# UNIVERSIDADE DE SÃO PAULO INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

# ANDRÉ AZEVEDO REIS TEIXEIRA

# Reconhecimento molecular na Doença de Chagas do ponto de vista do parasita e do hospedeiro

Versão corrigida da Tese

São Paulo

Data do Depósito na SPG: 03/10/2017

# ANDRÉ AZEVEDO REIS TEIXEIRA

# Reconhecimento molecular na Doença de Chagas do ponto de vista do parasita e do hospedeiro

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em Ciências Biológicas (Bioquímica)

Orientador (a): Prof. Dr. Ricardo José Giordano Coorientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Júlia Manso Alves

São Paulo 2017

### Ficha Catalográfica Elaborada pela Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

T266r

Teixeira, André Azevedo Reis Reconhecimento molecular na Doença de Chagas do ponto de vista do parasita e do hospedeiro / André Azevedo Reis Teixeira. - São Paulo, 2017.

117 p.

Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.

Orientador: Giordano, Ricardo Jose Coorientador: Alves, Maria Julia Manso

1. doença de Chagas. 2. Trypanosoma cruzi. 3. trans-sialidase. 4. phage diplay. 5. mapeamento de epítopos. I. T. II. Giordano, Ricardo Jose, orientador. III. Alves, Maria Julia Manso, coorientador.



## **Agradecimentos**

Apesar da competição ser um importante impulso em nosso progresso, a colaboração é que nos permite seguir em frente mesmo nos momentos mais difíceis. Este trabalho não seria possível sem o esforço, paciência e ajuda de muitas pessoas, às quais seguem meus sinceros agradecimentos:

Ao meu orientador Prof. Ricardo Giordano pelos ensinamentos, liderança, paciência e principalmente pela liberdade conferida para a realização dos projetos e amadurecimento científico. A minha coorientadora Prof<sup>a</sup>. Maria Júlia por dividir sua experiência e sabedoria na forma de tantas conversas e conselhos.

A toda equipe dos laboratórios (e agregados) onde o trabalho foi realizado, que fizeram os dias, dentro e fora do laboratório, mais felizes: Carlos, Fenny, Leila, Veronica, Lilian, Heloise, Luís, Luiza, Alexandre, Jussara, Laura, Renato, Chrislaine, Ana Zen, Dimitrius e outros que por ali passaram.

A equipe técnica: Celinha, Robertinho e Maria Luiza, que não só contribuíram com seu suporte, mas também com sua amizade ao longo destes anos.

Aos Prof. Walter Colli e Ivan Schumacher por dividirem seu conhecimento e sabedoria em muitas conversas sobre uma infinidade de assuntos: ciência, história, política, comida, carreira, e muitos outros.

A Fundação de Amparo à Pesquisa do Estado de São Paulo por acreditar no projeto e prover financiamento (processos 2012/13967-4 e 2016/06069-0), sabendo que isso só é possível devido a população do estado de São Paulo, cujo trabalho permite que continuemos nossas investigações.

To Renata Pasqualini and Wadih Arap for the opportunity to be part of their team and for the innumerable lessons I learned with them and with the other members of the lab: Christy, Amanda, Fernanda, Andrey, Bruce, Sara, Fortunato, Serena, Tracey, Daniela, Virginia, Marina. Of course, I cannot forget to thank Sara and Fortu for introducing me to the best burrito in the world on my birthday.

A família que encontrei em São Paulo: Nara, Carlos Eduardo e Pingo pelo acolhimento, carinho e suporte, transformando uma cidade tão fria em lugar muito melhor.

Aos meus pais Ana Isabel e José Luiz e minha irmã Bruna, por serem meus exemplos, pelo suporte incondicional e contínuo ao longo dos últimos 29 anos, cuja importância é impossível ser superestimada e sem o qual nada teria sido possível. A minha vó Lucy pelos muitos momentos alegres vividos ao seu lado. A todos meus primos, primas, tios e tias, que apesar da distância, sempre foram fontes de alegria, amizade e companheirismo em todos os momentos.

A minha namorada, melhor amiga, confidente, conselheira e companheira Camila, por estar comigo em todos momentos (bons e ruins), por todos os conselhos (errados e certos), todas as decisões tomadas juntas (melhores e piores) e pela oportunidade de compartilhar tudo isso.

	"– Permite-me, porém, uma objeção: tu que não possuis nada, que é que pretende dar?
	– Cada um dá o que tem. O guerreiro dá a sua força; o comerciante, a sua mercadoria; o mestre, a sua doutrina; o pescador, os seus peixes.
-	– Ótimo. E qual será o bem que tu poderás oferecer? Que aprendeste? Que sabes fazer?
-	– Sei pensar. Sei esperar. Sei jejuar."
-	- Só isso?
-	– Acho que é só isso."
	–Hermann Hesse, Siddhartha

### Resumo

TEIXEIRA, A.A. R. Reconhecimento molecular na Doença de Chagas do ponto de vista do parasita e do hospedeiro. 2017. Tese de Doutorado. Programa de Pós-Graduação em Ciências Biológicas (Bioquímica). Instituto de Química, Universidade de São Paulo, São Paulo.

A doença de Chagas, causada pelo parasita protozoário Trypanosoma cruzi, afeta milhões de pessoas, a maioria delas vivendo na América latina. Apesar dos avanços da medicina e da biotecnologia, ainda existem poucas opções de tratamento para indivíduos com a doença. Assim, é importante compreendermos os detalhes moleculares da infecção parasitária, para que novas alternativas terapêuticas e de diagnóstico possam ser desenvolvidas para esses pacientes. Neste trabalho estudamos esta doença em duas frentes, uma "do ponto de vista" do parasita, e a outra, da resposta do hospedeiro. Utilizando bioinformática, identificamos um peptídeo conservado (denominado TS9) presente nas proteínas de superfície gp85/transsialidases do parasita. Este peptídeo é capaz de promover adesão celular e, na sua forma sintética, inibe a entrada do *T. cruzi* na célula hospedeira. Análise da estrutura proteica revelou que o peptídeo TS9 encontra-se num domínio do tipo laminina-G, lado-a-lado com o peptídeo FLY, outro peptídeo conservado desta grande família, previamente descrito pelo nosso grupo. Juntos, eles formam um sítio de adesão a citoqueratinas e proteínas de filamento intermediário. Na segunda parte, investigamos os antígenos e epítopos reconhecidos pelas imunoglobulinas de pacientes portadores da doença nas suas diferentes formas clínicas: assintomática e cardiomiopatias, leve ou grave. Criamos uma biblioteca de phage display contendo, virtualmente, todos os fragmentos proteicos existentes no *T. cruzi*, que foi varrida contra imunoglobulinas para a construção de um mapa da resposta humoral dos pacientes com a doença de Chagas. Nossos resultados mostram que a resposta dos pacientes é complexa, e mais de dois mil epítopos foram mapeados. Muitos deles, como os antígenos B13, SAPA e FRA já foram previamente descritos, validando nosso método. Porém, um grande número de novos epítopos, inclusive contra proteína descritas como hipotéticas ou sem função conhecida, também foram encontrados. Seus papéis na infecção e resposta imune da doença merecem, portanto, atenção. Em resumo, as abordagens e técnicas utilizadas nesta tese são inovadoras, e permitiram a identificação de peptídeos e moléculas que poderão ser úteis para o desenvolvimento de novos métodos diagnósticos e terapêuticos para a doença de Chagas.

Palavras-chave: doença de Chagas, Trypanosoma cruzi, trans-sialidase, mapeamento de epítopos, phage diplay, IgG

### **Abstract**

TEIXEIRA, A.A. R. Molecular recognition in Chagas disease from the point of view of the parasite and the host. 2017. PhD Thesis - Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Chagas disease, caused by the protozoan parasite *Trypanosoma cruzi*, affects millions of people, most of them living in Latin America. Despite advances in medicine and biotechnology, there are still few treatment options for individuals with the disease. Thus, it is important to understand the molecular details of the parasitic infection, so that new therapeutic and diagnostic alternatives can be developed for these patients. In this work, we study this disease in two fronts, one "from the point of view" of the parasite, and the other, of the response of the host. Using bioinformatics, we identified a conserved peptide (called TS9) present in the surface proteins gp85 / trans-sialidases of the parasite. This peptide is capable of promoting cell adhesion and, in its synthetic form, inhibits the entry of *T. cruzi* into the host cell. Analysis of the protein structure revealed that the TS9 peptide is in a laminin-G-like domain, side-by-side with the peptide FLY, another conserved peptide of this large family, previously described by our group. Together, they form an adhesion site to cytokeratins and intermediate filament proteins. In the second part, we investigated the antigens and epitopes recognized by the immunoglobulins of patients with the disease in their different clinical forms: asymptomatic and cardiomyopathies, mild or severe. We created a phage display library containing virtually all existing protein fragments in T. cruzi. This library was screened against immunoglobulins for the construction of a humoral response map of patients with Chagas' disease. Our results show that the response of the patients is complex, and more than 2,000 epitopes have been mapped. Many of them, such as the B13, SAPA and FRA antigens have been previously described, validating our method. However, a large number of new epitopes, including many against proteins described as hypothetical or with no known function, were also found. Their roles in infection and immune response of the disease deserve, therefore, attention. In summary, the approaches and techniques used in this thesis are innovative and have allowed the identification of new peptides and molecules that may be useful for the development of new diagnostic and therapeutic methods for Chagas' disease.

