

# Universidad Autónoma del Estado de México

## Facultad de Ciencias

Tesis como requisito para la evaluación profesional de la carrera de Licenciado en Biotecnología presenta.

Predicción teórica de la actividad antigénica de los epítomos líderes pertenecientes a las proteínas de superficie de *Trypanosoma cruzi* mediante IEDB



Alumno: Gilberto Yong Macías

Directora: D. en CARN. Esvieta Tenorio Borroto

Codirector: D. en C. Juan Carlos Vázquez Chagoyán

(1° generación, 0810890)

# ÍNDICE

<b>I. INTRODUCCIÓN</b> .....	<b>- 5 -</b>
<b>OBJETIVO GENERAL</b> .....	<b>- 6 -</b>
<b>OBJETIVOS ESPECÍFICOS</b> .....	<b>- 7 -</b>
<b>HIPÓTESIS</b> .....	<b>- 7 -</b>
<b>JUSTIFICACIÓN</b> .....	<b>- 7 -</b>
<b>I. REVISIÓN DE LITERATURA</b> .....	<b>- 8 -</b>
1.1 <i>Chagas</i> .....	- 8 -
1.1 <i>TRYPANOSOMA CRUZI</i> .....	- 13 -
1. SISTEMA INMUNE .....	- 15 -
2.1 <i>Respuesta inmune adaptativa</i> .....	- 16 -
2.2 <i>Complejo Mayor de Histocompatibilidad</i> .....	- 19 -
2.2.1 MHC Clase I.....	- 20 -
2.2.2 MHC clase II .....	- 21 -
3. EPÍTOPOS.....	- 22 -
3.1 <i>Epítomos de linfocitos B</i> .....	- 23 -
3.2 <i>Epítomos de linfocitos T</i> .....	- 23 -
<b>II. MATERIAL Y MÉTODOS</b> .....	<b>- 24 -</b>
1. RECOLECCIÓN DE DATOS .....	- 24 -
1.1 <i>Identificar las secuencias de proteínas de superficie de Trypanosoma cruzi</i> .....	- 24 -
1.2 <i>Selección de alelos de Restricción del MHC I</i> .....	- 25 -
2. PREDICCIÓN DE EPÍTOPOS .....	- 25 -
2.1 <i>Aplicación del método de predicción consensus</i> .....	- 26 -
2.2 <i>Selección del núcleo o core del epítomo</i> .....	- 27 -
2.2.1 <i>Criterios para la determinación de la Longitud de cadena del núcleo de los epítomos</i> .....	- 27 -
2.2.2. <i>Selección del núcleo de los epítomos con alta afinidad</i> .....	- 27 -
3. SELECCIÓN DE EPÍTOPOS.....	- 28 -
3.1 <i>Análisis de frecuencias de epítomos</i> .....	- 28 -
3.1.1 <i>Determinación de la Hidrofobicidad del núcleo del epítomo</i> .....	- 28 -
3.2 <i>Análisis de la cobertura poblacional de los epítomos</i> .....	- 29 -
4. DIAGRAMA DE FLUJO.....	- 30 -
<b>III. RESULTADOS</b> .....	<b>- 31 -</b>
<b>IV DISCUSIÓN</b> .....	<b>4</b>
<b>V. CONCLUSIONES</b> .....	<b>9</b>
PERSPECTIVAS.....	9
<b>VI BIBLIOGRAFÍA</b> .....	<b>10</b>

## ÍNDICE DE FIGURAS

Ilustración 2: Zonas endémicas de Chagas y regiones con reportes de la enfermedad .....	- 9 -
Ilustración 3: Etapas del diagnóstico clínico de la enfermedad de Chagas.....	- 12 -
Ilustración 1: Ciclo de vida de Trypanosoma Cruzi .....	- 15 -
Ilustración 4 Esquema de un epítipo .....	- 16 -
Ilustración 5: Esquema de velocidad de respuesta del sistema inmune innato y adaptativo .....	- 17 -
Ilustración 6 Diagrama general de la metodología empleada .....	- 30 -
Ilustración 7 Falta de Reconocimiento de los alelos en el programa de Cobertura Poblacional.....	3

## ÍNDICE DE TABLAS

Tabla 1 Valores de afinidad del epítipo.....	- 26 -
Tabla 2 Proteínas antigénicas de superficie expresadas por Trypanosoma Cruzi .....	- 31 -
Tabla 3 Comportamiento de la afinidad al MCH II en linfocitos T de los epítomos seleccionados de las 13 proteínas antigénicas de T Cruzi. ....	- 32 -
Tabla 4 Comportamiento de la frecuencia del núcleo seleccionado en las proteínas del grupo A expresadas por Trypanosoma cruzi. ....	- 33 -
Tabla 5 Comportamiento de la frecuencia del núcleo seleccionado en las proteínas del grupo B expresadas en Trypanosoma cruzi .....	1
Tabla 6 Frecuencia del núcleo en Cruzipain del grupo C expresada en Trypanosoma cruzi .....	2
Tabla 7 Epítomos líderes de Trypanosoma cruzi.....	3

## I. INTRODUCCIÓN

Trypanosoma cruzi es un parásito que transmite la enfermedad de chagas, la cual se encuentra entre la lista de enfermedades infecciosas en riesgo de propagación. La organización Mundial de la salud (OMS) reporta que un estimado de entre 6 - 7 millones de personas portan la enfermedad alrededor del mundo siendo América Latina la más afectada. En México se estima que 1.1 millones son las personas infectadas, de acuerdo con las cifras oficiales (OMS), aunque existen otras estimaciones: 1 - 6 millones(Hotez, Dumonteil, Heffernan, & Bottazzi, 2013).

La enfermedad es considerada, un severo problema de salud en áreas rurales de México, América Central y Sudamérica, con manifestaciones clínicas y características epidemiológicas variables. Ahora, se reportan con mayor frecuencia casos provenientes de ciudades de Latinoamérica, presentándose como una transmisión enzoótica, involucrando a diferentes especies de vectores y mamíferos, entre ellos mapaches, zarigüeyas y perros domésticos, aunque los casos humanos autóctonos reportados son pocos(Bern & Montgomery, 2009).El mal de Chagas es una de las 17 enfermedades desatendidas generando poco interés de las empresas farmacéuticas, pues afecta principalmente a poblaciones de países pobres.

Desafortunadamente, los únicos fármacos disponibles para la enfermedad de Chagas son el Nifurtimox y el Benzonidazol, los cuales son relativamente tóxicos para los pacientes adultos y requieren una administración prolongada. El Benzonidazol es la primera opción para el tratamiento de la enfermedad de Chagas debido a que sus efectos secundarios son menores que el Nifurtimox. Sin embargo, se están buscando diferentes estrategias para superar la toxicidad del Benzonidazol, incluyendo horarios de administración más cortos o intermitentes, ya sea solo o en combinación con otros fármacos(Bermudez, Davies, Simonazzi, Real, & Palma, 2016).

Actualmente el desarrollo de vacunas recombinantes permite disminuir los efectos secundarios de los fármacos mencionados anteriormente y son más eficientes en contra de los diversos agentes infecciosos al generar una respuesta inmune a partir de moléculas llamadas epítomos, que identifican los linfocitos T o B en el organismo.

Gracias a nuevos algoritmos y a la capacidad de procesamiento masivo de datos podemos generar programas basados en *machine learning* para la predicción de los epítomos, los cuales se espera desencadenen una respuesta inmune con menores efectos secundarios.

En el mercado podemos encontrar 14 vacunas que han sido diseñadas contra la enfermedad de Chagas, todas parten de las proteínas antigénicas de superficie del parásito, sin embargo, la eficacia de ellas oscila en un 60 %. Una de las causas del rendimiento tan bajo es que *Trypanosoma cruzi* tiene la particularidad de presentarse en diferentes estadios dentro del torrente sanguíneo, lo que cambia la composición de sus proteínas de superficie, esto representa un reto para la identificación de epítomos, debido a esto tomamos como base un enfoque de predicción *consensus* tanto para la predicción de los epítomos, como para la posterior selección de estos.

El siguiente trabajo propone una metodología para identificar los epítomos líderes pertenecientes a las proteínas de superficie de los diferentes estadios del *Trypanosoma cruzi* con el fin de diseñar una vacuna que permita disminuir la tasa de propagación y prevalencia de esta enfermedad empleando técnicas bioinformáticas.

El estudio fue financiado por el proyecto PRODEP “Evaluación de las vacunas tcg2y tcg4 sobre las células de memoria central y efectos de ratones infectados con *Trypanosoma cruzi*”

### **Objetivo general**

Predecir la antigenicidad de los epítomos líderes pertenecientes a las proteínas de superficie en *Trypanosoma cruzi* para el diseño de una vacuna recombinante para América Latina.

## **Objetivos específicos**

1. Seleccionar los epítomos de mayor antigenicidad de las proteínas de superficie de *Trypanosoma cruzi* en diferentes estadios
2. Identificar los alelos representativos de América Latina que maximicen su cobertura.
3. Determinar las propiedades antigénicas e hidrofílicas de los epítomos líderes.

## **Hipótesis**

La predicción de los epítomos producidos en los diferentes estadios de *Trypanosoma cruzi* por medio de herramientas bioinformáticas, facilitara diseñar una vacuna recombinante contra este parásito.

## **Justificación**

En México se estima que 1.1 millones de personas presentan enfermedad de Chagas debido a la infección de *Trypanosoma Cruzi* (Hotez, Dumonteil, Heffernan, & Bottazzi, 2013), a pesar de que existen 14 vacunas registradas para su tratamiento se estima una eficacia de alrededor de un 90% si se aplica en etapas tempranas, cifra que va disminuyendo con la prevalencia de la enfermedad (Hotez et al., 2013), esto podría deberse a que dichas vacunas sólo presentan epítomos de un solo estadio del parásito y a la poca afinidad de los epítomos al MHC de los linfocitos.

Gracias al desarrollo y constante actualización de las bases de datos bioinformáticas se han realizado aproximaciones importantes en cuanto a la estructura y función de las proteínas, esto ha posibilitado grandes avances en la proteómica encaminada al diseño de vacunas recombinantes siendo más específicas contra agentes infecciosos (Toussaint, Maman et al., 2011). El propósito de este proyecto es predecir la actividad antigénica de los epítomos líderes que pertenecen a las proteínas de superficie de los diferentes estadios del *Trypanosoma cruzi* empleando para ello herramientas bioinformáticas como el desarrollada por

Immune epitope database “IEDB” con la perspectiva de obtener una vacuna más eficaz.

## I. Revisión de Literatura

### 1.1 Chagas

La enfermedad de Chagas, causada por la infección con el parásito protozoario *Trypanosoma cruzi*, es una enfermedad tropical desatendida que afecta hasta 10 millones de personas en todo el mundo. La enfermedad es endémica en vastas áreas del hemisferio occidental desde los Estados Unidos hasta Argentina, a pesar de eso muchos individuos infectados con *Trypanosoma cruzi* también residen en Canadá, Europa, y Australia.

Con el tiempo la deforestación avanza para hacer espacio a la agricultura y a la cría de ganado, los insectos y principalmente los triatomíneos se adaptan a las áreas rurales, formando áreas endémicas en América del Sur, Central y México. Más recientemente, a partir de la década de 1980, llamó la atención sobre la globalización de la enfermedad de Chagas a través de la migración de pacientes a EE. UU., Europa, Asia y Oceanía (Schmunis, 2007).

La migración también ha cambiado la distribución de la enfermedad de Chagas de un problema de salud solo en América Latina a un problema global. A partir de 2013, se estimaba que 36,7 millones de personas habían emigrado de América Latina y el Caribe y residían en otras partes del mundo, principalmente en América del Norte (Departamento de Asuntos Económicos y Sociales de las Naciones Unidas, 2013b).



