

UNIVERSIDAD DE LOS ANDES
FACULTAD DE CIENCIAS
DEPARTAMENTO DE BIOLOGÍA

**Análisis inmunoinformático y evolutivo de proteínas de
Trypanosoma cruzi para el diseño de una vacuna de
subunidades multiepítopes**

www.bdigital.ula.ve

Trabajo especial de grado presentado por
María Virginia Ramírez Montoya
Como requisito para optar al título de
Licenciada en Biología

Tutor: Prof. Ascanio Rojas
Tutor académico: Prof. Héctor Acosta
Mérida, Noviembre de 2019

A mis padres

www.bdigital.ula.ve

Reconocimiento-No comercial- Compartir igual

Agradecimientos

A mis padres, familiares, pareja, y amigos por su apoyo y comprensión. A mis tutores, los profesores Ascanio Rojas y Héctor Acosta, por su esfuerzo, tiempo y generosa labor como guías de esta investigación. A mis jurados, los profesores Wilfredo Quiñonez, Ana Cáceres y Amaranta Gómez, por sus pertinentes comentarios, correcciones y recomendaciones.

A Gisela Rodríguez, por su colaboración en la recolección de datos. Al Centro Nacional de Cálculo Científico, lugar en el que fue posible la elaboración de este trabajo. A la Universidad de Los Andes, la Facultad de Ciencias y a mis profesores, por los años de aprendizaje y crecimiento científico-humanístico.

www.bdigital.ula.ve

Resumen

Trypanosoma cruzi es el agente causal de la enfermedad de Chagas o tripanosomiasis americana, zoonosis que si bien afecta a millones de personas, no cuenta con estrategias de prevención diferentes al control vectorial. El estudio de proteínas de interés inmunológico de *T. cruzi* ha llevado a la determinación de posibles blancos para el diseño de vacunas preventivas y terapéuticas. En este sentido, la inmunoinformática y los análisis evolutivos representan útiles herramientas para la detección de posibles sitios específicos de reconocimiento inmune con el potencial de inducir respuestas protectoras. Mediante la evaluación de métodos inmunoinformáticos, en este trabajo se encontró que los predictores de epítopes lineales de células B, ABCPred y BepiPred 2.0, tuvieron mejor desempeño para el reconocimiento de epítopes de *T. cruzi* con un porcentaje promedio de detección de 94,35% y 79,57%, respectivamente. Mientras que, los predictores de péptidos de unión a MHC-I con mejor desempeño, juzgado por valores de AUC, fueron: SMM (0,968), NetMHCpan 3.0 (0,962), NetMHC 4.0 (0,957) y NetMHCpan 4.0 (0,955). Por otra parte, en esta investigación se aplicaron los modelos evolutivos FEL y MEME para la determinación de las presiones selectivas actuando en proteínas inmunogénicas del parásito, donde se detectaron sitios bajo selección positiva y negativa. Una de estas proteínas, HSP70, fue modelada para visualizar el grado de solapamiento de epítopes predichos y confirmados y los sitios bajo selección negativa. Por otro lado, se seleccionaron siete proteínas de *T. cruzi*: Cz, Tc80, TSA-1, TSSA, PPK, LYT-1 y Tc24, que constituyeron el grupo de proteínas problemas a partir de las cuales se diseñó el prototipo inmunoestimulante. Para esto, las proteínas fueron analizadas con los algoritmos de predicción mencionados y con NetMHCII para la detección de péptidos de unión a MHC-II. Adicionalmente, las mismas fueron modeladas para localizar los epítopes de células B predichos. Igualmente, estuvieron sujetas a las pruebas evolutivas con las que se determinaron los sitios bajo selección presentes en los péptidos predichos. Integrando los resultados derivados de este análisis, de las pruebas inmunoinformáticas y de los modelos de la estructura tridimensional de las proteínas, se seleccionó una serie de péptidos para la construcción del péptido quimérico. En total se emplearon 12 epítopes lineales de células B, 12 péptidos de unión a MHC-I, 13 péptidos de unión a MHC-II, la secuencia PADRE y péptidos de conexión, para la construcción de una quimera compuesta de 631 aminoácidos, que en conjunto con el adyuvante c-di-AMP representó el prototipo inmunoestimulante propuesto.

Palabras clave: Inmunoinformática, presión selectiva, respuesta inmunológica, modelo de proteínas, vacunas, enfermedad de Chagas, *Trypanosoma cruzi*.

INDICE GENERAL

I.	Introducción.....	1
	Enfermedad de Chagas y biología de su agente causal	1
	Respuesta inmunológica.....	2
	Genotipos de <i>T. cruzi</i> , distribución geográfica y patologías asociadas.....	3
	Perfil genético HLA en Venezuela.....	4
	Estrategias de vacunación.....	5
	La inmunoinformática.....	6
	Predicción de epítopes de células B.....	7
	Predicción de péptidos de unión a moléculas HLA.....	7
	Componentes del prototipo inmunoestimulante.....	8
	Enfoque evolutivo.....	10
	Justificación.....	10
II.	Hipótesis.....	11
III.	Objetivos.....	11
	Objetivo general.....	11
	Objetivos específicos.....	11
IV.	Metodología.....	12
	Evaluación de predictores.....	12
	Enfoque evolutivo.....	14
	Elección y análisis de proteínas problemas.....	14
	Modelado de proteínas de <i>T. cruzi</i>	15
	Diseño del prototipo inmunoestimulante.....	16
V.	Resultados.....	17
	Evaluación de predictores.....	17
	Enfoque evolutivo.....	20
	Elección y análisis de proteínas problemas.....	22
	Modelado de proteínas de <i>T. cruzi</i>	27
	Diseño del prototipo inmunoestimulante.....	40
VI.	Discusión.....	45
	Evaluación de predictores.....	45
	Enfoque evolutivo.....	46
	Elección y análisis de proteínas problemas.....	48

Modelado de proteínas de <i>T. cruzi</i>	49
Diseño del prototipo inmunoestimulante	50
VII. Bibliografía	54

www.bdigital.ula.ve

ÍNDICE DE FIGURAS

Figura 1. Representación del porcentaje de epítopes confirmados de <i>T. cruzi</i> que fueron predichos por proteína con cada predictor.....	18
Figura 2. Curva ROC de cada predictor de péptidos de unión a MHC-I evaluado.....	20
Figura 3. (A) Representación de los sitios bajo selección encontrados en los epítopes lineales de células B de proteínas de <i>T. cruzi</i> . (B) Señalamiento de los sitios bajo selección hallados en los epítopes de células T de proteínas de <i>T. cruzi</i> . Flechas negras, selección negativa; flechas blancas, selección positiva; HSP70, proteína de choque térmico 70; MASP, proteína de superficie asociada a mucina; CRA, antígeno citoplasmático repetitivo; TS, transialidasa; CA-2, antígeno de superficie 2.	22
Figura 4. Gráfico de Ramachandran del modelado de la proteína HSP70 antes (izquierda) y después del refinamiento (derecha).	28
Figura 5. Modelado de la proteína HSP70 de <i>T. cruzi</i>	29
Figura 6. Comparación de la estructura de las proteínas problemas, antes y después de los refinamientos. A, Cz; B, TSA-1; C, Tc24; D, TSSA; E, Tc80; F, PPDK; G, LYT-1.....	32
Figura 7. Gráficos de Ramachandran de los modelados de las proteínas problema antes (izquierda) y después de los refinamientos (derecha). A, Cz; B, TSA-1; C, Tc24; D, TSSA; E, Tc80; F, PPDK; G, LYT-1.	34
Figura 8. Vista ampliada del modelo de la proteína Tc80 donde se indican por color los epítopes lineales de células B y los sitios bajo selección negativa en ellos contenidos. Con óvalos rojos se señalan epítopes descartados para su incorporación en el péptido quimérico.	37
Figura 9. Modelos de las proteínas problema con el señalamiento por color de los epítopes lineales de células B y los sitios bajo selección negativa. Con óvalos rojos se indican epítopes descartados para su incorporación en el péptido quimérico. A, cruzipaina; B, LYT-1, C, TSA-1, D, Tc24; E, Tc80; F, TSSA; G, PPDK.	38
Figura 10. Secuencia aminoacídica del péptido quimérico. En naranja, péptidos de unión a MHC-I; en azul, péptidos de unión a MHC-II; en morado, epítopes lineales de células B; en verde, péptido PADRE; en negro, <i>linkers</i>	42
Figura 11. Resultados de las pruebas de alergenicidad del péptido quimérico.....	42
Figura 12. Comparación de la estructura del péptido quimérico antes y después de los refinamientos. ...	43
Figura 13. Gráficos de Ramachandran del modelo del péptido quimérico antes (izquierda) y después de los refinamientos (derecha).	43
Figura 14. Modelo en cintas del péptido quimérico indicando sus componentes.....	44
Figura 15. Modelo del péptido quimérico con esferas de superficie señalando los epítopes lineales de células B.....	45

INDICE DE TABLAS

Tabla 1. Cobertura promedio del total de epítopos confirmados de <i>T. cruzi</i> que fueron predichos por cada método.	19
Tabla 2. Valores de área bajo la curva (AUC) provenientes del análisis ROC aplicado a cada método de predicción de péptidos de unión a MHC-I.	20
Tabla 3. Proteínas a considerar en el diseño de la vacuna de subunidades multiepítopos	23
Tabla 4. Listado y categorización de los péptidos predichos de cada proteína problema.	24
Tabla 5. Porcentaje de aminoácidos de HSP70 en las regiones del gráfico de Ramachandran antes y después del refinamiento del modelo.	28
Tabla 6. Programas de modelado y refinamiento utilizados en cada proteína problema y características de los resultados.	30
Tabla 7. Porcentaje de aminoácidos de las proteínas problema en las regiones del gráfico de Ramachandran antes y después de los refinamientos.	31
Tabla 8. Péptidos incluidos en el diseño del péptido quimérico.	40
Tabla 9. Porcentaje de aminoácidos de HSP70 en las regiones del gráfico de Ramachandran antes y después del refinamiento del modelo.	44

www.bdigital.ula.ve

I. INTRODUCCIÓN

Enfermedad de Chagas y biología de su agente causal

La enfermedad de Chagas o tripanosomiasis americana es una zoonosis causada por el protozooario flagelado *Trypanosoma cruzi*, y es reconocida por la Organización Mundial de la Salud como una de las 16 enfermedades tropicales olvidadas que afecta de 8 a 10 millones de personas alrededor del mundo, principalmente en América Latina donde la enfermedad es endémica (Rassi Jr *et al.*, 2010).

Se estima que ocurren de 10.000 a 14.000 muertes anuales por esta enfermedad, y que entre 65 y 100 millones de personas están en riesgo de contraerla. Propensión que se deriva, en cierta parte, por la ocupación humana de zonas rurales donde la frecuencia de encuentro con un vector infectado es mayor. Sin embargo, se ha comprobado que la colonización de zonas urbanizadas por parte de los vectores está induciendo un aumento en el número de personas expuestas (Arriagada, 2014; Martínez-Campos *et al.*, 2015). Adicionalmente, debido a las altas tasas de migración, es sabido que la enfermedad se encuentra en creciente expansión más allá de América Latina, específicamente en Estados Unidos y varios países europeos y asiáticos (Schmunis and Yadon, 2010).

En América Latina la enfermedad de Chagas es considerada como la infección parasitaria con mayor carga económica debido a su prolongada cronicidad, generando una pérdida anual en la economía de 7,2 millardos de dólares (Martínez-Campos *et al.*, 2015). A pesar de ello, existe una actitud pasiva ante esta enfermedad provocada por el desconocimiento de su extensión, debiéndose mayormente a la escasa detección de casos, ya que solo suelen notificarse aquellos en los que las personas ya han desarrollado una patología concreta (Pérez-Aguilar *et al.*, 2018). Situación que genera la percepción errónea de su verdadero impacto, que refuerza la idea de una enfermedad silenciada, olvidada y que conlleva a una subestimación para el diseño y desarrollo de medidas de prevención y tratamiento (Pérez-Aguilar *et al.*, 2018).

A pesar del tiempo transcurrido desde su descubrimiento (1909) (Arriagada, 2014), aún no se dispone de una terapia eficaz para todas las fases clínicas de la enfermedad de Chagas. Los medicamentos utilizados actualmente para la eliminación de *T. cruzi* son de relativa eficacia y tampoco existen vacunas para prevenir la enfermedad. Adicionalmente, la producción de nuevas drogas para poblaciones percibidas en principio como relativamente pequeñas o de bajo nivel económico es desechada o postergada (Pérez-Aguilar *et al.*, 2018).

Por su parte, su agente causal, *T. cruzi*, es un parásito hemoflagelado perteneciente al filo Euglenozoa, orden Kinetoplastida, y familia Trypanosomatidae (Domingo, 2004). Se caracteriza por poseer un único flagelo y una estructura intracelular denominada mitocondrion, donde se almacena una compleja red de ADN extra-nuclear compactado llamado kinetoplasto. Adicionalmente, se caracteriza por presentar compartimientos intracelulares especializados tipo peroxisomas, denominados glicosomas, así como también mecanismos genéticos particulares (Michels *et al.*, 2006; Teixeira *et al.*, 2011).

Este organismo posee un ciclo de vida digenético, en el que desarrolla distintos estadios donde experimenta cambios morfológicos que varían de acuerdo a los hábitats que ocupa dentro del vector, triatomino pertenecientes a la familia Reduviidae y su huésped mamífero (Bontempi, 2015). *T. cruzi* presenta en total tres estadios: epimastigote, amastigote y tripomastigote (metacíclico y sanguíneo). El primero se encuentra en el vector y corresponde a una forma replicativa no infecciosa; la segunda es la forma replicativa dentro del huésped mamífero; y las últimas son formas altamente infecciosas presentes tanto en el vector como en el hospedero vertebrado (Bontempi, 2015).

Además de la vía vectorial, la enfermedad de Chagas puede transmitirse de manera congénita, por transfusión sanguínea, mediante el trasplante de órganos de personas infectadas y por el consumo de alimentos contaminados con heces de un triatomino infectado (Khatoon *et al.*, 2018). Una vez ocurrida la exposición, el periodo que transcurre desde que *T. cruzi* ingresa a su huésped y se disemina, generando un alto número de parásitos en sangre y tejidos, corresponde al primer periodo de la enfermedad, la fase aguda, de una duración aproximada de 2 a 4 meses, y que puede variar dependiendo de la cepa y el individuo (Elisei *et al.*, 2018). Posteriormente, el desarrollo de la infección continúa hacia la fase crónica, donde la parasitemia es mucho menor y la respuesta humoral es altamente persistente (Elisei *et al.*, 2018). En los seres humanos, la mayoría de personas infectadas en fase crónica, de 60 a 70% de los casos, son asintomáticos y se denominan pacientes indeterminados, por lo que la infección solamente puede determinarse mediante análisis serológicos (Bontempi, 2015). No obstante, en el resto de los casos, los individuos manifiestan un cuadro clínico severo caracterizado por afecciones cardíacas, digestivas, o ambas. La primera puede desencadenar la aparición de Cardiopatía Chagásica Crónica (CCC), que se caracteriza por provocar insuficiencia cardíaca congestiva, tromboembolias y muerte cardíaca súbita, y puede evidenciarse en un lapso que varía de 5 años hasta más de 20 años. En el caso de la forma digestiva, los individuos desarrollan los denominados megasíndromes (megaesófago y megacolon) (Bontempi, 2015).

Hasta el momento, las únicas drogas disponibles para el tratamiento etiológico de la enfermedad de Chagas son nitrofuranos y nitroimidazoles (Bahia *et al.*, 2014; Murcia *et al.*, 2013). Aunque estos compuestos exhiben efectos tripanocidas ninguno cumple con los preceptos de un buen fármaco según criterios de la OMS, debido a que son poco efectivos en los casos crónicos, los pacientes requieren largos periodos de tratamiento bajo vigilancia médica, son de alto costo y baja accesibilidad, producen importantes efectos secundarios y algunas cepas de *T. cruzi* son resistentes a ellos (Urbina, 2010).

Por otro lado, en cuanto a los mecanismos de prevención de la enfermedad, la mayoría de los esfuerzos se han centrado en el control del vector, lo que si bien ha conseguido una reducción importante de la incidencia, la erradicación de la enfermedad aún no ha sido alcanzada, es por ello que se sigue considerando una amenaza para la salud pública (Avenidaño-Rangel and Rey, 2016; Quijano-Hernandez and Dumonteil, 2011). Por esta razón, y debido a la ausencia de un tratamiento 100% eficaz e inocuo, es necesario el desarrollo de una estrategia preventiva que conlleve al control de esta enfermedad. En este contexto, la vacunación representa un mecanismo clave para suplir esta necesidad.

Respuesta inmunológica

Para desarrollar una vacuna efectiva, es fundamental la comprensión de las respuestas inmunitarias que se derivan de la infección con *T. cruzi*, ya que de esta manera sería posible la inducción de una respuesta eficaz a los antígenos suministrados. En relación, se ha establecido que la respuesta inmune innata, así como la respuesta inmune adaptativa, son mecanismos esenciales en el control de la parasitemia, donde se requieren los efectos combinados de diversas poblaciones celulares (Quijano-Hernandez and Dumonteil, 2011).

Desde las primeras investigaciones se ha destacado el papel que desempeñan los linfocitos CD4⁺ de la subpoblación Th1 en producir IL-2 e IFN- γ para modular la estimulación y expansión de linfocitos CD8⁺, células cruciales para la protección sistémica contra *T. cruzi* (Aujla *et al.*, 2008). Sin embargo, también se ha encontrado que una respuesta polarizada hacia el tipo Th1, está asociada a daño inflamatorio, y por ende, a un aumento en la susceptibilidad a la infección parasitaria. En concordancia, investigaciones señalan que durante la infección con *T. cruzi* debe producirse un balance entre la producción de citoquinas inflamatorias y anti-inflamatorias que permita controlar la parasitemia al mismo tiempo que evite el daño inflamatorio del huésped (Cai *et al.*, 2016; Tosello Boari *et al.*, 2012).

En este contexto, recientemente se ha demostrado el efecto de la familia de IL-17 en generar protección contra la infección con *T. cruzi* al mediar la respuesta inflamatoria evitando así el daño tisular colateral característicos (Tosello Boari *et al.*, 2012).

Debido a este nuevo aporte en la comprensión de la respuesta inmunitaria, otra investigación se encargó de dilucidar el rol de los linfocitos CD4⁺ Th17 como una de las poblaciones celulares encargadas de la producción de IL-17 (Cai *et al.*, 2016). Encontrando que durante la infección con *T. cruzi*, la IL-17 también induce el funcionamiento de la enzima NADPH oxidasa en macrófagos infectados, promoviendo así la eliminación de los parásitos (Cai *et al.*, 2016). Sin embargo, en la investigación no fue este el mecanismo que produjo el mayor aporte en la protección contra la parasitemia, ya que se demostró que las células Th17 son capaces de provocar la activación, mediante la secreción de IL-21, de los linfocitos T CD8⁺ (Cai *et al.*, 2016). No obstante, no solo se encontró que las células Th17 tienen esta propiedad, sino que aún más interesante, se determinó que tienen un efecto en la modulación de los linfocitos CD8⁺ aún más importante que el producido por las células Th1, a las que clásicamente se les atribuía el rol principal (Cai *et al.*, 2016). Este importante descubrimiento, tal como lo señalan los autores, hace que se considere a las células Th17, como un nuevo subconjunto de linfocitos T CD4⁺ a considerar en el diseño de una vacuna contra la enfermedad de Chagas (Cai *et al.*, 2016).

Genotipos de *T. cruzi*, distribución geográfica y patologías asociadas

Trypanosoma cruzi es un organismo que se caracteriza por presentar una importante plasticidad genética que se ve reflejada en su diversidad genotípica. Actualmente, su clasificación intraespecífica sigue una nomenclatura que divide los genotipos en siete unidades discretas de tipificación (UDTs): *T. cruzi* I (TcI), *T. cruzi* II (TcII), *T. cruzi* III (TcIII), *T. cruzi* IV (TcIV), *T. cruzi* V (TcV), *T. cruzi* VI (TcVI) y *T. cruzi* Bat (TcBat) (Jimenez *et al.*, 2019). En América, TcI ha sido reportada desde el norte de California hasta Argentina, por lo que es la UDT con mayor expansión geográfica (Breniere *et al.*, 2016; Dorn *et al.*, 2017; Shender *et al.*, 2016). También es el linaje predominante en los ciclos de transmisión selváticos, con similar importancia en los ciclos de transmisión domésticos. La infección en humanos por esta UDT se concentra principal, aunque no exclusivamente, en Norte y Centro América, así como en la región norte de América del Sur, siendo la cardiomiopatía su patología asociada (Breniere *et al.*, 2016; Jimenez *et al.*, 2019). Por su parte, TcII no ha tenido la misma capacidad de expansión, encontrándose casi exclusivamente en América del Sur, principalmente en Brasil, Colombia y Chile, presente en ciclos de transmisión domésticos (Breniere *et al.*, 2016). Esta UDT es responsable del desarrollo de megasíndromes aunque también se ha asociado con manifestaciones cardíacas (Jimenez *et al.*, 2019). TcIII y TcIV se encuentran generalmente en América del Sur, especialmente en Paraguay en el caso de TcIII y en la zona norte en el caso de TcIV. Aunque son linajes asociados a ciclos selváticos de transmisión, la presencia en humanos ha sido reportada, generando tanto cardiomiopatías como megasíndromes (Breniere *et al.*, 2016; Jimenez *et al.*, 2019). En cuanto a TcV y TcVI, son linajes encontrados principalmente en el sur de América, en países como Chile, Argentina y Paraguay, aunque también se ha documentado una presencia importante de TcV en Bolivia (Breniere *et al.*, 2016). Se encuentran restringidos a ciclos de transmisión doméstica con el potencial de causar megasíndromes y cardiomiopatías (Jimenez *et al.*, 2019). Por último TcBat, es un linaje que si bien era asociado exclusivamente a murciélagos, se ha encontrado infectando humanos en Brasil, Colombia, Ecuador y Panamá, reportándose fallecimientos en los dos primeros países (Breniere *et al.*, 2016; Jimenez *et al.*, 2019).

Ahora bien, en un estudio publicado para Venezuela por Carrasco *et al.* (2012) se encontró que el linaje predominante entre las distintos tipos de huésped era el TcI (94,1%), seguido por TcIV (3,1%) y TcIII (2,8%). Estando este último ausente en las muestras de humanos, donde el 79% de los individuos se

encontraron infectados con TcI y el 21% con TcIV. Estos resultados evidenciaron una incidencia importante de TcIV en el país, antes desconocida. Demostrando que es un linaje que también interviene en ciclos de transmisión domésticos, ya que los reportes indicaron una amplia distribución por todo el territorio nacional en regiones alejadas de ecotipos selváticos.

Como se ha visto, la identificación de linajes de *T. cruzi* es fundamental debido a que las patologías asociadas y los ciclos de transmisión varían de acuerdo a los genotipos, los que a su vez presentan una distribución geográfica definida. Aspectos de suma importancia a considerar en el diseño y la aplicación de estrategias de control y prevención de la enfermedad de Chagas, pues debe delimitarse el rango de efectividad de acuerdo a la región y a las UDTs presentes (Dorn *et al.*, 2017). Asimismo, la diversidad genética dentro de UDTs es otro factor relevante, ya que la alta variabilidad entre cepas de un mismo linaje ha sido reportada (Breniere *et al.*, 2016). La heterogeneidad de las cepas presentes en Venezuela es una característica no estudiada hasta la fecha, y considerando la diversidad genética encontrada en TcI (Breniere *et al.*, 2016), es fundamental evaluar en este estudio que exista una cobertura entre UDTs para garantizar la inclusión de las distintas cepas dentro de los linajes circulantes, y de esta manera garantizar teóricamente su efectividad.

Perfil genético HLA en Venezuela

En el diseño de un prototipo inmunoestimulante debe procurarse que tenga el potencial de generar protección contra determinada enfermedad con una cobertura poblacional adecuada, para esto, la consideración del perfil genético HLA es fundamental (Lasso *et al.*, 2016). Como se mencionó, durante la infección con *T. cruzi* el rol que cumplen los linfocitos T es primordial, y debido a que la activación de estas células depende de un proceso de presentación antigénica mediado por moléculas MHC-I y -II, es necesario el conocimiento del tipo y la representatividad de los haplotipos HLA en la población de estudio, a fin de asegurar el empleo de péptidos que sean reconocidos como epítopes por la variedad de alelos circulantes.

En Venezuela son escasos los estudios enfocados en la caracterización del perfil genético HLA de la población. Para el caso de los haplotipos HLA clase I solo destaca el trabajo realizado por Zoghbi *et al.* (2016), que determinaron la composición y frecuencia alélica en 765 individuos de 218 familias venezolanas, encontrando un total de 265 haplotipos HLA A-B, siendo los más frecuentes: HLA A*24-B*35 (11,98%), HLA A*02-B*51 (9,81%), HLA A*02-B*35 (8,30%), HLA A*02-B*39 (5,66%), HLA A*02-B*44 (5,28%), HLA A*11-B*35 (5,28%). Los autores explican que la alta frecuencia del haplotipo HLA A*24-B*35 obedece a la influencia amerindia en la población, ya que éstos representan alelos característicos de este grupo étnico. En el total de resultados también se evidencian alelos de origen europeo y africano, en correspondencia con los procesos de mestizaje que han tenido lugar en Venezuela.

Por otra parte, la composición de los haplotipos HLA clase II de la población venezolana no está determinada, no se encontraron reportes con una escala similar a la investigación anterior, ya que la mayoría de trabajos se han orientado únicamente en la caracterización de ciertos grupos indígenas, o de habitantes de zonas puntuales de algunos estados del país. Asimismo, gran parte de los estudios se han enfocado en la influencia de ciertos alelos en la susceptibilidad a enfermedades. Sin embargo, de estos trabajos puede obtenerse el listado de alelos identificados y su frecuencia, a fin de totalizarlos y tener una idea del perfil venezolano.

En relación, se puede señalar el trabajo conducido por (Pardo *et al.*, 2012), donde identificaron los haplotipos HLA clase I y II de una población mestiza del estado Zulia con una muestra de 40 individuos. Como en el estudio anterior, encontraron alta frecuencia de los supertipos HLA-A*02 (25%) y HLA-A*24 (13,75%), así como de HLA-B*35 (12,5%). También reportan altas frecuencias de

otros supertipos, como HLA-A*23 (16,25%), HLA-B*57 (21,25%). Adicionalmente y a diferencia del primer trabajo, los investigadores evaluaron los genes HLA-C, encontrando como predominante al alelo HLA-C*04 (22,5%). Ahora bien, en cuanto a los haplotipos clase II, estos autores hallaron una alta representatividad de los alelos HLA-DRB1*15 (22%), HLA-DRB1*03 (21,25%), HLA-DRB1*04 (12,5%), HLA-DQB1*06 (31%), HLA-DQB1*05 (25,2%) y HLA-DQB1*03 (23,7%).

Por otra parte, puede mencionarse el estudio de (Fortes Mdel *et al.*, 2007), donde determinaron en una muestra con representantes mestizos de todo el territorio nacional, las variaciones en la frecuencia alélica de haplotipos HLA entre 111 individuos sanos y 41 afectados por hepatitis autoinmune tipo 1. Para la clase II encontraron que en pacientes el alelo HLA-DRB1*03 tenía una representación del 26,83% y en individuos sanos del 8,41%, adicionalmente, estos autores hallaron una incidencia del alelo HLA-DRB1*1301 en individuos enfermos del 34,15%, y en individuos control del 7,48%. Por otro lado, reportaron una alta frecuencia del alelo DQB1*02 tanto en individuos sanos (32,38%) como enfermos (48,78%).

Estrategias de vacunación

A lo largo de los años este diseño ha llevado al estudio de una gran cantidad de antígenos que han sido aplicados mediante distintos tipos de vacunas elaboradas a partir de: parásitos completos, proteínas purificadas del parásito, proteínas recombinantes, ADN, y mediante virus y bacterias recombinantes (Quijano-Hernandez and Dumonteil, 2011). Para el primero de los casos se han utilizado parásitos muertos y vivos atenuados, siendo estos último los que consiguieron una mayor respuesta protectora contra la infección (Aujla *et al.*, 2008; Cai *et al.*, 2016). Sin embargo, este tipo de enfoque está en desuso debido al potencial riesgo que presentan la reversión a cepas virulentas y por el efecto que cepas atenuadas puedan tener en individuos inmunosuprimidos; asimismo, se suman los inconvenientes relacionados con su producción y suministro a gran escala (Bontempi, 2015; Quijano-Hernandez and Dumonteil, 2011). Por estos motivos, las investigaciones posteriores cambiaron de enfoque para la utilización de vacunas de subunidad, con proteínas purificadas a partir de cultivos de parásitos, como cruzipaina (Cz), proteínas del bastón paraflagelar y proteínas de secreción y excreción de tripomastigotes (Pham *et al.*, 2009; Zhumabekov *et al.*, 1995). Posteriormente, se procedió al uso de proteínas recombinantes como las proteínas GP90, GP82, GP56, Cz, algunos miembros de la familia de las transalidasas, entre otras (Quijano-Hernandez and Dumonteil, 2011; Vázquez-Chagoyán *et al.*, 2011). La utilización de estas vacunas no obtuvo resultados favorables debido a la baja inmunogenicidad que generaban, sin embargo, y como se tratará más adelante, se encontró que su aplicación conjugada con adyuvantes produce una mejora considerable en los resultados (Bontempi, 2015). Por otra parte, la vacunación con ADN, mediante virus y bacterias recombinante, también ha sido objeto de intensa experimentación, en la que se han utilizado nuevos antígenos, así como muchas de las proteínas mencionadas, como son: TSA1, ASP1, ASP2, KMP11, LYT1, Cz, PFR, CRP, entre otras. A pesar de las ventajas de su fabricación y de la generación de una respuesta celular importante, esta metodología acarrea consigo riesgos relacionados con eventos de recombinación e inserción en el genoma, por lo cual, el avance en la investigación para su uso en humanos comprende grandes restricciones (Bontempi, 2015; Quijano-Hernandez and Dumonteil, 2011).