Keywords: Chagas disease, *Trypanosoma cruzi*, trans-sialidase, epitope mapping, phage diplay, IgG

## Lista de abreviaturas e siglas

BRASIL Do inglês, biopanning and rapid analysis of selective interactive

ligands

BSA albumina de soro bovino (do inglês, *bovine serum albumin*)

CCC cardiomiopatia chagásica crônica

DAPI 4',6-diamidino-2-fenilindol

DMSO dimetilsulfóxido

LB meio de cultura lisogênico (do inglês, lysogeny broth)

IgG imunoglobulina gamma

kDa kilodaltons

mAb anticorpo monoclonal (do inglês, *monoclonal antibody*)

MEM meio mínimo essencial (do inglês, *minimal essential medium*)

Ni-NTA níquel-nitrilotriacético (do inglês, *nickel-nitrilotriacetic*)

ORF passo de leitura aberto (do inglês, *open reading frame*)

pb Pares de base

pIII Proteína III do bacteriófago M13

pVIII Proteína VIII do bacteriófago M13

PBS salina tamponada com fosfato (do inglês, phosphate buffered

saline)

PCR reação de polimerase em cadeia (do inglês, polymerase chain

reaction)

qPCR reação de polimerase em cadeia quantitativa (do inglês, quantitative

polymerase chain reaction)

SDS dodecil sulfato de sódio (do inglês, *sodium dodecyl sulphate*)

# Sumário

Agradecimentos	4	
Resumo	6	
Abstract		
Lista de abreviaturas e siglas	8	
Sumário	9	
Introdução: A doença de Chagas e o <i>Trypanosoma cruzi</i>		
Objetivos		
Capítulo 1	15	
Introdução	16	
Material e Métodos	18	
Resultados	25	
Discussão	42	
Conclusão	46	
Capítulo 2	47	
Introdução	48	
Material e Métodos	54	
Resultados	59	
Discussão	95	
Conclusão	99	
Conclusão da tese		
Referências Bibliográficas		
Anexo – Súmula curricular		

# Introdução: A doença de Chagas e o *Trypanosoma* cruzi

A doença de Chagas, também conhecida como tripanossomíase americana, é uma doença tropical causada pelo parasita *Trypanosoma cruzi*. A doença que talvez seja a mais negligenciada entre as chamadas "doenças negligenciadas" (THE LANCET, 2006) foi descoberta no início do século XX devido ao trabalho do médico mineiro Carlos Chagas (CHAGAS, 1909d). Enquanto trabalhava no norte do estado de Minas Gerais em uma campanha para o controle da malária, Chagas estudava a fauna da região (KROPF; SÁ, 2009). Foi então que identificou uma nova espécie de protozoário do gênero tripanosoma no sangue de macacos saguis (*Callithrix penicillata*) e, também interessado em entomologia, identificou o protozoário flagelado no interior de insetos hematófagos da região conhecidos como barbeiros (*Triatoma infestans*) (CHAGAS, 1909c), que tinham esse nome em decorrência de frequentemente picarem as pessoa na face. O primeiro caso a ser registrado da doença foi encontrado em uma garota de dois anos que sofria de febre e anemia (CHAGAS, 1909a, 1909b).

Tradicionalmente, a transmissão do parasita para mamíferos se dá pelas fezes do inseto barbeiro, que ao se alimentar de sangue expele o parasita que irá infectar o indivíduo. Apesar do *T. cruzi* não conseguir atravessar pele intacta para estabelecer a infecção, a ferida deixada pelo barbeiro ao se alimentar é utilizada como via de acesso para o interior do organismo. Centenas de mamíferos, incluindo animais domésticos e selvagens como cachorros, gatos, roedores, tatu, etc., podem servir como reservatórios naturais do parasita, mantendo um ciclo que independe de hospedeiros humanos (TEIXEIRA et al., 2011). Dessa maneira, ainda que fosse possível a vacinação da população e erradicação do inseto transmissor em regiões habitadas, a eliminação do parasita da natureza não seria viável.

Devido à forma de transmissão, a doença se limitou por muito tempo à América do Sul e Central, onde há ocorrência natural do vetor em contato com populações humanas, havendo evidência da existência de infecção do parasita em humanos da

região há pelo menos nove mil anos (AUFDERHEIDE et al., 2004). No entanto, devido a intensificação de fluxos migratórios ao redor do mundo e a procedimentos de transfusão de sangue e transplante de órgãos infectados, a doença hoje se espalha por quase todos os continentes (COURA; VIÑAS, 2010). Recentemente, casos de infecção oral da doença tem sido reportados devido à ingestão de alimentos contaminados como caldo de cana (BASTOS et al., 2010), açaí (PEREIRA et al., 2009; SOUZA-LIMA et al., 2013) e suco de goiaba (ALARCÓN DE NOYA et al., 2010). Estima-se que pelo menos 8 milhões de pessoas estão infectadas com o parasita e que mais de 10 mil mortes podem ser atribuídas à doença todos os anos ("WHO | What is Chagas disease?", [s.d.]). Ainda, a doença resulta em um impacto econômico global estimado de 7,19 bilhões de dólares por ano, devido aos custo de tratamento de pacientes e a perdas de produtividade associadas à morbidade (LEE et al., 2013).

A progressão clínica da doença de Chagas se divide em duas fases distintas: aguda e crônica. A fase aguda se dá nas primeiras semanas após a infecção, havendo alta parasitemia devido aos ciclos de replicação e ausência de uma resposta imune efetiva por parte do hospedeiro neste período. Alguns casos podem ser assintomáticos, no entanto é comum a presença de sintomas não específicos da doença como malestar generalizado, febre, dores musculares, dores de cabeça, dores nas articulações, vômito e anorexia (LARANJA et al., 1956). Na maioria dos casos, os sintomas são passageiros e não necessitam intervenção para que cessem. No entanto, em crianças com menos de dois anos e indivíduos imunocomprometidos, (casos de transplante de órgãos ou presença de co-morbidades como HIV) é comum a ocorrência de meningoencefalite seguida de morte ainda nesta fase da doença (CÓRDOVA et al., 2010; DIAZGRANADOS et al., 2009).

Após a fase aguda, inicia-se a fase crônica da doença de Chagas, onde os níveis de *T. cruzi* no sangue são mais baixos que na fase anterior. Cerca de dois terços dos indivíduos infectados permanecerão no que é chamado de forma indeterminada, não apresentando sintomas nem outras complicações relacionadas a doença pelo

resto de suas vidas. Já outro um terço dos infectados evoluirão para a chamada forma determinada, desenvolvendo problemas cardíacos e/ou digestivos. Menos frequente, a forma digestiva se desenvolve em 10-15% dos pacientes (RASSI JR; RASSI; MARINNETO, 17), e se manifesta na forma do aumento do cólon (megacólon) e/ou esôfago (megaesôfago). Problemas como malnutrição e constipação são comuns nesses casos devido à perda de função de parte do sistema digestivo. Essas alterações ocorrem devido à destruição dos nervos motores destes órgãos e à inflamação crônica na região (MENEGHELLI, 1985). Curiosamente, as megassíndromes são comuns na região central do Brasil e outras regiões mais ao sul, porém bastante raras no norte da América do Sul e na América Central onde a doença também é endêmica (MILES; FELICIANGELI; DE ARIAS, 2003), sendo possível que variações genéticas no parasita possam levar a diferentes manifestações clínicas.

Aforma cardíaca da doença de Chagas ocorre em 20-30% dos pacientes na fase crônica e também é conhecida como cardiomiopatia chagásica crônica. A inflamação persistente no miocárdio leva a fibrose, e o aumento das fibras de colágeno no interstício causa perda da função contrátil do órgão (ROSSI, 1991). Essas alterações levam ao aparecimento de sintomas como arritmias (ELIZARI; CHIALE, 1993), insuficiência cardíaca (MADY et al., 1994), tromboembolismo (SAMUEL et al., 1983) e morte súbita (BESTETTI et al., 1996; RASSI JR; RASSI; RASSI, 2001). Uma característica comum nestes pacientes é o aumento do coração (cardiomegalia). Acredita-se que tanto a persistência do parasita nos tecidos quanto o desenvolvimento de uma resposta imune contra autoantígenos são fatores que levam aos sintomas característicos da fase crônica. No entanto, o peso de cada um deste componentes ainda é disputado (BONNEY; ENGMAN, 2015; CUNHA-NETO et al., 1995, 2006; GIRONÈS; FRESNO, 2003; TARLETON, 2003; TEIXEIRA et al., 2011).

As opções de tratamento para pacientes com doença de Chagas ainda estão limitadas basicamente a duas drogas: benzinidazol e nifurtimox. Ambas as drogas apresentam elevada toxicidade, com importantes efeitos colaterais que podem

ser debilitantes para a saúde dos pacientes, que em 30% dos casos tem que interromper o tratamento (VIOTTI et al., 2009). Estes fármacos são mais eficazes se administrados durante a fase aguda da doença (MORENO et al., 2010), no entanto, como indivíduos infectados são frequentemente assintomáticos ou experimentam apenas sintomas inespecíficos durante os estágios iniciais, o diagnóstico correto da doença de Chagas é difícil até que seja tarde demais para o tratamento. Além destas limitações, existem diversos casos documentados de cepas de *T. cruzi* resistentes aos fármacos atualmente disponíveis (FILARDI; BRENER, 1987; MORENO et al., 2010). Dessa maneira, é urgente o desenvolvimento de novas alternativas terapêuticas para a doença de Chagas.

O fenômeno de reconhecimento molecular está presente virtualmente em todos os processos biológicos, inclusive na patogênese da doença de Chagas. Durante a infecção o parasita utiliza moléculas em sua superfície para que possa reconhecer e se ancorar em receptores presentes nas células do hospedeiro que serão invadidas. Sem a presença e função destas proteínas de superfície, o parasita não conseguiria continuar seu ciclo de vida no hospedeiro vertebrado, fazendo com que esse processo seja fundamental para sua sobrevivência. Ao mesmo tempo, o hospedeiro utiliza seu sistema imune para combater a infecção. Dentre as moléculas especializadas do sistema imune, as imunoglobulinas são produzidas com função de reconhecer o parasita com alta afinidade e especificidade e neutralizá-lo.