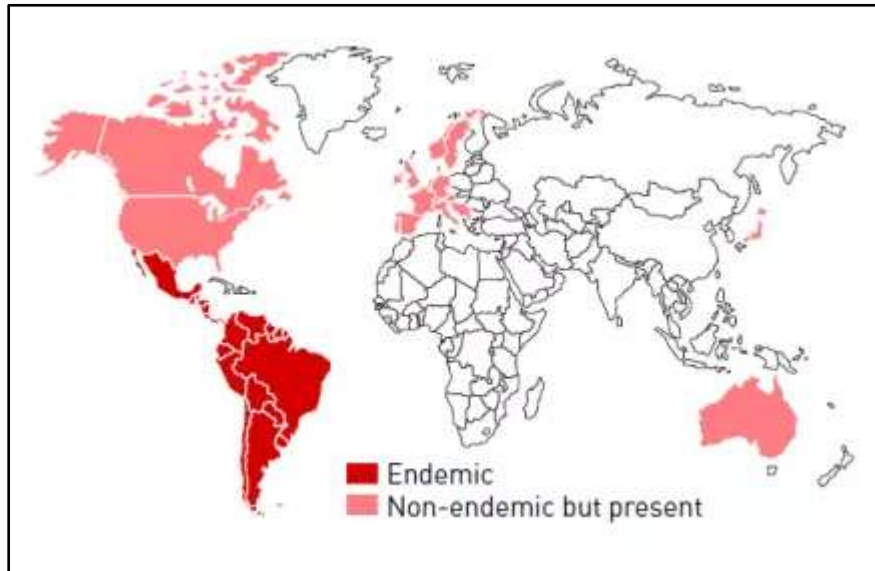


Ilustración 1: Zonas endémicas de Chagas y regiones con reportes de la enfermedad

Los mecanismos para la transmisión de la infección de Chagas a los humanos se pueden dividir en primarios y secundarios. Entre los mecanismos primarios están el vector, la transfusión de sangre, la transmisión oral, placentaria o congénita y el canal de parto en el momento del nacimiento. Los mecanismos secundarios se consideran menos frecuentes, entre estos están, los accidentes de laboratorio, manejo de animales infectados, ingestión de carne cruda de animales infectados, trasplantes de órganos de donantes infectados con *Trypanosoma cruzi*, transmisión sexual y, excepcionalmente, a través de infección inducida o criminal (Coura, 2007).

El principal modo de transmisión de *Trypanosoma cruzi* a los seres humanos es el vector transmitido, que ocurre sólo en América. Tradicionalmente es considerada una enfermedad de la pobreza; el riesgo de la enfermedad de Chagas se ha asociado con viviendas en áreas rurales, de mala calidad de construcción (p. Ej., Techo de palma, grietas en las paredes) y con mascotas domésticas y ganado en la casa.

La transmisión de Chagas se puede resumir en las siguientes formas:

1. Infección o enfermedad que se transmite entre animales salvajes.

2. Infección o enfermedad que se transmite de animales a humanos, ya sea cuando invaden el ecosistema silvestre o cuando los vectores o animales salvajes invaden áreas domésticas humanas.
3. Infección o enfermedad intercambiada entre animales y humanos
4. Infección o enfermedad que se transmite de humanos a animales, particularmente por vectores presentes en hogares a animales domésticos principalmente por triatominos (Coura, 2013).

Entre las características de esta enfermedad, quizás la más interesante es también la más discreta: la alta variabilidad en la infección, determinada en gran medida por la heterogeneidad genética tanto en la virulencia de cepas específicas de *Trypanosoma cruzi* como la susceptibilidad de un individuo específico a la infección, siendo el principal vector de la infección por Chagas el parásito *Triatoma infestans* que habita en una gran parte de Sudamérica, donde hasta hace poco se distribuía desde el sur de Argentina hasta el noreste de Brasil.

Este vector probablemente se originó en Bolivia, donde se encuentra en áreas domésticas, peridomésticos y silvestres, y fue transportado a Argentina, Chile, Paraguay, Uruguay y Brasil, donde se volvió exclusivamente doméstico. A través de la Iniciativa del Cono Sur, fue eliminada de Uruguay en 1997, de Chile en 1999 y de Brasil en 2006, según la certificación de la Organización Panamericana de la Salud (Coura & Dias, 2009), a pesar de estos esfuerzos la enfermedad aún prevalece en zonas rurales.

La falta de síntomas presentes y desconocimiento de la enfermedad Chagas ha generado que la segunda vía más frecuente de transmisión de la enfermedad sea a través de transfusiones sanguíneas. La enfermedad tiene 3 fases: aguda, indeterminada y crónica. La fase aguda ocurre inmediatamente después de la infección. Por lo general, es asintomático, pero produce fiebre y malestar en el 5% de la población.

La fase indeterminada es asintomática. Más de la mitad de los infectados permanecerán en esta fase de por vida y nunca experimentan secuelas a largo

plazo. Después de una década o más, entre el 20% y el 30% de las personas experimentan alguna enfermedad cardiovascular crónica de Chagas con secuelas que incluyen insuficiencia cardíaca, arritmias y tromboembolismo. Otro 15% a 20% experimentan secuelas digestivas crónicas que incluyen megaesófago y megacolon. La fase crónica de la enfermedad es una miocardiopatía inflamatoria que se desarrolla en aproximadamente un tercio de las personas infectadas con el parásito protozoario *Trypanosoma cruzi*.

La patogénesis de la enfermedad de Chagas sugiere que la fase crónica debe ser tratada más ampliamente para:

- A. Eliminar el *Trypanosoma cruzi* y prevenir nuevos focos inflamatorios y la extensión de lesiones tisulares.
- B. Promover la regeneración tisular para prevenir la fibrosis.
- C. Revertir la fibrosis existente.
- D. Prevenir la miocardiopatía, el megaesófago y el megacolon.
- E. Reducir o eliminar el bloqueo cardíaco y la arritmia.

(Stanaway & Roth, 2015).

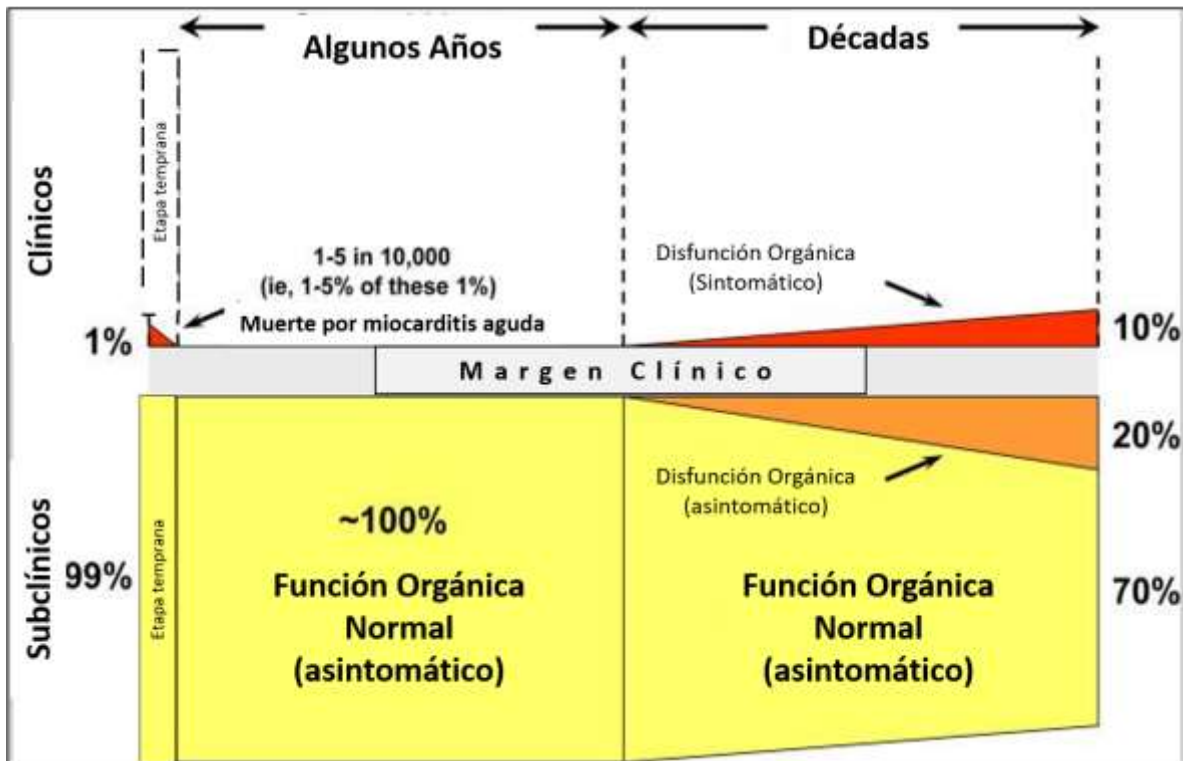


Ilustración 2: Etapas del diagnóstico clínico de la enfermedad de Chagas

Múltiples mecanismos de transmisión, junto con la naturaleza crónica, a menudo asintomática de la enfermedad y la falta de un tratamiento efectivo y accesible para los pacientes, significa que la prevalencia de la enfermedad continúa siendo alta (Organización Panamericana de la Salud, 2014). La morbilidad y mortalidad asociadas con la enfermedad de Chagas dan como resultado una carga económica mundial anual de US \$ 7,2 mil millones (Lee, Bacon, Bottazzi, & Hotez, 2013).

El diagnóstico de la infección aguda por *Trypanosoma cruzi* se puede realizar mediante análisis microscópicos de sangre completa o capa leucocitaria para detectar la presencia de trypomastigotes móviles son liberados de las células huésped después de las primeras rondas iniciales de replicación. En pacientes con Chagas crónico, el ensayo inmunoabsorbente ligado a enzimas, la hemaglutinación indirecta o la inmunofluorescencia indirecta pueden usarse para evaluar la presencia de IgG específica de *T. cruzi* en suero (Coura, 2007).

La administración adecuada de Nifurtimox y el Benzonidazol puede mejorar con éxito los síntomas de la enfermedad de Chagas en muchos individuos, especialmente en aquellos que experimentan la forma aguda de la enfermedad. Sin embargo, los tratamientos existentes para la enfermedad de Chagas tienen una eficacia limitada en ciertas poblaciones porque requieren administración repetida durante un período de tiempo prolongado, y a menudo producen efectos adversos graves como náuseas, vómitos, dolor gástrico, insomnio, dolor de cabeza, vértigos, mialgias, artralgias y convulsiones que muchas veces no pueden ser tolerados por todas las personas.

Varios otros medicamentos han demostrado ser efectivos *in vitro* contra *Trypanosoma cruzi*. Entre estos se encuentran el alopurinol, una hipoxantina anti-uricemia que se usa para tratar la gota y los antifúngicos que inhiben el ergosterol, como el ketoconazol, el itraconazol y el fluconazol 42-46. Estos no han demostrado ser efectivos *in vivo*, incluso durante la fase aguda de la enfermedad (Coura, 2007).

Actualmente, no hay una vacuna efectiva disponible para prevenir la infección por *Trypanosoma cruzi*, aunque muchos comparten la opinión de que las perspectivas son más alentadoras ahora que en el pasado.

### **1.1 *Trypanosoma cruzi***

*Trypanosoma cruzi* es un parásito protozoo que se alimenta de sangre y necesita de otro organismo que le sirva de huésped, su principal vector son los insectos hematófagos triatominos "chinche besucona" de acuerdo con su mecanismo de infección. Este comportamiento puede causar al hospedero enfermedades que varían dependiendo del organismo, por ejemplo, tripanosomiasis en humanos; "Chagas", durina y surra en caballos y una enfermedad similar a la brucelosis en el ganado.

Al triatomino le gustan los nidos de animales vertebrados como refugio, donde muerde y chupa sangre para alimentarse. Los triatominos infectados con protozoos, al estar en contacto con otros animales les transmiten tripanosomas cuando el triatomino los muerde y deposita sus heces en la superficie de la piel del huésped.

La penetración de las heces infectadas se facilita aún más por el rascado del área de la picadura por el huésped humano o animal.

*Trypanosoma cruzi* tiene un ciclo de vida complejo que consta de tres etapas de desarrollo (Ilustración 1), primero el vector ingiere los trypomastigotes metacíclicos de una fuente externa, después viajan por el sistema digestivo donde se transforman en epimastigotes en el intestino delgado, aquí se multiplica para avanzar al intestino grueso donde surge su segunda transformación en trypomastigotes metacíclicos.