Por estas razones, diversas investigaciones se han enfocado una vez más en el uso de vacunas de subunidad a través de la utilización de proteínas recombinantes, sumando la implementación de adyuvantes de nueva generación (Arce-Fonseca *et al.*, 2018; Bontempi, 2015; Carabarin-Lima *et al.*, 2014; Martinez-Campos *et al.*, 2015; Sanchez Alberti *et al.*, 2017). Pudiendo mencionarse el estudio realizado por Martinez-Campos *et al.* (2015), donde utilizaron la proteína recombinante de unión a calcio Tc24 del bolsillo flagelar de *T. cruzi* para el desarrollo de una vacuna con la adición del adyuvante MPL (Monofosforil Lípido A), consiguiendo una respuesta inmune polarizada hacia la

activación de células Th1, así como elevados niveles de anticuerpos IgG2a y de IFN- γ , produciendo una disminución estadísticamente significativa de la parasitemia, y por ende un aumento en la supervivencia de los ratones infectados. Por otra parte, otras investigaciones se han enfocado en el diseño de vacunas quiméricas utilizando distintas proteínas, o fragmentos de proteínas, al mismo tiempo; tal como se efectuó en el estudio conducido por Sanchez Alberti *et al.* (2017), quienes construyeron un antígeno quimérico al que denominaron Traspaina, que estaba formado por: el dominio N-terminal de la cruzipaina, la región central de la proteína 2 de superficie de amastigotes (ASP2), una transialidasa inactiva, y combinado con el adyuvante c-di-AMP. Formulación que produjo una respuesta humoral efectiva, así como una respuesta celular polarizada hacia la activación de células Th1 y Th17. Los resultados mostraron una reducción importante no solo de la parasitemia, sino también de la inflamación crónica.

Por otro lado, es pertinente mencionar el trabajo realizado por (Carabarin-Lima *et al.*, 2014), quienes amplificaron, clonaron y secuenciaron el gen codificante de la enolasa de *T. cruzi* (TcENO) a fin de expresar y purificar una TcENO recombinante. Los autores utilizaron ensayos *in silico* que demostraron que esta proteína poseía motivos característicos de enolasa, así como una región transmembrana. Además de esto, utilizando programas bioinformáticos de predicción de epítopes de linfocitos B y T, identificando regiones específicas en la proteína probablemente antigénicas, que sugerían el desarrollo de una respuesta humoral y celular por la exposición a esta proteína. En esta investigación, la presencia de anticuerpos anti-TcENO fue verificada en ratones infectados, así como en suero de pacientes con enfermedad de Chagas. Cuatro años después de esta investigación, los mismos autores condujeron un nuevo estudio donde emplearon TcENO en ratones a fin de estudiar su inmunogenicidad y la capacidad de generar protección contra la enfermedad. Detectando una respuesta inmune polarizada de tipo Th1, la reducción del número de parásitos en sangre, y un aumento en la supervivencia (Arce-Fonseca *et al.*, 2018).

Por otra parte, en lo que respecta a investigaciones netamente bioinformáticas, preliminares en el desarrollo de una vacuna preventiva contra la enfermedad de Chagas, puede señalarse el trabajo realizado por (Khatoon *et al.*, 2018), quienes diseñaron una vacuna de subunidades basada en la inmunogenicidad de los epítopes evaluados que según el análisis, serían capaces de generar respuestas humoral y celular. En total utilizaron un conjunto de 8 proteínas, que comprendía proteínas de secreción y de membrana que emplearon para construir un péptido combinado con el adyuvante RS09. Este tipo de enfoque ha sido aplicado para el diseño de vacunas contra una diversa gama de infecciones provocadas por parásitos, bacterias y virus (Backert and Kohlbacher, 2015; El-Manzalawy and Honavar, 2010; Khatoon *et al.*, 2017; Narula *et al.*, 2018).

La inmunoinformática

La inmunoinformática es una rama de la biología computacional dedicada a la resolución de problemas inmunológicos, que se ha enfocado principalmente en el desarrollo de predictores de epítopes utilizados en el diseño de vacunas, el inmunodiagnóstico, y en la inmunoterapia personalizada contra enfermedades como el cáncer (Backert and Kohlbacher, 2015). El avance de estas técnicas ha sido catalizado en gran medida por la necesidad de manejar la creciente cantidad de datos de interés inmunológico, lo que ha incentivado la creación de técnicas cada vez más sofisticadas y precisas que al mismo tiempo mejoran con la disponibilidad de más datos (Kazi *et al.*, 2018). La aplicación de estas metodologías representa un enfoque importante y necesario, ya que permite reducir el tiempo, los recursos y costos requeridos para la identificación experimental de epítopes mediante la disminución del repertorio de péptidos a evaluar (Backert and Kohlbacher, 2015; Kazi *et al.*, 2018). En vacunología, la inmunoinformática permite la predicción de epítopes a partir de datos genómicos, lo que propicia el diseño de vacunas basadas en epítopes. Este enfoque sostiene que la creación de una proteína quimérica

que contenga las regiones antigénicas representa un diseño racional que pudiese desencadenar respuestas inmunitarias humorales y celulares adecuadas contra determinado patógeno (Kazi *et al.*, 2018; Sanchez-Trincado *et al.*, 2017; Soria-Guerra *et al.*, 2015).

Predicción de epítopes de células B

Actualmente existen diversos métodos computacionales enfocados en la identificación de regiones proteicas con el potencial de actuar como epítopes lineales y conformacionales reconocidos por anticuerpos y receptores de células B (Backert and Kohlbacher, 2015; Kazi *et al.*, 2018). Este desarrollo comenzó en el siglo pasado con la creación de predictores basados en la categorización de aminoácidos de acuerdo a sus características fisicoquímicas, otorgándoles valores de propensión para actuar como epítopes (Backert and Kohlbacher, 2015; Kazi *et al.*, 2018; Sanchez-Trincado *et al.*, 2017). Métodos que obtuvieron resultados modestos de predicción y que llevaron a la creación de nuevos algoritmos con mejor desempeño. Así surgieron los métodos modernos basados en aprendizaje automatizado (*machine learning*), que varían de acuerdo a los datos utilizados para entrenar los algoritmos (Backert and Kohlbacher, 2015; Kazi *et al.*, 2018; Sanchez-Trincado *et al.*, 2017).

Estas herramientas han sido utilizadas en varias investigaciones de enfermedades infecciosas, enfocadas tanto en la inmunización como el serodiagnóstico. Tal es el caso del estudio realizado por Bastos *et al.* (2016), donde usando predictores seleccionaron péptidos que contenían epítopes de células B y T de la proteína GRA2 de *Toxoplasma gondii* que utilizaron para inmunizar ratones, consiguiendo una protección del 85% al exponerlos a una dosis letal del parásito. En cuanto al serodiagnóstico, puede mencionarse el estudio de Menezes-Souza *et al.* (2015), donde utilizaron un predictor de epítopes lineales en la proteína CatL para comprobar su eficacia en el serodiagnóstico de leishmaniasis, encontrando que el péptido seleccionado era específico, sensible y preciso en la detección de leishmaniasis visceral humana. Finalmente, Elisei *et al.* (2018) condujeron una investigación donde a partir del genoma completo de *Trypanosoma cruzi*, cepa CL Brenner, y utilizando un predictor de epítopes lineales, identificaron tres péptidos derivados de proteínas de superficie asociadas a mucinas que emplearon para el serodiagnóstico de este organismo, lo que produjo resultados importantes que mostraron su eficacia en la detección del parásito.

Predicción de péptidos de unión a moléculas HLA

Basado en el conocimiento de las interacciones fisicoquímicas que se presentan entre péptidos y moléculas MHC-I y -II se han desarrollado programas enfocados en la predicción de péptidos de unión dentro de determinada proteína. Estos algoritmos, aunque suelen denominarse predictores de epítopes de células T, realmente predicen la capacidad de unión de péptidos a moléculas HLA (Kazi *et al.*, 2018; Soria-Guerra *et al.*, 2015). Como en el caso anterior, el desarrollo de estos modelos comenzó en el siglo pasado a partir de datos que surgieron experimentalmente sobre péptidos conocidos de unión a HLA. Así, las predicciones se basaron en un primer momento en la detección de secuencias motivo encontradas en los péptidos, que se caracterizaban por contener residuos de anclaje (Backert and Kohlbacher, 2015; Kazi *et al.*, 2018; Soria-Guerra *et al.*, 2015). Luego, para considerar la contribución de los residuos totales, se desarrollaron nuevos métodos denominados “matrices de motivos”, que fueron superadas por las “matrices de afinidad cuantitativas”, que consideraban valores de afinidad de enlace (Desai and Kulkarni-Kale, 2014; Kazi *et al.*, 2018; Sanchez-Trincado *et al.*, 2017). Posteriormente, se desarrollaron nuevas metodologías que originaron los modelos de “relaciones de actividad estructural cuantitativa”, que tomaban en cuenta las interacciones entre cadenas laterales vecinas en la contribución individual de cada residuo. Estos avances son actualmente superados por los modelos basados en aprendizaje automatizado, que presentan algoritmos fundamentados en redes neuronales artificiales, máquinas de soporte vectorial (*support vector machines*), árboles de decisión o

cadenas ocultas de Markov, que han aumentado en gran medida la capacidad de detección con precisiones que alcanzan el 95% (Desai and Kulkarni-Kale, 2014; Kazi *et al.*, 2018; Sanchez-Trincado *et al.*, 2017). Hay que tomar en cuenta que uno de los principales desafíos en el desarrollo de modelos de predicción es la consideración de la inmensa variabilidad de las moléculas HLA. Sin embargo, debido a que se ha encontrado que distintas moléculas pueden unirse a conjuntos similares de péptidos permitiendo la categorización de alelos en súper tipos, la mayoría de predictores se han desarrollado de manera que permiten la predicción de acuerdo a los súper tipos de interés, e incluso muchas veces a nivel de alelos (Kazi *et al.*, 2018).

Entre los estudios donde se ha implementado este tipo de análisis *in silico* puede mencionarse el realizado por Egui *et al.* (2015), donde evaluaron la proteína TcCA-2 de *T. cruzi* con predictores de péptidos de unión a MHC-I (haplotipo HLA-A*02:01), probados en pacientes chagásicos crónicos y asintomáticos con la finalidad de encontrar posibles epítopes de células CD8⁺. Encontrando que cuatro de los ocho péptidos predichos fueron presentados y reconocidos por poblaciones de linfocitos en ambos tipos de pacientes. Los autores destacan la diferencia en las poblaciones de linfocitos T activados en cada caso: células T de memoria efectoras con características senescentes en pacientes crónicos, y células T vírgenes (*T naive*) en pacientes asintomáticos. Por lo que proponen la utilización de los epítopes como posibles biomarcadores de la progresión de la enfermedad, así como también plantean su potencial rol en inmunoterapia.

Otro de los estudios a mencionar fue el conducido por Duarte *et al.* (2015), donde se desarrolló un programa que compilaba los resultados de diversos predictores de péptidos de unión a MHC-I con la finalidad de hacer una evaluación en conjunto que permitiera una mejor elección de posibles epítopes de células CD8⁺. Para probar este programa los autores evaluaron 63 proteínas de *Leishmania braziliensis*, de donde seleccionaron ocho péptidos para una validación *in vivo*, encontrando que tres de los péptidos actuaron como péptidos de linfocitos CD8⁺, medido mediante la citotoxicidad inducida. Como en el apartado anterior, sobre los predictores de epítopes de células B, estos estudios son una muestra de la utilidad de la incorporación del análisis *in silico* en la selección de péptidos para facilitar el trabajo de validación experimental subsiguiente, lo que en último término representa el principal objetivo de esta investigación.

Componentes del prototipo inmunoestimulante

Además del conjunto de péptidos que puedan ser predichos con los algoritmos anteriormente descritos, un prototipo como el que se diseñó en este trabajo necesita de otros dos componentes: un adyuvante y secuencias de unión entre péptidos.

Los adyuvantes se definen como sustancias capaces de promover y dirigir una respuesta inmune adaptativa contra los antígenos empleados en una vacuna (Virgil and Lavelle, 2011). Por lo que se consideran de gran utilidad para el aumento de la inmunogenicidad de antígenos recombinantes, carentes de las moléculas de reconocimiento innato características de un patógeno (Reed *et al.*, 2013).

El empleo de adyuvantes tiene importantes beneficios en el diseño de vacunas, pues permite, entre otras cosas, la disminución en la concentración y el número de dosis efectivas, la disminución del tiempo de respuesta, así como la obtención del tipo de respuesta inmunitaria deseada para la eliminación del agente infeccioso (Reed *et al.*, 2013; Ruiz *et al.*, 2017). No obstante, dependiendo de la aplicación, los adyuvantes pueden o no actuar en la inmunomodulación, debido a que pueden ser usados únicamente como sistemas de envío que garanticen una presentación antigénica más efectiva. Sin embargo, también existen sustancias capaces de integrar todas estas cualidades (Reed *et al.*, 2013). Por lo tanto, la escogencia del tipo de adyuvante dependerá en gran medida del tipo de antígeno que se

utilice, por ejemplo, su uso en modulación inmune se hace redundante si la vacuna utiliza organismos completos (Nicholls *et al.*, 2010).

Por mucho tiempo después de su descubrimiento en 1920, los adyuvantes se elaboraban exclusivamente a base de sales de aluminio. No fue hasta 1990 cuando fue aprobado el empleo de emulsiones en humanos, específicamente, del adyuvante MF59, suministrado como parte de una vacuna contra la influenza en ancianos (Del Giudice *et al.*, 2018; Reed *et al.*, 2013). Actualmente, aún se desconoce en profundidad los mecanismos de acción de estos sistemas, sin embargo, es sabido que se caracterizan por desencadenar respuestas del tipo humoral. Lo que incentivó la formulación de sustancias con mecanismos de acción conocidos, capaces de inducir la producción de anticuerpos y que además fueran efectivas para la eliminación de organismos intracelulares (Reed *et al.*, 2013). En la mayoría de casos, se emplean moléculas inmunomodulatorias que actúan como ligando de receptores del sistema inmune innato como: receptores tipo Toll (TLR), NOD, RIG-1 y lectinas tipo C. El tipo de respuesta inmune adaptativa conseguida dependerá entonces del receptor o receptores activados (Del Giudice *et al.*, 2018).

Para aumentar la eficiencia de dicha respuesta se ha implementado la utilización de varios componentes en simultáneo para la creación de sistemas de adyuvantes. Entre los componentes más utilizados se encuentra el monofosforil lípido A (MPL), ligando del TLR4; CpG, ligando del TLR9; saponinas como el QS21, el dinucleótido cíclico c-di-AMP, antagonista de la molécula adaptadora STING, entre otras (Reed *et al.*, 2013; Sanchez Alberti *et al.*, 2017). Sus diferentes combinaciones han llevado a la creación de adyuvantes exitosos que están siendo implementados en vacunas ya aprobadas o que se encuentran en fases de prueba (Reed *et al.*, 2013). Como son las vacunas contra la influenza aviar, el virus de la hepatitis B, el virus del papiloma humano y el virus del herpes zoster. Por otra parte, la combinación del MPL con QS21 (adyuvante AS02) está también implementado en la vacuna RTS,S contra la malaria en fase de prueba (Reed *et al.*, 2013).

La capacidad de estos nuevos sistemas para modular y promover en simultáneo respuestas humorales y celulares, representa una ventaja para la creación de vacunas contra organismos como *T. cruzi*, en donde se requeriría la obtención de una respuesta polarizada tipo Th1/Th17, a fin de garantizar que se desencadenen reacciones de protección inmunitaria. Lo que motivó a que en esta investigación se llevase a cabo el estudio y la evaluación de los adyuvantes disponibles, a fin de escoger el que será implementado en el diseño final del prototipo inmunoestimulante en conjunto con el péptido quimérico.

Por otra parte, es sabido que la funcionabilidad de una vacuna basada en epítopes, va más allá de la elección de las regiones inmunogénicas que contenga. Depende también del modo en que éstas estén integradas, lo que ha llevado a la aplicación de distintas estrategias, en las que se ha probado desde el empleo de cocteles de epítopes hasta el empalme de éstos, con y sin secuencias de unión flanqueantes (Athanasίου *et al.*, 2017; Schubert and Kohlbacher, 2016). En tal sentido, diversas investigaciones han comprado que la adición de estas secuencias, o *linkers*, entre epítopes, ofrece ventajas para su procesamiento, presentación y reconocimiento (Athanasίου *et al.*, 2017; Livingston *et al.*, 2002; Schubert and Kohlbacher, 2016; Velders *et al.*, 2001). Sin embargo, también se ha evidenciado que agregarlos puede llevar a la formación de neo-epítopes, los que pueden tener efectos adversos que disminuyan la efectividad de la vacuna (Schubert and Kohlbacher, 2016). Por lo tanto, la elección de las secuencias de unión se considera un paso importante en su diseño, lo que fue tomado en cuenta durante la creación del péptido quimérico.

Enfoque evolutivo

Ahora bien, en adición al análisis inmunoinformático, esta investigación contempló un enfoque evolutivo destinado a la determinación de las presiones selectivas que actúan en las zonas de reconocimiento inmune presentes en las proteínas de *T. cruzi* seleccionadas como blancos inmunogénicos. Con relación a esto, es sabido que la coevolución huésped-patógeno se caracteriza por los cambios adaptativos recíprocos que se derivan de la interacción de las especies, donde la presión inmune del huésped y la evasión inmune del parásito representan los mecanismos claves de un proceso conocido como “carrera armamentista evolutiva” (Comas *et al.*, 2010). Dependiendo de los organismos patógenos, se ha encontrado que los genes que codifican antígenos tienden a ser muy variables, en gran medida, como consecuencia de una selección diversificadora que puede derivar en la evasión de la inmune. No obstante, existen organismos como *Mycobacterium tuberculosis* en donde se ha encontrado evidencias de una “hiper-conservación” de estos genes (Comas *et al.*, 2010).

En el diseño racional de vacunas, el estudio de la variabilidad de las proteínas portadoras de epítopes es fundamental para evitar la escogencia de aquellas cuya variabilidad no permita una detección inmunitaria (Raymond *et al.*, 2018). Por esta razón, las proteínas del parásito seleccionadas fueron evaluadas mediante la búsqueda de las presiones selectivas operando en los epítopes. Éstas estuvieron sujetas al análisis con distintos métodos que permitían localizar las regiones que se encuentren bajo selección positiva y negativa. Bajo este supuesto, las regiones que presentes indicios de selección positiva representarían posibles zonas de reconocimiento inmune donde la fijación de polimorfismos que generan cambios de residuos favorece la adecuación biológica o aptitud del parásito (*fitness*). Mientras que las regiones bajo selección negativa constituirían áreas conservadas de importancia para la proteína. Permitiendo, por lo tanto, la localización de zonas inmunogénicas específicas que sean accesibles al reconocimiento inmune (Echave *et al.*, 2016; Studer *et al.*, 2012). Se esperaba que mediante el solapamiento de los resultados de las pruebas inmunoinformáticas y evolutivas se localizasen péptidos bajo selección y potencialmente antigénicos, que puedan ser empleados en el diseño del prototipo inmunoestimulante.

Por lo anterior expuesto, tomando en cuenta las ventajas de utilizar vacunas de subunidad con adyuvantes de nueva generación y dada la utilidad de los análisis inmunoinformáticos preliminares y de los métodos para la determinación de presiones selectivas, la presente investigación tenía como objetivo la escogencia de diferentes proteínas de *T. cruzi* para el diseño de una vacuna de subunidades multiepítopes, tomando en cuenta las proteínas han sido evaluadas en investigaciones anteriores; así como también, considerando la variedad de adyuvantes disponibles. Lo que en sumatoria fue importante para el diseño de un prototipo con el potencial de desencadenar una respuesta celular y humoral que genere protección contra la enfermedad de Chagas.

Justificación

La enfermedad de Chagas afecta a millones de personas y representa una de las infecciones parasitaria más importante de América Latina (Perez-Aguilar *et al.*, 2018). Desde su descubrimiento, numerosas investigaciones se han enfocado en el entendimiento de la biología de su agente causal, *T. cruzi*, así como también de las respuestas inmunológicas que se desarrollan en el huésped mamífero producto de la infección. Los avances obtenidos han permitido identificar proteínas con potencial antigénico que podrían incorporarse en el diseño de una vacuna preventiva (Bontempi, 2015; Carabarin-Lima *et al.*, 2014; Elisei *et al.*, 2018; Khatoon *et al.*, 2018; Martinez-Campos *et al.*, 2015; Sanchez-Trincado *et al.*, 2017). A partir de este conocimiento, esta investigación se planteó analizar estas y otras proteínas de *T. cruzi* menos estudiadas, a fin de localizar las regiones proteicas clasificadas como epítopes por algoritmos de predicción cuyo desempeño fue previamente evaluado. Adicionalmente, se identificaron

las presiones selectivas actuando en estas regiones, enfoque ausente en la mayoría de las investigación *in silico* de determinación de epítopes, y que se realizó a fin de considerar el grado de conservación de estos epítopes. Del mismo modo, los resultados de las pruebas de selección aportaron información útil en relación a los sitios altamente conservados presentes en las proteínas a evaluar, los que pudo determinarse como zonas de interés terapéutico. Finalmente, el diseño de un péptido quimérico permitió la incorporación de diferentes epítopes de distintas proteínas del parásito, que con la ayuda del adyuvante representaría un prototipo inmunoestimulante con el potencial de ser utilizado como una vacuna de subunidades preventiva contra la enfermedad de Chagas. Esta investigación representó por lo tanto un estudio preliminar valioso para el desarrollo subsiguiente de investigaciones experimentales que permitirían determinar su eficacia.

II. HIPÓTESIS

El uso de herramientas inmunoinformáticas permitiría la elección de péptidos de *Trypanosoma cruzi* que al ser evaluados mediante métodos para la detección de presiones selectivas, podrían utilizarse como componentes en el diseño *in silico* de un péptido multiepítopes, que en compañía de un adyuvante adecuado, representaría un prototipo inmunoestimulante contra la enfermedad de Chagas.

III. OBJETIVOS

Objetivo general

Diseñar computacionalmente un péptido quimérico compuesto de posibles regiones de reconocimiento inmune de proteínas de *Trypanosoma cruzi*, seleccionadas mediante la aplicación de herramientas inmunoinformáticas y métodos para la detección de presiones selectivas.

Objetivos específicos

1. Evaluar la eficacia de los métodos de predicción de epítopes lineales de células B con la estimación de la proporción de epítopes confirmados que sean predichos.
2. Valorar el desempeño de los predictores de péptidos de unión a MHC-I mediante un análisis ROC-AUC.
3. Seleccionar del total de algoritmos evaluados, aquellos que serán utilizados en este estudio.
4. Comprobar la existencia de sitios bajo selección en epítopes proteicos de *T. cruzi*, mediante programas bioinformáticos.
5. Visualizar los epítopes lineales confirmados y predichos, y los sitios bajo selección en la estructura 3D de una proteína de *T. cruzi* utilizando herramientas de modelado.
6. Seleccionar proteínas de *T. cruzi* basándose en el potencial inmunogénico reportado en investigaciones *in silico* y experimentales previas a este estudio.
7. Hallar en las proteínas seleccionadas los posibles epítopes lineales de células B, mediante los predictores escogidos con la evaluación.
8. Determinar en las proteínas escogidas los posibles péptidos de unión a MHC-I, utilizando los métodos de predicción seleccionados durante la evaluación.
9. Establecer en las proteínas seleccionadas los posibles péptidos de unión a MHC-II, empleando los algoritmos de predicción disponibles.
10. Realizar el modelado de las proteínas seleccionadas, representando los epítopes lineales de células B predichos y los sitios bajo selección, con el uso de herramientas bioinformáticas.
11. Seleccionar el adyuvante que formará parte del prototipo inmunoestimulante.
12. Modelar el péptido quimérico mediante el ensamble de los epítopes predichos, utilizando algoritmos especializados.

13. Determinar la estructura y estabilidad tridimensional del péptido quimérico mediante programas especializados.

IV. METODOLOGÍA

Evaluación de predictores

Predictores de epítopes lineales de células B

Para la elección de los predictores que mejor desempeño presentan en la detección de epítopes proteicos de *T. cruzi*, se hizo una evaluación de los algoritmos disponibles utilizando la base de datos de epítopes lineales de células B del EIDB (*Immune Epitope Database*) (Vita *et al.*, 2015). De este repositorio fueron seleccionados únicamente los epítopes provenientes de ensayos experimentales con resultados positivos, donde la inmunogenicidad fue comprobada en personas chagásicas. Se realizó una prueba de tamaño de muestra haciendo uso del programa estadístico MiniTab versión 17 (Minitab, 2014) para conocer cuántas proteínas representarían una muestra significativa de la población total de la base de datos. Éstas que se analizaron con nueve predictores de epítopes lineales, siete de los cuales son métodos clásicos basados en diferentes propiedades fisicoquímicas que pueden ser utilizados simultáneamente en el servidor de BcePred (Saha and Raghava, 2004). Siendo los predictores: el índice de hidrofilia de Parker (Parker *et al.*, 1986), método de flexibilidad de Karplus (Karplus and Schulz, 1985), predictor de accesibilidad de Emini (Emini *et al.*, 1985), método de incidencia de giros β de Pellequer (Pellequer *et al.*, 1993), predictor de propensión antigénica de Kolaskar (Kolaskar and Tongaonkar, 1990), método de exposición a la superficie de Janin (Janin *et al.*, 1978) y el predictor de polaridad de Ponnuswamy (Ponnuswamy *et al.*, 1980). Los otros métodos usados son BepiPred 2.0 (Jespersen *et al.*, 2017), que se basa en un algoritmo de árboles de decisión (*random forest*) entrenado con epítopes anotados de estructuras 3D resueltas de antígeno-anticuerpo así como en una colección de epítopes lineales de la base de datos del IEDB, y ABCPred (Saha and Raghava, 2006), que aplica un método basado en redes neuronales recurrentes (RNN) entrenado con una base de datos curada de epítopes lineales.

La evaluación de cada uno de los predictores se realizó analizando la secuencia aminoacídica de cada proteína tomada de la muestra, que fueron obtenidas de la base de datos del Centro Nacional para la Información Biotecnológica (NCBI) (Geer *et al.*, 2010). En cada caso, se verificó si el predictor era capaz de señalar como epítopes a las regiones proteicas catalogadas como tal de manera experimental. De ocurrir la identificación, se determinó la cobertura del epítope predicho sobre la secuencia aminoacídica del dato experimental. De esta manera pudo determinarse un porcentaje de cobertura por proteína que permitió comparar el desempeño de cada predictor. Los resultados obtenidos fueron graficados con R (Team, 2013), donde se elaboró un mapa de calor con la librería *heatmap*.

Predictores de péptidos de unión a moléculas MHC-I

El método aplicado requería la búsqueda inicial de la base de datos de epítopes de *T. cruzi* determinados experimentalmente en personas chagásicas, disponible en el IEDB. Debido a que la representación de alelos diferentes al HLA-A2*01 en la base de datos era muy baja, se tomó éste como el alelo de referencia, necesario para análisis posteriores. La base de datos contaba con 34 proteínas y 70 epítopes en total, de cada proteína se obtuvo la secuencia aminoacídica presente del NCBI.

Para la escogencia de los predictores se tomó como referencia el estudio realizado por (Paul *et al.*, 2019), donde se determinó la eficiencia de predictores de péptidos de unión a MHC-I utilizando las proteínas del virus vaccinia como datos para la evaluación. Estos investigadores realizaron una

recopilación de los métodos de predicción gratuitos disponibles en la Web, de los 15 métodos seleccionados 7 fueron empleados en este estudio, ya que el resto de predictores no estaban disponibles para la fecha o fueron determinados defectuosos en el análisis. Consecuentemente, los métodos utilizados en este trabajo fueron: NetMHC 4.0 (Andreatta and Nielsen, 2015), NetMHCpan 4.0 (Jurtz *et al.*, 2017), NetMHCpan 3.0 (Nielsen and Andreatta, 2016), SMM (Peters and Sette, 2005), IEDB Consenso (Moutaftsi *et al.*, 2006), SYFPEITHI (Rammensee *et al.*, 1999) y PREDEP (Altuvia *et al.*, 1995). Los primeros tres son métodos basados en redes neuronales artificiales entrenados con datos de afinidad de péptidos conocidos, que se encargan de la predicción de este parámetro en péptidos desconocidos de longitud y propensión de unión a moléculas HLA-I diferente, lo que realizan mediante extrapolaciones de los datos de entrenamiento.

