilizando el compuesto PQQ acoplado a resinas con extractos de *T. cruzi* o Tc323 recombinante, para determinar si esta proteína queda retenida en la columna (Akaga *et al.*, 2016). También, sería plausible, contando con Tc323 purificada, ensayar experimentalmente su actividad deshidrogenasa (Day & Anthony, 1990; Vemuluri *et al.*, 2017). Los resultados de estos experimentos permitirían confirmar la presencia o ausencia de actividad ADH.

La estructura de propulsor beta está presente en proteínas con una amplia diversidad funcional, que unen o no distintas moléculas y ejercen diferentes actividades (Chen *et al.*, 2011; Mylemans *et al.*, 2021). Específicamente en el caso de las ADH, si bien los sitios de unión a PQQ toman la estructura de propulsor beta, tal plegamiento no implica unión a cofactor *per se*, como así tampoco de actividad deshidrogenasa. En el trabajo de Kasahara & Kato (2003) identifican en ratones, a una proteína involucrada en el metabolismo de la lisina como una enzima deshidrogenasa dependiente de PQQ. Los autores llegan a dicha conclusión al detectar secuencias correspondientes a motivos de unión a PQQ. En 2005, Felton & Anthony publican una crítica al mencionado trabajo, declarando que esa conclusión es errónea ya que las bases de datos de Smart y PFAM sólo identifican las hojas presentes en las estructuras de propulsor beta para identificar un dominio de unión a PQQ, característica de las enzimas ADH. En este sentido se pone de manifiesto que es necesario también evaluar la similitud a nivel de secuencia con proteínas de función reportada para poder inferir la capacidad de unión y actividad catalítica. En el caso de este trabajo de Tesina, se utilizó tanto la predicción de PFAM como la homología con otras proteínas para establecer que Tc323 presentaría una actividad de alcohol-deshidrogenasa dependiente del cofactor PQQ. Finalmente, es importante mencionar que los programas de predicción en general e I-TASSER en particular, pueden predecir la estructura de proteínas de hasta 1.500 aminoácidos de longitud, motivo por el cual no fue posible obtener un modelo de Tc323 completa que contiene 2.983 aminoácidos.

Los avances en los últimos años permitieron el desarrollo de nuevas herramientas para el modelado de las estructuras de proteínas. *AlphaFold* es un predictor basado en *Deep Learning* que permite obtener modelos con una alta precisión (Jumper *et al.*, 2021). El programa obtuvo el primer puesto en 2018 (AlphaFold 1) y 2020 (AlphaFold 2) según el *Critical Assessment of Techniques for Protein Structure Prediction (CASP)* (<https://predictioncenter.org/>), una evaluación para la predicción de estructura proteica. Esto muestra su potencial para la obtención de estructuras tridimensionales a partir de sólo la secuencia de aminoácidos. Aun así, para algunas especies esta tarea se complica generando modelos de baja confianza, aparentemente por el hecho de que hay baja representatividad de estos organismos en las bases de datos que utiliza *AlphaFold*. Este es el caso de *T. cruzi*. Otra limitación de este servidor es que, al igual que I-TASSER, el número máximo de aminoácidos que de los que puede predecir la estructura es 1.400 (para la versión *AlphaFold2_advanced* en

Google Colab), mucho menor a la longitud de Tc323. Aun así, la comunidad científica ha desarrollado herramientas de código abierto que acotan las secuencias utilizadas como referencia, mejorando las predicciones obtenidas para las proteínas del parásito (John, 2021).

Tc323 no presenta homólogos en otros parásitos patógenos para humanos como *T. brucei* y *Leishmania spp.* Para analizar con mayor detalle la presencia de este tipo de proteínas en los demás miembros de la familia *Trypanosomatidae*, se construyó un árbol filogenético a partir de Tc323 y las secuencias homólogas encontradas en la base de datos TriTrypDB. Este resultado evidenció que Tc323 y sus homólogas tienen origen en el momento en el que divergieron las especies de vida libre, como *B. saltans* y surgieron los tripanosomátidos, ya que está presente en especies cercanas a *T. cruzi*, pero también en *P. confusum*, un tripanosomátido divergente de los clados *Trypanosoma* y *Leishmania* (Skalický *et al.*, 2017; Flegontov *et al.*, 2013). Este análisis evolutivo también evidencia que los homólogos a Tc323 sufrieron eventos de pérdida, ya que están ausentes en la mayoría de los grupos de tripanosomátidos, a excepción de *T. cruzi* y especies muy relacionadas evolutivamente, como ejemplificamos en el siguiente esquema:



En relación con lo anterior, para cuantificar los eventos de pérdida en la evolución del ancestro de Tc323, es necesario considerar la cantidad de genomas secuenciados de tripanosomátidos con la que se cuenta actualmente. En NCBI se pueden encontrar 148 genomas pertenecientes a organismos de la familia *Trypanosomatidae* en los niveles de genoma

completo, cromosomas, *scaffold*, o *contigs*. Es necesario aclarar que hay especies, por ejemplo *T. cruzi*, cuyos genomas fueron secuenciados más de una vez, sobre todo si se considera la cantidad de cepas establecidas hasta el momento. Tomando esto en cuenta, se dispone del genoma secuenciado de 83 especies distintas dentro de esta familia (de las cuales, 74 están anotadas en la base de datos TriTrypDB y 24 corresponden a diferentes cepas de laboratorio de *T. cruzi*). Cabe aclarar que el ensamblado del genoma del parásito se dificulta debido a la alta presencia de secuencias repetitivas, por lo que es posible que se hayan perdido genes en el proceso (El-Sayed *et al.*, 2005). Por otro lado, en la búsqueda exhaustiva de proteínas homólogas a Tc323, los hallazgos se limitaron a 6 especies del grupo de los tripanosomátidos, resultando en una representatividad del 7% dentro de dicha familia. Lo anterior implica que, a lo largo de la evolución, el ancestro de Tc323 y sus homólogas se perdió en un 93% de las especies analizadas dentro de la familia *Trypanosomatidae*.

A la luz de estos resultados, se puede inferir que los homólogos de Tc323 son prescindibles para la mayoría de los tripanosomátidos, pero no para *T. cruzi* y unas pocas especies estrechamente relacionadas, no patógenas. En este sentido, surge la pregunta, ¿cuál es la característica común a *T. cruzi*, *T. grayi*, *T. rangeli*, y *T. conorhini* que pueda explicar una ventaja funcional de Tc323 y sus secuencias homólogas? Si bien se encuentran filogenéticamente relacionados, *T. rangeli*, una especie digenética no patogénica capaz de infectar diversos mamíferos, tiene un ciclo de vida que incluye la invasión de las glándulas salivales del insecto vector y se transmite a través de la saliva, mientras que *T. conorhini*, *T. grayi*, *T. theileri* y *T. cruzi* colonizan exclusivamente el sistema intestinal de los vectores y son transmitidos por las heces (Bradwell *et al.*, 2018; Kelly *et al.*, 2017). Además, *T. grayi*, a diferencia de los otros tripanosomas que infectan mamíferos, tiene como hospedador vertebrado a los cocodrilos (Kelly *et al.*, 2014). En estas observaciones, no es evidente una ventaja evolutiva común a Tc323 y sus homólogos y, son necesarios experimentos que ayuden a determinar cuál es el rol de esta proteína en el ciclo de vida de estos parásitos y de *T. cruzi* en particular. En base a nuestro conocimiento, en tripanosomas no se han descrito ADH-PQQ o enzimas cuya función dependa del cofactor PQQ, así como tampoco se ha demostrado que los alcoholes puedan ser utilizados como sustrato para aportar energía al metabolismo celular del parásito. Por el contrario, sí se ha relacionado el nivel de expresión de otras enzimas, como la ADH (TcADH) del tipo NAD(P)-dependiente, con la capacidad de detoxificación de