Na primeira parte do trabalho, estudamos a presença de motivos peptídicos conservados presentes na família de proteínas de superfície do parasita, as gp85/ trans-sialidase. Essas proteínas são importantes no processo de adesão celular, ou seja, no reconhecimento da célula hospedeira pelo parasita, e utiliza sequências peptídeos conservados para se ligar a citoqueratinas. Na segunda parte, mapeamos os epítopos e antígenos reconhecidos pelo sistema imune adaptativo humoral de humanos durante a infecção por *T. cruzi*. Construímos uma biblioteca de *phage diplay* utilizando o material genético do parasita e selecionamos os fagos reativos com o soro

dos pacientes chagásicos, seguido de sequenciamento de larga escala. A abordagem permitiu identificar milhares de sequências proteicas envolvidas no reconhecimento molecular do parasita pelo hospedeiro.

## **Objetivos**

Utilizar metodologias combinatoriais e computacionais para estudar aspectos do reconhecimento molecular presentes na interação do parasita com a célula hospedeira, e da resposta imune humoral de pacientes em diferentes formas clínicas da doença. Para isso vamos:

- Identificar regiões conservadas nas sequências das gp85/trans-sialidases que possam estar envolvidas no processo de adesão do parasita a célula hospedeira, assim como seu receptor e efeito na invasão das células hospedeiras pelo parasita.
- Construir uma biblioteca de phage display para identificar epítopos e antígenos do T. cruzi que são reconhecidos pelas imunoglobulinas do hospedeiro humano na doença de Chagas, em sua forma assintomática e com envolvimento cardíaco.

# Capítulo 1<sup>1</sup>.

Trypanosoma cruzi liga-se à citoqueratinas através de peptídeos conservados presentes no domínio laminina G de gp85/trans-sialidases

Figuras e parte do texto deste capítulo são reproduzidos do artigo Teixeira AAR, de Vasconcelos VdCS, Colli W, Alves MJM, Giordano RJ (2015) Trypanosoma cruzi Binds to Cytokeratin through Conserved Peptide Motifs Found in the Laminin-G-Like Domain of the gp85/Trans-sialidase Proteins. PLoS Negl Trop Dis 9(9): e0004099. https://doi.org/10.1371/journal.pntd.0004099

### Introdução

Para a continuidade do seu ciclo de vida dentro do hospedeiro vertebrado, o *T. cruzi* obrigatoriamente precisa invadir células para que possa se reproduzir e se proteger do sistema imune. O parasita é capaz de invadir célula fagocíticas e não-fagocíticas, empregando moléculas presentes em sua superfície para aderir a estas células e iniciar a invasão (ALVES; COLLI, 2007). Após se ancorar à membrana da célula hospedeira, há a formação do vacúolo parasitóforo, ou pela fusão de lisossomos com a membrana celular ou pela formação de uma invaginação na membrana e posterior fusão com lisossomos. O parasita secreta a proteína TcTox para escapar do vacúolo, tendo acesso ao citoplasma da célula (ANDRADE; ANDREWS, 2005) para se diferenciar na forma amastigota e iniciar sua replicação.

Diversas proteínas e receptores já foram descobertos no processo de adesão do parasita a célula hospedeira. Aparentemente, o *T. cruzi* tem a capacidade de explorar diferentes estratégias de reconhecimento molecular para que possa iniciar o processo de invasão, aumentando suas probabilidades de sucesso na infecção. O nosso grupo tem estudado uma grande família de glicoproteínas de superfície implicada na adesão celular e invasão por *T. cruzi* (ALVES et al., 1986; COLLI, 1993; GIORDANO et al., 1999). Esta molécula foi inicialmente identificada como uma proteína de superfície de 85 kDa expressa seletivamente pelas formas infecciosas do parasita e denominada Tc-85 (ABUIN et al., 1989; ALVES et al., 1986). Hoje sabemos que pertence a uma grande família de proteínas codificadas por múltiplos genes, coletivamente conhecidos como gp85/trans-sialidase (gp85/TS) multigene família (COLLI, 1993; CROSS; TAKLE, 1993; EL-SAYED, 2005; MATTOS et al., 2014). Com centenas de genes no genoma, todas as proteínas gp85/TS compartilham um domínio de sialidase com motivos de *ASP-box* (sequência SxDxG) e o motivo peptídico VTVxNVxLYNRPLN (CROSS; TAKLE, 1993).

Com base na similaridade de sequências, as proteínas codificadas por esta família de genes foram ainda subdivididas em oito grupos (FREITAS et al., 2011). As

proteínas pertencentes ao grupo I são as únicas proteínas que possuem domínios trans-sialidase enzimaticamente ativos, podendo remover resíduos de ácido siálico das glicoproteínas presentes na superfície celular dos mamíferos e transferí-los para glicoproteínas na membrana celular do parasita. Esta atividade de trans-sialidase é essencial para a invasão da célula hospedeira pelo *T. cruzi* (PEREIRA et al., 1996).

As proteínas Tc-85 pertencem ao grupo II da família gp85/TS não têm atividade enzimática e, portanto, não podem transferir ácido siálico, porém têm sido implicadas na adesão celular e invasão de células parasitas (MATTOS et al., 2014). Elas interagem com diferentes receptores presentes na matriz extracelular e superfície celular, como a laminina (GIORDANO et al., 1994, 1999; MARROQUIN-QUELOPANA et al., 2004), proteínas de filamentos intermediários como citoqueratina e vimentina (MAGDESIAN et al., 2001; TONELLI et al., 2010), fibronectina (OUAISSI et al., 1986), mucina (STAQUICINI et al., 2010) e receptor de procinetinina-2 (KHUSAL et al., 2015). Em estudos anteriores, foi mostrado que o peptídeo VTVTNVFLYNRPLN (denominado FLY, e derivado do motivo VTVxNVxLYNRPLN) presente na Tc-85 é um sítio de ligação de citoqueratina importante para a adesão e invasão de células pelos parasitas (MAGDESIAN et al., 2001; TONELLI et al., 2010). A ligação de Tc-85 à citoqueratina ativa a cascata de sinalização ERK1/2, resultando em um aumento no número de parasitas por célula (MAGDESIAN et al., 2007). Também, utilizando o phage display como um substituto para o peptídeo, foi mostrado que o peptídeo FLY pode contribuir para o tropismo de tecidos dirigindo o parasita para leitos vasculares específicos (TONELLI et al., 2010). Tomados em conjunto, esses estudos sugerem um papel importante para citoqueratinas e o peptídeo FLY presente em proteínas da família gp85/TS na adesão e invasão de células parasitas.

Aqui, expandimos estes estudos para explorar outros motivos peptídicos comuns encontrados nas proteínas gp85/TS, a fim de verificar o seu possível papel na adesão e invasão de células pelos parasitas. Por similaridade de sequência de proteínas pertencentes ao grupo II, identificamos vários peptídeos conservados

entre os membros desta família. Utilizando *phage display*, estes peptídeos foram expressos em bacteriófago e analisados quanto à sua capacidade para se ligar às células de mamífero. Os resultados mostraram que um dos peptídeos selecionados foi capaz de promover a adesão do fago a células de mamífero e, na forma sintética, inibir a invasão dessas pelo parasita. Curiosamente, este peptídeo também se liga à citoqueratina e vimentina, e é encontrado na estrutura tridimensional da proteína lado a lado com o peptídeo FLY, que compreende uma folha β anti-paralela que faz parte do domínio do tipo Laminina-G (LamG), presente nos membros da família gp85/TS. Por conseguinte, os resultados aqui apresentados fornecem outras evidências que suportam a afirmação de que o *T. cruzi* interage com proteínas do filamento intermediário, sendo o domínio LamG, dos membros da família gp85/TS, um elemento importante na adesão e invasão de células pelos parasitas.

### **Material e Métodos**

Cultura de células e de *T. cruzi*. Células epiteliais de *Macaca mulata* da linhagem de LLC-MK2 (ATCC CCL-7) (macaco rhesus) foram utilizadas para o cultivo do *T. cruzi* e os ensaios de infecção. As células foram mantidas em meio mínimo essencial (MEM) suplementado com 10% de soro fetal bovino (Life Technologies). A cepa Y do parasita, utilizada nos presentes experimentos, foi cultivada e mantida conforme descrito (ANDREWS; COLLI, 1982).

**Peptídeos sintéticos**. Os peptídeos foram sintetizados de acordo com as nossas especificações com um mínimo de 95% de pureza pela empresa Chinese Peptide Company (Hangzhou, China). Neste estudo, os peptídeos sintéticos utilizados foram: TS9 (FTLVASVTI) e *scramble* (LTVIFATVS). O peptídeo sintético FLY utilizado para os ensaios de competição possui a sequência original (GKKPSVTVTNVFLYNRPLN). descrita por MAGDESIAN et al. (2001).

Alinhamentos múltiplo de sequências e identificação de motivos conservados. A sequência primária das proteínas pertencentes a todos os grupos, I a VIII, da família

das gp85/trans-sialidases foram obtidas no banco de dados TriTrypDB (ASLETT et al., 2010) utilizando listas de genes publicadas por FREITAS et al. (2011). As sequências pertencentes ao grupo II das gp85/trans-sialidases foram então alinhadas utilizando o software "Clustal Ômega" (SIEVERS et al., 2011). Uma tabela contendo a frequência de cada um dos 20 aminoácidos em cada posição do alinhamento foi gerada utilizando o software UGENE (OKONECHNIKOV; GO LOSOVA; FURSOV, 2012). Para realizar a seleção dos peptídeos conservados nas proteínas gp85/trans-sialidases, foram calculadas as médias móveis com janelas de 6 a 19 posições de aminoácidos, utilizando-se a frequência do aminoácido mais prevalente na dada posição. Imagens das sequências do tipo "logo" foram geradas usando "WebLogo" (CROOKS et al., 2004). Para confirmar se os motivos conservados selecionados, TS1 a TS10, estavam também presentes em proteínas pertencentes a outros grupos, as sequências pertencentes a todos os grupos de gp85/trans-sialidases foram alinhadas utilizando "Clustal Ômega" e as regiões equivalentes aos peptídeos selecionados foram identificadas.