La transmisión natural a un nuevo huésped mamífero ocurre cuando las heces cargadas de parásitos contaminan las membranas mucosas orales o nasales, las conjuntivas o las heridas en la piel, incluidas las mordeduras del vector. Una vez en el huésped mamífero, los trypomastigotes metacíclicos ingresan a las células del huésped. La infección continúa cuando se transforman en amastigotes, lo que le permite reproducirse en el hospedero por fisión binaria y luego se transforman nuevamente en trypomastigotes.

Estas formas se liberan en el torrente sanguíneo a medida que la célula hospedera se rompe y luego están listas para invadir las células sanas (Machado et al., 2012).

Los trypomastigotes infectan células adyacentes no infectadas o se diseminan a través de los vasos linfáticos y sanguíneos para infectar células en sitios distantes. Aunque cualquier célula de mamífero nucleada puede ser parasitada, las comúnmente infectadas incluyen miocitos cardíacos, células periféricas esqueléticas y de músculo liso, células endoteliales y células de los sistema nervioso y reticuloendotelial y tejido adiposo.

Además de la transmisión vectorial, se ha informado de transmisión de *Trypanosoma cruzi* por transfusión de sangre o trasplante de órganos de donantes infectados, transmisión congénita y adquisición de la infección por vía oral.

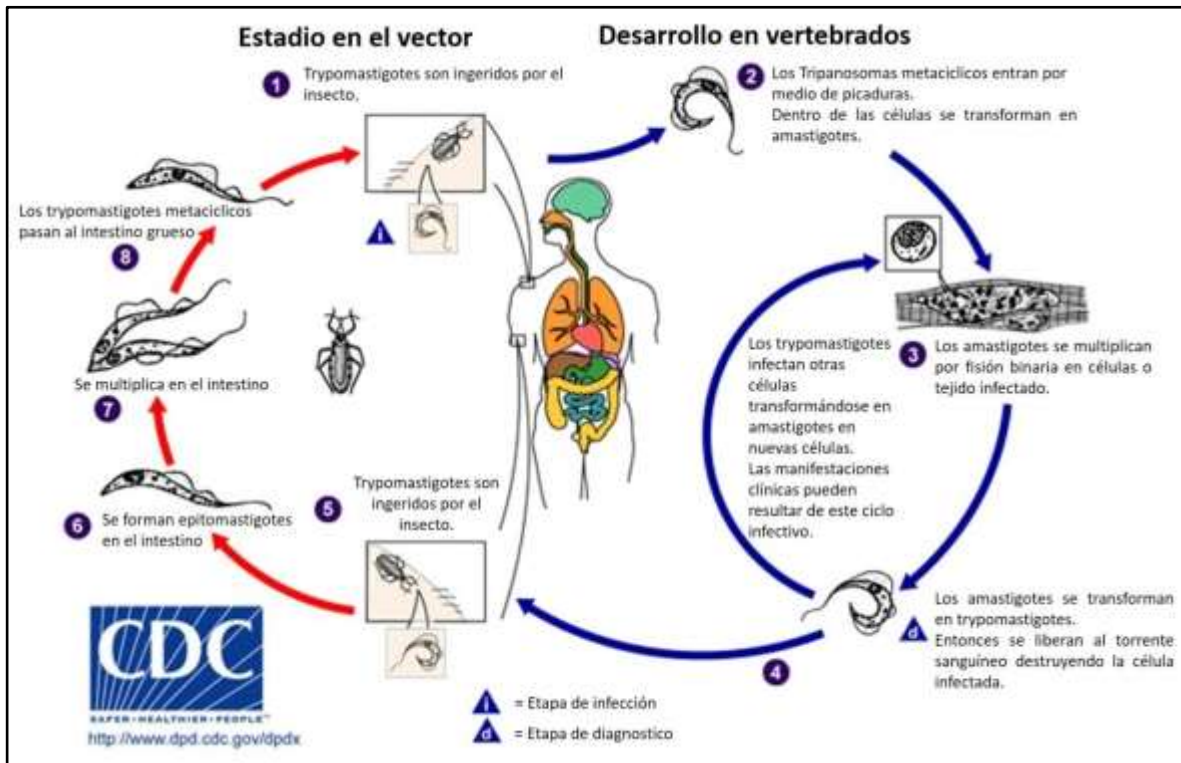


Ilustración 3: Ciclo de vida de *Trypanosoma Cruzi*

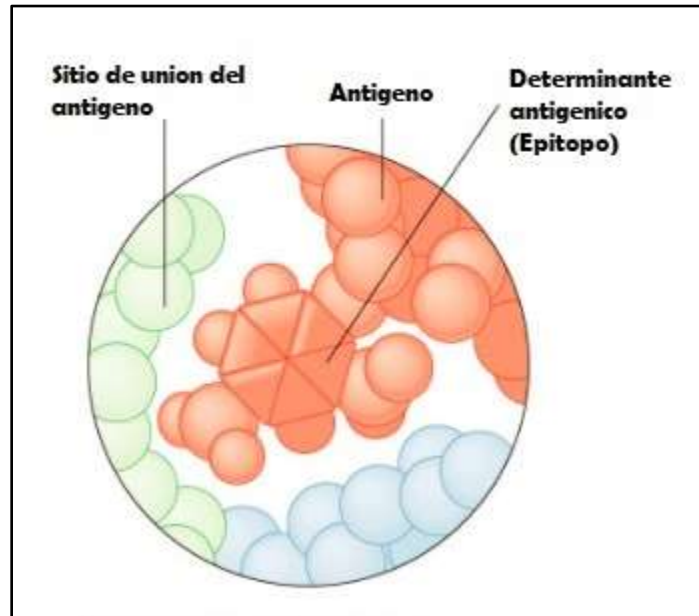
## 1. Sistema inmune

El trabajo del sistema inmune es proteger al organismo de las enfermedades, patógenos, agentes dañinos y células tumorales. Esto se logra al generar células y moléculas capaces de diferenciar las propias de las extrañas y eliminarlas. En los vertebrados el sistema inmune consta de dos partes: inmunidad innata e inmunidad adaptativa, las cuales trabajan en colaboración para proteger el cuerpo.

La respuesta innata del sistema inmune tiene la función de reaccionar rápidamente contra compuestos extraños en el cuerpo de un individuo sano, es capaz de identificar proteínas y péptidos específicos de organismos extraños. Un antígeno es una sustancia que desencadena la formación de anticuerpos y puede causar una respuesta inmunitaria, es generalmente una parte procesada de una proteína unida a una proteína del complejo mayor de histocompatibilidad (MHC).

La parte de una proteína responsable de una respuesta inmune es la región del antígeno que identifica el MHC, a menudo esta parte se denomina epítipo y la

proteína nativa de la que se originó el epítipo como antígeno. Los epítipos se definen originalmente como la parte de un antígeno que se une a una inmunoglobulina (Huang & Honda, 2006).



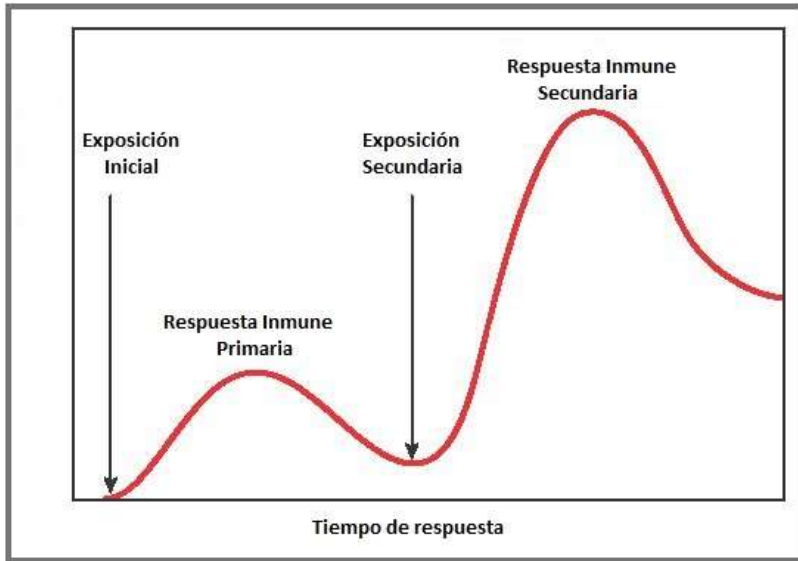
*Ilustración 4 Esquema de un epítipo*

Una gran parte de los estudios de inmunología son dedicados al descubrimiento de nuevos epítipos con el objetivo de desarrollar nuevas vacunas. Las moléculas del complejo mayor de histocompatibilidad (MHC) juegan un rol importante en la respuesta inmune adaptativa presentando los péptidos al sistema inmune.

## **2.1 Respuesta inmune adaptativa**

El sistema inmune adaptativo por otro lado es capaz de reconocer agentes externos de forma específica, eliminándolos y desarrollando memoria inmune, permitiendo al organismo efectuar una respuesta rápida y efectiva si el patógeno ingresa otra vez. Las vacunas tienen su efectividad en activar la respuesta inmune adaptativa mediante el cual el individuo adquiere inmunidad permanente con el fin de preservar esta inmunidad contra algún agente patógeno específico.





*Ilustración 5: Esquema de velocidad de respuesta del sistema inmune innato y adaptativo*

El sistema inmune adaptativo comprende un gran número de células y moléculas, las cuales participan en el reconocimiento de antígenos, patógenos y en el proceso de memoria. El sistema inmune adaptativo se compone de dos partes, el sistema inmune humoral con los linfocitos B y el sistema inmune celular donde los linfocitos T juegan un papel importante.

Los linfocitos B secretan anticuerpos, proteínas (gamma-globulinas) mientras circulan en el torrente sanguíneo y son capaces de reconocer patógenos y neutralizarlos. Otra función importante de las células es su capacidad de diferenciarse en células de memoria que tienen mayor longevidad y se activan rápidamente en caso de reconocer un mismo antígeno otra vez.

Al contrario de las células B, las células T no son capaces de reconocer antígenos libres, y su interacción con el MHC requiere de la activación del sistema inmune. Hay dos clases de moléculas de MHC, clase I y clase II, que son asociadas con diferentes grupos de células T:

- Linfocitos T citotóxicos (CTL). reconocen péptidos presentados por el MHC clase I presentado en todas las células nucleadas. CTL son células efectoras con la habilidad de reconocer y eliminar células infectadas, los CTL también son llamados células CD8 debido al receptor de proteína CD8 que expresan en su superficie.
- T helper también llamadas CD4 debido a que expresan el receptor CD4. Las células CD4 activan otras células del sistema inmune como las células B, CD8, macrófagos etc. Las células CD4 reconocen antígenos del ambiente extracelular encontradas en moléculas pertenecientes al MHC clase II.

Las vacunas que activan el sistema inmune adaptativo están desarrolladas a base del agente que causa la enfermedad, pero en un estado no patogénico. Éstas pueden estar constituidas por el agente causante de la enfermedad vivo, pero atenuado (disminuido en su capacidad de desencadenar la enfermedad), por el agente patógeno muerto, o por sus fracciones (antígenos).

Con la administración de la vacuna se desarrolla la respuesta inmune, el inconveniente con las vacunas tradicionales es el alto costo de producción y en algunos casos, la hipersensibilidad que puede tener la persona al usar un agente en estado atenuado o producir algún efecto adverso.

Para evitar tales efectos secundarios se han generado métodos que permiten identificar secuencias del genoma del agente infeccioso que generen una respuesta inmune en el organismo, esto resulta en un interesante reto de optimización combinatoria donde se debe seleccionar el conjunto mínimo de secuencias (epítomos) maximizando la cobertura en la población mundial, la cual está representada por sus frecuencias de alelos globales (Toussaint, Maman, Kohlbacher, & Louzoun, 2011).