xenobióticos, especies reactivas de oxígeno y otros metabolitos en cepas resistentes de *T. cruzi* a la droga Benznidazol (Campos *et al.*, 2009; Gonzalez *et al.*, 2017; Echave *et al.*, 2003). En este sentido y teniendo en cuenta los distintos ambientes que atraviesa el parásito durante su ciclo de vida, es plausible plantear la hipótesis que la actividad de Tc323 está involucrada en la detoxificación del espacio intracelular de compuestos reactivos, ya sea en el parásito como en las células que infecta, siendo este uno de los mecanismos que le asegura su supervivencia en el hospedero mamífero.

Con el objetivo de estudiar la presencia de una señal de localización, se emplearon distintos servidores para determinar la presencia de un péptido señal en la secuencia de aminoácidos de Tc323. Como resultado, se encontró que, de los diez servidores empleados, seis predicen la presencia de PS, mientras que los restantes sólo predicen la estructura de una alfa hélice. Analizando la secuencia y estructura de la región N-terminal de Tc323, se puede establecer una hipótesis que explique por qué los distintos servidores difieren en su predicción. Según el modelado estructural de los primeros 75 aminoácidos de la proteína, los aminoácidos de las posiciones 12 - 28 adoptan una estructura secundaria de alfa hélice. La tarea de los servidores de predicción es diferenciar esta región como HTM o como parte del PS. Para esto, se analiza la composición aminoacídica de los primeros 40 residuos de una proteína, ya que en general, un PS tiene una longitud menor a ese valor (Owji *et al.*, 2018). El PS es un segmento peptídico de alrededor de 20-40 aminoácidos, que está compuesto por tres regiones, una N-terminal, una región hidrofóbica intermedia (H) y una región C-terminal (von Heijne, 1985). En este modelo, la región N-terminal con 5-10 residuos de longitud, está compuesta en su mayoría por aminoácidos básicos o de carga positiva lisina, arginina o histidina (Peterson *et al.*, 2003). Esto contrasta con la composición de Tc323, que en esta región sólo cuenta con un único residuo básico (una arginina en la posición 6). Esto último podría explicar la ambigüedad encontrada en los resultados de los distintos servidores: al presentar baja proporción de residuos básicos, algunos algoritmos basados en secuencias consenso no predicen la presencia de PS con alta confianza. Aun así, es importante mencionar que estos programas tienen en cuenta la homología con otras proteínas, por ejemplo, para determinar finalmente si la secuencia corresponde o no a un PS. Esto muestra la capacidad predictiva de estos servidores, ya que no se basan simplemente en la composición aminoacídica de una secuencia, sino que también recurren a otros algoritmos para la decisión.

Dado que en otros microorganismos las enzimas ADH-PQQ pueden funcionar tanto ancladas a membrana como de manera soluble (Gómez-Manzo *et al.*, 2010; Sarmiento-Pavía & Sosa-Torres, 2021), se estudió con herramientas bioinformáticas cual es la localización posible de Tc323 en *T. cruzi*. Para alcanzar este objetivo se utilizaron servidores que se basan en distintas herramientas para realizar las predicciones. Como resultado, del total de nueve servidores utilizados, seis predicen que Tc323 participa de la vía de secreción con destino a vacuola/lisosoma, membrana plasmática, o espacio extracelular. Dentro de estos, tres servidores indican que la proteína se encontraría en la membrana plasmática. Si bien hay ambigüedad en los resultados, la predicción de PS y localización sub-celular para Tc323 son coherentes. En la mayoría de los casos, si una proteína participa de la vía de secreción sea cual sea su destino dentro de ella, es necesario que cuente con PS (Nothwehr & Gordon, 1990), salvo que contenga señales de direccionamiento independientes, que no se han identificado por PFAM. El consenso establecido sustenta la idea de que la proteína estudiada tiene dicha señal de direccionamiento y es dirigida a esa vía, tomando como destino la membrana plasmática de la célula o de alguna vesícula. Como limitación de los servidores utilizados es necesario considerar que se trabaja con una proteína de *T. cruzi*. Esto implica que organelas como los reservosomas o la vacuola contráctil no son tomados en cuenta para la predicción, ya que sólo analizan los compartimientos celulares compartidos por todos los eucariotas. En relación con lo anterior, la vacuola contráctil está siendo propuesta como partícipe de la vía de secreción, pudiendo ser un destino para Tc323 (Niyogi *et al.*, 2014). Estas predicciones concuerdan con los resultados obtenidos en otros grupos de trabajo, ya que se han identificado péptidos correspondientes a la proteína Tc323 en análisis del proteoma de fracciones enriquecidas en la vacuola contráctil (Ulrich *et al.*, 2011), reservosomas intactos (Sant'Anna *et al.*, 2009) y en la superficie de epimastigotes (Queiroz *et al.*, 2014). También se detectaron péptidos de Tc323 en vesículas secretorias provenientes de *T. cruzi* en el suero de pacientes infectados (Brossas *et al.*, 2017).

Resultados previos obtenidos en nuestro laboratorio refieren a la necesidad de permeabilizar la membrana plasmática de los parásitos (en sus estadios principales) para poder detectar a Tc323 mediante inmunofluorescencia (Figura 8), sugiriendo una ubicación intracelular de la proteína. Sin embargo, este hallazgo no es concluyente ya que experimentos más recientes mostraron que este paso no fue necesario para poder visualizarla y, se detectó

señal de la proteína tanto en parásitos permeabilizados como sin permeabilizar. Esto indica que Tc323 podría localizarse en la célula dependiendo de estímulos o condiciones fisiológicas que ameritan determinar.

Dado que nuestra hipótesis de trabajo propone a Tc323 de *T. cruzi* como antígeno para el diagnóstico de la Enfermedad de Chagas, es determinante identificar qué región de la proteína contiene al epítipo que reconoce el anticuerpo 6B6. Para esto, en nuestro laboratorio estamos llevando a cabo la expresión de manera recombinante de los seis dominios estructurales descritos en este trabajo. Aun así, es interesante analizar *in silico* la presencia putativa de epítopes específicos de células B en los dominios de Tc323. Los resultados obtenidos indican que el Dominio 6 cuenta con la mayor antigenicidad (60,40% de la secuencia predicha como antigénica, ver Tabla 6), en comparación al resto de los dominios. Si bien el modelado estructural de esta región de la proteína muestra que la zona antigénica no presenta estructura secundaria alguna, se ha demostrado que las regiones más inmunogénicas corresponden a regiones con plegamiento de lámina beta y de conformación aleatoria o *random coil*, lo que concuerda con el resultado obtenido (Rubinstein et al, 2008; Ofra et al., 2008). En relación con esto, es interesante comentar que se avanzó con la producción heteróloga del Dominio 6, y éste fue reconocido por plasmas de pacientes con ECC en Western Blot y ensayos ELISA.