Clonagem dos fagos. Para a inserção das sequências exógenas nos fagos, oligonucleotídeos sintéticos (Exxtend, Brasil) codificando ambas as fitas de cada peptídeo selecionado foram clonados no vector fUSE55 (SCOTT; SMITH, 1990) e transformados por eletroporação em células de bactéria *Escherichia coli* (cepa MC1061). As células transformadas foram cultivadas durante a noite em meio LB/ágar contendo o antibiótico tetraciclina para a seleção de clones contendo o plasmídeo de interesse. Clones foram expandidos em meio LB líquido e o DNA plasmideal foi extraído e sequenciado para verificar a presença da sequência correta do inserto. Os fagos produzidos por esses clones foram utilizados para infectas a bactéria *E. coli* cepa K91kan para produção em maior escala de fagos. Todos os fagos foram sequenciados novamente para confirmar a sequência de inserto e titulados por contagem de unidades transdutoras (TU) e PCR em tempo real (método qPhage) (DIAS-NETO et al., 2009).

Ensaio de ligação dos fagos à células. O método BRASIL foi utilizado para ensaios

de ligação dos fagos às células (GIORDANO et al., 2001). As células da linhagem LLC-MK2 foram tratadas rapidamente com solução de tripsina e EDTA para que se soltassem da superfície de cultura e ressuspendidas em meio completo. As células foram incubadas por duas horas à 37°C e 5% CO<sub>2</sub> para recuperação e, posteriormente, lavadas e ressuspendidas em meio sem soro contendo 1% BSA (albumina de soro bovino) para serem incubadas individualmente com os diferentes clones de fagos produzidos (10º células para 10¹¹ partículas virais) por três horas em gelo. Para ensaios de competição com peptídeos sintético, as células e os fagos foram incubados na ausência ou presença dos peptídeos em diferentes concentrações. Após a incubação, a mistura foi centrifugada sobre fase oleosa (dibutilftalato:ciclohexano 9:1 vol/vol) para separar os fagos ligados às células dos não ligados. O DNA foi então purificado e a quantificação de fagos ligados foram feitas por PCR em tempo real (DIAS-NETO et al., 2009).

Ensaio de ligação dos fagos às proteínas. Os ensaios de ligação de fagos às proteínas foram realizados como descrito por TONELLI et al., (2010). As citoqueratinas humanas recombinantes (Cell Sciences, vimentina bovina (Sigma-Aldrich Co) e gelatina de pele suína (Sigma-Aldrich Co) foram imobilizadas em placas de poliestireno de 96 poços de fundo plano (200 ng de proteína em 50 μl de solução salina tamponada com fosfato [PBS, "phosphate buffered saline"]) durante a noite a 4°C. Os poços foram lavados três vezes com PBS e bloqueados com solução de PBS 1% BSA por uma hora a temperatura ambiente. Os poços foram lavados novamente e os fagos, 10° TU diluídas em 50 μl de tampão de bloqueio, foram adicionados. Após duas horas de incubação a temperatura ambiente, os poços foram lavados 10 vezes com PBS e os fagos ligados foram então recuperados por infecção bacteriana e quantificados por contagem de colónias.

Cromatografia de afinidade. O peptídeo sintético TS9 (1 mg) foi conjugado covalentemente à uma matriz de agarose utilizando o kit CarboxyLink (Thermo Scientific) de acordo com as recomendações do fabricante. Um extrato proteico foi produzido

utilizando células LLC-MK2 cultivadas como descrito anteriormente. As células foram homogeneizadas utilizando solução de PBS 2% Nonidet P-40, centrifugadas a 10.000g durante 30 minutos e o sobrenadante recuperado e incubado com a resina durante a noite a 4°C. Após lavagem com 20 volumes de PBS e 5 volumes de solução de NaCl 1 M, a resina foi eluída sequencialmente com solução de PBS 1% SDS (primeira eluição) e 8M uréia (segunda eluição). As proteínas eluídas foram analisadas por SDS-PAGE e coradas com azul brilhante de Coomassie G250 (Sigma Aldrich) para visualização. As bandas de proteínas com massas moleculares calculadas de 49, 53, 59 e 66 kDa foram excisadas do gel, cortadas em pequenos pedaços, tratadas com ditiotreitol (DTT; 10 mM), iodoacetamida (100 mM) e depois digeridas com tripsina (5 μg/ml; Promega, Wl, EUA). Os peptídeos resultantes foram então analisados por HPLC-ESI-MS na Central Analítica do Instituto de Química da Universidade de São Paulo (SP). Os dados de espectrometria de massas foram analisados utilizando o software MASCOT (KOENIG et al., 2008) para identificação das proteínas correspondentes.

Western-blot. O extrato proteico de da célula LLC-MK2 foi separado SDS-PAGE e transferido para membranas de nitrocelulose. As membranas foram bloqueadas durante uma hora à temperatura ambiente com tampão de bloqueio Odyssey (LI-COR Biosciences) e incubadas com anticorpo monoclonal (mAb) anti-citoqueratina 18 (1 μg/ml) (Invitrogen, Clone DC-10). Após a incubação, a membrana foi lavada por cinco minutos com solução de PBS 0,1% Tween-20 por três vezes. A membrana foi então bloqueada com o anticorpo secundário conjugado com IRDye800 (LI-COR Biosciences). O desenvolvimento de fluorescência foi analisado utilizando-se o equipamento Odyssey Imaging System (LI-COR Biosciences).

Ensaio de citotoxicidade dos peptídeos. Células LLC-MK2 (5 x 10³) foram cultivadas em 100 μl de MEM contendo 10% de soro fetal bovino em placa de poliestireno de 96 poços. Vinte e quatro horas após semear as células, o meio foi substituído por MEM 2% soro fetal bovino contendo 1% DMSO, peptídeo TS9 ou peptídeo *scramble* (200 μM). As células foram então incubadas durante duas ou 48 horas. A viabilidade

celular após as incubações foi determinada utilizando-se o reagente WST-1 de acordo com as recomendações do fabricante (Roche). O procedimento também foi realizado utilizando-se tripomastigotos de cultura (3 x 10<sup>6</sup> por poço) em 100 µl de MEM 2% soro fetal bovino contendo 1% de DMSO e os peptídeos TS9 ou *scramble* (200 µM).

Ensaio de invasão celular com T. cruzi. Células LLC-MK2 (3 x 10<sup>4</sup> por poço) foram semeadas em placas de poliestireno de 24 poços contendo lamínulas de vidro de 13 mm de diâmetro 24 horas antes do experimento. Para o ensaio de modulação de invasão pelos peptídeos, concentrações crescentes de solução destes peptídeos diluídos em MEM contendo 2% soro fetal bovino 1% de DMSO foram pré-incubadas com a célula hospedeira por 15 minutos. Após esse período foram adicionados tripomastigotos derivados de cultura de células (3 x 106; multiplicidade de infecção igual a 100) às células. Após duas horas de infecção, as células foram lavadas dez vezes com MEM 2% soro fetal bovino para remover os tripomastigotos que não foram internalizados. As células foram cultivadas por mais 24 ou 48 horas antes da análise. O mesmo procedimento foi realizado para o ensaio de adesão celular, com exceção da temperatura de incubação das células com os tripomastigotos que foi modificada para 4°C e após as lavagens as células foram imediatamente fixadas para análise. No experimento utilizando a proteína recombinante LamG<sup>Tc85-11</sup>, as células foram incubadas com a proteína na concentração de 6,5 µM 15 minutos antes da infecção e então os tripomastigotos foram adicionados e incubados por duas horas a 37°C. As células foram lavadas e cultivadas por 24 horas. Ao fim de cada experimento, as amostras foram fixadas utilizando solução de PBS 4% paraformaldeído e depois coradas com DAPI e fotografadas com um microscópio de epifluorescência Nikon utilizando uma lente objetiva de 20x de aumento. O número de células totais e células infectadas foi quantificado em pelo menos 4 campos diferentes para cada tratamento e o número de parasitas por célula infectada foi quantificado em pelo menos 100 células infectadas.

**Modelagem molecular da estrutura da proteína Tc85-11.** Asequência de aminoácidos da proteína Tc85-11 foi obtida no banco de dados GenBank (código de identificação

AAD13347.1) e os domínios conservados presentes foram identificados utilizando o software "NCBI Conserved Domain Search" (CD-Search) (MARCHLER-BAUER et al., 2011). Um modelo estrutural tridimensional foi gerado utilizando o servidor "Iterative Threading ASSEmbly Refinement" (I-TASSER) em http://zhanglab.ccmb.med.umich. edu/I-TASSER/ (ROY; KUCUKURAL; ZHANG, 2010). A estrutura da proteína transsialidase de Trypanosoma rangeli (PDB id: 1WCS) foi utilizada como guia para gerar o modelo. A melhor estrutura gerada teve um "TM-score" igual a 0,50±0,15. Como referência, um "TM-score" maior que 0,5 indica um modelo de topologia correta (independente do comprimento da proteína) e um "TM-score" menor que 0,17 indica uma estrutura de semelhança aleatória (ZHANG; SKOLNICK, 2004).