Por otro lado, SMM es un método que a partir de datos provenientes de mediciones experimentales, construye matrices de afinidad cuantitativas que consideran la contribución por residuo para predecir la afinidad de péptidos. IEDB consenso, por su parte, es un algoritmo desarrollado para la plataforma del IEDB en el que tiene lugar la integración de tres métodos (SMM y versiones anteriores de NetMHC) a fin de generar una predicción combinada.

En cuanto a SYFPEITHI, corresponde al método de predicción más antiguo, y está basado en un algoritmo que analizando posiciones de aminoácidos en el péptido, tal como las posiciones de anclaje, otorga puntuaciones de acuerdo al tipo de residuo ocupante. Por último, e igual que en el caso anterior, PREDEP es un método que predice péptidos de unión a moléculas MHC-I a partir de matrices de puntuación por posición.

Cada uno de estos programas fue evaluado con las proteínas de la base de datos, para esto se introdujeron las secuencias acotando el alelo HLA-A*02:01 indicando predicciones con longitudes de péptidos de 9 y 10 mer, tamaño que corresponde al de los epítopes de la base de datos. El archivo de entrada para la evaluación contó con 53.076 péptidos, debido a que los métodos de predicción consideran todos los posibles péptidos de longitudes 9 y 10 mer que abarcaran la extensión completa de la proteína. Cada péptido recibió una puntuación por predictor, lo que determinó si era considerado por el método como epítope, dependiendo de los valores umbrales respectivos de cada algoritmo. Posteriormente, los resultados fueron agrupados y categorizados, identificando los epítopes experimentales del total de péptidos obtenidos. Estos epítopes, correspondieron al conjunto de verdaderos positivos (VP), mientras que el resto de epítopes representaron los falsos positivos (FP), datos necesarios para la construcción de la curva ROC-AUC empleada para evaluar el desempeño de los predictores.

Un análisis ROC (*Receiver Operating Characteristic*) es una metodología aplicada para evaluar y comparar clasificadores binarios en función de la precisión de predicción (Majnik and Bosnić, 2013). La representación gráfica que se obtiene de este análisis es una curva que muestra las variaciones en los valores de sensibilidad y especificidad que se derivan de tablas de contingencia que dependen de valores umbrales de discriminación (Kumar and Indrayan, 2011; Majnik and Bosnić, 2013). Entendiéndose como sensibilidad a la proporción de verdaderos positivos (VPR), es decir, el número de verdaderos positivos sobre el total de positivos ($VPR = VP / P$), siendo P igual a la sumatoria de verdaderos positivos y falsos negativos ($P = VP + FN$). Por otro lado, $1 -$ especificidad corresponde a la razón de falsos positivos (FPR), lo que indica la proporción de falsos positivos sobre el total de negativos ($FPR = FP / N$), siendo N igual a la sumatoria de verdaderos negativos y falsos positivos ($N = VN + FP$) (Kumar and Indrayan, 2011; Majnik and Bosnić, 2013). El desempeño de un predictor particular para un determinado valor umbral, representado por su sensibilidad y especificidad, se denota como un punto de la curva ROC. El punto con coordenadas (0,0) ($VPR = 0$, $FPR = 0$) representa un clasificador que nunca predice una clase positiva. Por otro lado, un punto en la coordenada (1,1)

representa la posición opuesta ($VPR = 1$, $FRP = 1$), donde el clasificador predice cada instancia como positiva. Por lo tanto, una clasificación perfecta estaría ubicada en el punto (0,1) ($VPR = 1$, $FPR = 0$) (Florkowski, 2008). Los predictores ubicados en la diagonal entre las coordenadas (0,0) y (1,1) se consideran con desempeño azaroso, por esto, los predictores útiles serán aquellos cuya curva se localice por encima de la diagonal (Florkowski, 2008). Del gráfico resultante puede obtenerse otro valor importante, el AUC o área bajo la curva, necesario para realizar comparaciones entre métodos de clasificación. Un valor máximo de $AUC = 1$ representa una condición ideal donde existe un poder de discriminación perfecto, sin predicciones FP ni FN. Por lo tanto, cuanto mayor sea el valor de AUC mejor es el desempeño del predictor. El valor mínimo de AUC es igual a 0,5, pues corresponde al área por debajo de la diagonal (Florkowski, 2008; Kumar and Indrayan, 2011; Majnik and Bosnić, 2013). Cada predictor seleccionado fue evaluado con un análisis ROC, disponible en el paquete estadístico SPSS versión 17 (Inc, 2008), de donde se obtuvo el gráfico y los valores AUC.

Enfoque evolutivo

De la base de datos de epítopes utilizada para realizar la evaluación de predictores se seleccionaron proteínas que contaran con suficientes secuencias genómicas en las bases de datos del NCBI y TriTryp (Aslett *et al.*, 2009). Lo que se determinó realizando un BLAST (Altschul *et al.*, 1990) de cada proteína, de donde se seleccionaron las secuencias con un porcentaje de coincidencia no menor al 90%, omitiendo duplicaciones por aquellas compartidas por ambos repositorios. En cada caso las secuencias fueron alineadas con MUSCLE (Robert, 2004), disponible en el programa MEGA (*Molecular Evolutionary Genetics Analysis*) versión 10 (Kumar *et al.*, 2018). Para establecer los codones específicos sujetos a selección se aplicaron métodos que se basan en la determinación de la tasa de sustituciones sinónimas (dS) y no sinónimas (dN) por sitio (ω). Entendiéndose como sustituciones sinónimas aquellas sustituciones nucleotídicas que generan nuevos codones que codifican para el mismo aminoácido, y las no sinónimas aquellas sustituciones que generan un remplazo del aminoácido. Suponiendo que los genes evolucionan bajo neutralidad, donde las tasa dS y dN son iguales, se determina que los sitios donde $dS > dN$ estarían sujetos a seleccionados purificadora, mientras que aquellos donde $dS < dN$ se encontrarían bajo selección positiva. Para ello, se aplicó el método de máxima verosimilitud de probabilidad de efectos fijos, FEL (*Fixed Effects Likelihood*) (Kosakovsky Pond and Frost, 2005) únicamente para la detección de selección negativa, ya que, si bien puede detectar selección positiva pervasiva, este método asume valores constantes de ω (dN / dS) entre linajes, siendo incapaz de identificar sitios bajo selección positiva episódica, para lo que requeriría considerar valores variables de ω entre ramas. Por tal razón, se aplicó el modelo de efectos mixtos de evolución, MEME (*Mixed Effects Model of Evolution*) (Murrell *et al.*, 2012), para la detección de selección positiva por sitio, tanto pervasiva como episódica. Los sitios bajo selección fueron considerados estadísticamente significativos si obtenían un valor de $p < 0,1$ en cada método. Cada prueba se utilizó desde el servidor de DataMonkey (Weaver *et al.*, 2018). Una vez determinados los sitios bajo selección se identificó si estaban contenidos en las regiones de reconocimiento inmune, lo que fue gráficamente representado.

Elección y análisis de proteínas problemáticas

Para escoger las proteínas a considerar en el diseño del péptido quimérico, se hizo una revisión bibliográfica sobre proteínas de *T. cruzi* catalogadas como blancos inmunológicos. De esta manera se obtuvo un conjunto de proteínas candidatas que fueron buscadas en las bases de datos del NCBI y TriTryp para verificar la representatividad de sus secuencias genómicas, y comprobar que fuese posible sobre estas realizar las pruebas de selección. Lo que llevó a la formación del grupo final de proteínas catalogado como “proteínas problema”. De cada una se recopilaron las secuencias nucleotídicas disponibles en los repositorios mencionados, las que seguidamente fueron alineadas con MUSCLE,

disponible en Mega X. A partir de estos alineamientos se aplicaron los métodos de determinación de presiones selectivas, MEME y FEL, en cada caso la detección de selección se consideró estadísticamente significativa si obtenía un valor de $p < 0,1$.

Posteriormente, se realizó el análisis de las proteínas con los programas de predicción de epítopes lineales de células B y de péptidos de unión a MHC-I seleccionados en la etapa de evaluación. Adicionalmente, el algoritmo NetMHCII-2.0 (Nielsen and Lund, 2009) fue elegido para la predicción de péptidos de unión a moléculas MHC-II. Método basado en redes neuronales recurrentes que cuenta con un sistema que corrige la redundancia de los datos de entrenamiento, e integra la contribución de residuos de anclaje y flanqueantes para un desempeño mejorado. Para la utilización de los predictores se buscó en UniProt la secuencia de aminoácidos de cada proteína, la que representó el archivo de entrada en cada método.

Para la obtención del listado preliminar de epítopes lineales de células B, se tomaron en cuenta únicamente los péptidos predichos simultáneamente por ABCPred y BepiPred 2.0 con una puntuación mayor o igual a 0,5, en un rango de 0 a 1. De este listado se escogieron los epítopes con las más altas puntuaciones, que no mostraron sitios polimórficos, donde se supone selección positiva, y presentaron un mayor número de sitios conservados o donde se infiere estén bajo selección negativa. Estos parámetros también fueron implementados para escoger los péptidos de unión a moléculas MHC-I y -II.

En el caso de MHC-I, el listado preliminar se obtuvo del análisis por proteína por alelo, eligiendo de los alelos disponibles aquellos presentes en la población venezolana. Del total de péptidos predichos por variante alélica de proteína se optó por aquellos con una puntuación menor o igual a 1, en un rango de 0 a 100, que fueron predichos simultáneamente por SMM, NetMHCpan 3.0, NetMHCpan 4.0 y NetMHC 4.0. Por otro lado, la selección del listado preliminar de péptidos de unión a MHC-II, se basó nuevamente en predicciones hechas por proteína por alelo, tomando en cuenta de cada conjunto aquellos péptidos con puntuación menores o iguales a 10, en un rango de 0 a 100. Como en el caso anterior, de los alelos disponibles en el análisis se seleccionaron aquellos con representación en la población venezolana.

Modelado de proteínas de *T. cruzi*

Modelado de HSP70 de *T. cruzi*

Para contar con una representación gráfica de la ubicación de epítopes lineales de células B, experimentales y predichos, y de los sitios bajo selección en la estructura tridimensional de una proteína, se utilizó como referencia la proteína de choque térmico 70 (HSP70) de *T. cruzi*, número de acceso P05456 en UniProt (Consortium, 2018). A partir de su secuencia, fue predicha su estructura tridimensional con LOMETS (Wu and Zhang, 2007), programa que integrando nueve diferentes servidores Webs obtiene un consenso de los mejores templados, que se evalúan de acuerdo a una puntuación “Z”, que dependiendo del umbral, permite determinar la calidad del modelado en relación a sus características estructurales y restricciones físicas. El modelo generado fue seguidamente analizado con RAMPAGE (Prisant *et al.*, 2003) para evaluar su estabilidad mediante la obtención del gráfico de Ramachandran, el que considerando los ángulos de torsión Φ (entre N – C α) y Ψ (entre C α – C), establece si la rotación de la proteína respeta o no los efectos estéricos entre los residuos, y si se encuentran, por lo tanto, en las regiones favorables, permitidas y no permitidas del diagrama (Ramachandran and Sasisekharan, 1968). Seguidamente, se realizó un refinamiento del modelo empleando ModRefiner (Xu and Zhang, 2011), programa que ejecuta la reconstrucción de los átomos de la cadena principal y cadenas laterales simultáneamente para obtener el modelo físicamente más estable según el algoritmo, que fue comprobado con una nueva validación en RAMPAGE que permitió

la comparación de los gráficos de Ramachandran antes y después del refinamiento. Por último, se utilizó Chimera 1.13.1 (Pettersen *et al.*, 2004) para la visualización y edición de la proteína resultante, en donde se ubicaron los epítopes experimentales y predichos, y los sitios bajo selección, diferenciados por color.

Modelado de las proteínas a considerar para el diseño del péptido quimérico

Para tener mayor certeza sobre la accesibilidad a los epítopes lineales de células B predichos, se procedió a su localización en la estructura tridimensional de las proteínas problema. Para esto se realizaron modelados por homología empleando diferentes servidores: LOMETS, I-TASSER (Yang *et al.*, 2015), Swiss-Model (Waterhouse *et al.*, 2018) y Phyre2 (Kelley *et al.*, 2015). Eligiendo, de acuerdo a la estructura resultante, los mejores modelos, lo que era determinado con la comparación de los gráficos de Ramachandran obtenidos de RAMPAGE. Al igual que LOMETS, el resto de programas se basan en planteados conocidos para, a partir de la secuencia proteica, construir un modelo que recibe una puntuación de acuerdo a parámetros relacionados con la estabilidad estructural y la similitud en la identidad con el planteadado. Adicionalmente, en el caso de I-TASSER y Phyre2, en ausencia de un planteadado conocido, estos algoritmos son capaces de construir modelos con el ensamblaje de fragmentos de estructuras resueltas de proteínas con secuencias que correspondan con la proteína de interés.

Luego de escoger los mejores modelos para cada proteína, se procedió a refinarlos empleando los programas Galaxy Refine (Heo *et al.*, 2013) y Galaxy Refine 2 (Lee *et al.*, 2019). Que funcionan aplicando perturbaciones repetitivas por segmentos de la estructura secundaria mediante dinámica molecular, para luego reconstruir favoreciendo la compactación y la estabilidad del modelo tanto a nivel de la cadena principal como de las cadenas laterales. Galaxy Refine se empleó para refinar todas las proteínas a excepción de TSSA, que por tener una longitud menor a 300 aminoácidos fue analizada con Galaxy Refine 2, que se especializa en proteínas de menor tamaño. Además de estos algoritmos, se probó refinar las proteínas con 3Drefine (Bhattacharya and Cheng, 2013) método que utiliza la optimización iterativa de la red de enlaces de hidrógeno en conjunto con la minimización de energía a nivel atómico para el refinamiento del modelo. Sin embargo, debido a que éste mostró un desempeño inferior a los primeros programas, su utilización fue descartada.

Cada uno de los refinados era evaluado con RAMPAGE, y se realizaron tantos como fuesen necesarios hasta obtener gráficos de Ramachandran sin variaciones. El modelo con mejor estructura de cada proteína fue seguidamente visualizado y editado con Chimera, donde se localizaron los epítopes lineales de células B y los sitios bajo selección negativa que en estos se encontraban.

Diseño del prototipo inmunoestimulante

Del conjunto de péptidos obtenidos anteriormente del análisis de las proteínas problema, se seleccionó un nuevo grupo formados por aquellos que contenían mayor número de sitios conservados. Luego, dependiendo de la categoría de los péptidos, se eligieron las secuencias de aminoácidos que se utilizaron como *linkers*. Además, se agregó al diseño de la quimera el péptido sintético universal de unión a moléculas HLA-DR conocido como PADRE (siglas en inglés de *pan DR epitope*). Una vez que se tuvieron todos estos componentes se procedió a su ensamblaje. La secuencia final fue después analizada con dos métodos, AllerTOP (Dimitrov *et al.*, 2013a) y AllergenFP (Dimitrov *et al.*, 2013b), los que basados en aprendizaje automatizado se encargan de predecir la existencia de alérgenos en proteínas.

Ahora bien, debido a las variaciones en la estructura secundaria de los péptidos, que se visualizó en las proteínas problema, se realizaron distintos ensamblajes que diferían entre sí en el orden de los péptidos

empleados, con la finalidad de obtener el mejor modelo. Luego de repetidos intentos, se seleccionó el modelo que expuso la mayor cantidad de epítopes lineales de células B, y que además tuviera una estructura que mantuviese el plegamiento que los epítopes tienen en la proteína de origen. El modelado se llevó a cabo en el programa Phyre2 en modo “intensivo”, opción donde el algoritmo integra el modelado por homología y el modelado por empalme de templados por fragmentos, con un método denominado *ab initio* que se encarga de predecir el plegamiento de una proteína por fragmentos sin el uso de templados, sino con simulaciones de Monte Carlo que consideran los campos de fuerza que se establecen por las interacciones atómicas (Kelley *et al.*, 2015). De todos los modelos resultantes del cambio de orden de los péptidos, se escogió el que presentó una conformación que expuso la mayor cantidad de epítopes lineales de células B, y que además fue estructuralmente más estable según el gráfico de Ramachandran obtenido de RAMPAGE. El modelo final fue refinado 11 veces con la herramienta Galaxy Refine hasta conseguir gráficos de Ramachandran sin variaciones. Por último, la visualización y edición se llevó a cabo en Chimera.

Por otra parte, la escogencia del adyuvante, que junto con el péptido de subunidades representa al prototipo inmunoestimulante, se basó en la revisión de investigaciones experimentales donde se emplearon adyuvantes que resultaron efectivos para la eliminación de organismos patógenos con ubicación intra y extracelular. De todos los candidatos se seleccionó el que se consideró más adecuado debido a la naturaleza del antígeno y al tipo de respuesta inmunológica inducida.

V. RESULTADOS

Evaluación de predictores

Predictores de epítopes lineales de células B

Del total de proteínas de la base de datos, 33 conformaron la muestra empleada para el estudio. Cada predictor fue evaluado con cada una de las proteínas, pudiendo observarse en la Figura 1 el desempeño de los mismo mediante el código de color presentado en la leyenda del mapa de calor, que contiene valores de 0 a 100 que hacen referencia al porcentaje de cobertura de cada predictor por proteína evaluada. Del mismo modo, en la Tabla 1 se encuentran los valores de cobertura promedio de cada predictor considerando el total de proteínas. De esta forma, se encontró que los predictores ABCPred y BepiPred 2.0 hicieron mejores predicciones en comparación con los predictores clásicos. Dados estos resultados, se escogieron los predictores ABCPred y BepiPred 2.0 como los métodos considerados en este estudio para analizar el conjunto de proteínas problema. Estos resultados concuerdan con lo descrito en la literatura sobre el desempeño deficiente de los predictores basados en escalas de propensión, que señalan superados por los métodos fundamentados en aprendizaje automatizado (Blythe and Flower, 2005; Potocnakova *et al.*, 2016; Yasser and Honavar, 2010). Lo que pudo comprobarse en la predicción de epítopes de células B de *T. cruzi*. La mejora en el poder predictivo de los métodos seleccionados explica su utilización en una amplia gama de investigaciones (Maura *et al.*, 2010; Park *et al.*, 2018; von Delft *et al.*, 2018; Won and Lee, 2018).

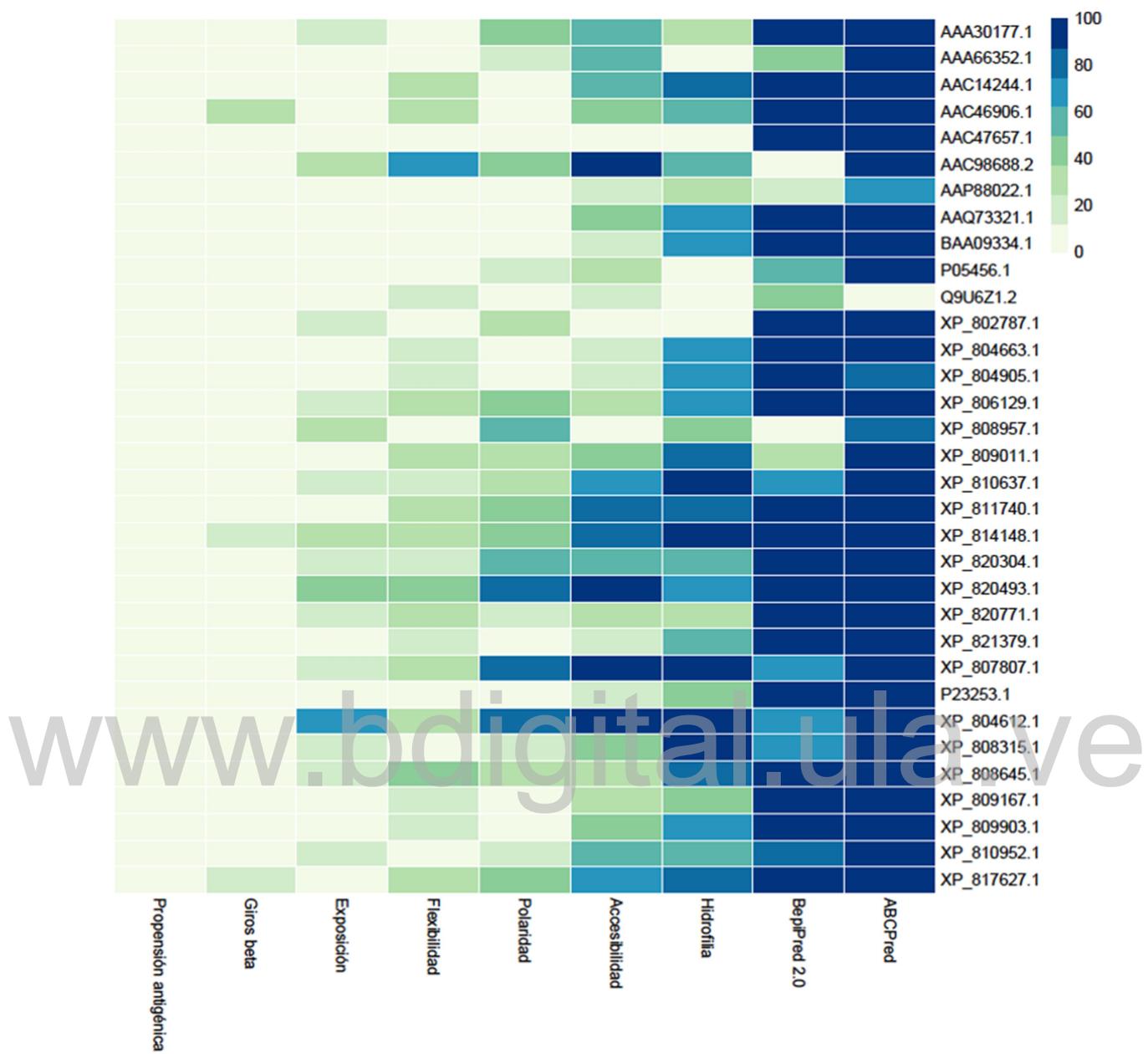


Figura 1. Porcentaje de epítopes confirmados de *T. cruzi* que fueron predichos por proteína con cada predictor.

Tabla 1. Cobertura promedio del total de epítopes confirmados de *T. cruzi* que fueron predichos por cada método.

Método	Cobertura promedio (%)
ABCPred	94,35
BepiPred 2.0	79,57
Hidrofilia de Parker	58,13
Accesibilidad de Emini	46,15
Polaridad de Ponnuswamy	26,14
Flexibilidad de Karplus	21,69
Exposición a la superficie de Janin	13,03
Incidencia de giros β de Pellequer	3,56
Propensión antigénica de Kolaskar	0,34

Predictores de péptidos de unión a moléculas MHC-I

Debido a que estos predictores varían en su desempeño dependiendo del alelo HLA, longitud de epítopes seleccionados y tipo de organismo estudiado, se buscó determinar qué métodos eran más eficientes en la predicción de péptidos de *T. cruzi* reconocidos por moléculas MHC-I del alelo HLA-A 02*01. Cabe destacar que la disponibilidad de alelos diferentes en la base de datos de epítopes de *T. cruzi* es escasa, lo que limitó la evaluación de los predictores, conllevando al uso del alelo HLA-A 02*01 por ser el que se encuentra mayormente representado en la base de datos.

Aplicando un análisis ROC se encontró que los predictores con mayor AUC fueron SMM (0,968), NetMHCpan 3.0 (0,962), NetMHC 4.0 (0,957), y NetMHCpan 4.0 (0,955), siendo estos los métodos escogidos para aplicar en esta investigación (Figura y Tabla 2). Se seleccionaron cuatro predictores debido a que la evaluación fue en base a un solo alelo, siendo posible que existan variaciones en las posiciones de desempeño al considerar alelos diferentes, de esta manera se espera que haya una consideración adecuada de los mejores predictores. Al igual que en este trabajo, otros estudios han categorizado a NetMHC y NetMHCpan entre los mejores métodos de predicción; de hecho, en una evaluación por alelos HLA-I hecha por Gowthaman *et al.* (2010) se encontró que para el alelo HLA-A 02*01 estos predictores tenían el mejor desempeño.

Por otra parte, Zhao and Sher (2018) determinaron menores valores de AUC para estos métodos analizando 32 alelos humanos HLA-I utilizando la base de datos completa del IEDB en 2016: SMM (0,856), NetMHCpan 3.0 (0,880), NetMHCpan 4.0 (0,872) y NetMHC 4.0 (0,860). Sin embargo, entre todos los predictores, estos poseían los más altos puntajes, superados únicamente por el predictor MHCFlurry (0,911), no disponible para el momento del estudio. Finalmente, en el estudio de (Paul *et al.*, 2019) tomado como referencia para la elección de los predictores a evaluar, los métodos presentaron los siguientes valores de AUC: SMM (0,938), NetMHCpan 3.0 (0,975), NetMHCpan 4.0 (0,979) y NetMHC 4.0 (0,974). Como en la investigación anterior, y contrario a lo obtenido en este trabajo, SMM obtuvo menores puntuaciones en comparación al resto de métodos. No obstante, debido a que en este trabajo se emplearon únicamente epítopes de *T. cruzi* para la evaluación de predictores, este programa fue incluido en la selección.

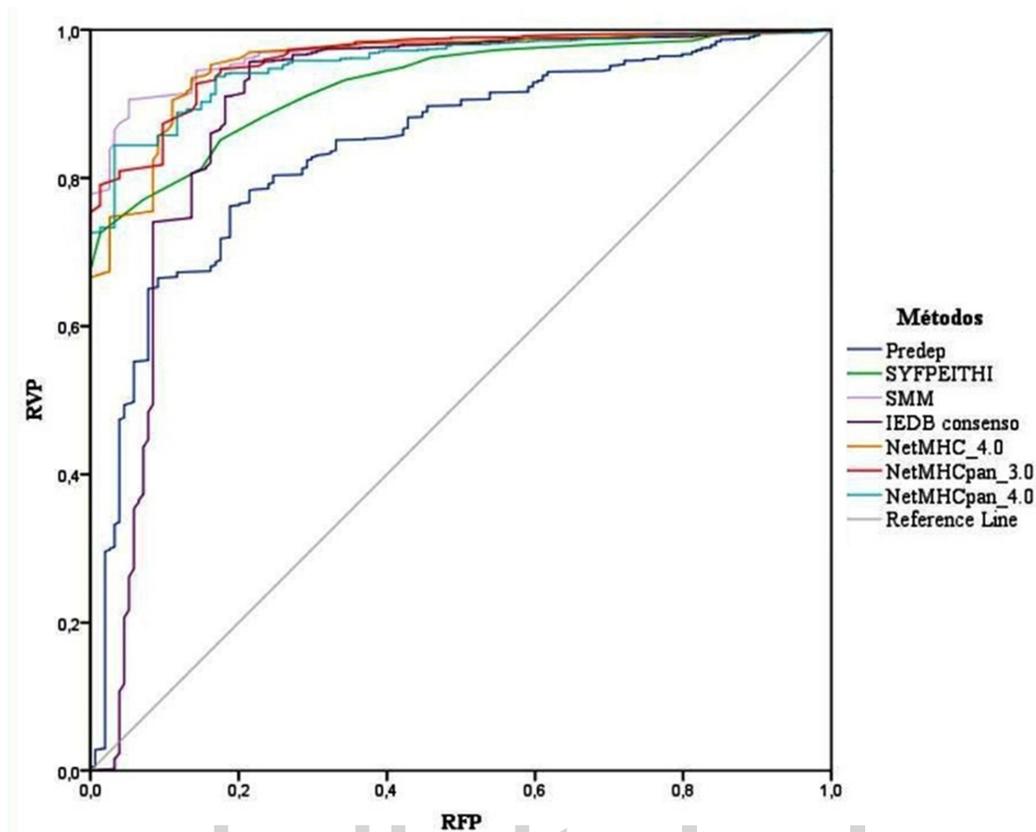


Figura 2. Curva ROC de cada predictor de péptidos de unión a MHC-I evaluado.

Tabla 2. Valores de área bajo la curva (AUC) provenientes del análisis ROC aplicado a cada método de predicción de péptidos de unión a MHC-I.