Respecto a la presencia de glicosilaciones en Tc323, se observó que varios de los sitios predichos para la *N*- y *O*-glicosilación se encontrarían en regiones antigénicas. Esta información es sumamente relevante al momento de elegir un sistema de producción heteróloga para los distintos dominios. En el uso diagnóstico es necesario que, además de contar con un plegamiento correcto, esté presente esta modificación postraduccional en Tc323 para el reconocimiento antigénico (Lisowska, 2002). Un sistema prometedor es el basado en el uso de *Leishmania tarentolae*, un tripanosomátido no patogénico. Este permitiría la expresión de antígenos de *T. cruzi* con estructuras y modificaciones postraducionales típicas de los tripanosomátidos, que mantendrían las propiedades antigénicas respecto al organismo nativo (Niborski et al., 2021; Murphy et al., 2020; Ferrer et al., 2019).

Sin embargo, si bien se determinó que varios sitios que pueden presentar esta modificación postraduccional, es importante mencionar que la secuencia de aminoácidos Asn-Xaa-Ser/Thr (una asparagina seguido de cualquier aminoácido menos prolina, y una serina

o treonina), conocida dentro de la glicobiología como *sequon*, es necesaria para que ocurra la modificación de una proteína por *N*-glicosilación (Stanley *et al.*, 2017). Sin embargo, esta condición no es suficiente ya que se ha reportado que alrededor de un 70% de las asparaginas dentro de estas secuencias consenso no están glicosiladas. Por otro lado, no se han encontrado aún evidencias que definan una secuencia consenso mínima para que ocurra la *O*-glicosilación (Brockhausen & Stanley, 2017), por lo que la presencia de un residuo serina o treonina permitiría esta modificación. Para esta última, el servidor utilizado otorga un score a cada uno de esos residuos en función de bases de datos glicoproteómicas para evaluar la probabilidad de que suceda la modificación (Steentoft, 2013). Luego, se consideran posibles sitios *O*-glicosilables a aquellos que superen al valor de corte establecido por la plataforma utilizada. Por lo tanto, es necesaria en este caso (como en todas las demás predicciones) evidencia experimental que permita confirmar la presencia de ambos tipos de glicosilación y posteriores estudios para revelar los sitios en donde ocurren.

Como conclusión, en este trabajo de Tesina se realizó un análisis *in silico* exhaustivo de las características estructurales, fisiológicas e inmunológicas de Tc323, cuyos resultados contribuyen al desarrollo de estrategias para validar a esta proteína como antígeno de diagnóstico, así como también para determinar su función en el ciclo de vida de *T. cruzi*.

5. Bibliografía

Abras, A., Gállego, M., Muñoz, C., Juiz, N. A., Ramírez, J. C., Cura, C. I., Tebar, S., Fernández-Arévalo, A., Pinazo, M. J., de la Torre, L., Posada, E., Navarro, F., Espinal, P., Ballart, C., Portús, M., Gascón, J., & Schijman, A. G. (2017). Identification of *Trypanosoma cruzi* Discrete Typing Units (DTUs) in Latin-American migrants in Barcelona (Spain). *Parasitology International*, 66(2), 83–88. <https://doi.org/10.1016/J.PARINT.2016.12.003>

Akagawa, M., Minematsu, K., Shibata, T., Kondo, T., Ishii, T., & Uchida, K. (2016). Identification of lactate dehydrogenase as a mammalian pyrroloquinoline quinone (PQQ)-binding protein. *Scientific Reports* 2016 6:1, 6(1), 1–19. <https://doi.org/10.1038/srep26723>

Akagawa, M., Minematsu, K., Shibata, T., Kondo, T., Ishii, T., & Uchida, K. (2016). Identification of lactate dehydrogenase as a mammalian pyrroloquinoline quinone (PQQ)-binding protein OPEN. Nature Publishing Group.

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* (Oxford, England), 33(21), 3387–3395. <https://doi.org/10.1093/bioinformatics/btx431>

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410.

Aluru, C., & Singh, M. (2021). Improved inference of tandem domain duplications. *Bioinformatics*, 37(Supplement_1), i133–i141. <https://doi.org/10.1093/BIOINFORMATICS/BTAB329>

Ambrish Roy, Alper Kucukural, Yang Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5: 725-738 (2010).

Anthony C. 2004. The quinoprotein dehydrogenases for methanol and glucose. *Arch Biochem Biophys.* 428(1):2-9. Review.

Anthony, C. (1996). Quinoprotein-catalysed reactions. *The Biochemical Journal*, 320 (Pt 3)(Pt 3), 697–711. <https://doi.org/10.1042/BJ3200697>

Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J. C., Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic acids research*, 38(Database issue), D457–D462. doi.org/10.1093/nar/gkp851

B. Yao, L. Zhang, S. Liang, C. Zhang. SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLoS ONE* (2012); 7(9):e45152.

Balouz, V., Agüero, F., & Buscaglia, C. A. (2017). Chagas Disease Diagnostic Applications: Present Knowledge and Future Steps. *Advances in parasitology*, 97, 1–45. doi.org/10.1016/bs.apar.2016.10.001

Barrias, E. S., de Carvalho, T. M. U., & de Souza, W. (2013). *Trypanosoma cruzi*: Entry into mammalian host cells and parasitophorous vacuole formation. *Frontiers in Immunology*, 4(AUG), 186. <https://doi.org/10.3389/FIMMU.2013.00186/BIBTEX>

Bayona, J. C., Nakayasu, E. S., Laverrière, M., Aguilar, C., Sobreira, T. J., Choi, H., Nesvizhskii, A. I., Almeida, I. C., Cazzulo, J. J., & Alvarez, V. E. (2011). SUMOylation pathway in *Trypanosoma cruzi*: functional characterization and proteomic analysis of target proteins. *Molecular & cellular proteomics : MCP*, 10(12), M110.007369.

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics: Jun*;4(6):1633-49.

Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*: 294(5): 1351-1362.