El alto grado de polimorfismo impone un gran desafío en el descubrimiento de epítomos. Afortunadamente, existen bases de datos de frecuencias alélicas en la web, como [allelefrequencies.net](http://allelefrequencies.net) (Gonzalez-Galarza, Christmas, Middleton, & Jones,

2011). Estas bases de datos ayudan a seleccionar los epítomos y así maximizar la cobertura poblacional de una vacuna recombinante.

## **2.2 Complejo Mayor de Histocompatibilidad**

En humanos, el MHC está constituido por al menos de 4 000 000 pares de bases y se considera que podrían ser hasta 7 000 000, estos codifican más de 200 genes ubicados en el brazo corto del cromosoma 6. En general, el MHC permite distinguir lo propio de lo extraño.

El complejo mayor de histocompatibilidad juega un papel crucial en el sistema inmune celular presentando péptidos derivados de patógenos en su superficie por el reconocimiento por los linfocitos T. Los complejos de histocompatibilidad de clase I y clase II difieren en su estructura, originando un sitio de reconocimiento distinto para los péptidos.

La región del brazo corto del cromosoma 6 que contiene los genes del MHC, posee la información de:

- Algunas glucoproteínas de la membrana plasmática involucradas en los mecanismos de presentación y procesamiento de antígenos a los linfocitos T: se agrupan en los genes de clase II (que codifican las proteínas MHC de clase II) y los genes de clase I (que codifican las proteínas MHC de clase I)
- citocinas y proteínas del sistema del complemento, importantes en la respuesta inmunológica, pero que no tienen nada que ver con los genes del MHC; estos genes se agrupan en la clase III.

Ambos tipos de moléculas participan en la respuesta inmunitaria, que permite la identificación de las moléculas propias y de las extrañas (invasoras), para eliminar estas últimas mediante diferentes mecanismos.

Debido a las miles de variantes del MHC existentes, el descubrimiento de nuevos epítomos es costoso a nivel de laboratorio con estudios de interacción péptidos-MHC, sin embargo, los métodos computacionales de predicción permiten ahorrar recursos en el descubrimiento de epítomos.

La región HLA-B es la región genética más polimórfica del genoma humano, seguida de la región del HLA-A. La región de genes MHC contiene una densidad y diversidad de genes extremadamente alta; la variación genética dentro de esta región juega un papel vital en la susceptibilidad a las enfermedades autoinmunes, infecciosas, y otras enfermedades (Vandiedonck & Knight, 2009).

Los métodos de predicción de unión de MHC cuentan en la actualidad con una exactitud muy alta y pueden predecir péptidos de unión muy precisos. Estos métodos pueden ser usados para predecir una amplia gama de alelos de MHC con diferentes tamaños de aminoácidos siendo esta una característica relevante para la unión con el MHC (Lundegaard, Lund, & Nielsen, 2012).

### **2.2.1 MHC Clase I**

Los péptidos presentados al MHC I son degradados por el proteosoma. Estos péptidos se componen usualmente entre 8- 11 aminoácidos. La unión péptido-MHC es un proceso selectivo muy riguroso, se estiman que solo 1 de 200 péptidos aleatorios se unirá a una molécula del MHC clase I. Cuando un péptido se une a una molécula de MHC clase I, el complejo péptido-MHC con la ayuda de las chaperonas es transportado a la superficie celular a través del complejo de Golgi.

Las moléculas del MHC clase I están compuestas de una cadena pesada  $\alpha$  y una  $\beta$ . La cadena  $\alpha$  consiste en tres dominios  $\alpha_1$ ,  $\alpha_2$  y  $\alpha_3$ . Los dominios  $\alpha_1$  y  $\alpha_2$  forman el sitio de unión. Para las moléculas del MHC clase I el sitio de unión se encuentra cercano al final de las cadenas, limitando el reconocimiento de péptidos a un rango de 8 a 11 aminoácidos.

En los humanos las moléculas del MHC son parte del sistema de antígeno leucocitario (HLA). Las cadenas pesadas del MHC clase I son codificadas en el cromosoma 6 por tres genes polimórficos: HLA - A, HLA - B, HLA - C. Existen cientos de versiones diferentes de HLA con los residuos más polimórficos localizados en el sitio de unión, resultando en un gran número de posibles moléculas de MHC con

diferentes uniones. El alto polimorfismo representa un desafío ante el desarrollo de nuevas vacunas y el descubrimiento de nuevos epítomos.

### **2.2.2 MHC clase II**

Al contrario del MHC clase I, el MHC clase II se une a moléculas provenientes del ambiente extracelular. Las moléculas del MHC II son ensambladas en el retículo endoplasmático y transportadas a través del aparato de Golgi al compartimiento del MHC II. Las proteínas obtenidas por endocitosis o fagocitosis son degradadas por proteasas y forman endosomas que eventualmente se fusionan con el compartimiento del MHC II. Después de reconocer los péptidos se forma el complejo del MHC II, el cual es transportado a la membrana plasmática para ser presentado a las células CD4.

Las células presentadoras de antígenos presentan péptidos antigénicos extracelulares procesados a los linfocitos T cooperadores (CD4+). El péptido antigénico se aloja en una hendidura formada por los dominios  $\alpha 1$  y  $\beta 1$ , mientras el reconocimiento del MHC-II por parte del linfocito T cooperador se hace en la cadena  $\beta 2$ .

El sitio de unión del MHC de clase II tiene espacios especiales que encajan con aminoácidos específicos del péptido de unión y son influenciados principalmente por la energía de unión (Bordner, 2010). La predicción del ligando de HLA clase II es más difícil que la predicción de clase I debido a la posición desconocida del núcleo de unión (core) debido a que los péptidos generalmente son más largos, derivando en algunas estructuras secundarias que bloquean el núcleo de unión.

Estructuralmente las células del MHC II son muy diferentes de las del MHC I, las moléculas del MHC II son heteros dímeros compuestos por cadenas  $\alpha$  y  $\beta$ , ambas con dominios transmembranales  $\alpha 2$  y  $\beta 2$ , ambos dominios están compuestos con sitios de unión que está disponible en ambos extremos y es receptivo a péptidos de longitud entre 15-20 residuos, o una proteína completa.

En humanos el MHC II se codifica en el cromosoma 6 igualmente. Tres diferentes loci codifican las cadenas  $\alpha$  y  $\beta$ : HLA-DR, HLA-DP y HLA-DQ, en el caso del HLA-DR, solo la cadena  $\beta$  codificada por el loci (DRB) se ha encontrado que es polimórfica, sin embargo, la fuerza de unión de las moléculas HLA-DR son determinadas por el DRB. Cada molécula específica puede denotarse por nombre del alelo ej. HLA.DRB1\*0101. A pesar de que HLA-DP y HLA-DQ, ambas cadenas  $\alpha$  y  $\beta$  muestran polimorfismo, una molécula puede ser identificada por los alelos que codifican ambas cadenas, ej. HLA-DPA1\*0302- DBP1\*0501.

### **3. Epítomos**

Las respuestas inmunitarias son específicas frente a los distintos antígenos y, de hecho, también frente a las diversas porciones de un solo complejo proteínico, de un polisacárido o de cualquier macromolécula. Los elementos de tales antígenos que son reconocidos específicamente por anticuerpos, linfocitos T citotóxicos o linfocitos B, se denominan determinantes o epítomos (Abbas, Lichtman, & Pillai, 2015). La identificación de epítomos en proteínas es de vital importancia para el desarrollo de vacunas basadas en péptidos sintéticos, pruebas de inmunodiagnóstico y producción de anticuerpos (Chen, Liu, Yang, & Chou, 2007).

Durante la búsqueda de epítomos para el desarrollo de vacunas se busca identificar los fragmentos más relevantes de una secuencia con ciertas cualidades inmunológicas para poder ser utilizados como epítomos. La importancia de la identificación de epítomos con un sitio de unión fuerte al MHC que contiene una proteína radica en el uso de péptidos como fármacos más específicos y con efectos secundarios menores o nulos. Para la unión de MHC de clase II, es intrínsecamente más difícil pasar de los datos de unión a péptidos a un motifs definido del núcleo de unión ya que se trata de un tramo continuo de nueve residuos de aminoácidos colocados en algún lugar en un péptido más grande usualmente en el intervalo de 13-25 residuos en longitud.

En el ser humano, las cadenas de MHC de clase II están codificadas por genes en los loci HLA-DR, -DQ y -DP. El conocimiento de los haplotipos es por lo tanto esencial para predecir los péptidos presentados por el HLA.

### **3.1 Epítomos de linfocitos B**

La interacción antígeno – anticuerpo es un evento clave en la respuesta inmune humoral contra la invasión por patógenos. Los epítomos de linfocitos B pueden ser definidos como clúster de aminoácidos accesibles que son reconocidos por anticuerpos secretados o receptores en linfocitos B y que son capaces de producir una respuesta inmune humoral o celular (Getzoff, Tainer, Lerner, & Geysen, 1988).

Los epítomos de linfocitos B son clasificados como lineales, regiones continuas de aminoácidos en una secuencia proteica y discontinuos, que son residuos separados en la secuencia, pero cercanos físicamente debido al plegamiento de la proteína (Chen et al., 2007).

La longitud mínima requerida para el plegamiento adecuado de un epítomo discontinuo en proteínas nativas es de al menos 400 aminoácidos (Potocnakova, Bhide, & Pulzova, 2016). Se cree que la mayoría de los determinantes antigénicos lineales identificados son partes de epítomos conformacionales. Usando una definición menos astringente, se ha encontrado que la mayoría de los epítomos discontinuos (más del 70%) están compuestos por 1 – 5 segmentos lineales de una longitud entre 1 – 6 aminoácidos (Rubinstein, Mayrose, & Pupko, 2009).

### **3.2 Epítomos de linfocitos T**

La unión de un péptido a una molécula del Complejo Mayor de Histocompatibilidad (CMH) es necesaria por su habilidad de activar la respuesta inmune de los linfocitos T. Los péptidos unidos a moléculas del MHC forman un complejo que es reconocido por los receptores de linfocitos T (TCRs).

Estos péptidos que son reconocidos por linfocitos T son conocidos como epítomos (Murphy, Travers et al. 2008). La fuerza de unión (afinidad) de un péptido con una

molécula del MHC es un factor importante que determina su potencial inmunogenicidad (Sette et al., 1994).

## **II. MATERIAL Y MÉTODOS**

El estudio se realizó en el Centro de Investigación y Estudios Avanzados en Salud Animal con el apoyo del área de investigación en biotecnología.

### **1. Recolección de Datos**

#### **1.1 Identificar las secuencias de proteínas de superficie de *Trypanosoma cruzi***

Se obtuvieron las secuencias de las proteínas de superficie relacionadas con *Trypanosoma cruzi* que muestran interacción con el hospedero registradas en la literatura, sólo se utilizaron las secuencias reportadas en UniProtKB debido a la calidad de sus secuencias, la cual es mayor a las reportadas en otras bases de datos.

La UniProt Knowledge Base (UniProtKB) es el centro para la recopilación de información funcional sobre proteínas, con anotaciones precisas y consistentes. Además de capturar los datos básicos obligatorios para cada entrada de UniProtKB (principalmente, la secuencia de aminoácidos, el nombre o la descripción de la proteína, los datos taxonómicos y la información de la cita), se agrega la mayor cantidad de información posible. Esta base de datos está disponible en la página (<https://www.uniprot.org>)

Este sitio tiene la particularidad de que sus secuencias son curadas a mano para evitar redundancias. Esta base de datos cuenta con 555,594 de secuencias revisadas manualmente y 90,050,711 revisadas de forma automática por lo que se considera confiable (“UniProt: the universal protein knowledgebase,” 2017).



## 1.2 Selección de alelos de Restricción del MHC I

Una vez obtenidas las secuencias de las proteínas, se seleccionaron los alelos de restricción del MHC I más frecuentes en la población latinoamericana, al ser la más afectada por *Trypanosoma cruzi*. Esto se realizó con el objetivo de identificar los alelos de restricción específicos que se emplearan en la predicción de los epítomos. La selección de los alelos se realizó en la Base de Datos de Frecuencia alélicas (<http://www.allelefrequencias.net>) que es una fuente central que proporciona, las frecuencias de alelos de restricción al MHC de diferentes áreas polimórficas en el Genoma Humano (Greenbaum et al., 2011).