Métodos	AUC
SMM	0,968
NetMHCpan 3.0	0,962
NetMHC 4.0	0,957
NetMHCpan 4.0	0,955
SYFPEITHI	0,928
IEDB Consenso	0,895
Predep	0,842

Enfoque evolutivo

Para determinar las presiones selectivas actuando en proteínas de interés inmunológico de *T. cruzi* se seleccionaron proteínas con epítopes disponibles en las bases de datos empleadas en este estudio. Éstas fueron analizadas con métodos que detectaron selección por sitio. Así se encontró que la mayoría de los epítopes de las proteínas se encuentran bajo selección positiva y negativa. Lo que también ha sido

encontrado en otros estudios, donde se evalúa el grado de conservación de antígenos de virus y bacterias (Comas *et al.*, 2010; Li *et al.*, 2019; Raymond *et al.*, 2018), en los que se plantea que los sitios bajo selección negativa forman parte de regiones conservadas donde la tasa de fijación de polimorfismos es menor, que podrían tener importancia en la función de la proteína, y por ende representar una ventaja para el reconocimiento inmune. Por otro lado, los sitios bajo selección positiva han sido relacionados como sitios de evasión, donde el polimorfismo favorecería la eficacia o adecuación biológica (*fitness*) del organismo (Comas *et al.*, 2010; Li *et al.*, 2019; Raymond *et al.*, 2018). A partir de los resultados se evidencia que ambas fuerzas evolutivas están actuando en las proteínas evaluadas del parásito (Figura 3). La magnitud de la respuesta inmunológica por epítipo sería un factor a considerar para entender el efecto de estas variaciones. En relación a esto, cabe destacar el caso de la región C-terminal de proteínas de superficie asociadas a mucina (MASP) evaluada (Figura 3A). El análisis de esta región se realizó tomando como referencia un estudio donde se caracterizó como una zona conservada (De Pablos *et al.*, 2016). En esta investigación realizaron la evaluación de la respuesta de anticuerpos a lo largo de la secuencia, encontrando que la afinidad de los anticuerpos provenientes de personas chagásicas era mayor para los últimos dos péptidos, que catalogaron como C₅ y C₆. En este trabajo se encontró que los tres sitios determinados bajo selección en la región C-terminal eran encontrados en ambos péptidos, donde se hallan dos residuos bajo selección negativa que podrían estar equilibrando el efecto del sitio polimórfico que comparte con el resto de péptidos con menor afinidad, lo que representaría un indicio del efecto directo de la presencia de sitios polimórficos y conservados en las regiones de interés inmunogénico.

www.bdigital.ula.ve

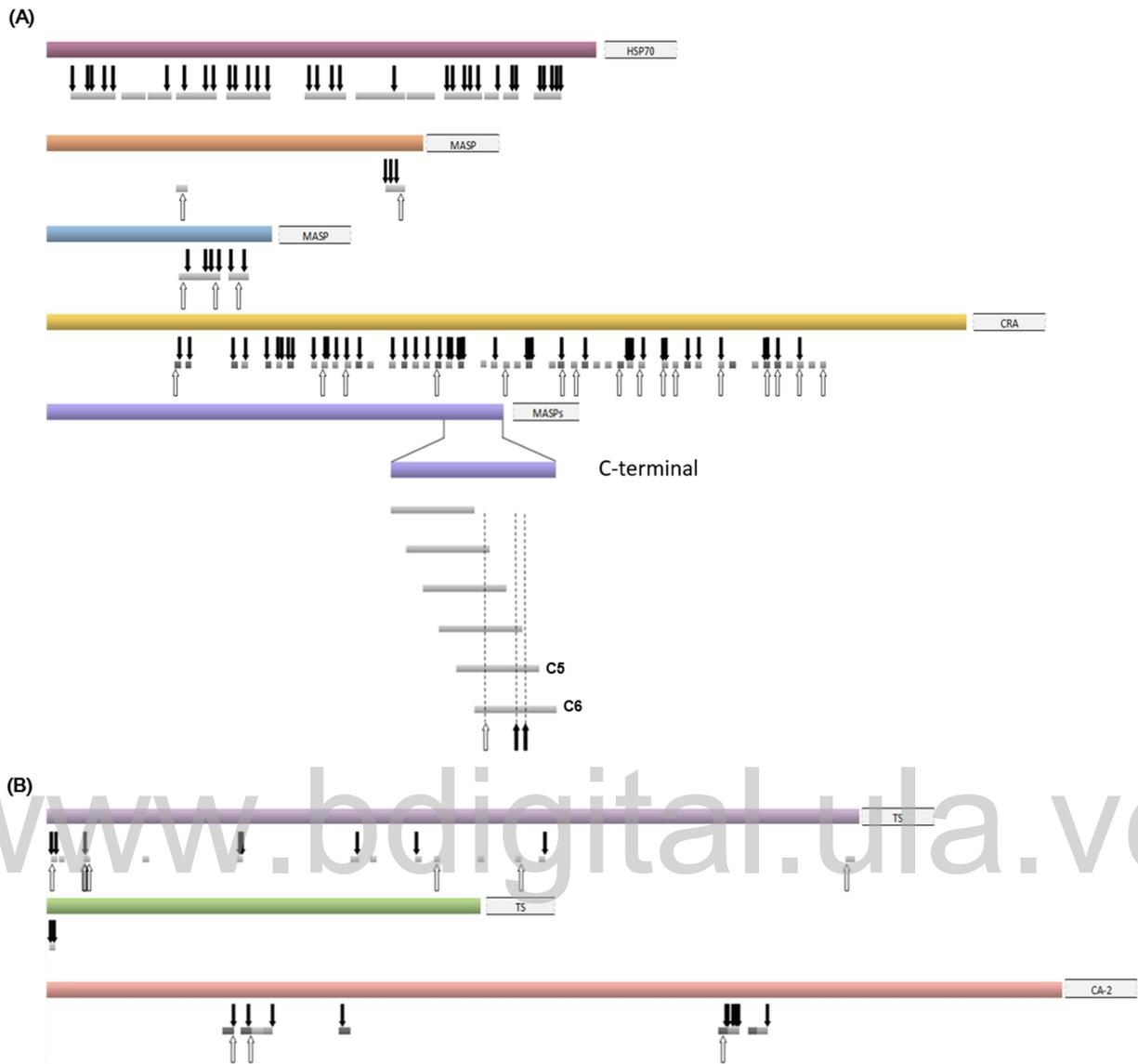


Figura 3. (A) Representación de los sitios bajo selección encontrados en los epítopes lineales de células B de proteínas de *T. cruzi*. (B) Señalamiento de los sitios bajo selección hallados en los epítopes de células T de proteínas de *T. cruzi*. Flechas negras, selección negativa; flechas blancas, selección positiva; HSP70, proteína de choque térmico 70; MASP, proteína de superficie asociada a mucina; CRA, antígeno citoplasmático repetitivo; TS, transialidasa; CA-2, antígeno de superficie 2.

Elección y análisis de proteínas problemáticas

En la Tabla 3, se muestra el listado de proteínas problema, seleccionadas de acuerdo a los criterios explicados anteriormente. Además, se agregó la proteína PPK, que si bien no ha sido objeto de investigación inmunológica en el ámbito de la vacunación, se consideró como un posible blanco debido a la ausencia de homología con proteínas humanas. Por otra parte, también se añadió la proteína LYT-1, que aunque posee inmunogenicidad comprobada, ha sido poco estudiada en comparación con el resto de proteínas catalogadas como blancos, que han sido empleadas con éxito en estrategias para la protección inmunológica contra la infección.

En este estudio el número de posibles epítopes a incluir en el péptido quimérico fue reducido mediante el análisis evolutivo de las proteínas problema, que llevó a que se evitara la escogencia de péptidos con

sitios polimórficos que pudiesen disminuir el rango de protección del prototipo inmunoestimulante. El total de péptidos por proteína se muestra en la Tabla 4, catalogados de acuerdo al potencial para actuar como epítopes lineales de células B o péptidos de unión a moléculas HLA, donde se indicó el tipo de alelo involucrado. Como puede verse, no todas las proteínas contenían péptidos de unión que cumplieran con los criterios para su elección en cada uno de los alelos evaluados. Del mismo modo, para cada péptido se indica su posición en la secuencia de la proteína y los sitios bajo selección negativa que contienen. Por último, la puntuación que recibe cada uno hace referencia al promedio de las puntuaciones recibidas por todos los métodos aplicados, a excepción de los valores que califican los péptidos de unión a MHC-II que representan el puntaje del único predictor empleado.

Tabla 3. Proteínas a considerar en el diseño de la vacuna de subunidades multiepítopes

Nº de acceso	Tipo	Nombre	Abreviatura	Referencias
XP_818578	Cistein peptidasa	Cruzipaina	Cz	Cerny <i>et al.</i> , 2016; Duschak & Couto, 2009; Fonseca <i>et al.</i> , 2005; Martínez, Campbella, Frasc & Cazzulo, 1991
XP_808304	Sensor de calcio	Proteína flagelar de unión a calcio de 24 kDa	Tc24	Martínez-Campos <i>et al.</i> , 2015; Villanueva-Lizama <i>et al.</i> , 2018
XP_814930	Transialidasa	Antígeno de superficie de tripomastigotes 1	TSA-1	Villanueva-Lizama <i>et al.</i> , 2018; Wizel, Garg & Tarleton, 1998
XP_820337	Protil oligopeptidasa	Protil oligopeptidasa de 80 kDa	Tc80	Bivona <i>et al.</i> , 2018
ACY54510	Proteína de adhesión	Antígeno pequeño de superficie de tripomastigotes	TSSA	Gonzalez <i>et al.</i> , 2015; Miyahira, 2010
XP_820203	Proteína auxiliar/glicólisis	Piruvato fosfato diquinasa	PPDK	Acosta <i>et al.</i> , 2004
Q9NG46	Proteína lítica	Proteína lítica 1	LYT-1	Ballesteros-Rodea <i>et al.</i> , 2012; Manning-Cela <i>et al.</i> , 2001

Tabla 4. Listado y categorización de los péptidos predichos de cada proteína problema.

Proteína	Tipo	Alelo	*Péptido	Inicio	Fin	Puntuación promedio	Sitios bajo selección negativa		
Cz	B	NA	AAAQERARVP	106	115	0,685	108,	114	
			RARVPVKVEV	111	120	0,575	114,	120	
			KVEVVGAPAA	117	126	0,623	120,	122	
			VQENNGAVYT	198	207	0,639	206		
	MHC-I	HLA-A*24:02	VYESAAEEAF	48	57	0,323	57		
		HLA-B*40:01	AEEAFRLSV	53	61	0,411	57,	58	
		HLA-B*58:01	SAIGNVECQW	151	160	0,128	154		
		HLA-A*02:01	FLAGHPLTNL	161	170	0,287	162,	164	
		HLA-A*02:01	GLMNNAFEWI	188	197	0,614	193,	194	
		HLA-B*58:01	GLMNNAFEW	188	196	0,530	193,	194	
		HLA-B*15:01	IVQENNGAVY	197	206	0,390	206		
		HLA-A*01:01	WTTQWGEEGY	306	315	0,307	311		
	MHC-II	DRB1_1501	AFRLSVFRENFLAR	56	70	5,500	57,	58,	63
		HLA-DQA10102-DQB10602	NGAVHFAAAQERARV	100	114	6,500	108,	110,	114
		DRB1_0301							
		HLA-DQA10101-DQB10501	GNVECQWFLAGHPLT	154	168	1,900	154,	162,	164
		DRB1_0401	NAFEWIVQENNGAVY	192	206	5,000	193,	194,	206
		HLA-DQA10501-DQB10301	TTSGHTVGAITIGHV	224	238	1,000	230,	232,	233
		HLA-DQA10201-DQB10202	IVTCGAETL TEEVFL	395	409	0,080	395,	405,	408
		DRB1_1301	RQRRHQPHYHSRHRRL	453	467	5,000	465		
LYT-1	B	NA	SHEGVSPLAK	85	94	0,592	89		
			ARRMRYGYEE	259	268	0,698	268		
			YEEGERDTLS	266	275	0,680	268		
			SPKRFLSTQP	289	298	0,662	291,	293	
			RSSRNYRNPH	300	309	0,705	309		
			HTANGTNGNS	309	318	0,659	309		
	MHC-I	HLA-A*03:01	RQMLFTGQK	231	239	0,307	235		
		HLA-B*27:05	GRMGSPKRF	285	293	0,361	291,	293	
		HLA-A*24:02	RYWEQSRAI	533	541	0,459	538		
		DRB1_0301	APPLMSDERRRQMLF	221	235	3,500	235		
MHC-II	DRB1_0401	SPKRFLSTQPPRSSR	289	303	1,000	291,	293		
	HLA-DQA10501-DQB10301	KSHPKRAAAVTSVPL	327	341	1,800	327			
	DRB1_1301	RSLMYRCNHS HHRST	383	397	1,700	393			
	HLA-DQA10102-DQB10602	HHRSTEGDVVRLAAK	393	407	1,900	393			
	HLA-DQA10101-DQB10501	ALQRRYWEQSRAILE	529	543	5,000	530,	538		
	HLA-DQA10201-DQB10202	WEQSRAILEQLENML	535	549	6,500	538			

Tabla 4 (cont.). Listado y categorización de los péptidos predichos de cada proteína problema.

Proteína	Tipo	Alelo	*Péptido	Inicio	Fin	Puntuación promedio	Sitios bajo selección negativa			
PPDK	B	NA	EDFEEAIGHM	156	165	0,642	159,	163		
			TGFPPEDPM	197	206	0,618	198,	199,	200	
			PQQIGQSLSL	293	302	0,650	300			
			VREGLITKEE	382	391	0,631	384			
			DADSAKEWSA	438	447	0,647	440,	443,	444	
			QFTLSGR	508	517	0,644	511			
	MHC-I	HLA-A*01:01	HLERFVYDSY	130	139	0,400	132,	133		
		HLA-B*27:05	RRFITMYADI	140	149	0,143	143,	144		
		HLA-A*24:02	RFITMYADI	141	149	0,385	143,	144		
		HLA-B*15:01	FQRKTGFPF	193	201	0,192	198,	199,	200	
		HLA-B*07:02	TPQQIGQSL	292	300	0,136	300			
		HLA-B*58:01	IGQSLSLRW	296	304	0,189	300			
		HLA-B*08:01	SARGKKVIM	446	454	0,360	453			
		HLA-A*02:01	YIVGTMIEV	763	771	0,176	766,	769		
		HLA-B*39:01	QQIGQSLSL	857	865	0,115	300			
		HLA-B*40:01	GEHGGDPTTI	857	866	0,124	857,	861,	865	
		HLA-A*03:01	VAAHAALK	889	897	0,368	890			
		MHC-II	HLA-DQA10501-DQB10301	LFSVRSGAAASMPGM	93	107	0,500	99,	100,	102
			HLA-DQA10201-DQB10202	YDSYRRFITMYADIV	136	150	0,300	143,	144	
			HLA-DQA10101-DQB10501	DSYRRFITMYADIVM	137	151	0,700	143,	144	
DRB1_0301	IREMILADTLEGRKA		601	615	0,500	601,	612			
DRB1_0401	LGIFRTMKGLPVTIR		629	643	0,500	636				
HLA-DQA10102-DQB10602	DYIVGTMIEVPRAAI		762	776	1,600	766,	769			
DRB1_1501	AGSFLRHYCSLGIYS		807	821	0,300	808,	816			
DRB1_1301	GELVRIAVQKGRRVK		834	848	0,400	836,	839,	843		
Tc24	B	NA	TKRAFDKARA	91	100	0,540	94,	95		
	MHC-I	HLA-B*07:02	RVRDITKRAF	86	95	0,476	94,	95		
	MHC-II	DRB1_1501	RRIELFKKFDKNETG	49	63	9,000	62			
		DRB1_0401	TKRAFDKARALGSKL	91	105	5,000	94,	95		
		HLA-DQA10501-DQB10301	ELDKNGTGSVTFDEF	175	189	5,500	179			
	HLA-DQA10102-DQB10602	LDKNGTGSVTFDEFA	176	190	6,500	179				

Tabla 4 (cont.). Listado y categorización de los péptidos predichos de cada proteína problema.

Proteína	Tipo	Alelo	*Péptido	Inicio	Fin	Puntuación promedio	Sitios bajo selección negativa		
Tc80	B	NA	FASEAEL RKK	52	61	0,625	59,	60	
			GAET D AAQDA	199	208	0,664	203		
			VEKGLGSEPL	269	278	0,658	278		
			LGSEPLVFKK	273	282	0,697	278		
			RSVLSQAALV	332	341	0,601	336		
			NAIPLPIGTI	368	677	0,543	369,	371,	374
			LDINDPQSSL	405	414	0,673	408		
			FKDDTVEGLL	417	426	0,650	419,	423	
	TVEGLLVDDF	421	430	0,624	423				
	MHC-I	HLA-A*01:01	LTDTAPSIF	105	113	0,183	111		
		HLA-B*08:01	FFYTRYPAL	184	192	0,120	187		
		HLA-B*15:01	YLGNEGSTF	293	301	0,309	300		
		HLA-A*03:01	VLSQAALVK	334	342	0,252	336		
		HLA-B*40:01	LEDPTLNAI	362	370	0,381	369		
		HLA-B*07:02	LPIGTITSF	372	380	0,314	374,	378	
		HLA-B*39:01	FKDDTVEGL	417	425	0,316	419,	423	
		HLA-A*01:01	SSESPLLLY	457	465	0,059	458		
		HLA-B*27:05	SRMVFLRDL	481	489	0,345	485		
		HLA-B*58:01	VANQAPELF	557	565	0,276	561,	564	
HLA-B*07:02		APELFRCVV	561	569	0,467	561,	564		
HLA-A*24:02	MYKF H KFTI	577	585	0,268	581,	584			
HLA-A*24:02	IYTFIAKNI	683	691	0,442	686,	690			
MHC-II	HLA-DQA10101-DQB10501	NEQNAFFEEYFASEA	42	56	0,080	43,	45,	56	
	DRB1_0401	APSIFLDPNTMSSDG	109	123	0,700	111,	121		
	DRB1_1501	RKKDFV S FKITSFLL	384	398	0,800	387			
	DRB1_0301	SLRVFKDDTVEGLLV	413	427	0,900	423			
	HLA-DQA10201-DQB10202	VFKDDTVEGLLVDDF	416	430	1,900	419,	423		
	HLA-DQA10102-DQB10602	PAFSSSRMVFLRDLG	476	490	1,800	478,	485		
	HLA-DQA10501-DQB10301	IEVAAGHGAGKPTSK	661	675	0,400	664			
TSSA	B	NA	AGEAPSPSGA	44	53	0,582	47,	50,	52
			PSPSGASSGE	48	57	0,618	52		
			GASSGEAEAS	52	61	0,645	52		
	MHC-I	HLA-B*07:02	APSPSGASS	47	55	0,812	47,	50,	52
	MHC-II	HLA-DQA10501-DQB10301	EKKAAAGEAPSPSGA	39	53	0,900	47		
		HLA-DQA10102-DQB10602	PLALTASALAYTALG	78	92	0,600	89		

Tabla 4 (cont.). Listado y categorización de los péptidos predichos de cada proteína problema.

Proteína	Tipo	Alelo	*Péptido	Inicio	Fin	Puntuación promedio	Sitios bajo selección negativa
TSA-1	B	NA	HNRRRVVGSS	80	89	0,663	89
			TEVQPTQPIN	295	204	0,603	203
			SPNNKGDLE	316	325	0,542	324
			DDGQKWEIPG	363	372	0,550	366
			IPGGVSSVAC	370	379	0,590	378
			VWLKSGPELP	424	433	0,697	424, 425, 433
	MHC-I	HLA-B*58:01	MAAGTLMVW	1	9	0,033	6, 8
		HLA-B*39:01	RHFYSAVLL	107	115	0,131	112
		HLA-A*24:02	FYSAVLLLL	109	117	0,274	112
		HLA-B*15:01	VMGNGLVF	335	343	0,087	336
		HLA-B*58:01	YSTDDGQKW	360	368	0,160	366
		HLA-A*01:01	YSDDALHLL	503	511	0,137	509
		HLA-B*58:01	LSYGAGGKW	656	664	0,082	662
		HLA-B*40:01	REYQVALML	683	691	0,032	689
		MHC-II	HLA-DQA10501-DQB10301	MVWFALSGGGVYCCC	7	21	0,800
HLA-DQA10201-DQB10202	KVFSLITYSTDDGQK		353	367	1,600	357, 366	
DRB1_0301	NLLYSDDALHLLQA		499	513	0,400	509	
HLA-DQA10102-DQB10602	HESTAVSLARLTEEL		517	531	1,000	527	
DRB1_1301	LLGASLRGNKRTKLI		640	654	0,800	647	

* En negrita se indican los sitios bajo selección negativa

Modelado de proteínas de *T. cruzi*

Modelado de HSP70 de *T. cruzi*

De todos los posibles modelos ofrecidos por LOMETS, se seleccionó aquel con mayor puntuación Z (29,92), el que presentó una mayor similitud con la proteína HSP70 humana. Éste fue el modelo seleccionado para el análisis con RAMPAGE, que se refinó con ModRefiner para la obtención del modelo final de la proteína, la comparación de los resultados derivados del gráfico de Ramachandran antes y después del refinamiento se encuentran en la Tabla 5 y la Figura 4, donde puede comprobarse el mejoramiento de la estructura por la reducción del número de aminoácidos en las zonas no permitidas y el consecuente aumento en regiones favorables del diagrama.

En el modelo (Figura 5), fueron identificadas en color azul las zonas en que ocurrió el solapamiento de epítopes, las que corresponden al 73,33% de epítopes experimentales que fueron predichos simultáneamente por ABCPred y BepiPred 2.0. El porcentaje de epítopes experimentales restantes, 26,67%, representa los epítopes confirmados no predichos y fueron identificados en color verde. Por otro lado, del total de predicciones, se encontró que los epítopes predichos solapados constituyeron el 75,15% del total de epítopes predichos. Por lo tanto, el 24,84% representó la proporción de epítopes predichos sin correspondencia con epítopes experimentales, los que se identificaron en color morado. De este porcentaje, el 16,77% de epítopes se encontró en la región de la proteína que contiene epítopes

experimentales (del aminoácido 30 hasta el 636 de un total de 680 residuos), mientras que el 8,07% restante se ubica en el extremo C-terminal de la proteína, que fue identificado como una región transmembranal. Todo esto indica la importancia del reconocimiento previo de los dominios de las proteínas problema, a fin de evitar la escogencia de epítopes predichos de células B sin posible acceso al reconocimiento inmune, que a su vez ejemplifica la ventaja de incluir otros análisis que integren o complementen los resultados predictivos. Finalmente, en el modelo también se observaron los sitios bajo selección negativa identificados en la proteína. Que permiten reconocer epítopes relevantes para este estudio, pues representan epítopes con menor variabilidad en las poblaciones.

Tabla 5. Porcentaje de aminoácidos de HSP70 en las regiones del gráfico de Ramachandran antes y después del refinamiento del modelo.

Regiones	Antes del refinado	Después del refinado
Favorables	649 (95,7%)	656 (96,8%)
Permitidas	20 (2,9%)	15 (2,2%)
No permitidas	9 (1,3%)	7 (1,0%)

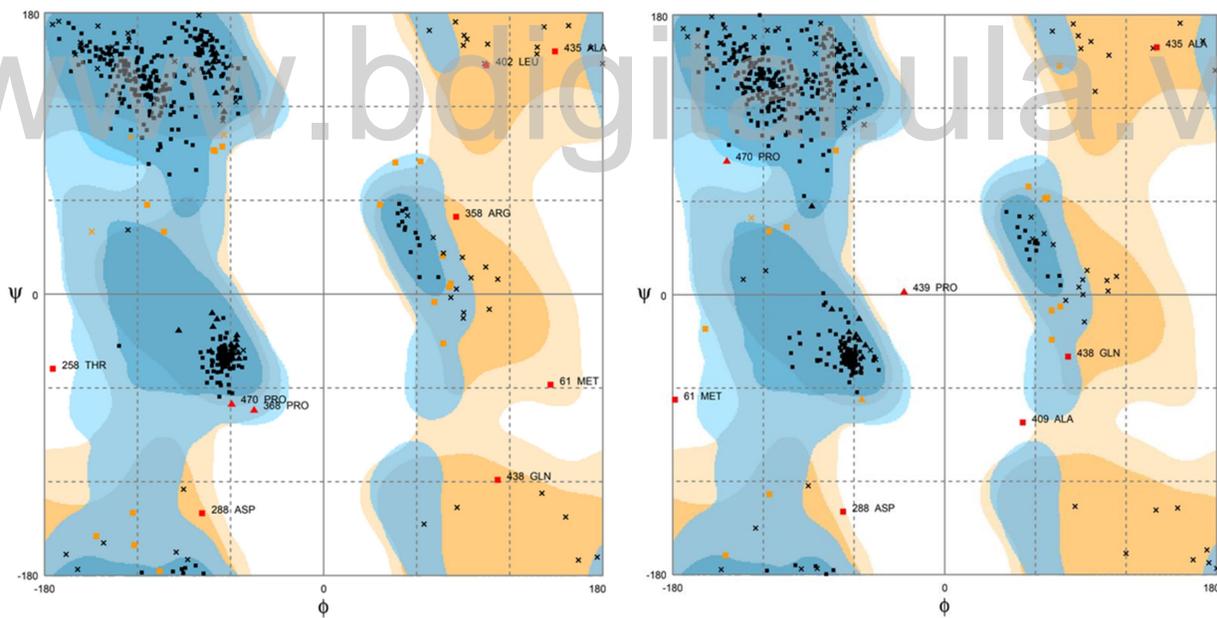


Figura 4. Gráfico de Ramachandran del modelado de la proteína HSP70 antes (izquierda) y después del refinamiento (derecha).

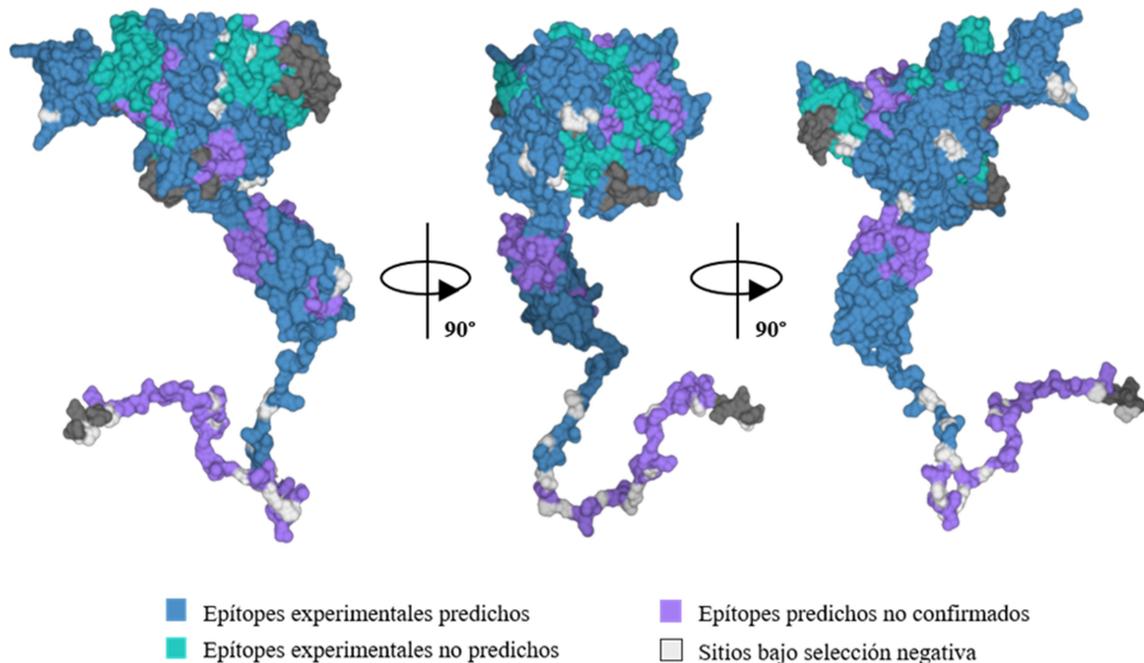


Figura 5. Modelado de la proteína HSP70 de *T. cruzi*.