Blum, T., Briesemeister, S., & Kohlbacher, O. (2009). MultiLoc2: Integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics*, 10(1), 274. <https://doi.org/10.1186/1471-2105-10-274>

Bradwell, K. R., Koparde, V. N., Matveyev, A. v., Serrano, M. G., Alves, J. M. P., Parikh, H., Huang, B., Lee, V., Espinosa-Alvarez, O., Ortiz, P. A., Costa-Martins, A. G., Teixeira, M. M. G., & Buck, G. A. (2018). Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma rangeli* to *Trypanosoma cruzi* strains of high and low virulence. *BMC Genomics* 2018 19:1, 19(1), 1–20. <https://doi.org/10.1186/S12864-018-5112-0>

Briesemeister, S., Rahnenführer, J., & Kohlbacher, O. (2010a). Going from where to why- interpretable prediction of protein subcellular localization. *Bioinformatics*, 26(9), 1232–1238. <https://doi.org/10.1093/bioinformatics/btq115>

Briesemeister, S., Rahnenführer, J., & Kohlbacher, O. (2010b). YLoc-an interpretable web server for predicting subcellular localization. *Nucleic Acids Research*, 38(SUPPL. 2), W497–W502. <https://doi.org/10.1093/nar/gkq477>

Brockhausen I, Stanley P. (2017). O-GalNAc Glycans. *Essentials of Glycobiology* [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 10. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK453030/> doi: 10.1101/glycobiology.3e.010

Brossas, J. Y., Gulin, J. E. N., Bisio, M. M. C., Chapelle, M., Marinach-Patrice, C., Bordessoules, M., Ruiz, G. P., Vion, J., Paris, L., Altcheh, J., & Mazier, D. (2017). Secretome analysis of *Trypanosoma cruzi* by proteomics studies. *PloS One*, 12(10). <https://doi.org/10.1371/JOURNAL.PONE.0185504>

Campos, F. M. F., Liarte, D. B., Mortara, R. A., Romanha, A. J., & Murta, S. M. F. (2009). Characterization of a gene encoding alcohol dehydrogenase in benzimidazole-susceptible and -resistant populations of *Trypanosoma cruzi*. *Acta Tropica*, 111(1), 56–63. <https://doi.org/10.1016/J.ACTATROPICA.2009.02.007>

Castro, E. (2009). Chagas' disease: Lessons from routine donation testing. *Transfusion Medicine*, 19(1), 16–23. <https://doi.org/10.1111/j.1365-3148.2009.00915.x>

Centros de Control y Prevención de enfermedades (2019). Chagas Disease – Biology. <https://www.cdc.gov/parasites/chagas/biology.html>

Chen, C. K. M., Chan, N. L., & Wang, A. H. J. (2011). The many blades of the β -propeller proteins: conserved but versatile. *Trends in Biochemical Sciences*, 36(10), 553–561. <https://doi.org/10.1016/J.TIBS.2011.07.004>

Cretu, C., Schmitzová, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E. I., Sharma, K., Will, C. L., Urlaub, H., Lührmann, R., & Pena, V. (2016). Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Molecular Cell*, 64(2), 307–319. <https://doi.org/10.1016/J.MOLCEL.2016.08.036>

da Silveira, J. F., Umezawa, E. S., & Luquetti, A. O. (2001). Chagas disease: recombinant *Trypanosoma cruzi* antigens for serological diagnosis. *Trends in Parasitology*, 17(6), 286–291. [https://doi.org/10.1016/S1471-4922\(01\)01897-9](https://doi.org/10.1016/S1471-4922(01)01897-9)

Daltro, R. T., Leony, L. M., Maron Freitas, N. E., Oliveira Silva, Â. A., Santos, E. F., Del-Rei, R. P., Felinto Brito, M. E., Brandão-Filho, S. P., Gomes, Y. M., Silva, M. S., Donato, S. T., Bezerra Jeronimo, S. M., de Góis Monteiro, G. R., Carvalho, L. P., Magalhães, A. S., Tonin Zanchin, N. I., Fiorani Celedon, P. A., & Neves Santos, F. L. (2019). Cross-Reactivity Using Chimeric *Trypanosoma cruzi* Antigens: Diagnostic Performance in Settings Where Chagas Disease and American Cutaneous or Visceral Leishmaniasis Are Coendemic. *Journal of Clinical Microbiology*, 57(8). <https://doi.org/10.1128/JCM.00762-19>

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*, 27(8), 1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>

Day, D. J., & Anthony, C. (1990). [33] Methanol dehydrogenase from *Methylobacterium extorquens* AM1. *Methods in Enzymology*, 188(C), 210–216. [https://doi.org/10.1016/0076-6879\(90\)88035-9](https://doi.org/10.1016/0076-6879(90)88035-9)

Deng, W., Wang, C., Zhang, Y., Xu Y., Zhang S., Liu Z., Xue Y.,(2016) GPS-PAIL: prediction of lysine acetyltransferase-specific modification sites from protein sequences. *Sci Rep* 6, 39787

Dong X., Yang Z. (2011). Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. *Biophysical Journal*, vol 101, 2525-2534.

Echave, P., Tamarit, J., Cabiscol, E., & Ros, J. (2003). Novel antioxidant role of alcohol dehydrogenase E from *Escherichia coli*. *The Journal of Biological Chemistry*, 278(32), 30193–30198. <https://doi.org/10.1074/JBC.M304351200>

Elias, M. C., da Cunha, J. P., de Faria, F. P., Mortara, R. A., Freymüller, E., & Schenkman, S. (2007). Morphological events during the *Trypanosoma cruzi* cell cycle. *Protist*, 158(2), 147–157. <https://doi.org/10.1016/j.protis.2006.10.002>

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Åslund, L., Attipoe, P., ... Andersson, B. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science (New York, N.Y.)*, 309(5733). <https://doi.org/10.1126/SCIENCE.1112631>

Feng, S. H., Zhang, W. X., Yang, J., Yang, Y., & Shen, H. Bin. (2020). Topology Prediction Improvement of α -helical Transmembrane Proteins Through Helix-tail Modeling and Multiscale Deep Learning Fusion. *Journal of Molecular Biology*, 432(4), 1279–1296. <https://doi.org/10.1016/j.jmb.2019.12.007>

Ferrer, E., Lares, M., Vietri, M., & Medina, M. (2013). [Comparison between immunological and molecular techniques for the diagnosis of Chagas disease]. *Enfermedades Infecciosas y Microbiología Clínica*, 31(5), 277–282. <https://doi.org/10.1016/J.EIMC.2012.09.007>

Ferrer, M. J., Wehrendt, D. P., Bonilla, M., Comini, M. A., Tellez-Iñón, M. T., & Potenza, M. (2019). Production of Recombinant *Trypanosoma cruzi* Antigens in *Leishmania tarentolae*. *Methods in Molecular Biology (Clifton, N.J.)*, 1955, 105–118. https://doi.org/10.1007/978-1-4939-9148-8_8

Flegontov, P., Votýpka, J., Skalický, T., Logacheva, M. D., Penin, A. A., Tanifuji, G., Onodera, N. T., Kondrashov, A. S., Volf, P., Archibald, J. M., & Lukeš, J. (2013). *Paratrypanosoma* is a novel early-branching trypanosomatid. *Current Biology: CB*, 23(18), 1787–1793. <https://doi.org/10.1016/J.CUB.2013.07.045>

Flores-Encarnación, M., Sánchez-Cuevas, M., & Ortiz-Gutiérrez, F. (2004). edigraphic.com Las PQQ-deshidrogenasas. Un novedoso ejemplo de quinoproteínas bacterianas. Revista Latinoamericana de MICROBIOLOGÍA. 46, 47–59.

Frank, K., & Sippl, M. J. (2008). High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics*, 24(19), 2172–2176. <https://doi.org/10.1093/bioinformatics/btn422>

Fritsche, C., Sitz, M., Weiland, N., Breitling, R., & Pohl, H. D. (2007). Characterization of the growth behavior of *Leishmania tarentolae*: a new expression system for recombinant proteins. *Journal of basic microbiology*, 47(5), 384–393.