Clonagem e expressão de domínio Tc85-11<sup>LamG</sup>. O DNA codificando o domínio LamG da proteína Tc85-11 (código de acesso no GenBank: AAD13347, aminoácidos 475 a 677) foi amplificado por PCR utilizando primers específicos (fita GACCATATGCTGGTGGCGGTATTGTCAAAC; fita senso: anti-senso GACGTCGACATTCAGTGGGCGGTTGTACAG) e clonado no vetor pET21 (Novagen) utilizando as enzimas de restrição Ndel e Sall (Fermentas | Thermo Fisher Scientific). O plasmídeo foi sequenciado para confirmar a presença de inserto e ausência de mutações. Para a produção da proteína recombinante LamG<sup>Tc85-11</sup>, a bactéria *E.* coli cepa Rosetta(DE3)pLysS contendo o plasmídeo de expressão foi cultivada em meio LB até atingir densidade ótica de 0,7 em 600 nm. Neste ponto, a expressão da proteína foi induzida pela adição de 0,8 mM de isopropil-p-D-1-tiogalactopiranosídeo (IPTG) durante 4h a 37°C. As células foram recuperadas por centrifugação e rompidas em tampão denaturante (100 mM Na, HPO, 10 mM Tris e 8M ureia, pH 8,0). A proteína recombinante purificada, que possui seis histidinas no seu carboxiterminal foi purificada utilizando resina de Ni-NTA (Qiagen). A resina foi lavada com tampão de lavagem denaturante (100 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM Tris e 8M ureia, pH 6,3) e a proteína ligada foi eluída com tampão de eluição denaturante (100 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM Tris e 8M ureia, pH 3,5). O pH das frações de eluição foram ajustados com uma solução tampão de 1M Tris pH 8,0. A proteína recombinante foi renaturada por diálise contra

PBS contendo concentrações decrescentes de ureia (4, 2, 1 e 0,5 M de ureia durante um período de 2h cada). A diálise final foi realizada contra PBS durante a noite.

Ensaio de ligação a proteínas. As proteínas (citoqueratinas, vimentina, gelatina e BSA) foram imobilizadas em placas de poliestireno de 96 poços (200 ng de proteínas em 50 μl de PBS; 4°C, durante a noite), bloqueadas com tampão de bloqueio Odyssey durante uma hora a 37°C e posteriormente incubadas com proteína recombinante LamG<sup>To85-11</sup> diluída em PBS (60 μg/ml) por uma hora. Os poços foram então lavados três vezes com PBS e a proteína ligada foi detectada utilizando anticorpo monoclonal anti-gp85 G1G8, gentilmente fornecido pela Dr. Eliciane Mattos, seguido por anticorpo secundário anti-lgG conjugado com IRDye680 (LI-COR Biosciences). As placas foram então analisadas utilizando o equipamento "Odyssey Imaging System" (LI-COR Biosciences). O anticorpo monoclonal G1G8 reconhece um epítopo neutralizante (sequência TGETPLEPFGFCFGA) dentro do domínio LamG da proteína Tc85-11 (GIORDANO et al., 1999). Para os ensaios de competição com o peptídeo sintético, a proteína recombinante LamG<sup>To85-11</sup> foi incubada na presença e ausência dos peptídeos FLY, TS9 e *scramble* (200 μM).

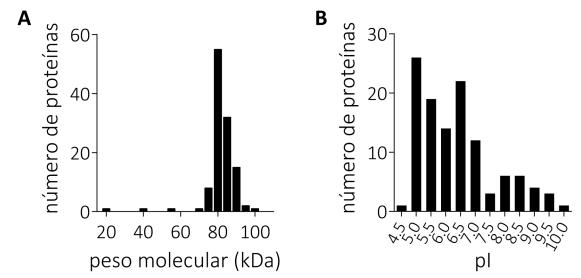


Figura 1. (**A**) Histograma representando a distribuição de peso molecular das 117 proteínas pertencentes ao grupo II das gp85/trans-sialidases. (**B**) Histograma representando a distribuição de ponto isoelétrico, calculado a partir do pK<sub>a</sub> da cadeia lateral dos aminoácidos, das 117 proteínas pertencentes ao grupo II das gp85/trans-sialidases.

### Resultados

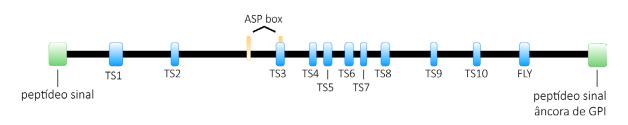
Identificação de peptídeos conservados em proteínas do grupo II da família das gp85/trans-sialidases.

Para identificar os motivos peptídicos conservados compartilhados pelos membros do grupo II da família gp85/TS, nós alinhamos as 117 sequências de proteínas pertencentes a esta família. As sequências individuais foram obtidas a partir da base de dados TryTryp (ASLETT et al., 2010) e consideradas como pertencentes ao grupo II da família gp85/TS de acordo com a classificação previamente descrita (FREITAS et al., 2011). Como esperado, a análise das sequências primárias mostrou que a maioria das proteínas do grupo II têm massa molecular variando de 80 a 90 kDa (mediana= 82; mínimo = 22; máximo = 102) e pontos isoelétrico ácidos (mediana= 6,2; mínimo = 4,7; máximo = 9,8) (Figura 1), há também a presença dos motivos ASP-box

Tabela 1. Peptídeos derivados de gp85/trans-sialidases selecionados para serem apresentados nos fagos.

Nome	Sequência	
TS1a	SPSLVSAGGVIAAFAE	
TS1b	FVSPSLVSAGGV	
TS2	KGNKVFLLAGS	
TS3	VYESRDMGTTWT	
TS4	RKVMLYTQRG	
TS5	LYLWVTDNNRSF	
TS6	LLYSDGNLHLLQ	
TS7	SLSRLTEEL	
TS8	SIPTAGLVAVLS	
TS9	FTLVASVTI	
TS10	KEYQVALMLQ	
FLY	VTVTNVFLYNRPLN	

e VTVxNVxLYNRPLN, também conhecido como FLY, comuns aos membros da família (Figura 2A). Para visualizar e quantificar os resultados do nosso alinhamento, geramos uma tabela contendo a frequência dos diferentes aminoácidos em cada posição no alinhamento. Em seguida, foi calculada a média móvel do aminoácido mais prevalente em cada posição (ex.: se na posição 105 o aminoácido mais comum é a alanina, que aparece em 70% das proteínas, o número 70 é utilizado para representar a conservação desta posição). Isto foi feito para minimizar as flutuações e para facilitar a visualização de regiões conservadas nos membros da família gp85/trans-sialidase (Figura 2B). Escolhemos uma média móvel de 9 aminoácidos porque aproxima o comprimento do peptídeo que pode ser exibido pelo bacteriófago. Além disso, análises semelhantes realizadas com janelas de 7 até 12 aminoácidos produziram perfis semelhantes.



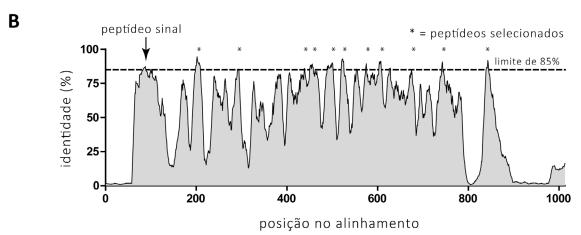


Figura 2. (**A**) Representação geral das proteínas gp85/trans-sialidases pertencentes ao grupo II. Estão indicados os peptídeos de sinal para a localização do retículo endoplasmático (ER), âncora de glicofosfatidilinositol (GPI) (caixas verdes) e as repetições típicas de sialidase chamadas *ASP-box* (caixas laranja) assim como os peptídeos conservados identificados neste estudo (caixas azuis), que incluem o motivo VTVxNVxLYNRLN denominado FLY. (**B**) Média móvel de nove aminoácidos para cada posição no alinhamento das proteínas do grupo II das gp85/trans-sialidases. O peptídeo sinal para a localização ER (seta) e os 10 peptídeos mais conservados (\*) estão indicados.

Com base nestes resultados, estabelecemos um limite mínimo de corte de 85% de identidade para selecionar os dez peptídeos mais conservados. Os peptídeos selecionados variaram de 9 a 19 aminoácidos de comprimento com identidades médias variando de 85,5% a 92,1% (Figura 2C). Os peptídeos selecionados foram denominados TS1 a TS10 (Tabela 1 e Figura 3). É importante notar que quase todas as proteínas analisadas têm um peptídeo sinal amino-terminal, que não é encontrado na proteína madura, e, portanto, estas regiões não foram consideradas para a seleção de peptídeos (Figura 2B seta).

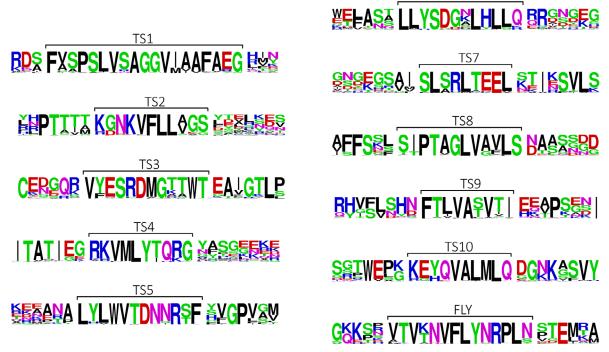


Figura 3. Representação no formato de *weblogo* dos peptídeos derivados das gp85/transsialidases (indicados pelas chaves). O tamanho da letra indica a conservação de aminoácidos em cada posição e a cor indica propriedades físico-químicas: aminoácidos polares (verde), hidrofóbicos (pretos), carregados positivamente (azul) ou carregados negativamente (vermelhos).

### O peptídeo TS9 liga-se a células de mamífero

Tendo identificado motivos peptídicos comuns presentes em todos os membros da família das gp85/TS, procuramos em seguida determinar se eles se ligavam a células de mamífero. Para isso, foi empregada a técnica de *phage display*, que usa um vírus bacteriófago como um substituto para o peptídeo (GIORDANO et al., 2001; TONELLI et al., 2010). Neste método, o genoma do fago é modificado por engenharia

genética para permitir a produção de partículas virais que possuem os peptídeos exógenos, derivados de gp85/TS, fusionados à proteína III (pIII) do capsídeo viral (Figura 4). A ligação dos fagos à célula foi avaliada utilizando o método BRASIL, que permite a separação de fagos ligados às células dos não-ligados em um único passo de centrifugação para posterior quantificação (GIORDANO et al., 2001).