Modelado de las proteínas a considerar para el diseño del péptido quimérico

Analizando los modelos obtenidos a partir de distintos métodos, se procedió a la escogencia del mejor para cada proteína. Los que fueron refinados repetidas veces hasta obtener la mejor estructura según el gráfico de Ramachandran. En la Tabla 6 se evidencia que el programa más utilizado fue Swiss-Model, empleado en cuatro de las siete proteínas. Del mismo modo, se muestra que el número de refinamientos varió para cada modelo generado. Si se comparan las proteínas TSA-1, Tc24 y Tc80, y si se toma en cuenta que mayores valores de QMEAN indican mejores modelos, puede verse que el número de refinamientos requeridos depende de la calidad de la estructura inicial. En estos casos, modelos con menores valores de QMEAN, como el de TSA-1 y Tc80, requirieron más refinamientos que el de Tc24, con mayor valor de QMEAN. Que también puede verse afectado por el tamaño de la proteína, debido a que como puede observarse, aunque para cruzipaina (25 kDa) se obtuvo un QMEAN de -2,95, inferior al de Tc80 de mayor tamaño (80 kDa), y al de Tc24 proteína de tamaño similar (24 kDa), fueron necesarios solo dos refinamientos para estabilizar este modelo, igual que para Tc24. El efecto del tamaño de la proteína se hace evidente en el modelado de PPK (monómero de 100 kDa) y TSSA (15 kDa), donde se observa que a pesar que la primera obtuvo una mejor puntuación, $Z = 45,95$, precisó de 5 refinamientos, a diferencia de TSSA, que con menor puntuación, $Z = 1,17$, se refinó en dos repeticiones. Por otra parte, LYT-1 resultó ser el caso más extremo, proteína que requirió de ocho refinamientos, lo que era de esperarse al considerar la ausencia de un único templado en el proceso de modelado. En este caso, el programa Phyre2, se encargó de hacer un modelo por agrupación de fragmentos, que disminuye su calidad inicial, y por ende, incrementa el número de pasos de refinamiento. La mejoría progresiva del modelo se hace particularmente notable en esta proteína si se comparan los valores derivados de los gráficos de Ramachandran (Tabla 7), debido a que como es de esperarse, las diferencias en estos valores son mayores conforme mayor es el número de repeticiones. Que resultó igualmente notable para TSA-1. Sin embargo, también destaca la mejoría del modelo de TSSA, que con únicamente dos repeticiones obtuvo un incremento del 31,1% de aminoácidos encontrados en las regiones favorables (Tabla 7), que como se mencionó anteriormente, pudo estar facilitado por el

pequeño tamaño de esta proteína. Los cambios en la estructura tridimensional de las proteínas pueden observarse en la Figura 6. Asimismo, en la Figura 7 se muestran los gráficos de Ramachandran de todas las proteínas obtenidas, antes y después de la etapa de refinamiento.

Posterior a la obtención de los modelos refinados, estos fueron editados en Chimera para localizar los epítopes lineales de células B predichos. En todos los casos se evaluó visualmente si los mismos se encontraban en la superficie, accesibles y libres de impedimentos estéricos impuestos por la misma proteína, que imposibilitarían el reconocimiento de anticuerpos y receptores de células B. Esta identificación puede verse en detalle en la Figura 8, donde se tiene una vista ampliada de la proteína Tc80 en la que además de indicarse todos los epítopes predichos, se señala uno de los descartados como componente del péptido quimérico. Por último, en la Figura 9 se observan los modelos de todas las proteínas, en cintas y en esferas de superficie, en éstos se señalaron todos los epítopes y se indicaron algunos de los excluidos, así como los sitios bajo selección negativa contenidos en estas regiones. Es importante señalar que en todos los casos, se realizó la búsqueda de los dominios proteicos, a fin de asegurar que ningún epítopo se localizara en una región transmembrana, en caso de que la misma hubiese sido modelada, como en el caso de HSP70. Sin embargo, para todas las proteínas, a excepción de LYT-1, se confirmó la ausencia de estas regiones en los modelos obtenidos.

Tabla 6. Programas de modelado y refinamiento utilizados en cada proteína problema y características de los resultados.

Proteína	Modelado			Refinamiento	
	Programa	Puntuación	Templado	Programa	N° de refinamientos
Cz	Swiss-Model	*QMEAN -2,95	Monómero	Galaxy Refine	2
LYT-1	Phyre2	Ensamblaje con 41,4% de confianza	c5od9B		8
TSA-1	Swiss-Model	*QMEAN -4,75	1ms4.1.A		6
Tc24	Swiss-Model	*QMEAN 0,26	3cs1.1		2
Tc80	Swiss-Model	*QMEAN -1,84	1vz3.1		4
PPDK	LOMETS	⁺ Z 45,95	1vbgA		5
TSSA	I-TASSER	⁺ Z 1,17	6igzG	Galaxy Refine 2	2

* Valores de QMEAN alrededor de 0 indican buenos modelos, mientras que valores iguales o menores a -4,0 indican modelos de menor calidad.

⁺ Valores de Z mayores a 1 indican modelos aceptables.

Tabla 7. Porcentaje de aminoácidos de las proteínas problema en las regiones del gráfico de Ramachandran antes y después de los refinamientos.

Proteína	Validación con RAMPAGE		
	Regiones	Antes del refinado	Después del refinado
Cz	Favorables	280 (92,1%)	291 (95,7%)
	Permitidas	20 (6,6%)	11 (3,6%)
	No permitidas	4 (1,3%)	2 (0,7%)
LYT-1	Favorables	375 (68,2%)	507 (92,2%)
	Permitidas	83 (15,1%)	35 (6,4%)
	No permitidas	92 (16,7%)	8 (1,5%)
TSA-1	Favorables	548 (88,2%)	598 (96,3%)
	Permitidas	53 (8,5%)	20 (3,2%)
	No permitidas	20 (3,2%)	3 (0,5%)
Tc24	Favorables	188 (98,9%)	189 (99,5%)
	Permitidas	2 (1,1%)	1 (0,5%)
	No permitidas	0 (0,0%)	0 (0,0%)
Tc80	Favorables	650 (94,6%)	676 (98,4%)
	Permitidas	28 (4,1%)	9 (1,3%)
	No permitidas	9 (1,3%)	2 (0,3%)
PPDK	Favorables	842 (92,4%)	877 (96,3%)
	Permitidas	48 (5,3%)	29 (3,2%)
	No permitidas	21 (2,3%)	5 (0,5%)
TSSA	Favorables	59 (65,6%)	87 (96,7%)
	Permitidas	19 (21,1%)	1 (1,1%)
	No permitidas	12 (13,3%)	2 (2,2%)

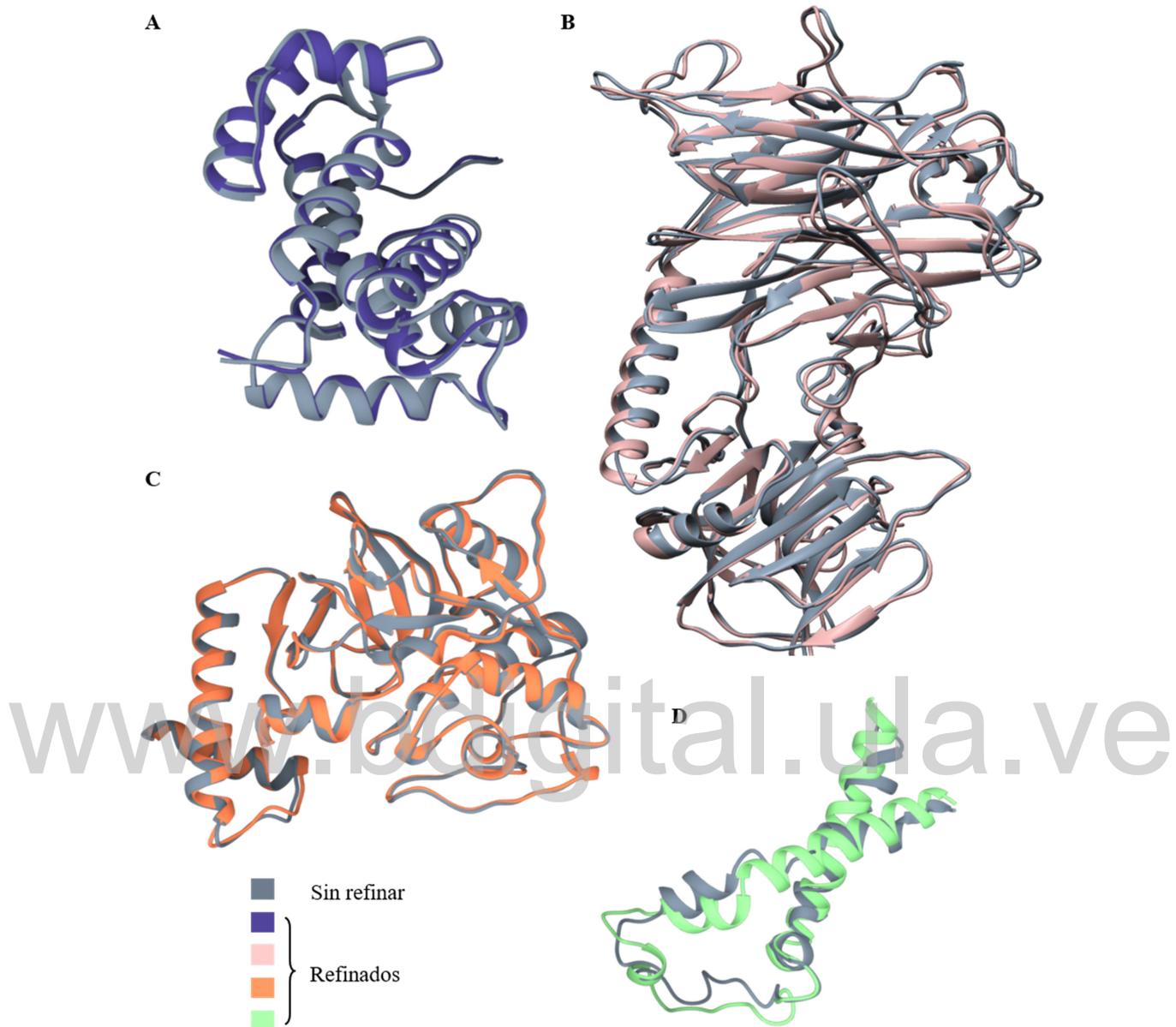
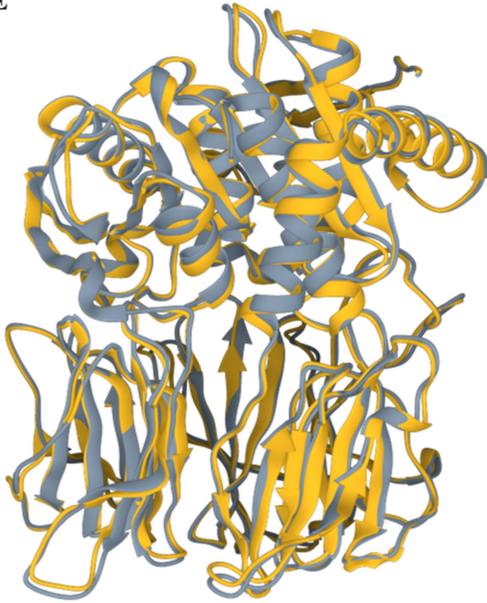


Figura 6. Comparación de la estructura de las proteínas problemáticas, antes y después de los refinamientos. A, Cz; B, TSA-1; C, Tc24; D, TSSA; E, Tc80; F, PPK; G, LYT-1.

E**F****G**

Sin refinar
 } Refinados

Figura 6 (cont.). Comparación de la estructura de las proteínas problemáticas, antes y después de los refinamientos. A, Cz; B, TSA-1; C, Tc24; D, TSSA; E, Tc80; F, PPK; G, LYT-1.

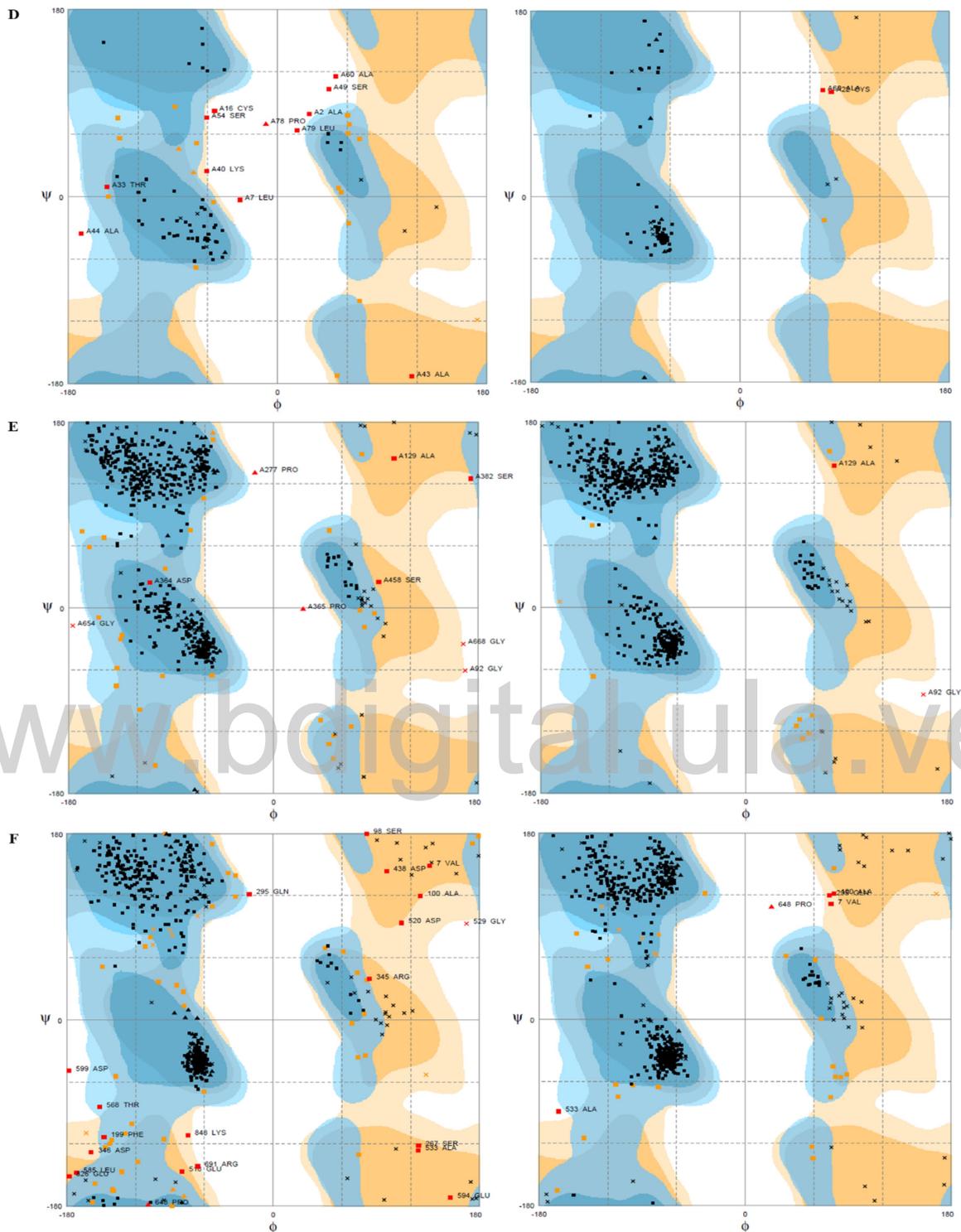


Figura 7 (cont.). Gráficos de Ramachandran de los modelos de las proteínas problema antes (izquierda) y después de los refinamientos (derecha). A, Cz; B, TSA-1; C, Tc24; D, TSSA; E, Tc80; F, PPK; G, LYT-1.

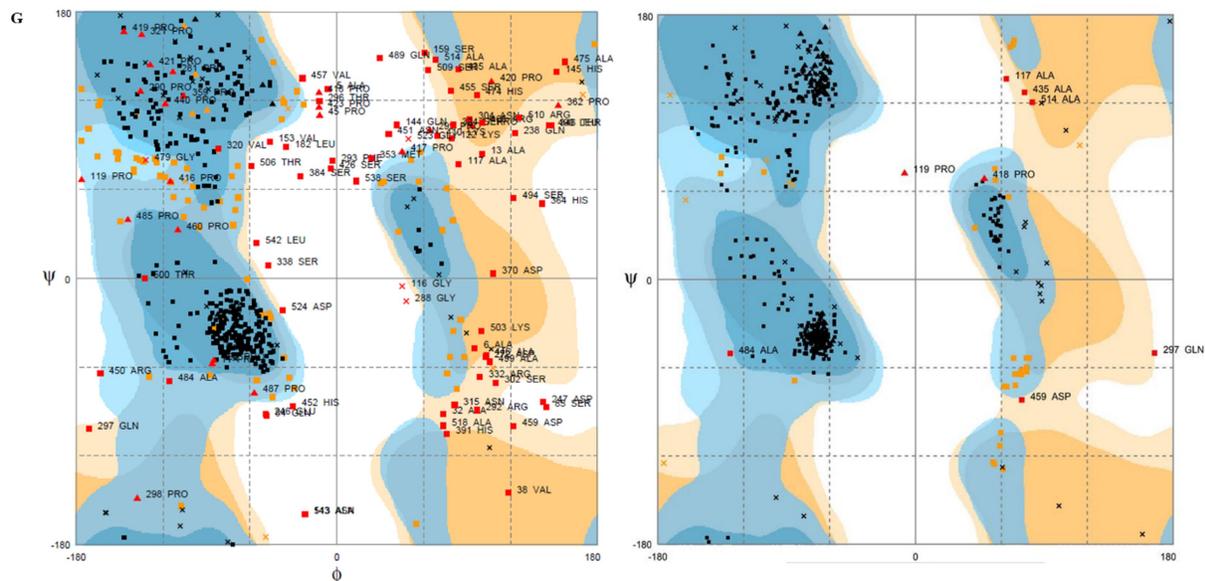
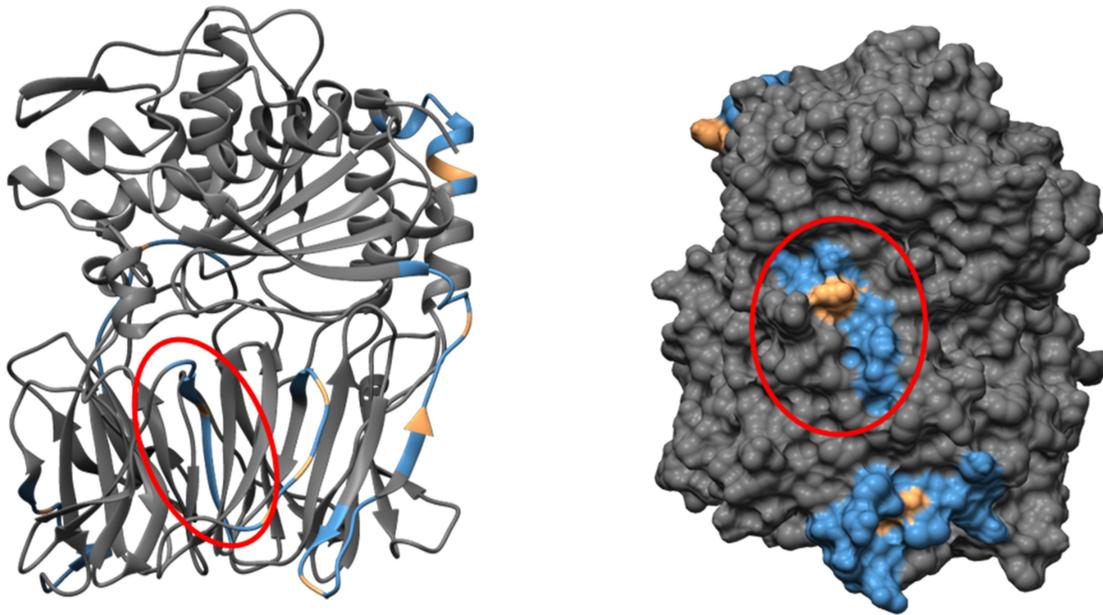


Figura 7 (cont.). Gráficos de Ramachandran de los modelados de las proteínas problema antes (izquierda) y después de los refinamientos (derecha). A, Cz; B, TSA-1; C, Tc24; D, TSSA; E, Tc80; F, PPK; G, LYT-1.

www.bdigital.ula.ve



■ Epítoto lineal de células B

■ Selección negativa

www.bdigital.ula.ve

Figura 8. Vista amplificada del modelo de la proteína Tc80 donde se indican por color los epítotoes lineales de células B y los sitios bajo selección negativa en ellos contenidos. Con óvalos rojos se señalan epítotoes descartados para su incorporación en el péptido quimérico.

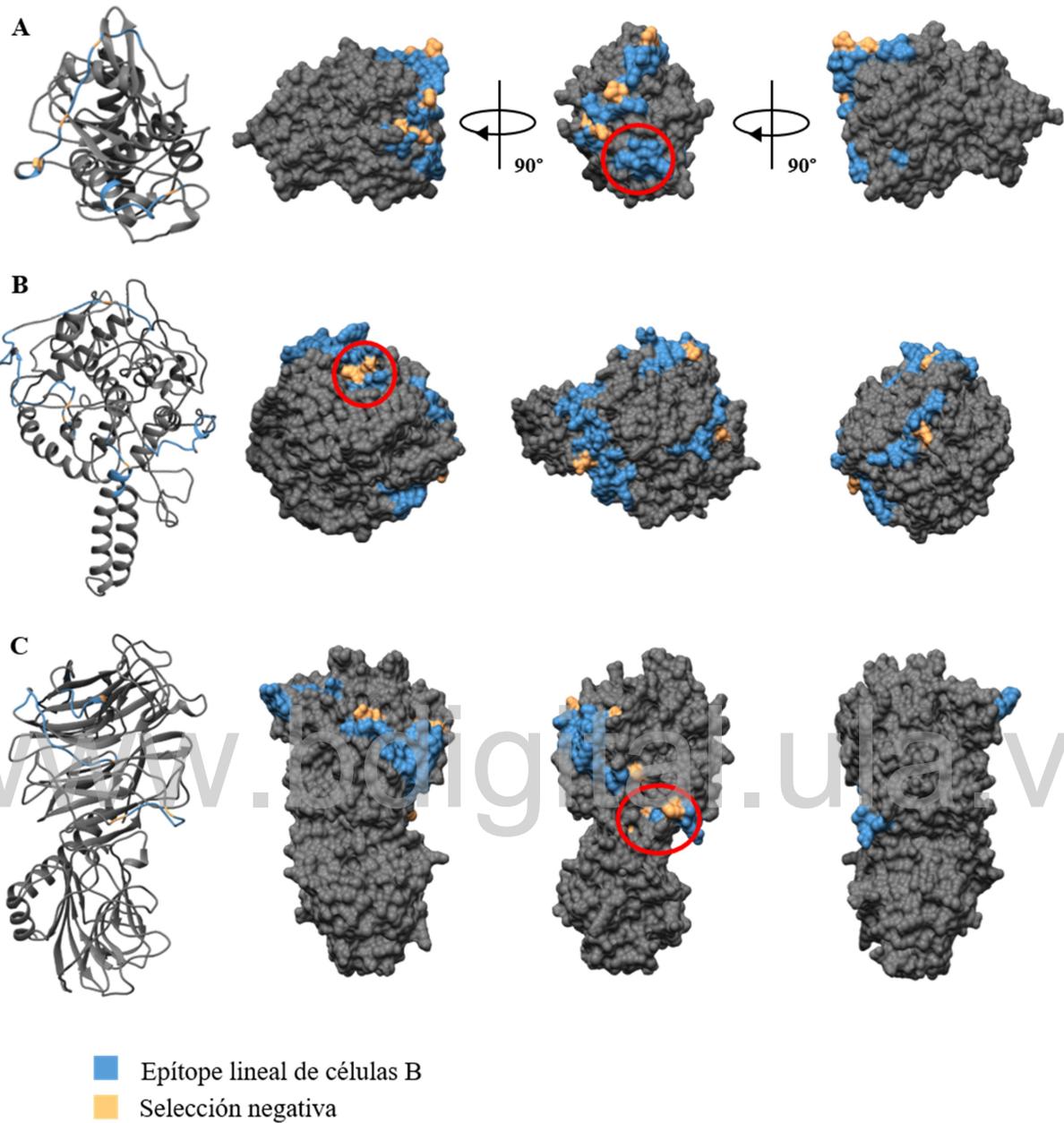
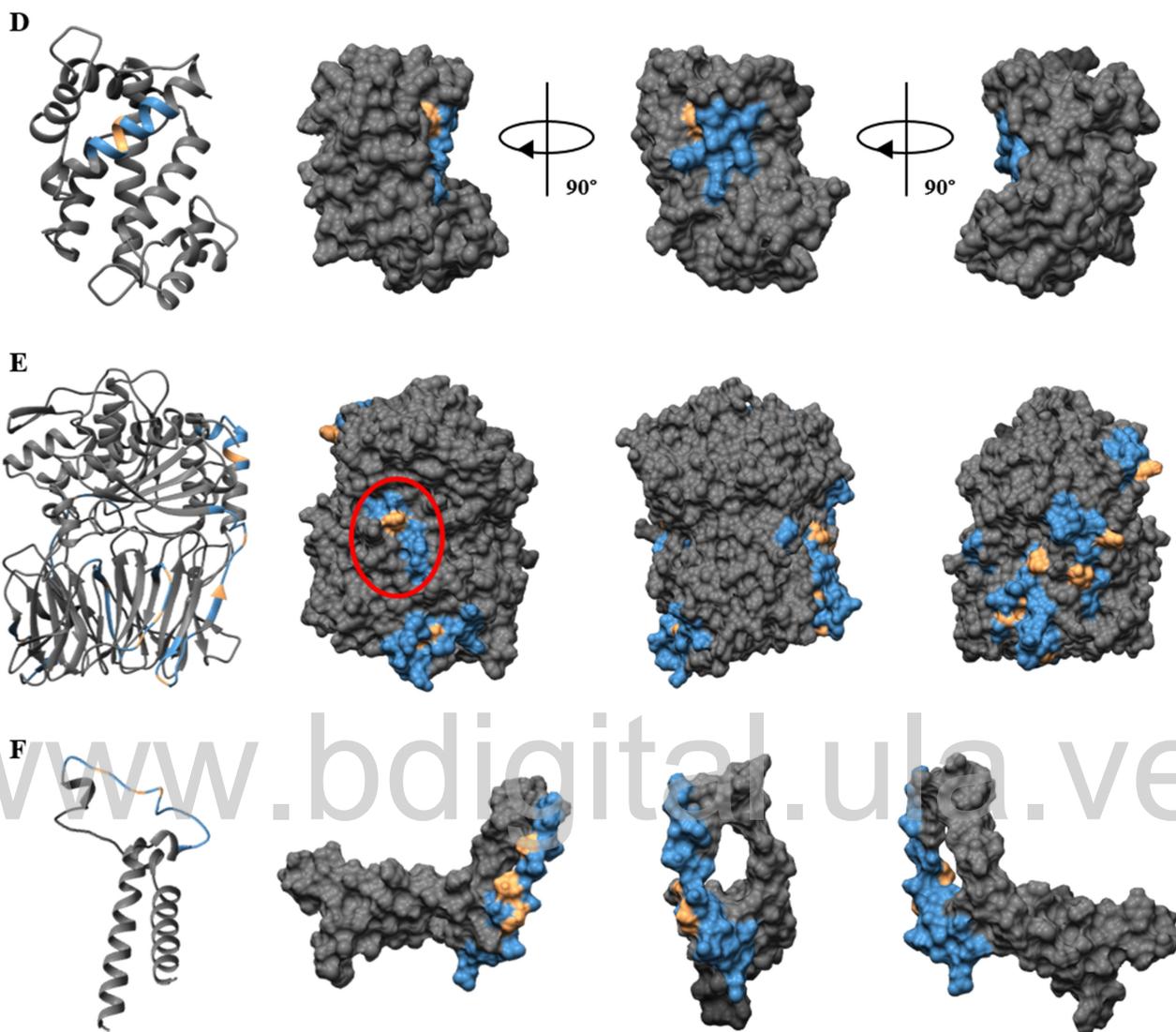


Figura 9. Modelos de las proteínas problema con el señalamiento por color de los epítotos lineales de células B y los sitios bajo selección negativa. Con óvalos rojos se indican epítotos descartados para su incorporación en el péptido quimérico. A, cruzipaina; B, LYT-1, C, TSA-1, D, Tc24; E, Tc80; F, TSSA; G, PPKK.



- Epítoto lineal de células B
- Selección negativa

Figura 9 (cont.). Modelos de las proteínas problema con el señalamiento por color de los epítotos lineales de células B y los sitios bajo selección negativa. Con óvalos rojos se indican epítotos descartados para su incorporación en el péptido quimérico. A, cruzipaina; B, LYT-1, C, TSA-1, D, Tc24; E, Tc80; F, TSSA; G, PPDK.

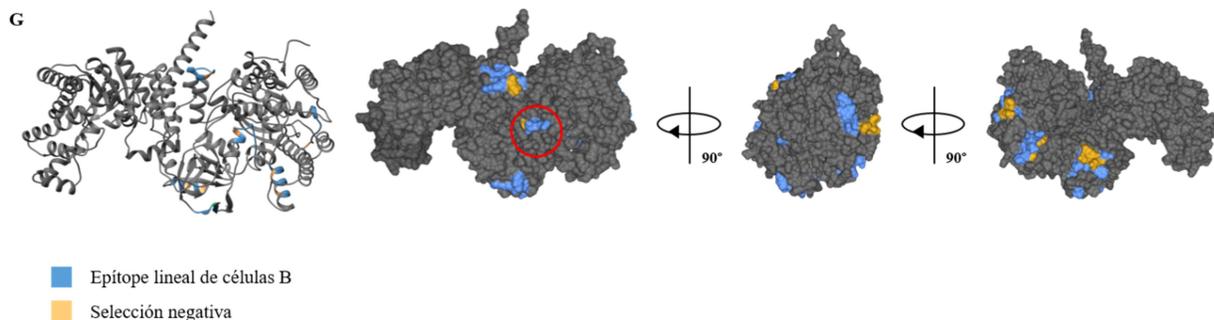


Figura 9 (cont.). Modelos de las proteínas problema con el señalamiento por color de los epítotos lineales de células B y los sitios bajo selección negativa. Con óvalos rojos se indican epítotos descartados para su incorporación en el péptido quimérico. A, cruzipaina; B, LYT-1, C, TSA-1, D, Tc24; E, Tc80; F, TSSA; G, PPK.