Garg, A., & Raghava, G. P. S. (2008). ESLpred2: Improved method for predicting subcellular localization of eukaryotic proteins. *BMC Bioinformatics*, 9(1), 1–10. <https://doi.org/10.1186/1471-2105-9-503>

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., Altermann, U., Angerer, P., Ansorge, S., Balasz, K., Bernhofer, M., Betz, A., Cizmadija, L., Do, K. T., Gerke, J., Greil, R., Joerdens, V., Hastreiter, M., Hembach, K., ... Rost, B. (2014). LocTree3 prediction of localization. *Nucleic Acids Research*, 42(W1). <https://doi.org/10.1093/nar/gku396>

Gomes, Y.M., Lorena, V.M., Luquetti, A.O. (2009). Diagnosis of Chagas disease: what has been achieved? What remains to be done with regard to diagnosis and follow up studies? *Mem. Inst. Oswaldo Cruz* 104 (Suppl. 1), 115e121.

Gómez-Manzo, S., Solano-Peralta, A., Saucedo-Vázquez, J. P., Escamilla-Marván, J. E., Kroneck, P. M. H., & Sosa-Torres, M. E. (2010). The membrane-bound quinohemoprotein alcohol. Dehydrogenase from *gluconacetobacter diazotrophicus* PAL5 carries a [2Fe-2S] Cluster. *Biochemistry*, 49(11), 2409–2415. https://doi.org/10.1021/B19015007/ASSET/IMAGES/B19015007.SOCIAL.JPEG_V03

Gonçalves, C. S., Ávila, A. R., de Souza, W., Motta, M. C. M., & Cavalcanti, D. P. (2018). Revisiting the *Trypanosoma cruzi* metacyclogenesis: morphological and ultrastructural analyses during cell differentiation. *Parasites & Vectors*, 11(1). <https://doi.org/10.1186/S13071-018-2664-4>

González, L., García-Huertas, P., Triana-Chávez, O., García, G. A., Murta, S. M. F., & Mejía-Jaramillo, A. M. (2017). Aldo-keto reductase and alcohol dehydrogenase contribute to benzimidazole natural resistance in *Trypanosoma cruzi*. *Molecular Microbiology*, 106(5), 704–718. <https://doi.org/10.1111/MMI.13830>

Görisch, H. (2003). The ethanol oxidation system and its regulation in *Pseudomonas aeruginosa*. [https://doi.org/10.1016/S1570-9639\(03\)00066-9](https://doi.org/10.1016/S1570-9639(03)00066-9)

Guindon, S., & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, 52(5), 696–704. <https://doi.org/10.1080/10635150390235520>

Guo, J., Zhang, Y., Li, H., Chu, H., Wang, Q., Jiang, S., Li, Y., Shen, H., Li, G., Chen, J., & Xu, C. (2018). Intramembrane ionic protein–lipid interaction regulates integrin structure and function. *PLOS Biology*, 16(11), e2006525. <https://doi.org/10.1371/JOURNAL.PBIO.2006525>

Gupta R., Jung E. and Brunak S.(2004). Prediction of N-glycosylation sites in human proteins. In preparation.

Hiller, K., Grote, A., Scheer, M., Münch, R., & Jahn, D. (2004). PrediSi: Prediction of signal peptides and their cleavage positions. *Nucleic Acids Research*, 32(WEB SERVER ISS.), W375. <https://doi.org/10.1093/nar/gkh378>

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, 35(SUPPL.2). <https://doi.org/10.1093/nar/gkm259>

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680–682. <https://doi.org/10.1093/bioinformatics/btq003>

Jefferis R. (2016). Posttranslational Modifications and the Immunogenicity of Biotherapeutics. *Journal of immunology research*, 2016, 5358272.

Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017 (Web Server issue). doi: 10.1093/nar/gkx352

Jianyí Yang, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, Yang Zhang. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, 12: 7-8 (2015).

Jing, H., Takagi, J., Liu, J. H., Lindgren, S., Zhang, R. G., Joachimiak, A., Wang, J. H., & Springer, T. A. (2002). Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure (London, England : 1993)*, 10(10), 1453–1464. [https://doi.org/10.1016/s0969-2126\(02\)00840-7](https://doi.org/10.1016/s0969-2126(02)00840-7)

John, R., & Id, W. (2021). A resource for improved predictions of *Trypanosoma* and *Leishmania* protein three-dimensional structure. *PLOS ONE*, 16(11), e0259871.

<https://doi.org/10.1371/JOURNAL.PONE.0259871>

Juiz, N. A., Estupiñán, E., Hernández, D., Garcilazo, A., Chadi, R., Morales Sanfurgo, G., Schijman, A. G., Longhi, S. A., & González, C. I. (2019). Association study between CCR2-CCR5 genes polymorphisms and chronic Chagas heart disease in Wichí and in admixed populations from Argentina. *PLOS Neglected Tropical Diseases*, 13(1), e0007033. <https://doi.org/10.1371/JOURNAL.PNTD.0007033>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 596:7873, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

Junqueira, Â. C. V. (2008). International meeting: new diagnostic tests are urgently needed to treat patients with Chagas disease. *Revista Da Sociedade Brasileira de Medicina Tropical*, 41(3), 315–319. <https://doi.org/10.1590/S0037-86822008000300020>

Kalichava, A., & Ochsenreiter, T. (2021). Ultrastructure expansion microscopy in *Trypanosoma brucei*. *Open Biology*, 11(10). <https://doi.org/10.1098/RSOB.210132>

- Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5), 1027–1036. <https://doi.org/10.1016/j.jmb.2004.03.016>
- Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2005). An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics*, 21(SUPPL. 1), 251–257. <https://doi.org/10.1093/bioinformatics/bti1014>
- Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Research*, 35(SUPPL.2), 429–432. <https://doi.org/10.1093/nar/gkm256>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Kaufer, A., Ellis, J., Stark, D., & Barratt, J. (2017). The evolution of trypanosomatid taxonomy. *Parasites & Vectors*, 10(1). <https://doi.org/10.1186/S13071-017-2204-7>
- Kelly, S., Ivens, A., Manna, P. T., Gibson, W., & Field, M. C. (2014). A draft genome for the African crocodylian trypanosome *Trypanosoma grayi*. *Scientific Data* 2014 1:1, 1(1), 1–7. <https://doi.org/10.1038/sdata.2014.24>
- Kelly, S., Ivens, A., Mott, G. A., O’Neill, E., Emms, D., Macleod, O., Voorheis, P., Tyler, K., Clark, M., Matthews, J., Matthews, K., & Carrington, M. (2017). An Alternative Strategy for Trypanosome Survival in the Mammalian Bloodstream Revealed through Genome and Transcriptome Analysis of the Ubiquitous Bovine Parasite *Trypanosoma* (*Megatrypanum*) *theileri*. *Genome Biology and Evolution*, 9(8), 2093. <https://doi.org/10.1093/GBE/EVX152>
- Kim, K. H., & Paetzel, M. (2011). Crystal Structure of *Escherichia coli* BamB, a Lipoprotein Component of the β -Barrel Assembly Machinery Complex. *Journal of Molecular Biology*, 406(5), 667–678. <https://doi.org/10.1016/J.JMB.2010.12.020>
- Klatt, S., Simpson, L., Maslov, D. A., & Konthur, Z. (2019). *Leishmania tarentolae*: Taxonomic classification and its application as a promising biotechnological expression host. *PLoS neglected tropical diseases*, 13(7), e0007424.