## 2. Predicción de Epítomos

Para la selección de los epítomos y el cálculo de su antigenicidad e hidrofobicidad se empleó el software Immune Epitope Database and Analysis (IEDB). Ofrece una fácil búsqueda de datos experimentales que caracterizan anticuerpos y epítomos de células T estudiados en seres humanos, primates y otras especies animales. Se incluyen epítomos implicados en enfermedades infecciosas, alergias, autoinmunidad y trasplante, también alberga herramientas para ayudar en la predicción y análisis de células B y epítomos de células T.

La Base de Datos del IEDB incorpora más de 120.000 epítomos curados, la mayoría de los cuales se extraen de publicaciones científicas y, a diferencia de SYFPEITHI, incluye también una gran cantidad de datos sobre péptidos sintéticos (Vita et al., 2015). La correcta identificación de los epítomos ayudará a desarrollar una mejor vacuna y nos dará información para entender cómo *Trypanosoma cruzi* se protege contra el sistema inmune. En IEDB MHC II se busca encontrar posibles epítomos que pueden activar una respuesta en el sistema inmune. Cada proteína que cumpla con los siguientes parámetros mencionados se filtrarán de acuerdo con la puntuación que obtengan en la categoría de consensus, esto para garantizar una alta afinidad al MHC II.

## 2.1 Aplicación del método de predicción consensus

Se realizó la predicción de la totalidad de los epítomos en las 13 proteínas seleccionadas, empleando el método de consenso. La precisión de las herramientas informáticas se incrementa con los años debido al aumento en el volumen de datos, mejorando los métodos basados en *machine learning*, para mejorar las predicciones es posible utilizar múltiples predictores al realizar una predicción por consenso.

Los métodos de consenso más utilizados son CONSENSUS, que está alojado en el sitio web de la IEDB(Moutaftsi et al., 2006). Para cada MHC clase II moléculas cuya unión se puede predecir por tres o más algoritmos, ese el enfoque para generar una predicción por consenso. Se seleccionaron los tres mejores métodos que dan el mejor rendimiento.

Para cada método, los péptidos examinados se clasifican por sus puntuaciones más altas. Para cada péptido se toman los tres rangos de diferentes métodos y se calcula la mediana de los tres, esta medida se toma como la puntuación de consenso. Implementando este enfoque de predicción se obtiene un mejor desempeño general. Los resultados se obtuvieron en unidades de IC50 nM para combinatory library y SMM\_Align. Luego se realizó una clasificación según la afinidad con el MHC II. Donde, un número más bajo indica mayor afinidad. (Tabla 2).

Además, se demostró que el tamaño de los datos con el cual se entrena el software y la correcta identificación del núcleo de unión son los dos factores que limitan el desempeño de las herramientas existentes.

*Tabla 1 Valores de afinidad del epítomo*

Alta Afinidad	IC50 <50 nM
Afinidad Intermedia	IC50 <500 nM
Baja Afinidad	IC50<5000

## **2.2 Selección del núcleo o core del epítipo**

La selección del núcleo o core del epítipo se llevó a cabo por el método SMM-Align, este método predice los valores cuantitativos de afinidad de unión del péptido: MHC, por lo que es ideal para el descubrimiento de epítipos de forma objetiva. El método ha sido entrenado y evaluado según por el conjunto de datos de referencia más grande disponible al público y abarca los nueve supertipos HLA-DR, así como tres alelos de ratón H2-IA.

El método SMM-Align, es utilizado para hacer una predicción cuantitativa de unión MHC clase II, siendo una extensión del método de matriz de estabilización. El método SMM-Align busca identificar una matriz de peso que reproduzca de manera óptima los valores de IC50 medidos para cada péptido en el conjunto de entrenamiento. Los resultados de este método se han comparado con el rendimiento predictivo de cinco métodos en términos de su capacidad para predecir la afinidad de unión de más de 4600 péptidos(Nielsen, Lundegaard, & Lund, 2007).

El método permite la identificación del motifs de unión MHC clase II en términos de una matriz de peso específico para cada posición. El resultado del método SMM-Align es IC50 que une los valores de afinidad, lo que permite obtener un valor directo del péptido.

### **2.2.1 Criterios para la determinación de la Longitud de cadena del núcleo de los epítipos**

La longitud de cadena en los epítipos puede ser variable, el núcleo debe tener una longitud de 8 - 11 aminoácidos que es la longitud promedio que permite al core o núcleo del receptor del MHC II para ser reconocido por los linfocitos T y mantenerse unido a él.

### **2.2.2. Selección del núcleo de los epítipos con alta afinidad**

Se creó una matriz ordenada de los epítipos previamente seleccionados por puntuación del método consenso, se obtuvo el núcleo de cada péptido según el

método SMM, el cual ha demostrado realizar una alineación mediante este método, ya que, se obtienen los mejores resultados de predicción en MHC clase II. El método SMM predice los valores de afinidad de los péptidos cuantitativos: MHC, lo que lo hace ideal para el descubrimiento de epítomos. El método ha sido entrenado y evaluado en el conjunto de datos de referencia más grande disponible para el público y cubre los nueve tipos de super-HLA-DR(Lundegaard, Lund, Kesmir, Brunak, & Nielsen, 2007).

Se obtuvo una matriz con las repeticiones de cada núcleo en su respectiva proteína con una longitud entre 8 y 10 aminoácidos. Con el fin de encontrar los epítomos más expresados en cada proteína y descartar los epítomos con baja frecuencia. Asumimos que un núcleo con una alta expresión asegura una mejor oportunidad de MHC II para detectar esta región en la proteína y desencadenar la respuesta inmune.

### **3. Selección de epítomos**

#### **3.1 Análisis de frecuencias de epítomos**

Una vez obtenidos los resultados de los núcleos de los epítomos se prosiguió al análisis de frecuencias, en este análisis buscamos las secuencias más repetidas en cada proteína, que cumplan con un score  $IC_{50} < 25$ . Se realizó el análisis a las 13 proteínas, para posteriormente identificar las secuencias compartidas entre proteínas.

##### **3.1.1 Determinación de la Hidrofobicidad del núcleo del epítomo.**

La determinación de la hidrofobicidad se realizó empleando el programa IEDB, durante la predicción de los epítomos mediante el método de SNN. La naturaleza de las interacciones hidrofóbicas representa una característica importante del plegamiento de una proteína, de manera que esto tiene un impacto significativo en la disponibilidad de los epítomos en una respuesta inmunitaria, pues éstos deben

estar dispuestos en la superficie para generar una respuesta inmune (Freeman et al., 2010).

### **3.2 Análisis de la cobertura poblacional de los epítomos**

Un epítomo dado provocará una respuesta sólo en individuos que expresan una molécula de MHC capaz de unirse a ese epítomo particular. Este fenómeno se conoce como restricción MHC de las respuestas de células T. Para asegurar que los epítomos tuvieran una cobertura poblacional significativa, utilizamos el Cálculo de Cobertura Poblacional del Recurso de Análisis de IEDB (Bui et al., 2006).

#### 4 diagrama de flujo

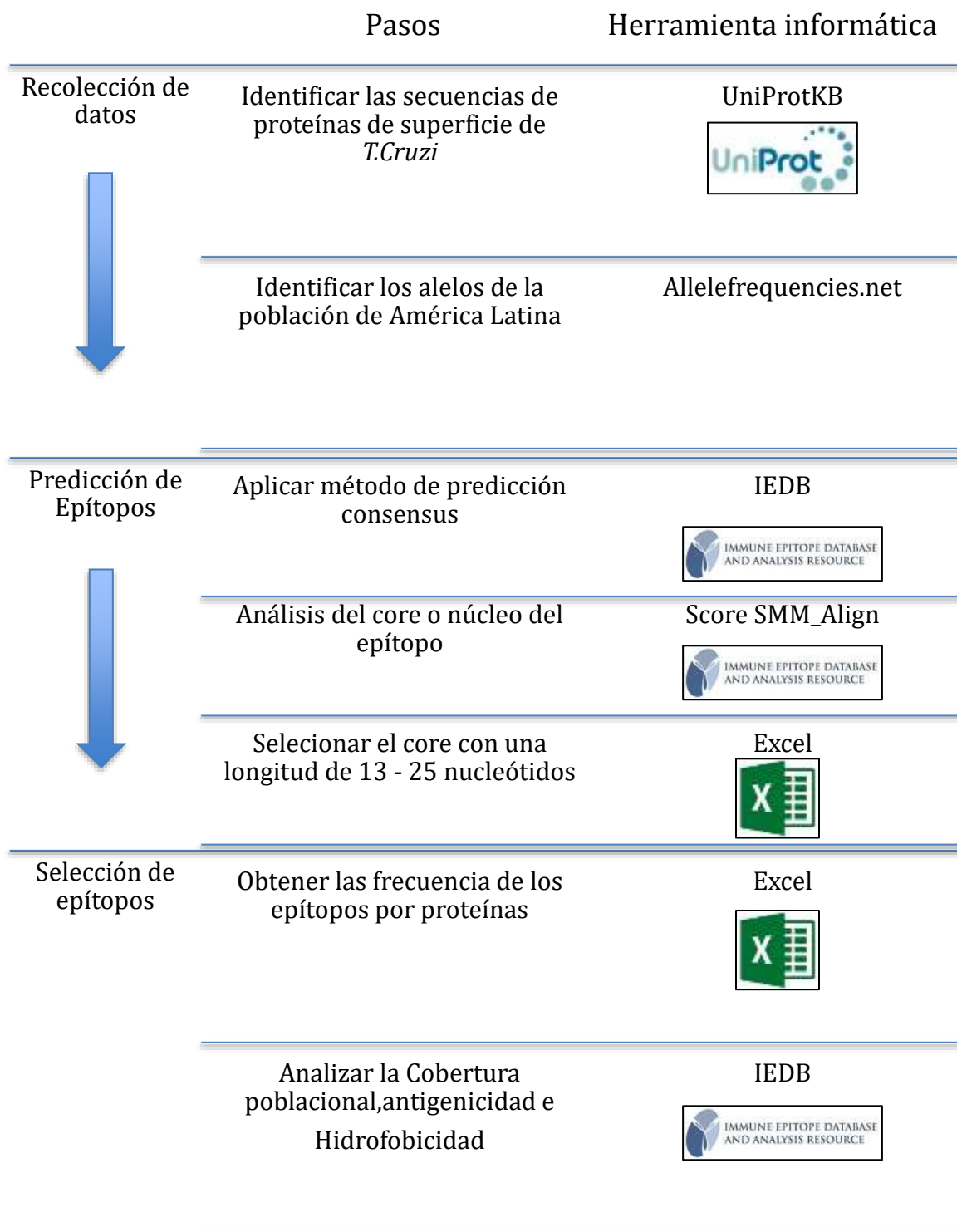


Ilustración 6 Diagrama general de la metodología empleada

### III. RESULTADOS

Con el uso de UniprotKB se encontraron 13 proteínas de superficie de *Trypanosoma cruzi* que tienen actividad de adhesión o que interactúan con las células del hospedero, su funcionamiento en la adhesión celular es un criterio importante para la identificación de los epítomos y la respuesta inmune, durante los diferentes estadios que presenta dentro del hospedero.

Se seleccionaron los alelos más frecuentes de la población latinoamericana por ser la más afectada por *Trypanosoma cruzi*; se obtuvieron los 98 haplotipos más representativos de esta región (México, centro y Suramérica). En la tabla 2 se muestran las proteínas de superficie de las diferentes etapas del *Trypanosoma* que se expresan dentro del cuerpo humano, estas se seleccionaron empleando el UniprotKB. En la fase de trypomastigotes se expresan 7 proteínas, en la de trypomastigotes metacíclicos se expresan 2 y en amastigotes 4 proteínas.