Diseño del prototipo inmunoestimulante

El péptido quimérico resultante estuvo formado por 12 epítotos lineales de células B, 12 péptidos de unión a MHC-I y 13 péptidos de unión a MHC-II (Tabla 8). Adicionalmente, las conexiones entre estos se hicieron con los *linkers* AAY y GPGPG. En la Tabla 8 se señalan en **negrita** los aminoácidos que no forman parte de los epítotos de células B, pero que se añadieron para mantener la integridad estructural del péptido. Del mismo modo, se incorporó a la quimera el péptido sintético PADRE con el *linker* HEYGAEALERAG por su capacidad para potenciar la respuesta celular (Athanasidou *et al.*, 2017). En total, el péptido quimérico estuvo formado por 631 aminoácidos, que indican una masa molecular aproximada de 65 kDa, valor que se aproxima al de otros prototipos diseñados (Khatoon *et al.*, 2018; Khatoon *et al.*, 2017).

En la Figura 10 se muestra la secuencia aminoacídica del péptido, que arrojó resultados negativos al ser analizada con los predictores de alergenicidad AllerTOP y AllergenFP (Figura 11). Asimismo, en la Figura 12 se muestra la comparación de la estructura tridimensional de los modelos antes y después de las etapas de refinamiento. Igualmente, los gráficos de Ramachandran y valores derivados se encuentran en la Figura 13 y la Tabla 9. Mientras que la estructura final del péptido donde se señalan sus componentes se muestran en las Figuras 14 y 15.

Por último, debido a la falta de inmunogenicidad intrínseca de las vacunas de subunidades, se seleccionó el adyuvante dinucleótido cíclico adenosín monofosfato, o c-di-AMP, como componente inmunomodulador del prototipo. Molécula implicada en diferentes procesos metabólicos en procariontas por su rol como segundo mensajero, que contiene patrones moleculares asociados a patógenos que actúan como señales de peligro reconocidas por el sistema inmune del huésped (Libanova *et al.*, 2012).

Tabla 8. Péptidos incluidos en el diseño del péptido quimérico.

Péptido	Tipo	Alelo	Proteína
---------	------	-------	----------

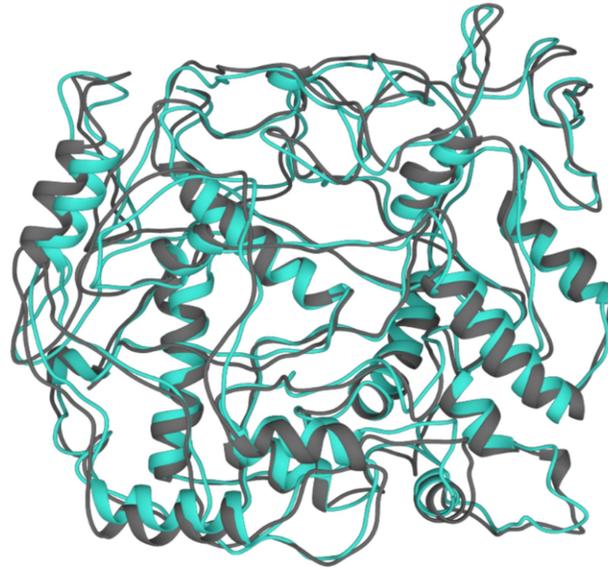
*PAAAQERARVPVKVEVVGAPAA			Cz
ARRMRYGYEE			LYT-1
*TGFPPEDPMT			PPDK
DADSAKEWSA			PPDK
*EDFEEAIGHMKEK			PPDK
QFTLSGRVFR			PPDK
*VRDITKRAFDKARALGSKLE	B	NA	Tc24
NAIPLPIG			Tc80
*FASEAELRKKIFESISNSQ			Tc80
FKDDTVEGLL			Tc80
DDGQKWEIPG			TSA-1
AGEAPSPSGA			TSSA
HLERFVYDSY		HLA-A*01:01	PPDK
FLAGHPLTNL		HLA-A*02:01	Cz
GLMNNAFEWI		HLA-A*02:01	Cz
YIVGTMIEV		HLA-A*02:01	PPDK
RFITMYADI		HLA-A*024:01	PPDK
MYKFHKFTI		HLA-A*024:01	Tc80
RVRDITKRAF	MHC-I	HLA-B*07:02	Tc24
APSPSGASS		HLA-B*07:02	TSSA
FQRKTGPPF		HLA-B*15:01	PPDK
GRMGSPKRF		HLA-B*27:05	LYT-1
RRFITMYADI		HLA-B*27:05	PPDK
AEFAFRLSV		HLA-B*40:01	Cz
MAAGTLMVW		HLA-B*58:01	TSA-1
IREMILADTLEGRKA		DRB1_0301	PPDK
NAFEWIVQENNGAVY		DRB1_0401	Cz
TKRAFDKARALGSKL		DRB1_0401	Tc24
GELVRIAVQKGRRVK		DRB1_1301	PPDK
AFRLSVFRENFLAR		DRB1_1501	Cz
GNVECQWFLAGHPLT		HLA-DQA10101-DQB10501	Cz
NEQNAFFEEYFASEA	MHC-II	HLA-DQA10101-DQB10501	Tc80
ALQRRYWEQSRAILE		HLA-DQA10101-DQB10501	LYT-1
DYIVGTMIEVPRAAI		HLA-DQA10102-DQB10602	PPDK
PAFSSSRMVFLRDLG		HLA-DQA10102-DQB10602	Tc80
KVFLITYSTDDGQK		HLA-DQA10201-DQB10202	TSA-1
MVWFALSGGGVYCCC		HLA-DQA10501-DQB10301	TSA-1
TTSGHTVGATITGHV		HLA-DQA10501-DQB10301	Cz

HLERFVYDSYAAYMVWFALSGGGVYCCCGPGPGAFRLSVFRENLFLARGPG
 PGFLAGHPLTNLAAYAGEAPSPSGAGPGPGDADSAKEWSAGPGPGGNVECQ
 WFLAGHPLTGPGPGRFITMYADIAAAYPAAQERARVPVKVEVVGAPAAGPG
 PGRVRDITKRAFAAYNAFEWIVQENNGAVYGPGPGTTSGHTVGATITGHVGP
 GPGEDFEEAIGHMKEKGPGPGNAIPLIGGPGPGMYKFHKFTIAAYFASEAE
 LRKKIFESISNSQGPGPGVRDITKRAFDKARALGSKLEGPGPGDDGQKWEIP
 GPGPGRRFITMYADIAAYTGFPPEDPMTGPGPGALQRRYWEQSRAILEGP
 GPGIREMILADTLEGRKAGGPGPGAEEAFRLSVAAYARRMRYGYEEGPGPGYI
 VGTMIEVAAYDYIVGTMIEVPRAAIGPPGPGTKRAFDKARALGSKLGPGPGGE
 LVRIAVQKGRRVKGPGPGNEQNAFFEEYFASEAGPPGPGFQRKTGFPFAAYPA
 FSSSRMVFLRDLGPGPGQFTLSGRVFRGPGPGGRMGSPKRFAAYKVFSLIT
 YSTDDGQKGPGPGFKDDTVEGLLGPGPGMAAGTLMVWHEYGAELERAG
 AKFVAAWTLKAAA

Figura 10. Secuencia aminoacídica del péptido quimérico. En naranja, péptidos de unión a MHC-I; en azul, péptidos de unión a MHC-II; en morado, epítopes lineales de células B; en verde, péptido PADRE; en negro, *linkers*.

The image shows two side-by-side screenshots of allergenicity prediction tools. On the left is 'Allertop v. 2.0' and on the right is 'AllergenFP v.1.0'. Both tools display a 3D protein structure with various regions highlighted in colors (orange, blue, purple, green, black). Below the structure, a binary sequence of 0s and 1s is shown. The main result for both is 'PROBABLE NON-ALLERGEN'. The left tool also provides the UniProtKB accession number Q86UU0, while the right tool provides Q9P281. Both mention a Tanimoto similarity index of 0.93.

Figura 11. Resultados de las pruebas de alergenicidad del péptido quimérico.



Sin refinar
 Refinado

Figura 12. Comparación de la estructura del péptido quimérico antes y después de los refinamientos.

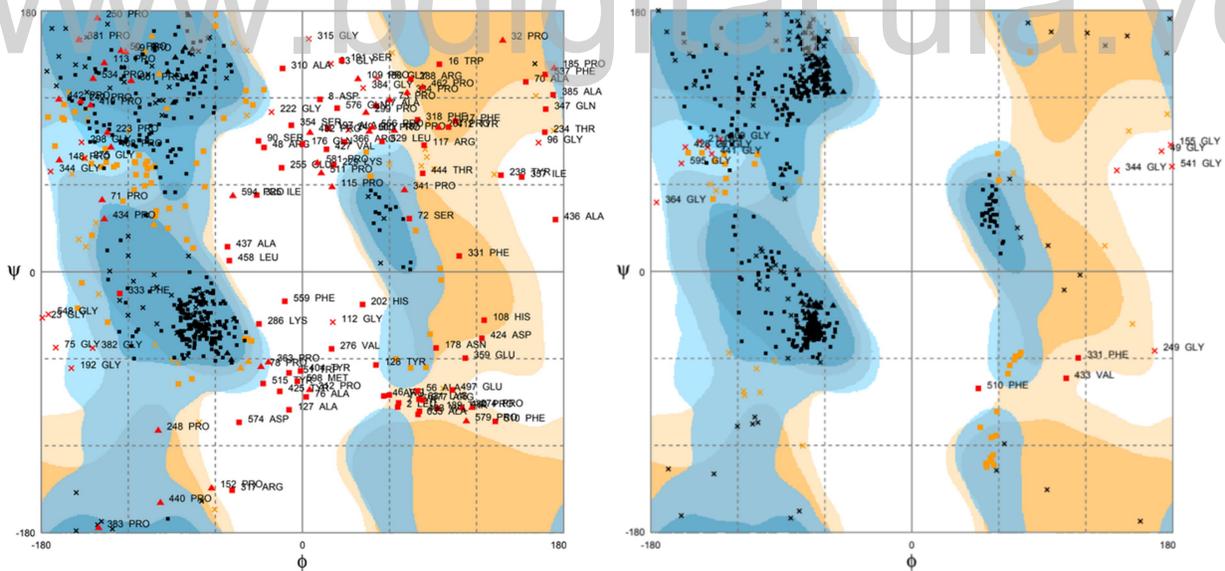


Figura 13. Gráficos de Ramachandran del modelo del péptido quimérico antes (izquierda) y después de los refinamientos (derecha).

Tabla 9. Porcentaje de aminoácidos de HSP70 en las regiones del gráfico de Ramachandran antes y después del refinamiento del modelo.

Regiones	Antes del refinado	Después del refinado
Favorables	409 (64,7%)	573 (90,7%)
Permitidas	99 (15,7%)	44 (7,0%)
No permitidas	124 (19,6%)	15 (2,4%)

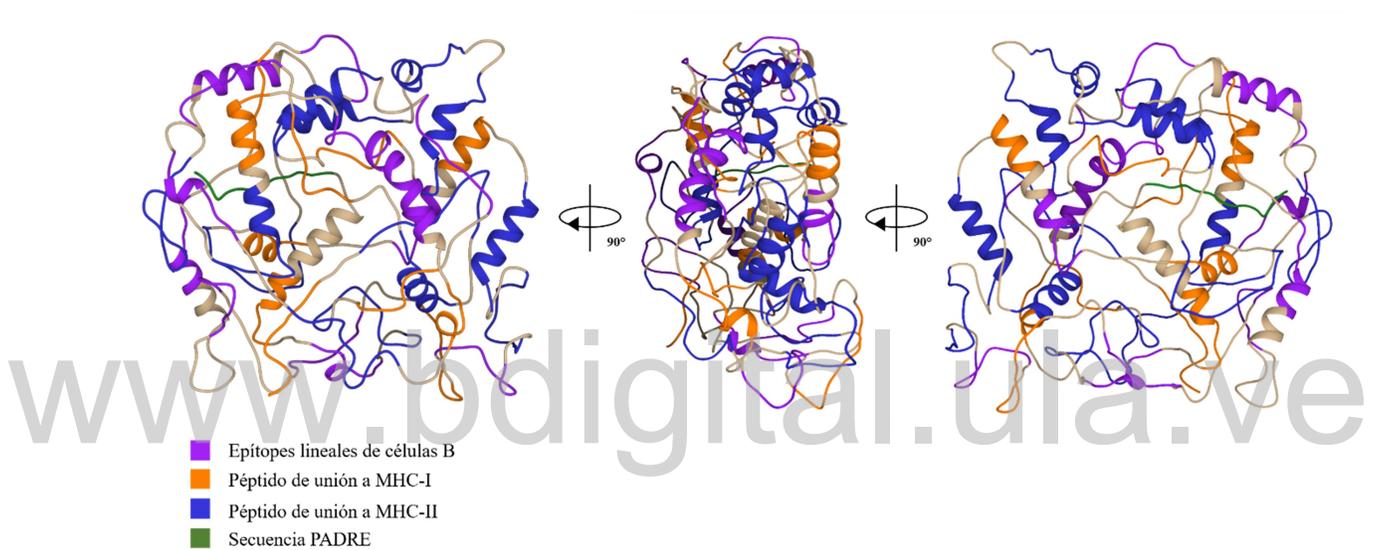


Figura 14. Modelo en cintas del péptido quimérico indicando sus componentes.

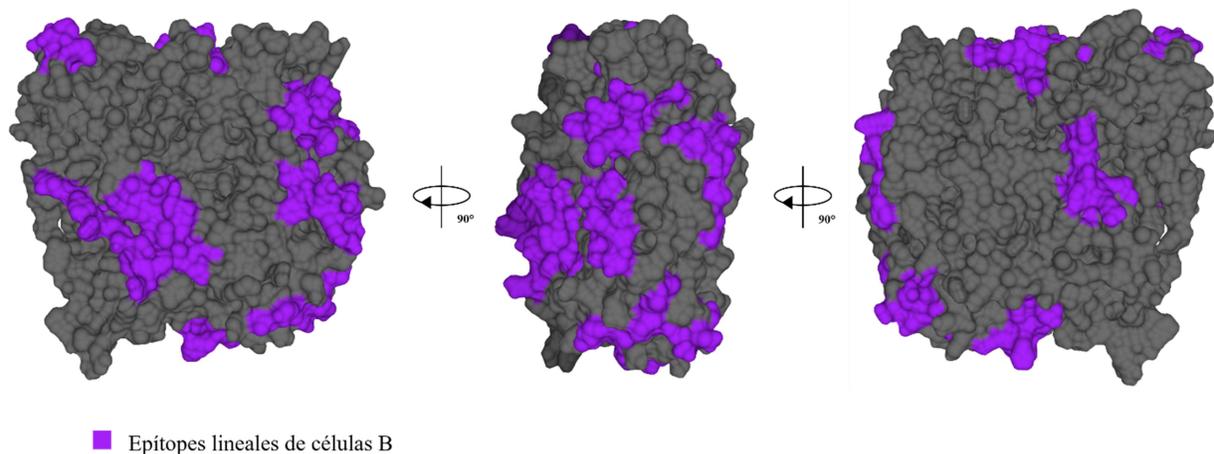


Figura 15. Modelo del péptido quimérico con esferas de superficie señalando los epítopes lineales de células B.

VI. DISCUSIÓN

Evaluación de predictores

El entendimiento de la etiología de una enfermedad, el monitoreo inmune, la inmunoterapia, el desarrollo de pruebas diagnósticas y el diseño de vacunas, son algunas de las razones que incentivan la identificación de epítopes (Sanchez-Trincado *et al.*, 2017). Tradicionalmente, esta identificación se ha basado completamente en métodos experimentales, laboriosos y costosos. Sin embargo, gracias a los métodos computacionales de predicción, la complejidad de estos procedimientos puede ser disminuida mediante la reducción del repertorio de péptidos a considerar (Kazi *et al.*, 2018; Sanchez-Trincado *et al.*, 2017). Por lo tanto, se vuelve necesaria la aplicación de los mejores algoritmos a fin de aumentar la probabilidad de obtención de verdaderos epítopes funcionales. Lo que motivó a que en esta investigación se llevara a cabo la evaluación previa de los programas disponibles para la utilización subsiguiente de los predictores con mejor desempeño, haciendo uso exclusivamente de epítopes de *T. cruzi* con soporte experimental. Los que fueron obtenidos del repositorio del IEDB, que en la actualidad representa la mayor base de datos de epítopes, razón por la que fue empleada en el estudio.

Por medio de la evaluación de predictores de epítopes lineales de células B, fue posible comparar los resultados de métodos clásicos, basados en escalas de propensión, y de algoritmos modernos, basados en aprendizaje automatizado (Figura y Tabla 1). Así se comprobó que el desempeño de estos últimos con respecto a los primeros, es bastante superior.

Si bien los métodos clásicos se fundamentan en propiedades fisicoquímicas importantes como la hidrofilia y la accesibilidad a la superficie, es evidente que las escalas de propensión tienen una eficiencia de categorización cuestionable para evaluar proteínas de *T. cruzi*. Lo que tal vez se deba a que estos programas fueron diseñados tomando en cuenta un conjunto reducido de epítopes confirmados para la época, los que podrían representar una muestra no significativa, y por lo tanto, tener propiedades que no sean simplemente extrapolables a otros epítopes. No obstante, estos resultados no se limitan al análisis de proteínas de este parásito, ya que los hallazgos de esta investigación están en concordancia con otros estudios, donde se ha encontrado que el poder de

predicción de estos métodos es generalmente superado por algoritmos modernos (Costa *et al.*, 2013). Las ventajas del aprendizaje automatizado recaen, principalmente, en el modo y la capacidad de análisis de una cantidad muy amplia de datos de entrenamiento. Lo que le permite a estos métodos considerar las múltiples diferencias de interrelación entre aminoácidos que puedan establecerse en cada uno de los epítopes de entrenamiento, contribuyendo a la generación de mejores predicciones (Backert and Kohlbacher, 2015; Sanchez-Trincado *et al.*, 2017). Cualidades con las que cuentan los programas ABCPred y BepiPred 2.0 y que fueron comprobadas con los porcentajes de cobertura encontrados.

En los resultados de evaluación de péptidos de unión a MHC-I, también se determinó que los métodos basados en aprendizaje automatizado, como NetMHC 4.0, NetMHCpan 3.0 y NetMHCpan 4.0, fueron más precisos en las predicciones, como lo demuestran sus valores de AUC (Figura y Tabla 2), que fueron superiores a los conseguidos por los algoritmos basados en matrices de puntuación, SYFPEITHY y PREDEP, los que solo contemplan la contribución de aminoácidos de anclaje para su análisis. Por otra parte, SMM, método que se consideraba superior a estos dos últimos, pero inferior a los predictores Net, fue hallado como el mejor predictor en este estudio. Sin embargo, es importante tomar en cuenta que la evaluación aquí realizada contemplaba únicamente la identificación de epítopes de *T. cruzi* de longitudes de 9 y 10 mer reconocidos por moléculas MHC-I del alelo HLA-A 02*01, lo que podría estar induciendo estas variaciones. La evaluación del poder predictivo utilizando un mayor rango de alelos sería necesaria para ratificar las posiciones de desempeño.

Si se toma en cuenta que los programas de predicción varían en los datos de entrenamiento y que las investigaciones de valoración de predictores varían en: el origen biológico de los datos de evaluación, las longitudes de epítopes y el conjunto de alelos contemplados. Se vuelve pertinente la escogencia de predictores en función de la representatividad de las proteínas del organismo problema en los datos de entrenamiento, y del desempeño predictivo de acuerdo a los alelos y longitudes de péptidos requeridas. En ese sentido, esta investigación puede ser tomada como referencia para la escogencia métodos de predicción de epítopes lineales de linfocitos B y de péptidos de unión a moléculas MHC-I de *T. cruzi*.

Vale destacar, que la evaluación de algoritmos de predicción de epítopes conformacionales de células B no se realizó debido a la imposibilidad de integrar este tipo de epítopes en un péptido quimérico, por lo que se decidió prescindir del empleo de estos métodos.

Enfoque evolutivo

El conjunto de interacciones que se establecen en la interacción entre vector-parásito-hospedador, puede generar “marcas evolutivas” en los genomas de las especies interactuantes, producto de eventos de coevolución. El descubrimiento y entendimiento de estas marcas puede explicar la capacidad de los parásitos para invadir múltiples tejidos en huéspedes y vectores mientras evaden el control inmunitario. En contraparte, este conocimiento puede ser empleado en el diseño de estrategias que permitan controlar los ciclos de transmisión (Early *et al.*, 2018).

En múltiples investigaciones se ha demostrado que la respuesta inmune del huésped ejerce una fuerte presión selectiva en los patógenos, lo que afecta directamente la diversidad de sus poblaciones (Berger *et al.*, 2010). En relación a esto, se ha visto que el establecimiento de una respuesta protectora provocada por el reconocimiento de determinados epítopes conduce a la eliminación de los individuos susceptibles y a la supervivencia de aquellos cuya variabilidad le confiera un aumento en su *fitness* (Berger *et al.*, 2010; Raymond *et al.*, 2018). Lo que conlleva a que los sitios polimórficos sean considerados como sitios de evasión inmune, en los que una selección balanceadora favorece el mantenimiento de la diversidad. Es por ello que existe interés en la identificación de regiones conservadas capaces de inducir mecanismos de protección inmunitaria (Raymond *et al.*, 2018).

En concordancia, se ha visto que el grado de conservación varía entre antígenos de una misma especie (Berger *et al.*, 2010; Comas *et al.*, 2010; Early *et al.*, 2018; Li *et al.*, 2019). Que corresponde a lo encontrado en el estudio de las presiones selectivas que actúan a nivel de las proteínas antigénicas evaluadas de *T. cruzi*. Donde resultó interesante la presencia de una gran cantidad de sitios conservados en los antígenos analizados, lo que podría explicarse, en cierta parte, mediante la “hipótesis de restricciones funcionales”, donde se plantea que la conservación de epítopes puede originarse por su localización en regiones proteicas donde los cambios de aminoácidos sean poco tolerados (Berkhoff *et al.*, 2005, 2006; Li *et al.*, 2019). Según esta hipótesis, las mutaciones no sinónimas en estos sitios, que si bien serían, en principio, beneficiosas para la evasión inmune, tendrían un efecto negativo en la función de la proteína. Por lo tanto, el costo en el *fitness* dependería de su importancia para la supervivencia y la reproducción del organismo en cuestión (Berkhoff *et al.*, 2005, 2006; Li *et al.*, 2019).

En otros organismos como *Mycobacterium tuberculosis*, *Salmonella enterica* y algunos virus, también se ha observado que ciertos epítopes presentan un alto grado de conservación (Comas *et al.*, 2010; Frank and Bush, 2007; Holt *et al.*, 2008; Li *et al.*, 2019). Para el primer caso, no solo se demostró que los genes que codifican epítopes están bajo selección purificadora, sino que además, se determinó que algunos de ellos están más conservados que los llamados “genes esenciales”. Los autores argumentan que además de evitar cambios funcionales indeseados en las proteínas, la ausencia de sitios bajo selección positiva llevaría a que se facilite el reconocimiento inmune por linfocitos T, evento que podría favorecer el establecimiento de una infección latente y la transmisión de la bacteria (Comas *et al.*, 2010). Mecanismo que podría compartir con *S. enterica*, organismo con similar ciclo de vida (Holt *et al.*, 2008). Por otro lado, la conservación de antígenos en virus como el del sarampión, rubeola, paperas e influenza C, ha sido explicada por la ausencia de presión inmune de sus huéspedes habituales, niños sin células de memoria protectoras (Frank and Bush, 2007).

Como se ha mencionado, la hipótesis de restricciones funcionales no podría ser la única causa del surgimiento de epítopes con sitios conservados, por lo que, podría sugerirse que en *T. cruzi* el reconocimiento inmune también podría traer, en cierta medida, un beneficio para el parásito. A pesar de la complejidad genética de este organismo, y de la amplia variedad de proteínas potencialmente antigénicas expresadas, se ha encontrado que existe una inmunodominancia importante en los epítopes de células CD8⁺ humanas, que se genera por la presentación predominante de un pequeño conjunto de epítopes de proteínas pertenecientes a la familia de las transalidasas (Martin *et al.*, 2006). Por lo tanto, tomando en cuenta que la inmunodominancia no es sinónimo de protección inmunológica (Dinter *et al.*, 2015), podría pensarse que el mantenimiento de la conservación de algunos epítopes es consecuencia del establecimiento de una respuesta inmunitaria ineficiente hacia estos antígenos, que previene la completa eliminación del parásito. Adicionalmente, considerando que el conjunto de epítopes inmunodominantes presentados tiende a variar de acuerdo al perfil genético HLA del hospedador y de acuerdo a las cepas (Martin *et al.*, 2006), es posible que los sitios polimórficos representen entonces sitios de diferenciación entre poblaciones del huésped, y al mismo tiempo, entre linajes del parásito.

Según Gupta *et al.* (1996), el mantenimiento de las cepas en la estructura poblacional de un agente infeccioso es un evento que puede originarse en organismos que experimentan reproducción tanto sexual como asexual, debido al desequilibrio de ligamiento que puede formarse entre genes que codifican ciertos determinantes antigénicos dominantes. En su modelo, los autores proponen que la estructura por cepas es el resultado de la presentación de epítopes no solapados, epítopes conservados en una misma cepa, pero divergentes entre cepas, lo que podría sugerir la existencia de una respuesta inmune importante contra un antígeno polimórfico dominante, en donde la respuesta contra la forma presente en una no genera protección contra otras variantes. En el caso contrario, donde se establezca

una respuesta inmune beneficiosa para el huésped y dirigida hacia un determinante conservado dominante, entre cepas, no se espera entonces una estructura poblacional por cepas sino el mantenimiento, por exclusión competitiva, de la cepa con un mayor tamaño efectivo o con mayor tasa de reproducción. Según este modelo, la dinámica de una población de patógenos que está estructurada por cepas debido a una respuesta inmune dominante hacia antígenos polimórficos, puede estar influenciada o favorecida por respuestas más débiles a otros determinantes, ya sean conservados o variables (Gupta *et al.*, 1996). Sin embargo, esto no descarta la posibilidad de ocurrencia de cambios en la estructura poblacional por efecto de la presión inmune del huésped. En relación a esto, los autores sugieren que aunque las variaciones de aminoácidos en un determinante antigénico dominante generan una estructura poblacional por cepas, se ha comprobado que la exposición repetida a epítopes conservados puede inducir el establecimiento de una respuesta inmune protectora contra todas las cepas de una población (Gupta and Day., 1994; Gupta *et al.*, 1996).

Bajo esta premisa se podría asumir, que la escogencia de epítopes conservados no involucrados en eventos de inmunodominancia que favorezcan a *T. cruzi*, podría representar una estrategia para diseñar un prototipo que induzca una respuesta inmunitaria capaz de generar protección contra, idealmente, cada cepa de este organismo. Asunción que fue implementada en esta investigación y que llevó a la escogencia de epítopes bajo selección negativa para la creación del péptido quimérico.

Elección y análisis de proteínas problemas

La efectividad de una vacuna depende, en gran medida, de la adecuada selección de los antígenos. En relación, décadas de investigación han permitido identificar proteínas de *T. cruzi* con roles importantes para la activación de la respuesta inmunológica. Lo que ha llevado a catalogarlas como blancos prometedores para el diseño de vacunas preventivas y terapéuticas contra la enfermedad de Chagas (Beaumier *et al.*, 2016; Rodríguez-Morales *et al.*, 2015). Potencialidad que fue considerada en este estudio para la elección de las proteínas problema que estarían sujetas a los análisis inmunoinformáticos y a pruebas evolutivas.