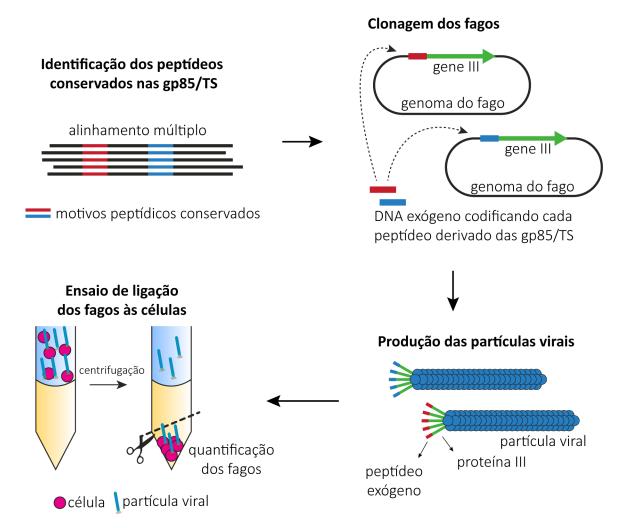


Figura 4. Fluxograma resumindo o processo para seleção *in silico*, apresentação em bacteriófago e o ensaio funcional *in vitro* (método BRASIL) para a identificação de peptídeos conservados derivados da família das gp85/trans-sialidases com propriedades de ligação celular.

Como é conhecido, a pIII está envolvido na infecção da bactéria pelo bacteriófago e é frequentemente inativado por peptídeos com mais de 16 aminoácidos. Para resolver este problema, o peptídeo TS1 foi dividido em dois peptídeos mais curtos denominados TS1a e TS1b. Estes peptídeos tem 16 e 12 aminoácidos de extensão,

respectivamente, e continham uma sobreposição de 10 aminoácidos em sequência para aumentar a probabilidade de preservar um potencial sítio de ligação celular presente no peptídeo TS1 original. Dessa maneira, todos os peptídeos selecionados foram exibidos com sucesso no bacteriófago.

Em seguida, utilizamos a linhagem de células epiteliais LLC-MK2 para o ensaio de ligação celular, uma vez que esta é rotineiramente utilizada para cultura de *T. cruzi* e estudos de infecção (ABUIN et al., 1989; ALVES et al., 1986). Dentre os fagos testados em nossos ensaios de ligação, apenas o fago contendo o peptídeo TS9 ligouse consistentemente a células LLC-MK2, em experimentos independentes (Figura 5A). Não foi observada ligação do fago de Fd-tet, utilizado como controle negativo pois não possui nenhum peptídeo exógeno em sua estrutura. Já o fago expressando o peptídeo FLY se ligou às células como esperado (MAGDESIAN et al., 2001; TONELLI et al., 2010). Para confirmar se a ligação do fago TS9 às células LLC-MK2 é mediada pelo peptídeo apresentado na pIII, realizamos um teste de competição do fago TS9 com o peptídeo sintético de mesma sequência. As células incubadas com concentrações crescentes do peptídeo sintético TS9 tiveram reduzida a ligação ao fago (Figura 5B).

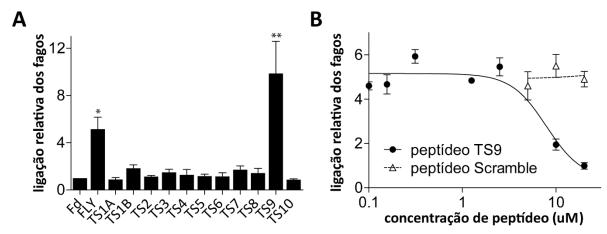


Figura 5. (**A**) Quantificação da ligação do fago às células LLC-MK2 utilizando a metodologia BRASIL e PCR quantitativo (qPCR). Os resultados de ligação do fago foram normalizados relativamente ao fago de controle Fd-tet (fago sem peptídeo exógeno). A média e o erro padrão da média de quatro amostras são mostradas. Observou-se uma ligação significativa das células para o fago FLY e TS9 (\* p <0,05 e \*\*\* p <0,001, ANOVA, N=4). (**B**) Efeito do peptídeo sintético TS9 e do peptídeo controle *scramble* na ligação do fago TS9 às células LLC-MK2. Os resultados de ligação do fago foram normalizados em relação ao fago de controle Fd-tet. São apresentadas as médias ± SEM de um experimento representativo, realizado em triplicata.

Nenhum efeito foi observado quando o peptídeo controle *scramble* foi utilizado. Este peptídeo controle tem a mesma composição de aminoácidos do peptídeo TS9, no entanto estes se encontram em uma ordem diferente (sequência LTVIFATVS). Em resumo, identificou-se um novo peptídeo de ligação celular conservado nos membros do grupo II da família das gp85/TS.

### O peptídeo TS9 é conservado em todos os grupos de gp85/trans-sialidases

Em nossa primeira análise onde foram selecionados os peptídeos testados, levamos em consideração apenas a sequências das proteínas pertencentes ao grupo II da família das gp85/TS. Para determinar se o peptídeo que se ligou às células, TS9, está presente em todos os grupos da família, alinhamos as 505 sequências de proteínas obtidas a partir da base de dados TryTryp, pertencentes a todos os oito diferentes grupos da família. O alinhamento revelou que o motivo contendo o peptídeo TS9 é conservado em todos os grupos da família gênica (Figura 6). Apenas os motivos peptídicos TS5, TS9 e FLY são consistentemente conservados em todos os grupos da família. Os motivos TS9 e FLY co-ocorrem em todos os membros da família, com exceção de alguns membros truncados que não tem FLY apenas ou ambos os motivos.

### A citoqueratina é um receptor para o peptídeo TS9

Tendo mostrado que o peptídeo TS9 liga-se às células LLC-MK2, buscamos determinar a qual receptor presente na superfície celular esse se ligava. Para este fim, realizamos uma cromatografia de afinidade onde o peptídeo TS9 foi imobilizado junto a uma matriz formando a fase estacionária. Para melhor mimetizar as condições existentes no fago, o peptídeo foi ligado pela porção carboxi-terminal, já que o mesmo é fusionado ao amino-terminal da proteína III do capsídeo viral. O extrato proteico obtido a partir de células LLC-MK2 foi passado pela coluna contendo a resina e, após extensivas lavagens, as proteínas ligadas foram eluídas sequencialmente com dodecilsulfato de sódio a 1% (SDS) e ureia 8M (em tampão fosfato) e todas as frações analisadas por eletroforese em gel de poliacrilamida-SDS (SDS-PAGE).

Observamos que as proteínas com massas moleculares calculadas de 49, 53,



Figura 6. Conservação dos motivos peptídicos selecionados nos diferentes grupos, I a VIII, da família das gp85/trans-sialidases.

59 e 66 kDa foram enriquecidas no eluato de uréia 8M, em comparação ao extrato total (Figura 7A). Para identificação dessas proteínas, as bandas do gel foram excisadas e submetidas a espectrometria de massas. A citoqueratina-18 e a citoqueratina-8 (KRT18 e KRT8) foram as proteínas com as maiores pontuações identificadas como possíveis ligantes para o peptídeo TS9 presente na região do gel correspondente a massas moleculares de 49 e 53 kDa (Figura 7A, proteínas indicadas pelas setas A e B). Foram 29 peptídeos encontrados para a citoqueratina 18 no espectro de massas (pontuação igual a 809) e 28 para citoqueratina 8 (pontuação de 723). Também identificamos no experimento outra proteína de citoesqueleto: tubulina (59 kDa, seta C) e também a proteína HSP90 (66 kDa, seta D) como sendo possíveis ligantes para o peptídeo TS9. Dada a natureza do ensaio, é possível que alguns dos resultados sejam devidos a uma associação secundária, já que se sabe que citoqueratinas 8 e 18 formam complexos e também HSP90 se liga a esses filamentos (FOSTINIS et al., 1992). Uma vez que citoqueratinas 8 e 18 são ambos constituintes de filamentos intermediários e foram

previamente identificados pelo nosso grupo como receptores para o peptídeo FLY (MAGDESIAN et al., 2001), decidimos concentrar nosso estudo na interação entre o peptídeo TS9 a essas proteínas. Confirmados os dados da espectrometria de massas, os experimentos de *western blot* mostraram a presença da citoqueratinas 18 no eluato (Figura 7B).

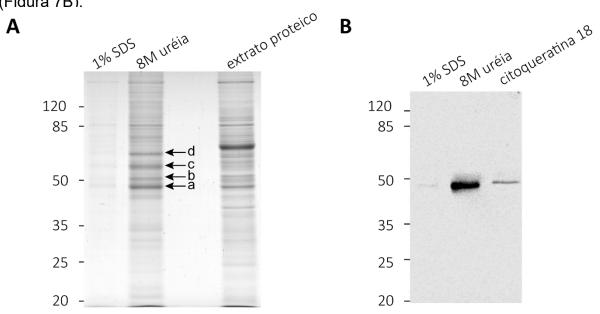


Figura 7. (**A** e **B**). Cromatografia de afinidade do extrato de células LLC-MK2 contra o peptídeo TS9 imobilizado, eluído sequencialmente com solução de 1% SDS e 8M ureia, e analisada por *SDS-PAGE* seguido por coloração com *Coomassie brilliant blue* (**A**) ou *Western blot* usando anticorpo anti-citoqueratina 18 (**B**). Em (**A**), as proteínas com massas calculadas de 49, 53, 59 e 66 kDa, identificadas por coloração de *Coomassie* (setas a-d, respectivamente), foram eluídas do gel e analisadas por espectrometria de massas. Em (B) verifica-se a imunorreatividade do anticorpo anti-citoqueratina 18 com os eluatos de 1% SDS e 8M ureia. A citoqueratina 18 recombinante foi utilizado como controle positivo.