Tabla 2 Proteínas antigénicas de superficie expresadas por *Trypanosoma Cruzi*

Etapa	Proteína	Código de identificación en UniprotKB
Trypomastigotes	Gp90	Q8M369
	Tc85	O77209
	Gp85	Q03877
	Gp82	A9XBF2
	Gp83	P90604
	Ts-1	Q27793
	Cruzipain	P25779
Trypomastigotes Metacíclicos	gp63	Q27793
	SA85-1.1	P18269
Amastigotes	ASP-1	Q8MVR2
	ASP-2	P90605
	ASP-3	B5U6T8
	ASP-4	B5U6T9

En la Tabla 3 se muestran la cantidad de epítomos con mayor afinidad en cada una de las 13 proteínas obtenidas del método consensus. Los epítomos se clasificaron acorde a su puntuación ICM, en el caso del experimento se utilizaron valores menores al 25% para procurar obtener una alta afinidad al Complejo Mayor de Histocompatibilidad, se analizaron un total de 194,174 candidatos a epítomos, de los cuales sólo 66,700 cumplen con un valor de  $ICM < 25$ . En esta tabla se puede observar que el mayor número de péptidos con un  $ICM < 25$  pertenece a la proteína Cruzipain de la fase trypomastigotes con 8778 péptidos y el menor número corresponde a ASP-1 de la fase de amastigotes con 2469 péptidos.

*Tabla 3 Comportamiento de la afinidad al MCH II en linfocitos T de los epítomos seleccionados de las 13 proteínas antigénicas de T. Cruzi.*

Etapa	Proteína	Péptidos Obtenidos por el método consensus	Péptidos con $IC_{50} < 25$
Trypomastigotes	Gp90	17002	5448
	Tc85	19302	5844
	Gp85	17502	5775
	Gp82	17802	6435
	Gp83	4252	1326
	Ts-1	12677	4661
	Cruzipain	11327	8778
Trypomastigotes	gp63	13227	4620
Metacíclicos	SA85-1.1	18425	5988
Amastigotes	ASP-1	14677	2469
	ASP-2	17302	5754
	ASP-3	14077	5183
	ASP-4	16602	4419

Al analizarlos núcleos de los 66,700 péptidos obtenidos por el método de SMM\_Align, se encontraron secuencias de los núcleos con mayores frecuencias, esto se traduce en una mayor probabilidad de que los linfocitos T puedan detectar dichos epítomos por el sistema inmune. Se compararon las frecuencias de núcleos entre las proteínas para obtener las repeticiones máximas globales (Tabla 4 ,5,6).



Se obtuvieron dos grupos de proteínas que comparten núcleos similares con una alta probabilidad de ser candidatos a vacunas. El grupo de proteínas A incluyen (Asp-3, Asp-2, Gp85, Gp90, Tc85, Sa85) y el grupo B(Asp-3, Gp82, Gp83), además la proteína Cruzipain no comparte similitud con las demás proteínas, esto indica que su actividad biológica tiene un comportamiento diferente y por lo tanto se incluye en el grupo C. Dentro del grupo A se incluyen proteínas de las tres fases, en el grupo B se incluyen proteínas de las fases trypomastigote y amastigotes.

En la Tabla 4 correspondiente al grupo A, el núcleo de mayor frecuencia se repite 150 veces y se incluyen en las proteínas Gp85, Gp90, Tc85, Sa85. Por otra parte, en la Tabla 5 correspondiente al grupo B, el núcleo de mayor frecuencia se repite 59 veces y se incluyen en las proteínas Gp82 y Gp83. En la Tabla 6, son 2 núcleos con una frecuencia de 63 en la Cruzipain.

*Tabla 4 Comportamiento de la frecuencia del núcleo seleccionado en las proteínas del grupo A expresadas por Trypanosoma cruzi.*

smm_unicos	Asp-3	Asp-2	Gp85	Gp90	Tc85	Sa85	Frecuencia
NVFLYNRPL	0	19	20	24	20	20	103
YLWVTDNNR	0	15	17	18	16	16	82
LYLWVTDNN	0	7	8	7	9	8	39
VSAGGVIAA	0	0	13	12	13	13	51
SAGGVIAAF	0	0	10	11	10	10	41
LVSAGGVIA	0	0	38	35	38	40	151
VFLYNRPLN	0	0	4	6	4	4	18
SLLMIVDCE	0	0	9	9	14	15	47
VTNVFLYNR	0	9	12	0	12	12	45
TVTNVFLYN	0	12	14	0	14	14	54
LWVTDNNRS	0	0	9	10	10	10	39
VTVTNVFLY	0	26	25	0	25	25	101
TNVFLYNRP	0	5	5	0	5	5	20
SLSRLTEEL	0	0	5	5	7	8	25
PSLVSAGGV	0	5	11	0	11	4	31
GSLLMIVDC	0	0	1	2	1	1	5
QVALMLQGN	28	0	0	5	5	7	45

En la tabla 4 se representa el Heatmap la frecuencia de los núcleos de las diferentes proteínas siendo el valor 0 color rojo y verde los valores con mayor frecuencia.

Tabla 5 Comportamiento de la frecuencia del núcleo seleccionado en las proteínas del grupo B expresadas en *Trypanosoma cruzi*

smm_unicos	Asp-3	Gp82	Gp83
FLLVGSYYL	33	58	58
VALLPTAVS	19	41	41
WRAYTVFNRR	26	38	38
IHLLVGKAT	6	37	37
IEWGAPTSL	17	30	30
LVGSYYLIR	37	33	33
YYLIRDATG	11	29	29
RFGVALLPT	4	29	29
LIRDATGRR	34	27	27
TGSDVFLLV	5	26	26
RAYTVFNRE	4	26	26
FNRESSKDR	13	25	25
IVADISSDD	58	15	15
VFLLVGSYY	24	22	22
FGVALLPTA	22	22	22
LLVGKATQS	6	19	19
STGSDVFLLV	29	17	17
GVALLPTAV	16	16	16
YTVFNRESS	5	14	14
LPTAVSTGS	4	14	14
ISSDDWRAY	41	13	13
LLVGSYYLI	38	13	13
DGWDIHLLV	27	13	13
WDIHLLVGK	2	13	13
LVGKATQST	6	11	11
YLIRDATGR	22	10	10
SIVADISSD	1	5	5
KDRFGVALL	4	9	9
DIHLLVGKA	1	7	7
DDGWDIHLL	14	6	6
DRFGVALLP	1	6	6
VGSYYLIRD	1	5	5

VADISSDDW	1	4	4
RWNDDGWDI	36	4	4
GRRWNDDGW	7	4	4
WNDDGWDIH	7	2	2
AYTVFNRES	5	2	2
TVFNRESSK	2	2	2
WVVAQAALW	26	40	40
VVAQAALWR	25	25	25
KELGGTLWV	14	2	2
LWVVAQAAL	13	52	52
GGTLWVVA	13	6	6
YHGNKSLLC	13	29	29
PYHGNKSL	10	4	4
GTLWVVAQA	10	8	8
WIQAFFSKG	10	24	24
IQAFFSKGK	10	37	37
VRIVGPPYH	9	24	24
TLWVVAQA	9	17	17
PVRIVGPPY	8	9	9
AQALWRKKA	6	11	11
LGTLWVVA	5	23	23
AKQGEPIEW	4	4	4
WISQRDPVR	4	31	31
AAQALWRKK	2	2	2
MTGEWQAF	1	39	39
GEWQAFFS	1	7	7
WRKKAGWCF	1	22	22

En la tabla 5 se representan el Heatmap la frecuencia de los núcleos de las diferentes proteínas, siendo el color más opaco según incremente la frecuencia.

*Tabla 6 Frecuencia del núcleo en Cruzipain del grupo C expresada en Trypanosoma cruzi*

smm_unicos	Cruzipain
LLAAVLVV	63
LVPAATASL	63
YNDSAAVPY	49
LHAAANPHA	48
FLAGHPLTN	47
VRSSVPLNK	45
YWIIKNSWT	42
FRENFLAR	41
FRLSVFREN	41
VVGAPAAVD	41
LSVFRENLF	39
LMNNAFEWI	39
LLTSGVSA	38
FGVTPFSDL	37
FVQMSCTDA	37
LLRGSVEFF	37
LAAVLVVMA	35
LHAEETLTS	35
VLVVMACLV	35
WLAVNGPVA	34

En adición a lo anterior, el análisis de la cobertura de la población en relación con el polimorfismo del MHC se dificulta debido al hecho de que los diferentes tipos de HLA se expresan a frecuencias dramáticamente variable en diferentes grupos étnicos. Por lo tanto, sin una cuidadosa consideración, podría resultar una vacuna o un diagnóstico con una cobertura parcial de la población objeto de estudio. En nuestro estudio se observó que al realizar el análisis de cobertura de población en IEDB, la página no cuenta con los alelos totales para realizar el cálculo de forma adecuada. La ilustración 7, muestran los 14 alelos que no se encuentran dentro de la base de datos, para el cálculo de la cobertura poblacional de los epítomos seleccionados.

## Population Coverage Calculation Result

Note: Following allele(s) are not available and therefore not included in the calculation, please note that allele names are case sensitive:  
 HLA-DQA1\*05:01/DQB1\*03:01, HLA-DQA1\*01:02/DQB1\*06:02, HLA-DPA1\*01/DPB1\*04:01, HLA-DPA1\*01:03/DPB1\*02:01, HLA-DQA1\*01:01/DQB1\*05:01, HLA-DQA1\*03:01/DQB1\*03:02, HLA-DRB3\*01:01, HLA-DRB4\*01:01, HLA-DRB5\*01:01, HLA-DQA1\*04:01/DQB1\*04:02, HLA-DPA1\*03:01/DPB1\*04:02, HLA-DPA1\*02:01/DPB1\*01:01, HLA-DRB1\*07:01 HLA-DPA1\*02:01/DPB1\*05:01,

Ilustración 7 Falta de Reconocimiento de los alelos en el programa de Cobertura Poblacional

Por su parte la tabla 7, muestra los 10 epítomos que cumplen con las características necesarias para ser candidatos a vacunas recombinantes en el tratamiento contra *Trypanosoma cruzi* de la población en Latinoamérica. De estos dos pertenecen al grupo A, cuatro al grupo B y C. Destacando que los epitopes de mayor frecuencia (63) se encuentran en el grupo B.

Tabla 7 Epítomos líderes de *Trypanosoma cruzi*

	Frecuencia Promedio	Score SMM_Align(IC50<25.)	Consensus (IC50<25.)
NVFLYNRPL	20.6	0.05	0.08
YLWVTDNNR	16.4	0.01	0.01
FLLVGSYYL	49.7	0.03	0.03
VALLPTAVS	33.7	0.31	0.98
WRAYTVFNR	34.0	0.28	0.28
IHLLVGKAT	26.7	2.66	2.66
LLLA AVLVV	63	0.04	0.04
LVPAATASL	63	0.83	0.83
YNDSAAVPY	49	0.39	1.44
LHAAANPHA	48	0.58	0.42

Grupo A
  Grupo B
  Grupo C

## IV DISCUSIÓN

El ciclo de infección de *Trypanosoma cruzi* puede ser subdividida en tres etapas: adhesión y reconocimiento, señalización e invasión, este estudio se fundamenta en las 13 proteínas de superficie de que tienen actividad de adhesión, esto implica el reconocimiento de moléculas presentes en la superficie de las células del parásito y del huésped (de Souza, de Carvalho et al. 2010). Estas proteínas se encuentran clasificadas en la base de datos UniProtKB según la fase del ciclo de vida de *Trypanosoma cruzi* en la cual es expresada (Tabla 1). Como resultado de nuestro análisis estas se dividieron en 3 grupos proteínas del grupo A, B y C.