Kringelum, J. V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS computational biology*, 8(12), e1002829. <https://doi.org/10.1371/journal.pcbi.1002829>

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, 41(W1), W22–W28. <https://doi.org/10.1093/nar/gkt389>

Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J., & Laustsen, A. H. (2018). Basics of Antibody Phage Display Technology. *Toxins*, 10(6). <https://doi.org/10.3390/TOXINS10060236>

Leiria Campo, V., Braga Martins-Teixeira, M., & Carvalho, I. (2016). *Trypanosoma cruzi* Invasion into Host Cells: A Complex Molecular Targets Interplay. *Mini Reviews in Medicinal Chemistry*, 16(13), 1084–1097. <https://doi.org/10.2174/1389557516666160607230238>

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., & Gascuel, O. (2018). Renewing Felsenstein’s phylogenetic bootstrap in the era of big data. *Nature*, 556(7702), 452–456. <https://doi.org/10.1038/s41586-018-0043-0>

Li, A., Xue, Y., Jin, C., Wang, M., & Yao, X. (2006). Prediction of Nepsilon-acetylation on internal lysines implemented in Bayesian Discriminant Method. *Biochemical and biophysical research communications*, 350(4), 818–824.

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>

Lima, L., Espinosa-Álvarez, O., Ortiz, P. A., Trejo-Varón, J. A., Carranza, J. C., Pinto, C. M., Serrano, M. G., Buck, G. A., Camargo, E. P., & Teixeira, M. M. G. (2015). Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses

supporting Tcbat as an independent DTU (discrete typing unit). *Acta Tropica*, 151(1), 166–177.
<https://doi.org/10.1016/J.ACTATROPICA.2015.07.015>

Lingg N, Zhang P, & Song Z (2012) The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol J* 7(12):1462–1472.

Lisowska, E. (2002). The role of glycosylation in protein antigenic properties. *CMLS, Cell. Mol. Life Sci.* 59, 445–455.

Lisowska, E. (2002). The role of glycosylation in protein antigenic properties. *Cellular and Molecular Life Sciences CMLS* 2002 59:3, 59(3), 445–455. <https://doi.org/10.1007/S00018-002-8437-3>

Lowe, E. C., Baslé, A., Czjzek, M., Firbank, S. J., & Bolam, D. N. (2012). A scissor blade-like closing mechanism implicated in transmembrane signaling in a *Bacteroides* hybrid two-component system. *Proceedings of the National Academy of Sciences*, 109(19), 7298–7303.
<https://doi.org/10.1073/PNAS.1200479109>

Lucchese, G., Stufano, A., Trost, B., Kusalik, A., & Kanduc, D. (2006). Peptidology: short amino acid modules in cell biology and immunology. *Amino Acids* 2006 33:4, 33(4), 703–707.
<https://doi.org/10.1007/S00726-006-0458-Z>

Messenger, L. A., & Bern, C. (2018). Congenital Chagas disease: current diagnostics, limitations and future perspectives. *Current opinion in infectious diseases*, 31(5), 415–421.
<https://doi.org/10.1097/QCO.0000000000000478>

Ministerio de Salud de la Nación, (2021). Chagas.
<https://www.argentina.gob.ar/salud/glosario/chagas>

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419.
<https://doi.org/10.1093/nar/gkaa913>

Murcia, L., Carrilero, B., Saura, D., Iborra, M. A., & Segovia, M. (2013). [Diagnosis and treatment of Chagas disease]. *Enfermedades Infecciosas y Microbiología Clínica*, 31 Suppl 1(SUPPL.1), 26–34. [https://doi.org/10.1016/S0213-005X\(13\)70111-3](https://doi.org/10.1016/S0213-005X(13)70111-3)

Murphy, N., Rooney, B., Bhattacharyya, T., Triana-Chavez, O., Krueger, A., Haslam, S. M., O'Rourke, V., Pańczuk, M., Tsang, J., Bickford-Smith, J., Gilman, R. H., Tetteh, K., Drakeley, C., Smales, C. M., & Miles, M. A. (2020). Glycosylation of *Trypanosoma cruzi* TcI antigen reveals recognition by chagasic sera. *Scientific Reports* 2020 10:1, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-73390-9>

Mylemans, B., Voet, A. R., & Tame, J. R. (2021). The Taming of the Screw: the natural and artificial development of β -propeller proteins. *Current Opinion in Structural Biology*, 68, 48–54. <https://doi.org/10.1016/J.SBI.2020.11.009>

Nakayasu, E. S., Burnet, M. C., Walukiewicz, H. E., Wilkins, C. S., Shukla, A. K., Brooks, S., Plutz, M. J., Lee, B. D., Schilling, B., Wolfe, A. J., Müller, S., Kirby, J. R., Rao, C. V., Cort, J. R., & Payne, S. H. (2017). Ancient Regulatory Role of Lysine Acetylation in Central Metabolism. *mBio*, 8(6), e01894-17. <https://doi.org/10.1128/mBio.01894-17>

Niyogi, S., Mucci, J., Campetella, O., & Docampo, R. (2014). Rab11 regulates trafficking of trans-sialidase to the plasma membrane through the contractile vacuole complex of *Trypanosoma cruzi*. *PLoS Pathogens*, 10(6). <https://doi.org/10.1371/JOURNAL.PPAT.1004224>

Nothwehr, S. F., & Gordon, J. I. (1990). Targeting of proteins into the eukaryotic secretory pathway: signal peptide structure/function relationships. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 12(10), 479–484. <https://doi.org/10.1002/BIES.950121005>

Nunes, M. C. P., Dones, W., Morillo, C. A., Encina, J. J., & Ribeiro, A. L. (2013). Chagas disease: an overview of clinical and epidemiological aspects. *Journal of the American College of Cardiology*, 62(9), 767–776. <https://doi.org/10.1016/J.JACC.2013.05.046>

Niborski, L. L., Potenza, M., Chirivi, R. G. S., Simonetti, L., Ossowski, M. S., Grippo, V., May, M., Staquicini, D. I., Parodi-Talice, A., Robello, C., Comini, M. A., Alonso, G. D., Raats, J. M. H. Gómez, K. A. (2021). Recombinant antibody against *Trypanosoma cruzi* from patients with chronic Chagas heart disease recognizes mammalian nervous system.

Ofran, Y., Schlessinger, A., & Rost, B. (2008). Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes.