Varias propiedades resultaron relevantes de cada proteína e influyeron en su selección. Cruzipaina, principal cisteína proteasa, fue elegida debido a que participa en eventos de nutrición e invasión celular, se localiza tanto a nivel lisosomal como en la membrana plasmática, es secretada, y además se expresa en todos los estadios y cepas del parásito (Cerny *et al.*, 2016; Duschak and Couto, 2009). Además, es capaz de estimular respuestas humorales y celulares en modelos animales y humanos (Cerny *et al.*, 2016; Duschak and Couto, 2009; Martinez *et al.*, 1991). En concordancia, se ha visto que personas en distintas etapas de la enfermedad poseen anticuerpos contra esta proteína, así como células CD4⁺ y CD8⁺ efectoras, lo que evidencia la inducción de memoria inmunológica (Cerny *et al.*, 2016; Fonseca *et al.*, 2005). Por otro lado, Tc24 es una proteína que actúa como un sensor de calcio, que puede ser secretada y que se ubica en la membrana plasmática durante todas las fases del ciclo de vida de *T. cruzi*, principalmente en el bolsillo flagelar (Martinez-Campos *et al.*, 2015). Su estudio como antígeno ha revelado que participa en la estimulación de células B y linfocitos T. De igual forma, su empleo para la inmunización de modelos animales ha comprobado su eficacia para generar protección, igualmente, su rol terapéutico ha sido demostrado (Martinez-Campos *et al.*, 2015; Villanueva-Lizama *et al.*, 2018). Otra proteína con las mismas propiedades es TSA-1, miembro de la familia de las transalidasas que se expresa en la superficie de tripomastigotes sanguíneos (Villanueva-Lizama *et al.*, 2018; Wizel *et al.*, 1998).

Por otra parte, Tc80, es una prolil olipeptidasa expresada en tripomastigotes sanguíneos y amastigotes que participa en la invasión celular (Bivona *et al.*, 2018). Su inclusión entre las proteínas problema estuvo determinada por una reciente investigación conducida por Bivona *et al.* (2018), en la que se

designó como un nuevo candidato para vacunación, ya que favoreció la producción de anticuerpos que intervinieron en la inhibición enzimática de la proteína y en la neutralización de la infección mediante la lisis de tripomastigotes guiada por complemento. Que se presentó en conjunto con una respuesta celular de linfocitos CD4⁺ y CD8⁺ epítipo-específica.

TSSA, por su parte, pertenece a la familia de proteínas similares a mucinas que se expresa en la superficie de tripomastigotes sanguíneos y tiene un rol en la adhesión celular, prerequisite para la internalización subsiguiente del parásito (Gonzalez and Buscaglia, 2015). Fue elegida debido a que es una proteína que induce una potente respuesta humoral y es reconocida por linfocitos CD8⁺, contribuyendo, por lo tanto, al establecimiento de una respuesta citotóxica que es fundamental para la eliminación del parásito (Takayama *et al.*, 2010).

Por otro lado, PPK es una enzima auxiliar de la glicólisis ubicada en los glicosomas que contribuye al balance ATP/ADP y como enlace entre rutas catabólicas y de biosíntesis (Acosta *et al.*, 2004). El hecho de que PPK no sea homóloga a ninguna proteína expresada por el hospedador, hace que sea considerada en este estudio. Lo que fue además reafirmado al encontrar que los algoritmos de predicción detectaron que posee péptidos de unión a moléculas MHC-I y -II con puntuaciones similares a las péptidos predichos de proteínas con inmunogenicidad comprobada. Lo que no descarta la posibilidad de que PPK sea procesada y reconocida por células presentadoras de antígeno, y en el mejor de los casos, participe en la activación de una respuesta celular.

Por último, LYT-1, es una proteína que presenta actividad lítica en condiciones ácidas y que tiene un rol importante en el proceso de infección y en los cambios de estadio de *T. cruzi*. Su presencia ha sido demostrada en el núcleo, kinetoplasto, membrana, vesículas, y además es secretada (Ballesteros-Rodea *et al.*, 2012; Manning-Cela *et al.*, 2001). Asimismo, es una proteína que no presenta homología con ninguna otra conocida (Ballesteros-Rodea *et al.*, 2012). A pesar de estas características, la respuesta inmunológica contra esta proteína no ha sido tan estudiada si se compara con las primeras cinco proteínas descritas. No obstante, se ha resaltado su potencialidad para desencadenar, en modelo murino, respuestas citotóxicas mediadas por linfocitos CD8⁺. Lo que condujo a que LYT-1 fuera propuesta como posible candidata en el diseño de vacunas contra la enfermedad de Chagas (Ballesteros-Rodea *et al.*, 2012).

Como puede verse, todas las proteínas mencionadas se escogieron en base a la importancia del desarrollo de una respuesta humoral y celular para la protección inmunológica contra *T. cruzi* dada su condición de parásito intra y extra celular. Por lo que se espera que producto de la integración de los epítipes predichos de estas proteínas en el péptido quimérico, se consiga el diseño de una molécula que induzca el reconocimiento inmune y la estimulación de poblaciones celulares que se diferencien en células efectoras y de memoria que sean efectivas en eventos de reinfección para la eliminación del parásito. Asimismo, se espera que al incluir epítipes conservados, la cepa del organismo no intervenga en la efectividad del prototipo. Finalmente, el empleo simultáneo de programas con algoritmos independientes para la predicción de péptidos de relevancia inmunológica, se realizó con la finalidad de contar con varias predicciones positivas que aumenten la probabilidad de escoger un péptido adecuado.

Modelado de proteínas de *T. cruzi*

En ausencia de una estructura experimentalmente resuelta, el modelado por homología representa una herramienta útil para predecir la conformación tridimensional de proteínas, que ha sido aplicada en una amplia gama de investigaciones enfocadas en caracterización enzimática, diseño de fármacos, eventos de reconocimiento molecular ligando-receptor, entre muchos otros (Harish Vagadia *et al.*, 2016; Vinothkanna *et al.*, 2018). En este estudio, esta técnica permitió visualizar la estructura de ciertas proteínas de interés de *T. cruzi*.

En relación, con el modelado de HSP70 se logró la localización de los epítopes predichos y con confirmación experimental. Como se observa en la Figura 5, la mayoría de epítopes experimentales fueron predichos tanto por ABCPred como por BepiPred 2.0. A pesar que del total de predicciones un 24,84% correspondió a epítopes sin confirmar, se evidenció que la mayoría están dispuestos en la superficie de la proteína, lo que indica que deben contar con ciertas características propias de un epítope de células B, como el de una composición de aminoácidos con alto nivel de hidrofiliidad, lo que pudo ser la causa de predicciones erróneas.

Por otro lado, considerando la importancia de ubicar los epítopes predichos en las proteínas problema, en este estudio se procuró la obtención del mejor modelo posible mediante pasos sucesivos de refinamiento, a fin de localizar estas regiones con mayor precisión y contar con una mejor aproximación de la estructura real de las proteínas. En cuatro de las siete proteínas se encontró que los modelos hechos por Swiss-Model tenían mejor calidad, por lo que la escogencia del resto de programas para modelar las tres proteínas restantes se debió a la ausencia de plantados (Tabla 6). No obstante, tanto Swiss-Model, I-TASSER y Phyre2, se encuentran catalogados entre los mejores algoritmos disponibles para la fecha según el comité de evaluación crítica de métodos de modelado de proteínas, o CASP por sus siglas en inglés (Moult *et al.*, 2018). Es relevante destacar que de las proteínas problema, únicamente ha sido resuelta por cristalografía de rayos X la estructura de cruzipaina y Tc24, plantados utilizados para la realización de sus modelos. Para el resto se emplearon plantados de proteínas de otros organismos, los que si bien variaban en su porcentaje de identidad, permitieron la construcción de modelos estructuralmente estables (Tabla 7). Siendo el modelo de LYT-1 el que presenta un menor porcentaje de aminoácidos en las regiones favorables y una mayor proporción en regiones no permitidas del gráfico de Ramachandran. A pesar de esto, el modelo obtenido en este estudio puede considerarse como la primera aproximación de la estructura tridimensional de esta proteína única de *T. cruzi*.

Por último, debe resaltarse que si bien la selección de epítopes a incluir en el péptido quimérico se basó en su disposición en la estructura de las proteínas, esto no descarta que en el parásito, el reconocimiento de anticuerpos de estas regiones se vea impedido por: efectos estéricos impuestos por proteínas adyacentes, modificaciones postraduccionales de las proteínas problema, o impedimentos estéricos provocados por las mismas proteínas al considerarlas como estructuras dinámicas. Lo que llevó a la escogencia de varios epítopes, a fin de aumentar las probabilidades de reconocimiento.

Diseño del prototipo inmunoestimulante

El objetivo de esta investigación fue el diseño de un prototipo formado por elementos adecuados que en conjunto tuvieran el potencial de generar protección contra la enfermedad de Chagas. Por esta razón, la elección de sus componentes siempre estuvo en función de la respuesta inmunológica que se deseaba promover.

En este sentido, el adyuvante c-di-AMP se seleccionó por su capacidad para actuar como ligando de la proteína STING, ubicada en el retículo endoplasmático y responsable de la activación de la cascada de señalización TBK3-IRF3, que genera la transcripción de genes que codifican citoquinas pro inflamatorias como IFN tipo I y TNF (Ebensen *et al.*, 2019). Eventos que llevan a que se desencadene una respuesta balanceada tipo Th1/Th2/Th17, que convierte al c-di-AMP en una molécula versátil por su capacidad para promover la activación de distintas poblaciones celulares que de manera sinérgica estimulan la producción de anticuerpos y la activación de células con capacidad citotóxica (Libanova *et al.*, 2012).

Debido a esto, el adyuvante ha sido empleado con éxito en varios protocolos de vacunación contra organismos intra y extra celulares (Matos *et al.*, 2017; Schulze *et al.*, 2017a; Temizoz *et al.*, 2018). De

hecho, éste ha sido probado en inmunizaciones contra *T. cruzi* en modelos murinos, donde se ha resaltado su efectividad para conseguir respuestas protectoras debido a las propiedades antes descritas de la molécula, y especialmente, por su participación en una respuesta inmune que incluye la activación de células Th17 (Matos *et al.*, 2017; Sanchez Alberti *et al.*, 2017). Linfocitos con una capacidad de modular positivamente células CD8⁺, que supera a la de las células Th1 (Cai *et al.*, 2016). Atributos que en conjunto incentivaron la elección del c-di-AMP como adyuvante del prototipo inmunoestimulante.

Por otra parte, la secuencia PADRE incluida en el péptido quimérico, corresponde a un epítipo sintético con alta afinidad por 15 de las 16 variantes alélicas de la molécula HLA-DR, que fue específicamente diseñado para ser inmunogénico en humanos (Alexander *et al.*, 1998). El mismo ha sido parte del diseño de vacunas contra organismos como *Leishmania infantum* (Athanasiou *et al.*, 2017), *Toxoplasma gondii* (El Bissati *et al.*, 2016) en donde se ha comprobado que es capaz de promover la activación de la respuesta celular. De igual manera, su presencia mejoró la eficacia de una vacuna compuesta por péptidos del antígeno Her2/neu asociado a tumores, mediante la expansión de linfocitos CD4⁺ y CD8⁺, cuya contribución de IFN- γ estimulaba poblaciones celulares con efectos antígeno-específicos (Ghaffari-Nazari *et al.*, 2015). Eventos que se originan gracias al procesamiento y presentación de PADRE, lo que se consigue por el reconocimiento de la región flanqueante que le precede. *Linker* con el que es habitualmente asociado y que contiene aminoácidos de clivaje tanto para la degradación proteosomal como lisosomal (Alexander *et al.*, 1998). Sitios que también están presentes en las secuencias de conexión de epítopos AAY y GPGPG que fueron empleadas en la quimera (Livingston *et al.*, 2002; Velders *et al.*, 2001).

En cuanto al modelado del péptido multiepítopos, puede verse que las etapas de refinamiento permitieron la obtención de un modelo estructuralmente más estable (Figura 13 y Tabla 9). En el que fue posible la localización de sus componentes (Figura 14), donde destaca la posición de los epítopos lineales de células B, los que se evidenciaron expuestos en la superficie (Figura 15). Representando su potencialidad para ser reconocidos por anticuerpos y receptores de células B.

La implementación de péptidos con sitios conservados y completamente ausentes de sitios polimórficos se aplicó como una estrategia para prevenir eventos de evasión. En relación a esto, se ha visto que el potencial de ciertas proteínas para ser usadas como blancos de vacunación, se ve comprometido por la existencia de variabilidad entre cepas que cuentan con determinantes antigénicos no solapados, imposibilitando el surgimiento de reconocimiento cruzado (Early *et al.*, 2018; Thompson *et al.*, 2018). Esta variabilidad es, sin embargo, una fuerte señal de detección inmunológica, lo que indica que naturalmente la respuesta inmune está enfocada hacia estas zonas, lo que provoca un efecto de apantallamiento de las regiones conservadas, haciéndolas, por lo tanto, menos inmunogénicas (Early *et al.*, 2018; Thompson *et al.*, 2018).

En concordancia, se ha evidenciado que en proteínas relevantes para el género *Plasmodium* como AMA1 expuestas a la detección humoral, existe una organización particular de regiones conservadas que se encuentran rodeadas por sitios muy polimórficos (Guy *et al.*, 2018). Las primeras representan zonas importantes para el funcionamiento de la proteína por contener sitios de unión para invasión, generando restricciones para la fijación de polimorfismos. Lo que si bien las convierte en blancos de vacunación atractivos, no descarta que presenten inmunogenicidad deficiente (Guy *et al.*, 2018). Hecho que ha llevado a la búsqueda y localización de epítopos capaces de inducir respuestas de protección dentro de regiones conservadas en una amplio rango de investigaciones (Cazorla *et al.*, 2009; Dutta *et al.*, 2013; Guy *et al.*, 2018; Ntumngia *et al.*, 2016). Enfoque que fue también aplicado en este trabajo por medio de la asociación de análisis inmunoinformáticos y pruebas evolutivas. Así, fue posible

determinar del conjunto de posibles epítopes, péptidos con sitios conservados, teóricamente capaces de generar protección ante las diferentes cepas de *T. cruzi*.

Es importante mencionar, que además de esta estrategia, en otros estudios se han desarrollado protocolos de inmunización que comprenden el empleo de proteínas de todas las variantes alélicas que pueda presentar un antígeno dado, a fin de conseguir una respuesta inmunológica que abarque la variabilidad de las cepas del organismo patógeno (Lourenço *et al.*, 2017). Lo que si bien ha conseguido cierta efectividad, presenta inconvenientes dada la diversidad intrínseca de muchos agentes infecciosos.

Por otra parte, dada la inmunodominancia de algunos sitios polimórficos sobre sitios conservados, en otras investigaciones se han aplicado métodos relacionados con “técnicas de enfoque de epítopes”, que tienen por objetivo dirigir el reconocimiento inmune hacia sitios conservados en los que la eficacia para generar protección haya sido comprobada (Guy *et al.*, 2018). En uno de estos estudios, por medio de inmunizaciones con cuatro variantes de la proteína AMA1 de *plasmodium falciparum* consiguieron amplificar la respuesta a epítopes conservados mediante la expansión desigual de células B de memoria (Dutta *et al.*, 2013). Los autores explican que la inmunodominancia de los sitios polimórficos de cada variante puede verse diluida por la presencia combinada de estas regiones y por la existencia de regiones conservadas entre variantes. De esta manera, la exposición a las proteínas puede aumentar la concentración de epítopes conservados que son ahora capaces de competir por el reconocimiento (Dutta *et al.*, 2013). Resultado que puede también conseguirse bajo la exposición repetida a sitios conservados (Gupta and Day, 1994). Lo que podría aplicarse como método de inmunización en caso de que se compruebe la eficacia de los péptidos incluidos en la quimera. Y de esta manera, sobreponerse a posibles eventos de inmunodominancia provenientes de epítopes polimórficos, o incluso conservados no efectivos, dentro de las proteínas problema.

Ahora bien, la exclusión de este tipo de epítopes dentro de la misma quimera es otra de las ventajas que trae el diseño de este prototipo. Debido a que de esta forma fue posible evitar el uso de epítopes inmunodominantes presente en dos de las proteínas problema, Cz y TSA-1. En una investigación dirigida por (Cazorla *et al.*, 2010), donde se analizó la respuesta humoral y celular inducida por Cruzipaina, se demostró que existe una inmunodominancia por el dominio C-terminal no catalítico de la enzima. Con respecto a esto, los autores encontraron que al inmunizar ratones con los dominios N- y C- terminal se presentaba disparidad en la respuesta de anticuerpos, obteniendo un título 15 veces menor al emplear el dominio N-terminal. Resultado que éstos comprueban con los reportes que indican que el suero de la mayoría de personas chagásicas reacciona contra el dominio C-terminal (Cazorla *et al.*, 2010). No obstante, a pesar del bajo título, se encontró que por medio de las inmunizaciones los anticuerpos dirigidos al dominio N-terminal, eran capaces de producir una reducción significativa de la parasitemia y de la tasa de infectividad de tripomastigotes. Contrariamente, los anticuerpos contra el dominio C-terminal fueron incapaces de generar una reducción significativa de la parasitemia y de la invasión celular. Resultados que fueron similares a los obtenidos al inmunizar con la proteína recombinante (Cazorla *et al.*, 2010). Lo que indica la existencia de una inmunodominancia hacia epítopes no efectivos para el control de la infección, hecho que los investigadores interpretan como un mecanismo de escape inmunológico, donde el reconocimiento del dominio C-terminal enmascara y protege al dominio catalítico N-terminal. Hipótesis que fue apoyada por los resultados obtenidos al evaluar la respuesta celular, donde evidenciaron el mismo patrón de protección (Cazorla *et al.*, 2010).

Este hallazgo permitió evitar la inclusión de péptidos predichos y conservados que comprendieran el dominio C-terminal: 306 WTTQWGEEGY 315, 395 IVTCGAETLTEEVFL 409 y 453 RQRRHQPYHSRHRRL 467 (Tabla 4). Con la finalidad de promover la aparición de respuestas efectivas para la eliminación del parásito. Coincidentemente, se encontró que algunos de los péptidos seleccionados de Cz han sido exitosamente evaluados en otras investigaciones, como el péptido 224

TTSGHTVVGATITGHV 238, que contiene el epítipo TSGHTVVGATI incorporado en el péptido quimérico Traspaina desarrollado por Sanchez Alberti *et al.* (2017), y el péptido 188 GLMNNAFEWI 197, presentado por moléculas MHC-I del alelo HLA-A*02:01, capaz de activar células CD8⁺ en personas chagásicas (Fonseca *et al.*, 2005). Por otro lado, mediante los métodos inmunoinformáticos aplicados se predijo que el mismo péptido puede ser reconocido por moléculas MHC-I HLA-B*58:01, lo que indica que pudiese existir promiscuidad entre al menos dos alelos del tipo HLA-I.

Adicionalmente, en otra investigación se comprobó que la región próxima al C-terminal (aminoácidos desde la posición 618 a la 835) de TSA-1 es capaz de inducir una fuerte respuesta celular y humoral que es ineficiente para la eliminación del parásito y para la generación de protección por inmunización de un modelo murino (Wrightsmann *et al.*, 1994). Fenómeno que encubre el rol de epítopes sub-dominantes más efectivos ubicados en regiones diferentes de la proteína. Hecho que llevó a que en este trabajo se prescindiera de péptidos localizados en la región dominante. La determinación de proteínas con sub-dominancia inmunogénica ha sido el objetivo de varias investigaciones, donde se ha intentado dilucidar el rol de epítopes ausentes en proteínas miembro de las grandes familias de proteínas de superficie de *T. cruzi*, como transialidasas y mucinas, responsables de permitir la persistencia del parásito al desviar la respuesta inmunológica del huésped (Rosenberg *et al.*, 2016). Por lo tanto, en este trabajo se espera que la creación de una quimera formada por posibles epítopes sub-dominantes represente un prototipo inmunoestimulante que promueva la expansión de células que generen respuestas de protección.

El uso de péptidos multiepítopes ha sido una metodología propuesta por varios investigadores que señalan las ventajas de potenciar respuestas dirigidas a epítopes que promueven respuestas inmunológicas favorables. En este sentido, destaca el estudio dirigido por (Athanasίου *et al.*, 2017), quienes integrando epítopes de tres proteínas inmunogénicas de *Leishmania infantum* diseñaron una quimera que en asociación con el adyuvante MPLA, fue capaz de inducir la activación de células dendríticas que generaron la proliferación de linfocitos T, CD4⁺ y CD8⁺. Asimismo, protocolos de inmunización con ese péptido consiguieron niveles significativos de protección en el modelo utilizado. Por otra parte, además de la prevención de enfermedades infecciosas, el uso de péptidos multiepítopes ha sido una estrategia implementada en el estudio del tratamiento y prevención de ciertos tipos de cáncer, como el cáncer de cuello uterino generado por el virus del papiloma humano (Fernandes de Oliveira *et al.*, 2015). En una investigación fue específicamente comprado que además de un efecto preventivo, un péptido compuesto de epítopes de las proteínas E6 y E7 del virus, capaz de fomentar respuestas humorales y celulares, fue apto para producir la regresión tumores malignas (Fernandes de Oliveira *et al.*, 2015).

Desde esta perspectiva y tomando en cuenta que algunas de las proteínas problemas, como Cz, Tc24 y TSA-1, han sido eficaces en el tratamiento y control experimental de las lesiones musculares características de la enfermedad de Chagas (Barry *et al.*, 2019; Biter *et al.*, 2018; Cerny *et al.*, 2016), no se descarta que el prototipo diseñado en este estudio presente un efecto terapéutico además de preventivo. No obstante, es importante resaltar que la efectividad de este diseño depende de la cobertura poblacional que pueda conseguir, así como también, del grado real de conservación que presenten los epítopes si se considera la extensa variabilidad encontrada entre y dentro de los linajes del parásito. Por lo tanto, desde un enfoque experimental, además de necesitar la verificación del efecto inmunológico de los péptidos seleccionados, sería ideal para el diseño de una vacuna con cobertura nacional, poder contar con una muestra significativa de las cepas circulantes de *T. cruzi* en Venezuela, así como también del perfil genético HLA de sus habitantes. De esta forma, mediante la secuenciación de estos organismos, podría verificarse la ausencia de sitios bajo selección positiva en los péptidos, e incluso, podrían identificarse nuevas regiones de interés dentro de las proteínas. Del mismo modo, el

tener una idea más precisa de la diversidad y frecuencia de los alelos de las moléculas HLA, llevaría a la integración de epítopes que sean reconocidos por los alelos más representativos de la población. Y aún más importante, sería la identificación de péptidos promiscuos reconocidos por distintas variantes, que sean capaces de la activación de poblaciones celulares que generen protección.

Para un trabajo de esta magnitud, los análisis inmunoinformáticos y las pruebas evolutivas resultan imprescindibles. Por lo tanto, se considera que una estrategia que contemple integrar el uso de estas herramientas junto con metodologías experimentales pertinentes, que lleve a la creación de un prototipo compuesto por un péptido multiepítopes y un adyuvante adecuado, como el seleccionado en este estudio, representa una alternativa prometedora para la prevención y el tratamiento de la tripanosomiasis americana.

VII. BIBLIOGRAFÍA

- Acosta, H., Dubourdieu, M., Quiñones, W., Cáceres, A., Bringaud, F., and Concepción, J.L. (2004). Pyruvate phosphate dikinase and pyrophosphate metabolism in the glycosome of *Trypanosoma cruzi* epimastigotes. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 138, 347-356.
- Alexander, J., Fikes, J., Hoffman, S., Franke, E., Sacci, J., and Appella, E. (1998). The optimization of helper T lymphocyte (HTL) function in vaccine development. *Immunologic research* 18, 79-92.
- Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J mol Biol* 215, 403-410.
- Altuvia, Y., Schueler, O., and Margalit, H. (1995). Ranking potential binding peptides to MHC molecules by a computational threading approach. *Journal of molecular biology* 249.
- Andreatta, M., and Nielsen, M. (2015). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32, 511-517.
- Arce-Fonseca, M., Gonzalez-Vazquez, M.C., Rodriguez-Morales, O., Graullera-Rivera, V., Aranda-Fraustro, A., Reyes, P.A., Carabarin-Lima, A., and Rosales-Encina, J.L. (2018). Recombinant Enolase of *Trypanosoma cruzi* as a Novel Vaccine Candidate against Chagas Disease in a Mouse Model of Acute Infection. *Journal of immunology research* 2018, 8964085.
- Arriagada, K. (2014). Tipificación de linajes de *Trypanosoma cruzi* en individuos con enfermedad de Chagas cardíopatas y no cardíopatas. In Facultad de Ciencias Veterinarias y Pecuarias (Santiago, Chile: Universidad de Chile).
- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B., Carrington, M., and Gardner, M. (2009). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic acids research* 38.
- Athanasίου, E., Agallou, M., Tastsoglou, S., Kammona, O., and Kiparissides, C. (2017). A Poly(Lactic-co-Glycolic) Acid Nanovaccine Based on Chimeric Peptides from Different *Leishmania infantum* Proteins Induces Dendritic Cells Maturation and Promotes peptide-specific IFN-Producing CD8⁺ T cells essential for the protection against experimental visceral Leishmaniasis. *Frontiers in immunology* 8.
- Aujla, S., Chan, Y., Zheng, M., Fei, M., Askew, D., Pociask, D., and Husain, S. (2008). IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nature medicine* 14.
- Avendaño-Rangel, F., and Rey, K. (2016). Avances y desafíos en el control de la enfermedad de Chagas en Venezuela y un estudio de caso en el estado Mérida. *Ciencia y Salud* 6, 155-164.
- Backert, L., and Kohlbacher, O. (2015). Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome medicine* 7, 119.

- Bahia, M., Diniz, L., and Mosqueira, V. (2014). Therapeutical approaches under investigation for treatment of Chagas disease. *Expert opinion on investigational drugs* 23, 1225-1237.
- Ballesteros-Rodea, G., Santillán, M., Cruz-Aguilar, M., Márquez-Dueñas, C., Lugo-Caballero, C., Martínez-Calvillo, S., and Manning-Cela, R. (2012). The alternative products of *Trypanosoma cruzi* LYTI have different localization patterns. *Veterinaria Mexico* 43, 29-43.
- Barry, M., Versteeg, L., Wang, Q., Pollet, J., Zhan, B., Gusovsky, F., Bottazzi, M., Hotez, P., and Jones, K. (2019). A therapeutic vaccine prototype induces protective immunity and reduces cardiac fibrosis in a mouse model of chronic *Trypanosoma cruzi* infection. *PLoS neglected tropical diseases* 13.
- Bastos, L., Macedo Jr., A., Silva, M., Santiago, F., Ramos, E., Santos, F., Pirovani, C., Goulart, L., Mineo, T., and Mineo, J. (2016). Toxoplasma gondii-Derived Synthetic Peptides Containing B- and T-Cell Epitopes from GRA2 Protein Are Able to Enhance Mice Survival in a Model of Experimental Toxoplasmosis. *Frontiers in Cellular and Infection Microbiology* 6.
- Beaumier, C.M., Gillespie, P.M., Strych, U., Hayward, T., Hotez, P.J., and Bottazzi, M.E. (2016). Status of vaccine research and development of vaccines for Chagas disease. *Vaccine* 34, 2996-3000.
- Berger, C.T., Carlson, J.M., Brumme, C.J., Hartman, K.L., Brumme, Z.L., Henry, L.M., and Heckerman, D. (2010). Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *Journal of experimental medicine* 207, 61-75.
- Berkhoff, E., De Wit, E., Geelhoed-Mieras, M., Boon, A., Symons, J., Fouchier, R., and Rimmelzwaan, G.F. (2005). Functional constraints of influenza A virus epitopes limit escape from cytotoxic T lymphocytes. *Journal of virology* 79, 11239-11246.
- Berkhoff, E., De Wit, E., Geelhoed-Mieras, M., Boon, A., Symons, J., Fouchier, R., and Rimmelzwaan, G.F. (2006). Fitness costs limit escape from cytotoxic T lymphocytes by influenza A viruses. *Vaccine* 24, 6594-6596.
- Bhattacharya, D., and Cheng, J. (2013). 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. *Proteins: Structure, Function, and Bioinformatics* 81, 119-131.
- Biter, A.B., Weltje, S., Hudspeth, E.M., Seid, C.A., McAtee, C.P., Chen, W.H., Pollet, J.B., Strych, U., Hotez, P.J., and Bottazzi, M.E. (2018). Characterization and Stability of *Trypanosoma cruzi* 24-C4 (Tc24-C4), a Candidate Antigen for a Therapeutic Vaccine Against Chagas Disease. *Journal of pharmaceutical sciences* 107, 1468-1473.
- Bivona, A.E., Sanchez Alberti, A., Matos, M.N., Cerny, N., Cardoso, A.C., Morales, C., Gonzalez, G., Cazorla, S.I., and Malchiodi, E.L. (2018). *Trypanosoma cruzi* 80 kDa prolyl oligopeptidase (Tc80) as a novel immunogen for Chagas disease vaccine. *PLoS neglected tropical diseases* 12, e0006384.
- Blythe, M.J., and Flower, D.R. (2005). Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein science : a publication of the Protein Society* 14, 246-248.
- Bontempi, I. (2015). Desarrollo de una estrategia inmunomoduladora para el control de la infección generada por *Trypanosoma cruzi*. In Facultad de Bioquímica y Ciencias Biológicas (Santa Fe, Argentina: Universidad Nacional del Litoral).
- Breniere, S.F., Waleckx, E., and Barnabe, C. (2016). Over Six Thousand *Trypanosoma cruzi* Strains Classified into Discrete Typing Units (DTUs): Attempt at an Inventory. *PLoS neglected tropical diseases* 10, e0004792.
- Cai, C.W., Blase, J.R., Zhang, X., Eickhoff, C.S., and Hoft, D.F. (2016). Th17 Cells Are More Protective Than Th1 Cells Against the Intracellular Parasite *Trypanosoma cruzi*. *PLoS pathogens* 12, e1005902.
- Carabarin-Lima, A., Rodriguez-Morales, O., Gonzalez-Vazquez, M.C., Baylon-Pacheco, L., Reyes, P.A., Arce-Fonseca, M., and Rosales-Encina, J.L. (2014). In silico approach for the identification of

immunological properties of enolase from *Trypanosoma cruzi* and its possible usefulness as vaccine in Chagas disease. *Parasitology research* 113, 1029-1039.