En el grupo A, el núcleo **LVSAGGVIA** tiene una frecuencia de 151 en las proteínas (Gp85, Gp90, Tc85 y Sa85). La molécula Tc85 es abundante en trypomastigotes y forma parte de la familia de trans-sialidase como la Gp85, Gp82 y TSA-1. La Gp85 es una glicoproteína que se expresan en la membrana de trypomastigotes, sin embargo, no se encuentran presentes en la forma epimastigote del parásito. Se cree que Gp85/TS participa de los procesos de adhesión a la célula hospedera e invasión debido a la interacción con múltiples ligandos como laminina, fibronectina, colágeno, citoqueratina(CK) y algunas otras moléculas de superficie y proteínas extracelulares (Tonelli, Giordano et al. 2010)

La proteína Gp90, es una glicoproteína de la familia de las trans-sialidasas y participa en los procesos de invasión celular y no activa la liberación del Ca<sup>2+</sup>. Esta es específica de la superficie de trypomastigote metacíclico y se expresa en forma diferencial en las distintas cepas de *Trypanosoma cruzi*. En otras palabras, la expresión de Gp90 se correlaciona de manera inversa con la capacidad de *Trypanosoma cruzi* para invadir células mamíferas. (Bayer-Santos, Cunha-e-Silva et al. 2013)

Las trans-sialidasas son un amplio grupo de proteínas que presentan masas moleculares desde los 60 a más de 220 kDa. Son las glicoproteínas de superficie más abundantes relacionadas con el proceso de invasión celular en *Trypanosoma cruzi*. y presentan la característica de ser glicoproteínas que se anclan a un tallo glicosilfosfatidilinositol (GPI).

En el grupo B, el núcleo **FLLVGSYYL** tiene una frecuencia de 58 en las proteínas (Asp-3, Gp82, Gp83). La Gp82 también ha sido asociada con el proceso de invasión de *Trypanosoma Cruzi* a células epiteliales de la mucosa gástrica, ya que posee la capacidad de unirse a la mucina gástrica a través de la región media del extremo carboxilo terminal de la molécula de Gp82 (Neira, Silva et al. 2003). La unión de esta proteína a la membrana plasmática del parásito ocurre a través de un anclaje GPI, y produce liberación de Ca<sup>2+</sup> y la activación de vías de señalización como otras proteínas (Gp30 y Gp 35/50).

En el grupo C, los núcleos **LLLA AVL VV** y **LVPAATASL**, tienen una frecuencia de 63 en la Cruzipain, que es una cisteína proteasa lisosomal de *Trypanosoma cruzi* que también puede encontrarse en menor medida en la membrana plasmática parasitaria. Se expresa como una mezcla compleja de isoformas en los diferentes estadios de desarrollo de todas las cepas del parásito. Esta glicoproteína participa en la internalización del *Trypanosoma cruzi* en las células mamíferas, lo que ha sido demostrado con inhibidores específicos de la enzima que interfieren en la invasión celular y la replicación del parásito (Sbaraglini, Bellera et al. 2016). Además, Cruzipain genera una fuerte respuesta inmune en individuos infectados. Estas características hacen a esta proteína un blanco potencial de drogas terapéuticas. (Salas-Sarduy, Landaburu et al. 2017)

Las moléculas de adhesión de *Trypanosoma Cruzi* han sido motivo de innumerables estudios realizados por diversos grupos de investigadores, no sólo por importancia que implica conocer sobre su funcionamiento, sino también como una posible alternativa de blanco terapéutico para una enfermedad que actualmente no posee un tratamiento que elimine de forma efectiva al parásito.

La adhesión a las células hospederas es el primer paso para la invasión celular (Burleigh 2005, Maeda, Cortez et al. 2012). Una vez que el trypomastigote metacíclico ingresa al hospedero mamífero, invade primariamente a células presentes en el sitio de inoculación como macrófagos, fibroblastos y células

epiteliales, en un proceso que involucra la activación de diferentes moléculas de superficie tanto en las células del hospedero mamífero como en las del propio parásito (de Souza, de Carvalho et al. 2010).

Posteriormente ocurre la activación de vías de señalización específicas que determinan la internalización del parásito a través la formación de una vesícula endocítica, conocida con el nombre de vesícula parasitófora. En esta etapa es vital para *Trypanosoma cruzi* que las formas trypomastigotes logren sobrevivir a un estado altamente oxidativo que se genera al interior de dicha vesícula. Finalmente, los parásitos escapan hacia el citoplasma de la célula hospedera, estableciendo una infección intracelular.

Los 10 epitopes seleccionados a los cuales se les determinó la cobertura con los haplotipos de Latinoamérica, fueron recopilados en [www.allelefrequencias.net](http://www.allelefrequencias.net), estos datos son de vital importancia para el descubrimiento de nuevas vacunas, ya que, gracias a ellos es posible determinar nuevos epítomos para poblaciones específicas como en este estudio, en el cual se procuró abarcar a la población de américa latina para la identificación de epítomos. En este análisis un obstáculo se presenta de manera inobjetable: la frecuencia de los haplotipos es, por lo general, baja, lo que dificulta los análisis estadísticos pertinentes.

Sin embargo, una gran ventaja de la página [www.allelefrequencias.net](http://www.allelefrequencias.net) es la cantidad de información que recaba, y un claro ejemplo de su utilidad se encuentra en IEDB. La predicción de epítomos usa métodos matemáticos que permiten hacer frente a los distintos tipos de Haplotipos, este reconocimiento de las diferentes variables alélicas, nos permite desarrollar vacunas con un enfoque poblacional específico, como en el caso de los 98 haplotipos que se obtuvieron a partir de la base de datos de esta página y permitió realizar la predicción sin ningún inconveniente, obteniendo así 194,174 candidatos a epítomos que están enfocados para la población de América latina.

La secuencia del alelo HLA es muy diversa y pertenece a seis diferentes clases (A, B, C, E, F, G), un total de 11406 alelos han sido identificado a partir de octubre de

2016 (lanzamiento de la base de datos IPDIMGT / HLA3.26.0). Las moléculas de MHC de clase I humana (HLA-A, HLA-B, HLA-C) son altamente polimórficas.

Presentan péptidos antigénicos al TCR expresado por CTL. El polimorfismo HLA es el resultado de la selección natural para lograr patógenos específicos inmunidad. El HLA altamente diverso, en el genoma humano juega un papel importante en la interacción huésped-patógeno al mediar en respuestas inmunes celulares innatas y adaptativas (Archbold, Macdonald et al. 2009). Los alelos HLA se han asociado con la gravedad, la persistencia variada de los resultados de la enfermedad, la aparición y la transmisión de varias enfermedades infecciosas.

En contraposición el Análisis de cobertura global al no contar en su base de datos con los haplotipos necesarios no permitió corroborar los datos, dado que es una base de datos que no recaba datos de forma libre, su actualización depende de la frecuencia que decida el programador. A pesar de que esto el estudio poblacional que se desarrolló con la selección de los haplotipos de latino américa garantiza la unión con el MHC II, debido a que el programa de predicción de epítomos si considera los haplotipos faltantes.

El método de predicción consensus en conjunto con smm align demostró ser en combinación una forma práctica de realizar un análisis global de las proteínas de *Trypanosoma cruzi*, ya que es más efectivo y tiene una mayor confiabilidad en comparación a los métodos convencionales. El beneficio que aporta el método consensus se debe a que toma en cuenta tres rangos de diferentes métodos, lo cual disminuye los errores individuales de cálculo de cada uno (Moutaftsi et al., 2006).

El método smm align al aplicarlo de forma subsecuente permite analizar el core de cada epítopo pre seleccionado, lo cual le otorga un score mayor en lugar de analizarlo de forma individual (Nielsen, Lundegaard, & Lund, 2007). La ventaja de la utilización de este método es la integración de un parámetro importante que es la hidrofobicidad, este parámetro fue tomado en cuenta en este método al calcularlo con el software del programa IEDB, por lo tanto no es necesario realizar un análisis posterior de hidrofobicidad.



Los datos reflejan una relación entre las actividades que desempeñan las proteínas y sus secuencias de epítomos, es decir las funciones que cumplen están ligadas a sus secuencias. Esta relación generó los 3 grupos de proteínas. El grupo integrado por las proteínas Asp-3, Gp83 y Gp8 que presentó la mayor frecuencia de repeticiones de epítomos y un nivel alto de afinidad al MHC II, a consecuencia de esto proponemos este grupo como mejor candidato para una vacuna.

Los epítomos de las proteínas Asp-3Asp-2Gp85Gp90Tc85Sa85 tienen una frecuencia menor al grupo A, a pesar de abarcar un mayor número de proteínas, esto suponemos repercutirá en un reconocimiento más ineficiente por parte del sistema inmune, por ello, se requiere un análisis conformacional de todas las proteínas para poder considerar este criterio con mayor fuerza, ya que, seguramente será un factor determinante en la selección de epítomos en un estudio posterior.

Finalmente, Cruzipain tiene un papel de gran importancia en el proceso de invasión del parásito hacia el hospedero, ya que es la proteína encargada de regular la señalización entre ambos, a pesar del rol que desempeña, los resultados de este estudio muestran que la frecuencia de sus epítomos no es suficiente en comparación a las de las otras proteínas, por lo tanto, su efecto como vacuna requeriría más estudios para definir su utilización de forma efectiva.

El presente estudio demostró una propuesta para analizar por el método de consenso las proteínas de superficie con el fin de identificar posibles candidatos a vacunas, al aplicar la metodología propuesta se obtuvieron 10 epítomos como candidatos a vacunas que abarcan la población global de América latina, a pesar de ello el estudio no toma en cuenta el plegamiento de la proteína una vez sintetizada, lo cual repercute en que el sitio de unión pueda ser accesible para su reconocimiento por parte del MHC II, el aporte esencial es dar a conocer los métodos bioinformáticos disponibles y el ahorro económico que representa contra los métodos tradicionales en el desarrollo de vacunas.

## V. CONCLUSIONES

1. En este estudio se seleccionaron 10 epítomos con mayor antigenicidad de las 13 las proteínas de superficie de *Trypanosoma cruzi*. Tomando en consideración la frecuencia de repetición en ellas del núcleo de la secuencia
2. Se identificaron 98 haplotipos representativos de América Latina que interactúan con el MCH I, donde su cobertura poblacional era de más del 80%
3. Las propiedades antigénicas e hidrofílicas de los epítomos líderes se realizaron por el método consensus y se tomaron en consideración para la selección de los 10 epitopes.
4. Los métodos consensus y SMM\_Alignincluidos en la herramienta bioinformática IEDB nos permitió determinar las secuencias de los epítomos líderes de las proteínas de superficie en *Trypanosoma cruzi*,
5. . La combinación del método consensus para identificación de epítomos favorece una mejor predicción y en conjunto al método SMM\_align que analiza las propiedades hidrofílicas nos permiten obtener un resultado certero para su posterior análisis in vitro.

## Perspectivas

- Aplicar la misma metodología en distintos patógenos para identificar los grupos de proteínas que son más susceptibles a usarse como futuras vacunas recombinantes
- Evaluar la factibilidad de las proteínas candidatos para su síntesis.
- Evaluar las proteínas en su forma cuaternaria para contrastar los datos obtenidos.
- Corroborar los datos de forma in vitro por medio de análisis inmunológicos.
- Realizar el mismo procedimiento para haplotipos en ratones para su posterior uso en humanos.

## VI BIBLIOGRAFÍA

Abbas, A. K., Lichtman, A. H. H., & Pillai, S. (2015). *Inmunología celular olecular + StudentConsult*. Elsevier Health Sciences Spain. Retrieved from <https://books.google.com.mx/books?id=kzcTBwAAQBAJ>

Bermudez, J., Davies, C., Simonazzi, A., Real, J. P., & Palma, S. (2016). Current drug therapy and pharmaceutical challenges for Chagas disease. *Acta Trop*, 156, 1–16. <https://doi.org/10.1016/j.actatropica.2015.12.017>

Bern, C., & Montgomery, S. P. (2009). An estimate of the burden of Chagas disease in the United States. *Clin Infect Dis*, 49(5), e52-4. <https://doi.org/10.1086/605091>

Bordner, A. J. (2010). Towards universal structure-based prediction of class II MHC epitopes for diverse allotypes. *PLoS One*, 5(12), e14383. <https://doi.org/10.1371/journal.pone.0014383>

Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*, 7, 153. <https://doi.org/10.1186/1471-2105-7-153>

Chen, J., Liu, H., Yang, J., & Chou, K. C. (2007). Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids*, 33(3), 423–428. <https://doi.org/10.1007/s00726-006-0485-9>

Coura, J. R. (2007). Chagas disease: what is known and what is needed--a background article. *Mem Inst Oswaldo Cruz*, 102 Suppl, 113–122.