Journal of Immunology (Baltimore, Md. : 1950), 181(9), 6230–6235.
<https://doi.org/10.4049/JIMMUNOL.181.9.6230>

Opperdoes, F. R., Butenko, A., Flegontov, P., Yurchenko, V., & Lukeš, J. (2016). Comparative Metabolism of Free-living Bodo saltans and Parasitic Trypanosomatids. Journal of Eukaryotic Microbiology, 63(5), 657–678. <https://doi.org/10.1111/JEU.12315>

Organización Mundial de la Salud, (2021). Chagas disease: control and elimination. <https://www.who.int/chagas/resources/en>

Oubrie, A., Rozeboom, H. J., Kalk, K. H., Huizinga, E. G., & Dijkstra, B. W. (2001). Crystal structure of quinohemoprotein alcohol dehydrogenase from Comamonas testosteroni: structural basis for substrate oxidation and electron transfer. The Journal of Biological Chemistry, 277(5), 3727–3732. <https://doi.org/10.1074/JBC.M109403200>

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. European Journal of Cell Biology, 97(6), 422–441. <https://doi.org/10.1016/J.EJCB.2018.06.003>

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. European Journal of Cell Biology, 97(6), 422–441. <https://doi.org/10.1016/J.EJCB.2018.06.003>

Pérez-Mazliah, D., Ward, A. I., & Lewis, M. D. (2021). Host-parasite dynamics in Chagas disease from systemic to hyper-local scales. Parasite Immunology, 43(2). <https://doi.org/10.1111/PIM.12786>

Peterson, J. H., Woolhead, C. A., & Bernstein, H. D. (2003). Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. The Journal of Biological Chemistry, 278(46), 46155–46162. <https://doi.org/10.1074/JBC.M309082200>

Queiroz, R. M. L., Charneau, S., Bastos, I. M. D., Santana, J. M., Sousa, M. v., Roepstorff, P., & Ricart, C. A. O. (2014). Cell surface proteome analysis of human-hosted Trypanosoma cruzi life stages. Journal of Proteome Research, 13(8), 3530–3541. <https://doi.org/10.1021/PR401120Y>

Rassi Jr., A., Rassi, A., Marin-Neto, J.A., 2010. Chagas disease. Lancet 375, 1388e1402.

- Ren, J., Gao, X., Jin, C., Zhu, M., Wang, X., Shaw, A., Wen, L., Yao, X., & Xue, Y. (2009). Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *Proteomics*, 9(12), 3409–3412.
- Roy A., Kucukura A., Zhang Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5: 725-738.
- Rozeboom, H. J., Yu, S., Mikkelsen, R., Nikolaev, I., Mulder, H. J., & Dijkstra, B. W. (2015). Crystal structure of quinone-dependent alcohol dehydrogenase from *Pseudogluconobacter saccharoketogenes*. A versatile dehydrogenase oxidizing alcohols and carbohydrates. *Protein Science : A Publication of the Protein Society*, 24(12), 2044–2054.
- Rubinstein, N. D., Mayrose, I., Halperin, D., Yekutieli, D., Gershoni, J. M., & Pupko, T. (2008). Computational characterization of B-cell epitopes. *Molecular Immunology*, 45(12), 3477–3489. <https://doi.org/10.1016/J.MOLIMM.2007.10.016>
- Saha S., Raghava G.P.S. (2004) BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. In: Nicosia G., Cutello V., Bentley P.J., Timmis J. (eds) *Artificial Immune Systems. ICARIS 2004. Lecture Notes in Computer Science*, vol 3239. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30220-9_16
- Saha, S., & Raghava, G. P. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, 65(1), 40–48. <https://doi.org/10.1002/prot.21078>
- Sant’Anna, C., Nakayasu, E. S., Pereira, M. G., Lourenço, D., de Souza, W., Almeida, I. C., & Cunha-e-Silva, N. L. (2009). Subcellular proteomics of *Trypanosoma cruzi* reservosomes. *Proteomics*, 9(7), 1782–1794. <https://doi.org/10.1002/PMIC.200800730>
- Sarmiento-Pavía, P. D., & Sosa-Torres, M. E. (2021). Bioinorganic insights of the PQQ-dependent alcohol dehydrogenases. *JBIC Journal of Biological Inorganic Chemistry* 2021 26:2, 26(2), 177–203. <https://doi.org/10.1007/S00775-021-01852-0>
- Sarmiento-Pavía, P. D., & Sosa-Torres, M. E. (2021). Bioinorganic insights of the PQQ-dependent alcohol dehydrogenases. *JBIC Journal of Biological Inorganic Chemistry* 2021 26:2, 26(2), 177–203. <https://doi.org/10.1007/S00775-021-01852-0>

Savojardo, C., Martelli, P. L., Fariselli, P., & Casadio, R. (2018). DeepSig: Deep learning improves signal peptide detection in proteins. *Bioinformatics*, 34(10), 1690–1696. <https://doi.org/10.1093/bioinformatics/btx818>

Savojardo, C., Martelli, P. L., Fariselli, P., Profiti, G., & Casadio, R. (2018). BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Research*, 46(W1), W459–W466. <https://doi.org/10.1093/nar/gky320>

Sbicego S, Schnauffer A, & Blum B (1998) Transient and stable transfection of Leishmania by particle bombardment. *Molecular and Biochemical Parasitology* 94 (1998) 123–126

Schmidt FR (2004) Recombinant expression systems in the pharmaceutical industry. *Appl Microbiol Biotechnol* 65: 363-372.

Schrödinger, L., & DeLano, W. (2020). PyMOL. <http://www.pymol.org/pymol>

Shijman, A., Burgos, J & Marcet, P. (2017). Molecular Tools and Strategies for Diagnosis of Chagas Disease and Leishmaniasis. *Frontiers in Parasitology Volume: 1 Molecular and Cellular Biology of Pathogenic Trypanosomatids*. Pp. 394-453 (60).

Sillitoe, I., Bordin, N., Dawson, N., Waman, V. P., Ashford, P., Scholes, H. M., Pang, C. S. M., Woodridge, L., Rauer, C., Sen, N., Abbasian, M., Le Cornu, S., Lam, S. D., Berka, K., Varekova, I. H., Svobodova, R., Lees, J., & Orengo, C. A. (2021). CATH: Increased structural coverage of functional space. *Nucleic Acids Research*, 49(D1), D266–D273. <https://doi.org/10.1093/nar/gkaa1079>

Silva-dos-Santos, D., Barreto-de-Albuquerque, J., Guerra, B., Moreira, O. C., Berbert, L. R., Ramos, M. T., Mascarenhas, B. A. S., Britto, C., Morrot, A., Serra Villa-Verde, D. M., Garzoni, L. R., Savino, W., Cotta-de-Almeida, V., & Meis, J. de. (2017). Unraveling Chagas disease transmission through the oral route: Gateways to *Trypanosoma cruzi* infection and target tissues. *PLOS Neglected Tropical Diseases*, 11(4), e0005507. <https://doi.org/10.1371/JOURNAL.PNTD.0005507>

Simpson, A. G. B., Stevens, J. R., & Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends in Parasitology*, 22(4), 168–174. <https://doi.org/10.1016/J.PT.2006.02.006>

Skalický, T., Dobáková, E., Wheeler, R. J., Tesařová, M., Flegontov, P., Jirsová, D., Votýpka, J., Yurchenko, V., Ayala, F. J., & Lukeš, J. (2017). Extensive flagellar remodeling during the complex life cycle of Paratrypanosoma, an early-branching trypanosomatid. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44), 11757–11762. https://doi.org/10.1073/PNAS.1712311114/SUPPL_FILE/PNAS.1712311114.SM02.AVI

Song, W. S., Cho, S. Y., Hong, H. J., Park, S. C., & Yoon, S. il. (2017). Self-Oligomerizing Structure of the Flagellar Cap Protein FliD and Its Implication in Filament Assembly. *Journal of Molecular Biology*, 429(6), 847–857. <https://doi.org/10.1016/J.JMB.2017.02.001>

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>

Stanaway, J.D., Roth, G., 2015. The burden of Chagas disease: estimates and challenges. *Glob. Heart* 10, 139e144.