Carrasco, H.J., Segovia, M., Llewellyn, M.S., Morocoima, A., Urdaneta-Morales, S., Martinez, C., Martinez, C.E., Garcia, C., Rodriguez, M., Espinosa, R., *et al.* (2012). Geographical distribution of *Trypanosoma cruzi* genotypes in Venezuela. *PLoS neglected tropical diseases* 6, e1707.

Cazorla, S., Frank, F., Becker, P., Mirkin, G., Corral, R., Guzman, C., and Malchiodi, E. (2010). Redirection of the Immune Response to the Functional Catalytic Domain of the Cystein Proteinase Cruzipain Improves Protective Immunity against *Trypanosoma cruzi* Infection. *The Journal of infection diseases* 202, 136–144.

Cazorla, S.I., Frank, F.M., and Malchiodi, E.L. (2009). Vaccination approaches against *Trypanosoma cruzi* infection. *Expert review of vaccines* 8, 921-935.

Cerny, N., Sánchez Alberti, A., Bivona, A.E., De Marzi, M., Frank, F.M., Cazorla, S., and Malchiodi, E. (2016). Coadministration of cruzipain and GM-CSF DNAs, a new immunotherapeutic vaccine against *Trypanosoma cruzi* infection. *Human vaccines & immunotherapeutics* 12, 438-450.

Comas, I., Chakravarti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., and Gagneux, S. (2010). Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nature genetics* 42, 498-503.

Consortium, U. (2018). UniProt: the universal protein knowledgebase. *Nucleic acids research* 46.

Costa, J.G., Faccendini, P.L., Sferco, S., Lagier, C., and Marcipar, I. (2013). Evaluation and comparison of the ability of online available prediction programs to predict true linear B-cell epitopes. *Protein and peptide Letters* 20, 724-730.

De Pablos, L.M., Diaz Lozano, I.M., Jercic, M.I., Quinzada, M., Gimenez, M.J., Calabuig, E., Espino, A.M., Schijman, A.G., Zulantay, I., Apt, W., *et al.* (2016). The C-terminal region of *Trypanosoma cruzi* MASPs is antigenic and secreted via exovesicles. *Scientific reports* 6, 27293.

Del Giudice, G., Rappuoli, R., and Didierlaurent, A.M. (2018). Correlates of adjuvanticity: A review on adjuvants in licensed vaccines. *Seminars in immunology* 39, 14-21.

Desai, D.V., and Kulkarni-Kale, U. (2014). T-cell epitope prediction methods: an overview. *Methods in molecular biology* 1184, 333-364.

Dimitrov, I., Flower, D., and Doytchinova, I. (2013a). AllerTOP-a server for in silico prediction of allergens. *BMC bioinformatics* 14.

Dimitrov, I., Naneva, L., Doytchinova, I., and Bangov, I. (2013b). AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30, 846-851.

Dinter, J., Duong, E., Lai, N.Y., Berberich, M., Kourjian, G., Bracho-Sanchez, E., and Le Gall, S. (2015). Variable processing and cross-presentation of HIV by dendritic cells and macrophages shapes CTL immunodominance and immune escape. *PLoS pathogens* 11.

Domingo, M. (2004). Ubicación subcelular, purificación, caracterización y estudios de interacción proteína-proteína de la enolasa de epimastigotes de *Trypanosoma cruzi* In Departamento de Biología, Facultad de Ciencias (Mérida, Venezuela: Universidad de Los Andes).

Dorn, P.L., McClure, A.G., Gallaspy, M.D., Waleckx, E., Woods, A.S., Monroy, M.C., and Stevens, L. (2017). The diversity of the Chagas parasite, *Trypanosoma cruzi*, infecting the main Central American vector, *Triatoma dimidiata*, from Mexico to Colombia. *PLoS neglected tropical diseases* 11, e0005878.

Duarte, A., Queiroz, A.T., Tosta, R., Carvalho, A.M., Barbosa, C.H., Bellio, M., de Oliveira, C.I., and Barral-Netto, M. (2015). Prediction of CD8+ Epitopes in *Leishmania braziliensis* Proteins Using EPIBOT: In Silico Search and In Vivo Validation. *PloS one* 10, e0124786.

Duschak, V., and Couto, A. (2009). Cruzipain, the major cysteine protease of *Trypanosoma cruzi*: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. A review. *Current medicinal chemistry* 16, 3174-3202.

- Dutta, S., Dlugosz, L., Drew, D., Ge, X., and Ababacar, D. (2013). Overcoming antigenic diversity by enhancing the immunogenicity of conserved epitopes on the malaria vaccine candidate apical membrane antigen-1. *PLoS pathogens* 9.
- Early, A.M., Lievens, M., MacInnis, B.L., Ockenhouse, C.F., Volkman, S.K., Adjei, S., Agbenyega, T., Ansong, D., Gondi, S., Greenwood, B., *et al.* (2018). Host-mediated selection impacts the diversity of *Plasmodium falciparum* antigens within infections. *Nature communications* 9, 1381.
- Ebensen, T., Delandre, S., Prochnow, B., Guzmán, C., and Schulze, K. (2019). The Combination Vaccine Adjuvant System Alum/c-di-AMP Results in Quantitative and Qualitative Enhanced Immune Responses Post Immunization. *Frontiers in Cellular and Infection Microbiology* 9.
- Echave, J., Spielman, S., and Wilke, C. (2016). Causes of evolutionary rate variation among protein sites. *Nature* 17.
- El-Manzalawy, Y., and Honavar, V. (2010). Recent advances in B-cell epitope prediction methods. *Immunome research* 6 Suppl 2, S2.
- El Bissati, K., Chentoufi, A., Krishack, P., Zhou, Y., Woods, S., and Dubey, J. (2016). Adjuvanted multi-epitope vaccines protect HLA-A*11:01 transgenic mice against *Toxoplasma gondii*. *JCI Insight* 1.
- Elisei, R.M.T., Matos, C.S., Carvalho, A., Chaves, A.T., Medeiros, F.A.C., Barbosa, R., Marcelino, A.P., Dos Santos Emidio, K., Coelho, E.A.F., Duarte, M.C., *et al.* (2018). Immunogenomic screening approach to identify new antigens for the serological diagnosis of chronic Chagas' disease. *Applied microbiology and biotechnology* 102, 6069-6080.
- Emini, E., Hughes, J., Perlow, D., and Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide 55, 836-839.
- Fernandes de Oliveira, M., Gallote, M., Muniz, A., Marques, A., Soriano, A., de Oliveira, M., Soares Schanoski, A., Lopes, R., de Souza, L., de Oliveira, M., *et al.* (2015). Design, Immune Responses and Anti-Tumor Potential of an HPV16 E6E7 Multi-Epitope Vaccine. *PloS one* 10.
- Florkowski, C.M. (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical Biochemist* 29.
- Fonseca, S.G., Moins-Teisserenc, H., Clave, E., Ianni, B., Nunes, V.L., Mady, C., Iwai, L.K., Sette, A., Sidney, J., Marin, M.L., *et al.* (2005). Identification of multiple HLA-A*0201-restricted cruzipain and FL-160 CD8+ epitopes recognized by T cells from chronically *Trypanosoma cruzi*-infected patients. *Microbes and infection* 7, 688-697.
- Fortes Mdel, P., Machado, I.V., Gil, G., Fernandez-Mestre, M., Dagher, L., Leon, R.V., Bianco, N.E., and Tassinari, P. (2007). Genetic contribution of major histocompatibility complex class II region to type 1 autoimmune hepatitis susceptibility in Venezuela. *Liver international : official journal of the International Association for the Study of the Liver* 27, 1409-1416.
- Frank, S.A., and Bush, R.M. (2007). Barriers to antigenic escape by pathogens: trade-off between reproductive rate and antigenic mutability. *BMC evolutionary biology* 7.
- Geer, L., Marchier-Bauer, A., Geer, R., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S. (2010). The NCBI BioSystems database. *Nucleic acids research* 34.
- Ghaffari-Nazari, H., Tavakkol-Afshari, J., Jaafari, M., Tahaghoghi-Hajghorbani, S., Masoumi, E., and Jalali, S. (2015). Improving multi-epitope long peptide vaccine potency by using a strategy that enhances CD4+ T help in BALB/c mice. *PloS one* 10.
- Gonzalez, N., and Buscaglia, C. (2015). Mapping antigenic motifs in the trypomastigote small surface antigen from *Trypanosoma cruzi*. *Clinical and vaccine immunology : CVI* 22, 304-312.
- Gowthaman, U., Chodiseti, S.B., Parihar, P., and Agrewala, J.N. (2010). Evaluation of different generic in silico methods for predicting HLA class I binding peptide vaccine candidates using a reverse approach. *Amino acids* 39, 1333-1342.

- Gupta, S., and Day, C. (1994). A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasit Immunology* 16, 361-370.
- Gupta, S., and Day, P. (1994). A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasit Immunology* 16, 361-370.
- Gupta, S., Maiden, M., Feavers, I.M., Nee, S., May, R., and Anderson, R. (1996). The maintenance of strain structure in populations of recombining infectious agents. *Nature medicine* 2.
- Guy, A.J., Irani, V., Richards, J.S., and Ramsland, P.A. (2018). Structural patterns of selection and diversity for *Plasmodium vivax* antigens DBP and AMA1. *Malaria journal* 17, 183.
- Harish Vagadia, B., Vanga, S.K., Singh, A., and Raghavan, V. (2016). Effects of thermal and electric fields on soybean trypsin inhibitor protein: A molecular modelling study. *Innovative Food Science & Emerging Technologies* 35, 9-20.
- Heo, L., Park, H., and Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic acids research* 41.
- Holt, K.E., Parkhill, J., Mazzoni, C.J., Roumagnac, P., Weill, F.X., Goodhead, I., and Dolecek, C. (2008). High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nature genetics* 40.
- Inc, S. (2008). SPSS Statistics for Windows, Version 17.0 (Chicago).
- Janin, J., Wodak, S., Levitt, M., and Maigret, B. (1978). Conformation of amino acid side-chains in proteins. *Journal of molecular biology* 125, 357-388.
- Jespersen, M.C., Peters, B., Nielsen, M., and Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic acids research* 45, W24-W29.
- Jimenez, P., Jaimes, J., Poveda, C., and Ramirez, J.D. (2019). A systematic review of the *Trypanosoma cruzi* genetic heterogeneity, host immune response and genetic factors as plausible drivers of chronic chagasic cardiomyopathy. *Parasitology* 146, 269-283.
- Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *Journal of immunology research* 199, 3360-3368.
- Karplus, P., and Schulz, G. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften*. 72.
- Kazi, A., Chuah, C., Majeed, A.B.A., Leow, C.H., Lim, B.H., and Leow, C.Y. (2018). Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathogens and global health* 112, 123-131.
- Kelley, L., Mezulis, S., Yates, C., Wass, M.N., and Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10.
- Khatoon, N., Ojha, R., Mishra, A., and Prajapati, V.K. (2018). Examination of antigenic proteins of *Trypanosoma cruzi* to fabricate an epitope-based subunit vaccine by exploiting epitope mapping mechanism. *Vaccine* 36, 6290-6300.
- Khatoon, N., Pandey, R.K., and Prajapati, V.K. (2017). Exploring *Leishmania* secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Scientific reports* 7, 8285.
- Kolaskar, A., and Tongaonkar, P. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters* 276, 172-174.
- Kosakovsky Pond, S., and Frost, S.D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 22, 1208-1222.
- Kumar, R., and Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics* 48, 277-287.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35, 1547-1549.

- Lasso, P., Beltran, L., Guzman, F., Rosas, F., Thomas, M.C., Lopez, M.C., Gonzalez, J.M., Cuellar, A., and Puerta, C.J. (2016). Promiscuous Recognition of a *Trypanosoma cruzi* CD8+ T Cell Epitope among HLA-A2, HLA-A24 and HLA-A1 Supertypes in Chagasic Patients. *PloS one* *11*, e0150996.
- Lee, G., Won, J., Heo, L., and Seok, C. (2019). GalaxyRefine2: simultaneous refinement of inaccurate local regions and overall protein structure. *Nucleic acids research*.
- Li, Z.T., Zarnitsyna, V.I., Lowen, A.C., Weissman, D., Koelle, K., Kohlmeier, J.E., and Antia, R. (2019). Why Are CD8 T Cell Epitopes of Human Influenza A Virus Conserved? *Journal of virology* *93*.
- Libanova, R., Becker, P., and Guzman, C. (2012). Cyclic di-nucleotides: new era for small molecules as adjuvants. *Microbial biotechnology* *5*, 168-176.
- Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, A., and Sette, A. (2002). A Rational Strategy to Design Multiepitope Immunogens Based on Multiple Th Lymphocyte Epitopes. *The Journal of Immunology* *12*.
- Lourenço, J., Spensley, K., Wikramaratna, K., Penman, B., Walker, A., Smith, A., Pybus, O., Jean, L., and Gupta, S. (2017). Reverse immunodynamics: a new 1 method for identifying targets of protective immunity. *bioRxiv*.
- Majnik, M., and Bosnić, Z. (2013). ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis* *17*, 531-558.
- Manning-Cela, R., Cortes, A., Gonzalez-Rey, E., Van Voorhis, W.C., Swindle, J., and Gonzalez, A. (2001). LYT1 protein is required for efficient in vitro infection by *Trypanosoma cruzi*. *Infection and immunity* *69*, 3916-3923.
- Martin, D.L., Weatherly, D.B., Laucella, S.A., Cabinian, M.A., Crim, M.T., Sullivan, S., and Sette, A. (2006). CD8+ T-Cell responses to *Trypanosoma cruzi* are highly focused on strain-variant trans-sialidase epitopes. *PLoS pathogens* *2*.
- Martinez-Campos, V., Martinez-Vega, P., Ramirez-Sierra, M.J., Rosado-Vallado, M., Seid, C.A., Hudspeth, E.M., Wei, J., Liu, Z., Kwityn, C., Hammond, M., *et al.* (2015). Expression, purification, immunogenicity, and protective efficacy of a recombinant Tc24 antigen as a vaccine against *Trypanosoma cruzi* infection in mice. *Vaccine* *33*, 4505-4512.
- Martinez, J., Campetella, O., Frasch, J., and Cazzulo, J. (1991). The major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* is antigenic in human infections. *Infection and immunity* *59*, 4275-4277.
- Matos, M.N., Cazorla, S., Schulze, K., Ebensen, T., Guzman, C., and Malchiodi, E. (2017). Immunization with Tc52 or its amino terminal domain adjuvanted with c-di-AMP induces Th17+Th1 specific immune responses and confers protection against *Trypanosoma cruzi*. *PLoS Negl Trop Dis*.
- Maura, R., Fernández, S., Reyes, G., Perez, J., Reyes, F., García, M., and Sierra, G. (2010). Evaluation of the potential of *Mycobacterium smegmatis* as vaccine Candidate against tuberculosis by in silico and in vivo studies. *Vaccine* *19*, 20-26.
- Menezes-Souza, D., de Oliveira Mendes, T.A., de Araujo Leao, A.C., de Souza Gomes, M., Fujiwara, R.T., and Bartholomeu, D.C. (2015). Linear B-cell epitope mapping of MAPK3 and MAPK4 from *Leishmania braziliensis*: implications for the serodiagnosis of human and canine leishmaniasis. *Applied microbiology and biotechnology* *99*, 1323-1336.
- Michels, P., Bringaud, F., Herman, M., and Hannaert, V. (2006). Metabolic functions of glycosomes in trypanosomatids. *Biochimica et Biophysica Acta* *1763*, 1463-1477.
- Minitab, I. (2014). MINITAB release 17: statistical software for windows, M. Inc, ed. (USA).
- Moutaftsi, M., Peters, B., Pasquetto, V., Tschärke, D., Sidney, J., and Bui, H.H. (2006). A consensus epitope prediction approach identifies the breadth of murine TCD8- cell responses to vaccinia virus. *Nature Biotechnology* *24*, 817-819.

- Murcia, L., Carrilero, B., Saura, D., Iborra, M., and Segovia, M. (2013). Diagnosis and treatment of Chagas disease. *Enfermedades infecciosas y microbiología clínica* *31*, 26-34.
- Murrell, B., Wertheim, J., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky, S. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PloS Genetics* *8*.
- Narula, A., Pandey, R.K., Khatoon, N., Mishra, A., and Prajapati, V.K. (2018). Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* *61*, 4-15.
- Nicholls, E.F., Madera, L., and Hancock, R.E. (2010). Immunomodulators as adjuvants for vaccines and antimicrobial therapy. *Annals of the New York Academy of Sciences* *1213*, 46-61.
- Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome medicine* *8*.
- Nielsen, M., and Lund, O. (2009). NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC bioinformatics* *10*.
- Ntumngia, F., Thomson-Luque, R., de Torres, L., Gunalan, K., Carvalho, L., and Adams, J. (2016). A novel erythrocyte binding protein of *Plasmodium vivax* suggests an alternate invasion pathway into Duffy-positive reticulocytes. *MBio* *7*.
- Pardo, T., Fernández, M., Fleitas, H., Borjas, L., Zabala, W., Reyes, F., Quintero, J., Delgado, W., and Aranguren, J. (2012). Caracterización molecular a través de marcadores de HLA clase I y II de la población San José de Heras, estado Zulia, Venezuela. *Ciencia* *20*, 98-108.
- Park, S., Won, G., Kim, J., Kim, H.B., and Lee, J.H. (2018). Potent O-antigen-deficient (rough) mutants of *Salmonella Typhimurium* secreting *Lawsonia intracellularis* antigens enhance immunogenicity and provide single-immunization protection against proliferative enteropathy and salmonellosis in a murine model. *Veterinary research* *49*, 57.
- Parker, J., Guo, D., and Hodges, R. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *The Biochemical journal* *25*, 5425-5432.
- Paul, S., Croft, N.P., Purcell, A.W., Tschärke, D.C., Sette, A., Nielsen, M., and Peters, B. (2019). Benchmarking predictions of MHC class I restricted T cell epitopes.
- Pellequer, J., Westhof, E., and Van Regenmortel, M. (1993). Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunology letters. Immunology* *36*.
- Perez-Aguilar, M., Rondon-Mercado, R., Goncalves, L., and Acosta, H. (2018). Neuroimmunoendocrine System During Infection by *Trypanosoma cruzi*: Mechanisms of Immunoregulation. *Microbiology* *14*, 581-596.
- Peters, B., and Sette, A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinformatics* *6*, 132.
- Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., and Ferrin, T. (2004). Chimera—A visualization system for exploratory research and analysis. *Journal of computational biology : a journal of computational molecular cell biology* *25*, 1605–1612.
- Pham, N., Badovinac, V., and Harty, J. (2009). A default pathway of memory CD8 T cell differentiation after dendritic cell immunization is deflected by encounter with inflammatory cytokines during antigen-driven proliferation. *The Journal of Immunology* *183*, 2337-2348.
- Ponnuswamy, P., Prabhakaran, M., and Manavalan, P. (1980). Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochimica et Biophysica Acta* *2*, 301-316.
- Potocnakova, L., Bhide, M., and Pulzova, L.B. (2016). An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *Journal of immunology research* *2016*, 6760830.

- Prisant, M., Richardson, J., and Richardson, D. (2003). Structure validation by C α geometry: phi, psi and C β deviation. *Proteins* 50.
- Quijano-Hernandez, I., and Dumonteil, E. (2011). Advances and challenges towards a vaccine against Chagas disease. *Human vaccines* 7, 1184-1191.
- Ramachandran, G.T., and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. In *Advances in protein chemistry*. Academic Press 23.
- Rammensee, H., Bachmann, J., Emmerich, N., Bachor, O., and Stevanović, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213-219.
- Rassi Jr, A., Rassi, A., and Marin-Nieto, J. (2010). Chagas disease. *The Lancet* 375, 1388-1402.
- Raymond, D., Bajica, G., Ferdmanb, J., Suphaphiphatb, P., Settembreg, E., Moodyc, M., Schmidta, A., and Harrison, S. (2018). Conserved epitope on influenza-virus hemagglutinin head defined by a vaccine-induced antibody. *PNAS* 115.
- Reed, S.G., Orr, M.T., and Fox, C.B. (2013). Key roles of adjuvants in modern vaccines. *Nature medicine* 19, 1597-1608.
- Robert, C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32, 1792-1795.
- Rodriguez-Morales, O., Monteon-Padilla, V., Carrillo-Sanchez, S.C., Rios-Castro, M., Martinez-Cruz, M., Carabarin-Lima, A., and Arce-Fonseca, M. (2015). Experimental Vaccines against Chagas Disease: A Journey through History. *Journal of immunology research* 2015, 489758.
- Rosenberg, C., Zhang, W., Bustamante, J., and Tarleton, R.L. (2016). Long-Term Immunity to *Trypanosoma cruzi* in the Absence of Immunodominant trans-Sialidase-Specific CD8⁺ T Cells. *Infection and immunity* 84, 2627-2638.
- Ruiz, J.T., Lujan, L., Blank, M., and Shoenfeld, Y. (2017). Adjuvants- and vaccines-induced autoimmunity: animal models. *Immunologic research* 65, 55-65.
- Saha, S., and Raghava, G.P. (2004). BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties (Berlin Heidelberg).
- Saha, S., and Raghava, G.P. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65, 40-48.
- Sanchez-Trincado, J.L., Gomez-Perosanz, M., and Reche, P.A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of immunology research* 2017, 2680160.
- Sanchez Alberti, A., Bivona, A.E., Cerny, N., Schulze, K., Weissmann, S., Ebensen, T., Morales, C., Padilla, A.M., Cazorla, S.I., Tarleton, R.L., *et al.* (2017). Engineered trivalent immunogen adjuvanted with a STING agonist confers protection against *Trypanosoma cruzi* infection. *NPJ vaccines* 2, 9.
- Schmunis, G., and Yadon, Z. (2010). Chagas disease: a Latin American health problem becoming a world health problem. *Acta Tropica* 115, 14-21.
- Schubert, B., and Kohlbacher, O. (2016). Designing sting of beads vaccines with optimal spacers. *Genome medicine* 8.
- Schulze, K., Ebensen, T., Babiuk, L., Gerdt, V., and Guzman, C. (2017a). Intranasal vaccination with an adjuvanted polyphosphazenes nanoparticle-based vaccine formulation stimulates protective immune responses in mice. *Nanomedicine* 13, 2169–2178.
- Shender, L.A., Lewis, M.D., Rejmanek, D., and Mazet, J.A. (2016). Molecular Diversity of *Trypanosoma cruzi* Detected in the Vector *Triatoma protracta* from California, USA. *PLoS neglected tropical diseases* 10, e0004291.
- Soria-Guerra, R.E., Nieto-Gomez, R., Govea-Alonso, D.O., and Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *Journal of biomedical informatics* 53, 405-414.

- Studer, R., Dessailly, B., and Orengo, C. (2012). Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *The Biochemical journal* *449*, 581-594.
- Takayama, E., Ono, T., Carnero, E., Umemoto, S., Yamaguchi, Y., Kanayama, A., and Miyahira, Y. (2010). Quantitative and qualitative features of heterologous virus-vector-induced antigen-specific CD8⁺ T cells against *Trypanosoma cruzi* infection. *International journal for parasitology* *40*, 1549-1561.
- Team, R.C. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria).
- Teixeira, A., Hecht, M., Guimaro, M., Sousa, A., and Nitz, N. (2011). Pathogenesis of Chagas' Disease: Parasite Persistence and Autoimmunity. *Clinical Microbiology* *24*, 592-630.
- Temizoz, B., Kuroda, E., and Ishii, K. (2018). Combination and inducible adjuvants targeting nucleic acid sensors. *Curr Opin Pharmacol* *41*, 104-113.
- Thompson, C., Lourenço, J., Walters, A., Obolski, U., Edmans, M., Palmer, D., Kooblall, K., Carnell, G., O'Connor, D., and Gupta, S. (2018). A naturally protective epitope of limited variability as an influenza vaccine target. *Nature communications* *9*.
- Tosello Boari, J., Amezcua Vesely, M.C., Bermejo, D.A., Ramello, M.C., Montes, C.L., Cejas, H., Gruppi, A., and Acosta Rodriguez, E.V. (2012). IL-17RA signaling reduces inflammation and mortality during *Trypanosoma cruzi* infection by recruiting suppressive IL-10-producing neutrophils. *PLoS pathogens* *8*, e1002658.
- Urbina, J. (2010). Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches. *Acta Tropica* *115*, 55-68.
- Vázquez-Chagoyán, J., Gupta, S., and Garg, N. (2011). Vaccine development against *Trypanosoma cruzi* and Chagas disease. *Advances in parasitology* *75*, 121-146.
- Velders, M., Weijzen, S., Gretchen, E., Elmishad, A., Kloetzel, P., Higgins, T., Ciccarelli, R., Evans, M., and Man, S. (2001). Defined Flanking Spacers and Enhanced Proteolysis Is Essential for Eradication of Established Tumors by an Epitope String DNA Vaccine. *The Journal of Immunology* *10*.
- Villanueva-Lizama, L.E., Cruz-Chan, J.V., Aguilar-Cetina, A.D.C., Herrera-Sanchez, L.F., Rodriguez-Perez, J.M., Rosado-Vallado, M.E., Ramirez-Sierra, M.J., Ortega-Lopez, J., Jones, K., Hotez, P., *et al.* (2018). *Trypanosoma cruzi* vaccine candidate antigens Tc24 and TSA-1 recall memory immune response associated with HLA-A and -B supertypes in Chagasic chronic patients from Mexico. *PLoS neglected tropical diseases* *12*, e0006240.
- Vinothkanna, A., Karthikeyan, B., Vijayan, R., and Vekar, S. (2018). Assessment of anti-arthritis potential of traditionally fermented ayurvedic polyherbal product chandanasava by molecular modelling, docking and dynamics approaches. *International Journal of Computational Biology and Drug Design* *11*, 346-368.
- Virgil, E., and Lavelle, E. (2011). Trends in vaccine adjuvants. *Expert Reviews* *10*, 539-550.
- Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., *et al.* (2015). The immune epitope database (IEDB) 3.0. *Nucleic acids research* *43*, D405-412.
- von Delft, A., Donnison, T.A., Lourenco, J., Hutchings, C., Mullarkey, C.E., Brown, A., Pybus, O.G., Klenerman, P., Chinnakannan, S., and Barnes, E. (2018). The generation of a simian adenoviral vectored HCV vaccine encoding genetically conserved gene segments to target multiple HCV genotypes. *Vaccine* *36*, 313-321.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F., de Beer, T., Rempfer, C., Bordoli, L., *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* *46*.

- Weaver, S., Shank, S., Spielman, S., Li, M., Muse, S., and Kosakovsky Pond, S. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution* *35*, 773-777.
- Wizel, B., Garg, N., and Tarleton, R.L. (1998). Vaccination with trypanomastigote surface antigen 1-encoding plasmid DNA confers protection against lethal *Trypanosoma cruzi* infection. *Infection and immunity* *66*, 5073-5081.
- Won, G., and Lee, J.H. (2018). Antigenic and functional profiles of a *Lawsonia intracellularis* protein that shows a flagellin-like trait and its immuno-stimulatory assessment. *Veterinary research* *49*, 17.
- Wrightsmann, R., Dawson, B., Fouts, D., and Manning, J. (1994). Identification of immunodominant epitopes in *Trypanosoma cruzi* Trypomastigote Surface Antigen-1 Protein That Mask Protective Epitopes. *Journal of immunology* *153*.
- Wu, S., and Zhang, Y. (2007). LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic acids research* *35*, 3375-3382.
- Xu, D., and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal* *101*, 2525-2534.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods* *12*.
- Yasser, E., and Honavar, V. (2010). Recent advances in B-cell epitope prediction methods. *Immunologic research* *6*, 1-9.
- Zhao, W., and Sher, X. (2018). Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLoS Comput Biol* *14*, e1006457.
- Zhumabekov, T., Corbella, P., Tolaini, M., and Kioussis, D. (1995). Improved version of a human CD2 minigene based vector for T cell-specific expression in transgenic mice. *Journal of immunological methods* *185*, 133-140.
- Zoghbi, N., Tassinari, P., Fortes, M., Paredes, M., Gamez, L., and Palacios, M. (2016). Perfil genético HLA en familias venezolanas: útil herramienta biomédica. *Comunidad y Salud* *1*.