Coura, J. R. (2013). Chagas disease: control, elimination and eradication. Is it possible? *Mem Inst Oswaldo Cruz*, 108(8), 962–967. <https://doi.org/10.1590/0074-0276130565>

Coura, J. R., & Dias, J. C. (2009). Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. *Mem Inst Oswaldo Cruz*, 104 Suppl, 31–40. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/19753455>

Getzoff, E. D., Tainer, J. A., Lerner, R. A., & Geysen, H. M. (1988). The Chemistry and Mechanism of Antibody Binding to Protein Antigens. In F. J. Dixon (Ed.), *Advances in Immunology* (Vol. 43, pp. 1–98). Academic Press. [https://doi.org/https://doi.org/10.1016/S0065-2776\(08\)60363-6](https://doi.org/https://doi.org/10.1016/S0065-2776(08)60363-6)

Gonzalez-Galarza, F. F., Christmas, S., Middleton, D., & Jones, A. R. (2011). Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res*, 39(Database issue), D913-9. <https://doi.org/10.1093/nar/gkq1128>

Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., & Sette, A. (2011).

Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics*, 63(6), 325–335. <https://doi.org/10.1007/s00251-011-0513-0>

Hotez, P. J., Dumonteil, E., Heffernan, M. J., & Bottazzi, M. E. (2013). Innovation for the “bottom 100 million”: eliminating neglected tropical diseases in the Americas. *Adv Exp Med Biol*, 764, 1–12. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23654053>

Huang, J., & Honda, W. (2006). CED: a conformational epitope database. *BMC Immunol*, 7, 7. <https://doi.org/10.1186/1471-2172-7-7>

Lee, B. Y., Bacon, K. M., Bottazzi, M. E., & Hotez, P. J. (2013). Global economic burden of Chagas disease: a computational simulation model. *Lancet Infect Dis*, 13(4), 342–348. [https://doi.org/10.1016/s1473-3099\(13\)70002-1](https://doi.org/10.1016/s1473-3099(13)70002-1)

Lundegaard, C., Lund, O., Kesmir, C., Brunak, S., & Nielsen, M. (2007). Modeling the adaptive immune system: predictions and simulations. *Bioinformatics*, 23(24), 3265–3275. <https://doi.org/10.1093/bioinformatics/btm471>

Lundegaard, C., Lund, O., & Nielsen, M. (2012). Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy? *Expert Rev Vaccines*, 11(1), 43–54. <https://doi.org/10.1586/erv.11.160>

Machado, F. S., Dutra, W. O., Esper, L., Gollob, K. J., Teixeira, M. M., Factor, S. M., ... Garg, N. J. (2012). Current understanding of immunity to *Trypanosoma cruzi* infection and pathogenesis of Chagas disease. *Semin Immunopathol*, 34(6), 753–770. <https://doi.org/10.1007/s00281-012-0351-7>

Moutaftsi, M., Peters, B., Pasquetto, V., Tschärke, D. C., Sidney, J., Bui, H.-H., ... Sette, A. (2006). A consensus epitope prediction approach identifies the breadth of murine TCD8<sup>+</sup>-cell responses to vaccinia virus. *Nat Biotech*, 24(7), 817–819. [https://doi.org/http://www.nature.com/nbt/journal/v24/n7/suppinf/nbt1215\\_S1.html](https://doi.org/http://www.nature.com/nbt/journal/v24/n7/suppinf/nbt1215_S1.html)

Nielsen, M., Lundegaard, C., & Lund, O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*, 8, 238. <https://doi.org/10.1186/1471-2105-8-238>

Potocnakova, L., Bhide, M., & Pulzova, L. B. (2016). An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *J Immunol Res*, 2016, 6760830. <https://doi.org/10.1155/2016/6760830>

Rubinstein, N. D., Mayrose, I., & Pupko, T. (2009). A machine-learning approach for predicting B-cell epitopes. *Mol Immunol*, 46(5), 840–847. <https://doi.org/10.1016/j.molimm.2008.09.009>

Schmunis, G. A. (2007). Epidemiology of Chagas disease in non-endemic

countries: the role of international migration. *Mem Inst Oswaldo Cruz*, 102 Suppl, 75–85.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W. M., ... Chisari, F. V. (1994). The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol*, 153(12), 5586–5592.

Stanaway, J. D., & Roth, G. (2015). The Burden of Chagas Disease: Estimates and Challenges. *Global Heart*, 10(3), 139–144.  
<https://doi.org/https://doi.org/10.1016/j.gheart.2015.06.001>

Toussaint, N. C., Maman, Y., Kohlbacher, O., & Louzoun, Y. (2011). Universal peptide vaccines - optimal peptide vaccine design based on viral sequence conservation. *Vaccine*, 29(47), 8745–8753.  
<https://doi.org/10.1016/j.vaccine.2011.07.132>

UniProt: the universal protein knowledgebase. (2017). *Nucleic Acids Research*, 45(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>

Vandiedonck, C., & Knight, J. C. (2009). The human Major Histocompatibility Complex as a paradigm in genomics research. *Brief Funct Genomic Proteomic*, 8(5), 379–394. <https://doi.org/10.1093/bfgp/elp010>

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., ... Peters, B. (2015). The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*, 43(Database issue), D405–12. <https://doi.org/10.1093/nar/gku938>

Archbold, J. K., W. A. Macdonald, S. Gras, L. K. Ely, J. J. Miles, M. J. Bell, R. M. Brennan, T. Beddoe, M. C. Wilce, C. S. Clements, A. W. Purcell, J. McCluskey, S. R. Burrows and J. Rossjohn (2009). "Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition." *J Exp Med*206(1): 209-219.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.

Burleigh, B. A. (2005). "Host cell signaling and *Trypanosoma cruzi* invasion: do all roads lead to lysosomes?" *Sci STKE*2005(293): pe36.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." *Int J Cell Biol*2010.

Maeda, F. Y., C. Cortez and N. Yoshida (2012). "Cell signaling during *Trypanosoma cruzi* invasion." *Front Immunol*3: 361.

- Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." *Infect Immun*71(1): 557-561.
- Salas-Sarduy, E., L. U. Landaburu, J. X. Karpiak, K. P. Madauss, J. J. Cazzulo, F. Agüero and V. E. Alvarez (2017). "Novel scaffolds for inhibition of Cruzipain identified from high-throughput screening of anti-kinetoplastid chemical boxes." *Sci Rep*7(1): 12073.
- Sbaraglini, M. L., C. L. Bellera, L. Fraccaroli, L. Larocca, C. Carrillo, A. Talevi and C. D. Alba Soto (2016). "Novel Cruzipain inhibitors for the chemotherapy of chronic Chagas disease." *Int J Antimicrob Agents*48(1): 91-95.
- Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.
- Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.
- Burleigh, B. A. (2005). "Host cell signaling and *Trypanosoma cruzi* invasion: do all roads lead to lysosomes?" *Sci STKE*2005(293): pe36.
- de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." *Int J Cell Biol*2010.
- Maeda, F. Y., C. Cortez and N. Yoshida (2012). "Cell signaling during *Trypanosoma cruzi* invasion." *Front Immunol*3: 361.
- Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." *Infect Immun*71(1): 557-561.
- Salas-Sarduy, E., L. U. Landaburu, J. X. Karpiak, K. P. Madauss, J. J. Cazzulo, F. Agüero and V. E. Alvarez (2017). "Novel scaffolds for inhibition of Cruzipain identified from high-throughput screening of anti-kinetoplastid chemical boxes." *Sci Rep*7(1): 12073.
- Sbaraglini, M. L., C. L. Bellera, L. Fraccaroli, L. Larocca, C. Carrillo, A. Talevi and C. D. Alba Soto (2016). "Novel Cruzipain inhibitors for the chemotherapy of chronic Chagas disease." *Int J Antimicrob Agents*48(1): 91-95.
- Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.

Burleigh, B. A. (2005). "Host cell signaling and *Trypanosoma cruzi* invasion: do all roads lead to lysosomes?" *Sci STKE*2005(293): pe36.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." *Int J Cell Biol*2010.

Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." *Infect Immun*71(1): 557-561.

Salas-Sarduy, E., L. U. Landaburu, J. X. Karpiak, K. P. Madauss, J. J. Cazzulo, F. Agüero and V. E. Alvarez (2017). "Novel scaffolds for inhibition of Cruzipain identified from high-throughput screening of anti-kinetoplastid chemical boxes." *Sci Rep*7(1): 12073.

Sbaraglini, M. L., C. L. Bellera, L. Fraccaroli, L. Larocca, C. Carrillo, A. Talevi and C. D. Alba Soto (2016). "Novel Cruzipain inhibitors for the chemotherapy of chronic Chagas disease." *Int J Antimicrob Agents*48(1): 91-95.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." *Int J Cell Biol*2010.

Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." *Infect Immun*71(1): 557-561.

Salas-Sarduy, E., L. U. Landaburu, J. X. Karpiak, K. P. Madauss, J. J. Cazzulo, F. Agüero and V. E. Alvarez (2017). "Novel scaffolds for inhibition of Cruzipain identified from high-throughput screening of anti-kinetoplastid chemical boxes." *Sci Rep*7(1): 12073.

Sbaraglini, M. L., C. L. Bellera, L. Fraccaroli, L. Larocca, C. Carrillo, A. Talevi and C. D. Alba Soto (2016). "Novel Cruzipain inhibitors for the chemotherapy of chronic Chagas disease." *Int J Antimicrob Agents*48(1): 91-95.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a

conserved peptide motif to the vasculature in vivo." PLoS Negl Trop Dis4(11): e864.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." Parasit Vectors6: 127.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." Int J Cell Biol2010.

Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." Infect Immun71(1): 557-561.

Salas-Sarduy, E., L. U. Landaburu, J. X. Karpiak, K. P. Madauss, J. J. Cazzulo, F. Agüero and V. E. Alvarez (2017). "Novel scaffolds for inhibition of Cruzipain identified from high-throughput screening of anti-kinetoplastid chemical boxes." Sci Rep7(1): 12073.

Sbraglini, M. L., C. L. Bellera, L. Fraccaroli, L. Larocca, C. Carrillo, A. Talevi and C. D. Alba Soto (2016). "Novel Cruzipain inhibitors for the chemotherapy of chronic Chagas disease." Int J Antimicrob Agents48(1): 91-95.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." PLoS Negl Trop Dis4(11): e864.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." Parasit Vectors6: 127.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." Int J Cell Biol2010.

Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." Infect Immun71(1): 557-561.

Salas-Sarduy, E., L. U. Landaburu, J. X. Karpiak, K. P. Madauss, J. J. Cazzulo, F. Agüero and V. E. Alvarez (2017). "Novel scaffolds for inhibition of Cruzipain identified from high-throughput screening of anti-kinetoplastid chemical boxes." Sci Rep7(1): 12073.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." PLoS Negl Trop Dis4(11): e864.



Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." *Int J Cell Biol*2010.

Neira, I., F. A. Silva, M. Cortez and N. Yoshida (2003). "Involvement of *Trypanosoma cruzi* metacyclic trypomastigote surface molecule gp82 in adhesion to gastric mucin and invasion of epithelial cells." *Infect Immun*71(1): 557-561.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.

de Souza, W., T. M. de Carvalho and E. S. Barrias (2010). "Review on *Trypanosoma cruzi*: Host Cell Interaction." *Int J Cell Biol*2010.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.

Bayer-Santos, E., N. L. Cunha-e-Silva, N. Yoshida and J. Franco da Silveira (2013). "Expression and cellular trafficking of GP82 and GP90 glycoproteins during *Trypanosoma cruzi* metacyclogenesis." *Parasit Vectors*6: 127.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.

Tonelli, R. R., R. J. Giordano, E. M. Barbu, A. C. Torrecilhas, G. S. Kobayashi, R. R. Langley, W. Arap, R. Pasqualini, W. Colli and M. J. Alves (2010). "Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo." *PLoS Negl Trop Dis*4(11): e864.