Stanley P, Taniguchi N, Aebi M. N-Glycans. 2017. In: Varki A, Cummings RD, Esko JD, *et al.*, editors. *Essentials of Glycobiology* [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 9. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK453020/> doi: 10.1101/glycobiology.3e.009

Steenftoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. *EMBO J*, 32(10):1478-88, May 15, 2013.(doi: 10.1038/emboj.2013.79. Epub 2013 Apr 12)

Stevens, J. R. (2008). Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite* (Paris, France), 15(3), 226–232. <https://doi.org/10.1051/PARASITE/2008153226>

Takagi, Y., Sato, M., Naya, M., & Sato, C. (2022). Differentiating *Trypanosoma cruzi* in a Host Mammalian Cell Imaged in Aqueous Liquid by Atmospheric Scanning Electron Microscopy. *Microbiology Spectrum*, 10(1). <https://doi.org/10.1128/SPECTRUM.01413-21>

Takeda K, Ishida T, Yoshida M, Samejima M, Ohno H, Igarashi K, Nakamura N. 2019. Crystal structure of the catalytic and cytochrome b domains in a eukaryotic pyrroloquinoline quinone-dependent dehydrogenase. *Appl Environ Microbiol* 85:e01692-19.

Teixeira, D. E., Benchimol, M., Crepaldi, P. H., & de Souza, W. (2012). Interactive Multimedia to Teach the Life Cycle of *Trypanosoma cruzi*, the Causative Agent of Chagas Disease. *PLOS Neglected Tropical Diseases*, 6(8), e1749. <https://doi.org/10.1371/JOURNAL.PNTD.0001749>

Pons, T., Gómez, T., Chinaa, G., and Valencia, A. (2003) “ Beta-propellers: Associated Functions and their Role in Human Diseases”, *Current Medicinal Chemistry*. 10: 505. <https://doi.org/10.2174/09298670333368204>

Toyama, H., Mathews, F. S., Adachi, O., & Matsushita, K. (2004). Quinohemoprotein alcohol dehydrogenases: structure, function, and physiology. *Archives of Biochemistry and Biophysics*, 428(1), 10–21. <https://doi.org/10.1016/J.ABB.2004.03.037>

Toyama, H., Mathews, F. S., Adachi, O., & Matsushita, K. (2004). Quinohemoprotein alcohol dehydrogenases: structure, function, and physiology. *Archives of biochemistry and biophysics*, 428(1), 10–21. <https://doi.org/10.1016/j.abb.2004.03.037>

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., & Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research*, 43(W1), W401–W407. <https://doi.org/10.1093/NAR/GKV485>

Ulrich, P. N., Jimenez, V., Park, M., Martins, V. P., Atwood, J., Moles, K., Collins, D., Rohloff, P., Tarleton, R., Moreno, S. N. J., Orlando, R., & Docampo, R. (2011). Identification of contractile vacuole proteins in *Trypanosoma cruzi*. *PloS One*, 6(3). <https://doi.org/10.1371/JOURNAL.PONE.0018013>

Valente, M., Castillo-Acosta, V. ctor M., Vidal, A. E., & Gonzalez-Pacanowska, D. (2019). Overview of the role of kinetoplastid surface carbohydrates in infection and host cell invasion: prospects for therapeutic intervention. *Parasitology*, 146(14), 1743–1754. <https://doi.org/10.1017/S0031182019001355>

Vemuluri, V. R., Shaw, S., Autenrieth, C., & Ghosh, R. (2017). A rapid procedure for the in situ assay of periplasmic, PQQ-dependent methanol dehydrogenase in intact single bacterial

colonies. Journal of microbiological methods, 137, 46–49.
<https://doi.org/10.1016/j.mimet.2017.03.013>

von Heijne, G. (1985). Signal sequences. The limits of variation. Journal of Molecular Biology, 184(1), 99–105. [https://doi.org/10.1016/0022-2836\(85\)90046-4](https://doi.org/10.1016/0022-2836(85)90046-4)

Wan, S., Mak, M. W., & Kung, S. Y. (2012). MGOASVM: Multi-label protein subcellular localization based on gene ontology and support vector machines. BMC Bioinformatics, 13(1), 1–16. <https://doi.org/10.1186/1471-2105-13-290>

Wang, W., Peng, D., Baptista, R. P., Li, Y., Kissinger, J. C., & Tarleton, R. L. (2021). Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease. PLOS Pathogens, 17(1), e1009254. <https://doi.org/10.1371/JOURNAL.PPAT.1009254>

White, S., Boyd, G., Mathews, F. S., Xia, Z. xiang, Dai, W. wen, Zhang, Y. fan, & Davidson, V. L. (2002). The active site structure of the calcium-containing quinoprotein methanol dehydrogenase. Biochemistry, 32(48), 12955–12958. <https://doi.org/10.1021/BI00211A002>

Xu D. & Zhang Y. (2011). Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. Biophysical Journal, vol 101, 2525-2534.

Yang, J., Yan, R., Roy A., Xu D., Poisson J., Zhang Y. (2015). The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8

Yao B., Zhang L., Liang S., C. Zhang C. SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PLoS ONE (2012); 7(9):e45152.

Zalucki, Y. M., & Jennings, M. P. (2017). Signal peptidase I processed secretory signal sequences: Selection for and against specific amino acids at the second position of mature protein. Biochemical and Biophysical Research Communications, 483(3), 972–977. <https://doi.org/10.1016/J.BBRC.2017.01.044>

Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9:40

Zhang Y., & Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins*, 57(4), 702–710.

Zhang, W. X., Pan, X., & Shen, H. Bin. (2020). Signal-3L 3.0: Improving Signal Peptide Prediction through Combining Attention Deep Learning with Window-Based Scoring. *Journal of Chemical Information and Modeling*, 60(7), 3679–3686. <https://doi.org/10.1021/acs.jcim.0c00401>

Zhao J. (2007). Sumoylation regulates diverse biological processes. *Cellular and molecular life sciences : CMLS*, 64(23), 3017–3033. <https://doi.org/10.1007/s00018-007-7137-4>

Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., Liu, Z., Zhao, Y., Xue, Y., & Ren, J. (2014). GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic acids research*, 42(Web Server issue), W325–W330.

Zingales, B., Andrade, S. G., Briones, M. R. S., Campbell, D. A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A. M., Machado, C. R., Miles, M. A., Romanha, A. J., Sturm, N. R., Tibayrenc, M., & Schijman, A. G. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memorias Do Instituto Oswaldo Cruz*, 104(7), 1051–1054. <https://doi.org/10.1590/S0074-02762009000700021>

Zingales, B., Miles, M. A., Campbell, D. A., Tibayrenc, M., Macedo, A. M., Teixeira, M. M. G., Schijman, A. G., Llewellyn, M. S., Lages-Silva, E., Machado, C. R., Andrade, S. G., & Sturm, N. R. (2012). The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 12(2), 240–253. <https://doi.org/10.1016/J.MEEGID.2011.12.009>