Para confirmar que o peptídeo TS9 se liga à citoqueratina, utilizou-se novamente a técnica de *phage display*. Como existem vários genes que codificam diferentes proteínas da citoqueratina, e as células muitas vezes expressam mais de um gene, expandimos nossa validação para incluir outros membros recombinantes desta família, bem como vimentina, outra proteína de filamento intermediário. Os resultados do nosso ensaio de ligação mostram que ambos os fagos TS9 e FLY se ligam a todas as proteínas de filamento intermediário que foram testadas (Figura 8). A interação é específica uma vez que não foi observada ligação do fago TS9 a gelatina ou albumina sérica bovina (BSA), e o fago controle Fd-tet não se ligou a nenhuma das proteínas

testadas. O perfil de ligação do fago TS9 está também de acordo com os resultados da cromatografia de afinidade em que pelo menos duas proteínas de citoqueratina diferentes foram eluídas da coluna peptídica. Em resumo, estes dados corroboram os resultados anteriores mostrando que as proteínas da família das gp85/TS podem interagir com diferentes citoqueratinas e proteínas de filamento intermediário expressas pelas células do hospedeiro (MAGDESIAN et al., 2001, 2007; TONELLI et al., 2010).

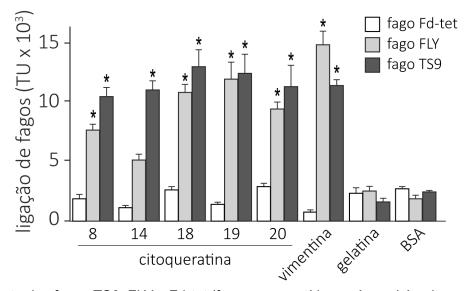


Figura 8. Ligação dos fagos TS9, FLY e Fd-tet (fago sem peptídeo exógeno) às citoqueratinas humanas recombinantes 8, 14, 18, 19, 20, vimentina, gelatina e albumina de soro bovino (BSA). A média e o erro padrão da média de um experimento representativo realizado em triplicata está representada (\* indica p <0,05, ANOVA).

### Os peptídeos TS9 e FLY formam um único sítio de ligação ao domínio LamG

Os peptídeos, TS9 e FLY, são separados um do outro por mais de 100 aminoácidos na sequência primária das proteínas gp85/TS e, portanto, não podem formar um sítio de ligação linear. Dessa forma, questionamos se eles poderiam estar localizados próximos uns dos outros na estrutura tridimensional da proteína, constituindo assim um sítio de ligação conformacional. A arquitetura típica das proteínas pertencentes à família das gp85/TS é composta de dois domínios: um sialidase no amino-terminal (NCBI CDD id: cl21531) e um laminina G (LamG) no carboxi-terminal (NCBI CDD id: cl00102). Notamos que ambos os peptídeos TS9 e FLY estão contidos dentro do domínio LamG, estando o peptídeo TS9 localizado no início do domínio, enquanto FLY está próximo ao carboxi-terminal (Figura 9A).



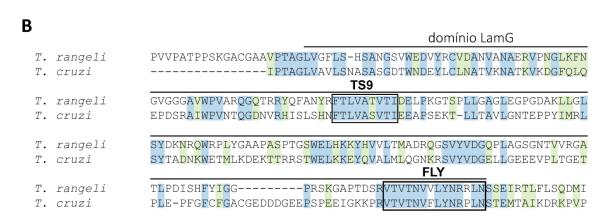


Figura 9. (**A**) Representação esquemática dos dois domínios conservados (sialidase e LamG) presentes nos membros da família gp85/trans-sialidases. A localização dos peptídeos TS9 e FLY está indicada. (**B**) Alinhamento da sequência do domínio LamG da proteína Tc85-11 de *T. cruzi* e de uma sialidase de *T. rangeli* (sequência retirada do registro do PDB 1WCS, aminoácidos 426-624). Os aminoácidos idênticos são destacados em azul e as mudanças para aminoácidos semelhantes em verde. Os peptídeos TS9 e FLY são mostrados (caixas).

Para que pudéssemos avaliar a relação entre TS9 e FLY, buscamos estruturas de sialidase de tripanossomatídeos disponíveis em bancos de dados, bem como um modelo que geramos para um membro da família gênica. Selecionamos a proteína Tc85-11 (GenBank id: AAD13347) como um membro representativo porque é uma proteína de ligação celular pertencente ao grupo II da família das gp85/TS e foi anteriormente utilizada em estudos de nosso grupo (GIORDANO et al., 1999; MARROQUIN-QUELOPANA et al., 2004). Pesquisando no banco de dados de estrutura de proteínas RCSB-PDB (BERMAN et al., 2000), identificamos a sialidase de *Trypanosoma rangeli* como sendo um homólogo próximo da proteína Tc85-11 com uma estrutura terciária resolvida por difração de raios-X (ID PDB: 1WCS, PARIS et al., 2005). Esta proteína partilha 50% de semelhança de sequência com Tc85-11 e contém ambos os peptídeos TS9 e FLY com substituições que estão também presentes em outros membros da família gp85/TS de *T. cruzi* (Figura 9B). Para construir o modelo Tc85-11 usamos o servidor I-Tasser (ROY; KUCUKURAL; ZHANG, 2010) indicando a estrutura 1WCS

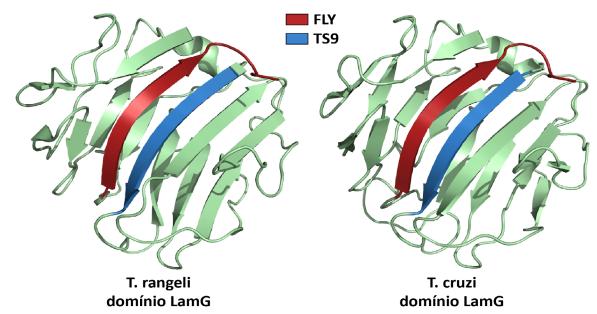


Figura 10. Representação das estruturas tridimensionais do domínio LamG da proteína sialidase de *T. rangeli* (PDB 1WCS) e da estrutura modelada da proteína Tc85-11 de *T. cruzi*. As posições dos peptídeos TS9 (azul) e FLY (vermelho) são destacadas na estrutura. como modelo.

A análise das estruturas tridimensionais da sialidase de *T. rangelli* e do modelo da Tc85-11 indicou que, embora os peptídeos TS9 e FLY estejam afastados na sequência primária, estão próximos um do outro na conformação nativa da proteína, constituindo estruturas de folha β antiparalelas (Figura 10). Estes dados sugerem que os peptídeos TS9 e FLY podem compreender um único sítio de ligação conformacional não-linear para citoqueratinas presentes em todos os membros da família gp85/TS e implicam ainda o domínio LamG como importante para a adesão do parasita.

### O domínio LamG de gp85/trans-sialidases liga-se a citoqueratinas.

Para confirmar que o domínio LamG de gp85/TS se liga às citoqueratinas, como sugerido pela presença dos dois motivos peptídicos conservados (TS9 e FLY), produzimos em *E. coli* como proteína recombinante o domínio LamG de Tc85-11 (denominado Tc85-11<sup>LamG</sup>) (Figura 11A). A proteína purificada foi reconhecida por um anticorpo monoclonal (mAb) anti-gp85 (Figura 11B), possibilitando o uso deste mAb em ensaios de ELISA, onde os possíveis receptores – citoqueratinas, vimentina e controle – foram imobilizados e posteriormente incubados com o domínio recombinante que foi

então detectado.

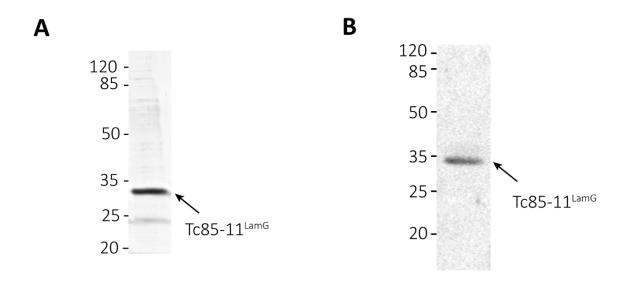


Figura 11. (**A**) Proteína Tc85-11<sup>LamG</sup> recombinante produzida e purificada analisada por SDS-PAGE e corada com *Coomassie brilliant blue* e (**B**) imunorreatividade com um anticorpo monoclonal anti-gp85/trans-sialidases.

De modo semelhante ao fago TS9, a proteína Tc85-11<sup>LamG</sup> liga-se a todas as citoqueratinas imobilizadas que testamos (Figura 12A). A ligação foi específica para citoqueratinas e não se observou interação com gelatina, utilizada como controle negativo. Para determinar se os peptídeos TS9 e FLY desempenham um papel na ligação do domínio LamG à citoqueratina, realizamos ensaios de competição de ligação na presença dos peptídeos sintéticos. A ligação de Tc85-11<sup>LamG</sup> a citoqueratina 18 foi inibida pelos peptídeos sintéticos FLY e TS9, mas não pelo peptídeo controle *scramble* (Figura 12B). Em experimentos independentes (N=3) o peptídeo TS9 inibiu a ligação de Tc85-11<sup>LamG</sup> a KRT18 em média 81% e o peptídeo FLY 70%. A combinação de ambos os peptídeos não resultou em aumento significativo nos níveis inibitórios. Estes resultados corroboram os dados que sugerem que os peptídeos TS9 e FLY compreendem um local de ligação conformacional não-linear para citoqueratinas.

### O peptídeo TS9 inibe a adesão celular e reduz a infecção celular pelo *T. cruzi*

Os membros da família gp85/TS já foram anteriormente associados à adesão e invasão celular por nosso grupo (ABUIN et al., 1989; ALVES et al., 1986; GIORDANO

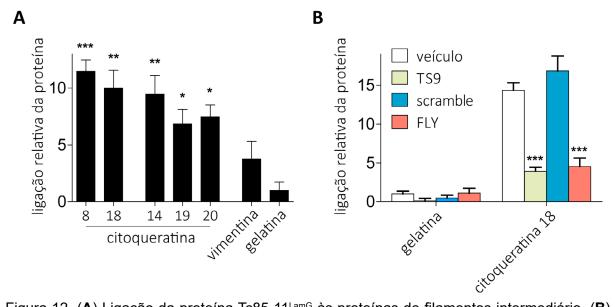


Figura 12. (**A**) Ligação da proteína Tc85-11<sup>LamG</sup> às proteínas de filamentos intermediário. (**B**) Ensaio de competição de ligação entre peptídeos sintéticos - *scramble*, TS9 e FLY - e proteína Tc85-11<sup>LamG</sup> à proteína citoqueratina 18. Os resultados foram normalizados pelo controle negativo (gelatina). A média e erro padrão da média de um experimento representativo realizado em triplicata é apresentado (\* p <0,05, \*\* p <0,01 \*\*\*, p <0,001, ANOVA).

et al., 1994; MATTOS et al., 2014) e outros (ARAYA et al., 1994; CORTEZ et al., 2014; SANTOS; GARG; TARLETON, 1997). Tendo mostrado que o domíno Tc85-11<sup>LamG</sup> e o peptídeo TS9 derivado destas proteínas, ligam-se a células hospedeiras e interagem com citoqueratinas, testamos se estes eram capazes de modular a infecção de células de mamíferos pelo parasita. As células epiteliais LLC-MK2 foram pré-incubadas com o peptídeo sintético TS9 (200 µM) durante 15 minutos antes da adição de formas tripomastigotas à cultura celular. Após duas horas de infecção, as células foram lavadas para remover parasitas extracelulares, e posteriormente cultivadas durante 48 horas. As células foram então fixadas e o número de células infectadas e parasitas (amastigotas) por células infectada foi quantificado. O peptídeo TS9 reduziu a infecção por *T. cruzi* em 78% (Figura 13A) e também reduziu o número de parasitas por célula em 69% (Figura 13 B e C). Adicionalmente, a infecção celular foi reduzida de um modo dependente da concentração utilizada (Figura 13D). O tratamento de células hospedeiras e parasitas apenas com o veículo dimetilsulfóxido (DMSO) ou com o peptídeo controle scramble na mesma concentração não teve efeito sobre a infecção das células pelo parasita.

Para determinar se a redução na infecção era devida a um efeito tóxico do peptídeo ao parasita ou hospedeiro, as células foram incubadas com o peptídeo TS9 (200 µM) durante o mesmo período de tempo utilizado nos ensaios de infecção (duas horas para tripomastigotos ou 48h para células LLC-MK2). Após o período, a viabilidade celular foi medida pelo ensaio de conversão do sal WST-1 a formazan. Nenhum efeito na viabilidade celular foi observado, sugerindo que o peptídeo por si só não é tóxico para qualquer tipo de célula (Figura 14).

A fim de avaliar o efeito do peptídeo TS9 na adesão de células de *T. cruzi*, repetiu-se o ensaio de infecção a uma temperatura mais baixa: 4°C. Ao se realizar o ensaio nesta temperatura, uma condição que permite que os parasitas adiram mas não entrem na célula, observamos que o peptídeo TS9 reduziu em 58% o número de parasitas ancorados à superfície das células epiteliais (Figura 13E). Não se observou qualquer efeito com o peptídeo controle *scramble*. Estes resultados sugerem que o peptídeo TS9 afeta a adesão do *T. cruzi* às células e consequentemente reduz a infecção celular.

### O domínio LamG de gp85/trans-sialidases inibe a infecção por *T. cruzi*

Tendo mostrado que um peptídeo derivado do domínio LamG de gp85/transsialidase modula a invasão celular, determinamos em seguida o efeito de todo o domínio no processo. As células LLC-MK2 e os tripomastigotos foram incubadas na presença ou ausência de proteína recombinante Tc85-11<sup>LamG</sup> (6,5 μM) durante duas horas. As células foram então lavadas e cultivadas durante 24 horas antes de serem fixadas e coradas para análise por microscópio de fluorescência. Houve uma redução de 60% no número de células infectadas pela proteína Tc85-11<sup>LamG</sup> (Figura 13F), confirmando a importância deste domínio no processo de infecção de células parasitas.

É relevante notar que o domínio LamG encontrado em proteínas gp85/transsialidase está também presente na lectina Concanavalina-A (família de proteínas, pfam13385 lectina concanavalina A/superfamília das glucanases). A concanavalina A liga-se a carboidratos contendo grupos  $\alpha$ -D-manosil e  $\alpha$ -D-glucosil e é possível

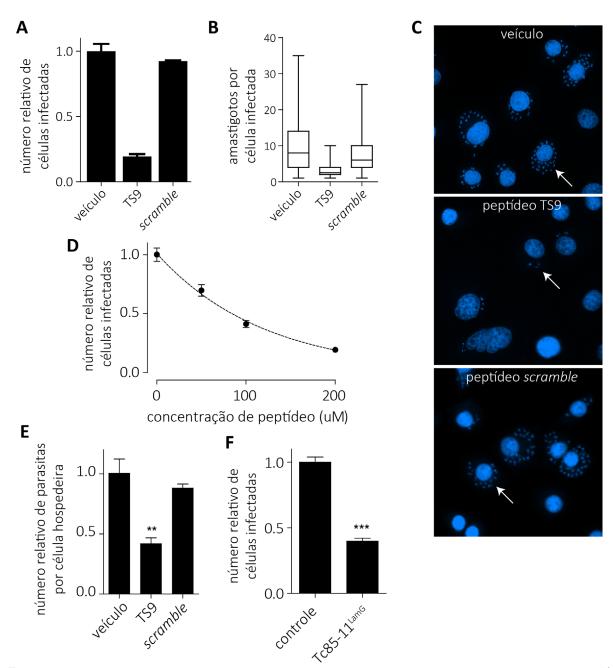


Figura 13. O peptídeo TS9 inibe a adesão e a invasão das células pelos parasitas. (A e B) Efeito do peptídeo TS9 e do peptídeo de controle scramble (200 µM) na infecção da célula hospedeira LLC-MK2 pelo T. cruzi. Em (A) o número de células infectadas é indicado. A proporção de célula infectadas no tratamento apenas com veículo foi fixado em 1 e os outros valores são relativos a esse número. São apresentadas as médias e o erro padrão da média de um experimento representativo realizado em triplicata. Em (B) o número de parasitas por células hospedeiras infectadas é mostrado. Os dados são apresentados em um diagrama de caixas nas quais as caixas definem os percentis 25 e 75, com uma linha na mediana e whiskers definindo os valores máximo e mínimo encontrados no experimento (\*\*\* p <0,001, ANOVA, Dunnet pós-teste). (C) Imagens representativas de campos de microscopia utilizados para a quantificação da invasão de parasitas. As células foram coradas com DAPI. As setas indicam células infectadas com amastigotas no citoplasma. (D) Efeito da dose do peptídeo TS9 na infecção da célula hospedeira LLC-MK2 pelo parasita. (E) Ensaio de infecção realizado a 4°C para evitar que o parasita entre nas células hospedeiras na presença ou ausência dos peptídeos TS9 ou scramble (200 µM). (F) Efeito da proteína Tc85-11LamG na infecção da célula hospedeira pelo parasita.

considerar a possibilidade de que os domínios LamG de *T. cruzi* tenham também essa propriedade. Alguns membros da família gp85/trans-sialidase, em especial os pertencentes ao grupo I, têm a capacidade de transferir moléculas de ácidos siálico do terminal de uma cadeia de carboidratos para outra, sendo possível que o domínio LamG dessas proteínas e seus peptídeos conservados poderiam ligar-se aos carboidratos, facilitando este processo. Analisamos então as estruturas da concanavalina-A ligada a uma molécula de trimanosídeo e o domínio LamG da sialidase de *T. rangeli*. Por comparação das estruturas tridimensionais, o sítio de ligação de carboidratos na concanavalina-A não se sobrepõe ao sítio de ligação da citoqueratina das folhas β antiparalelas compreendido pelos peptídeos TS9 e FLY (Figura 15). Além disso, o sítio de ligação da citoqueratina tem uma área de superfície relativamente grande composta principalmente de aminoácidos hidrofóbicos, em comparação com o sítio de ligação de carboidratos menores na concanavalina-A, que é rico em resíduos hidrofílicos. Dessa maneira, uma possível ligação de açúcares à proteína seria improvável de ter participação das folhas formadas por TS9 e FLY.

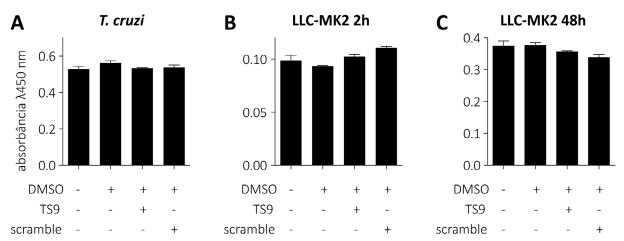
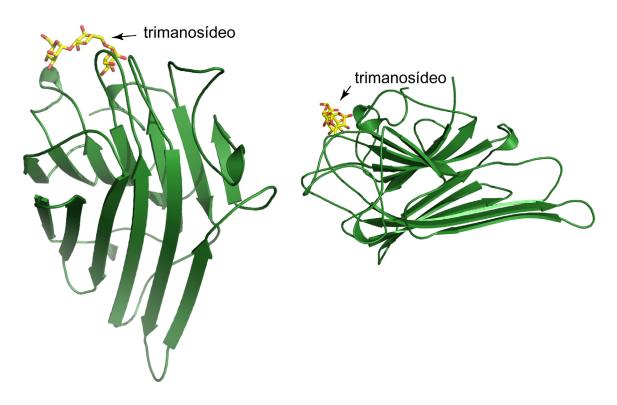


Figura 14. Efeito da toxicidade dos peptídeos TS9 e scramble nas células hospedeiras e de *T. cruzi*. Tripomastigotos e células LLC-MK2 foram incubadas com os peptídeos TS9 ou scramble (200 uM) durante 2 h (A e B) ou 48 horas (C) e analisadas para a viabilidade celular. As médias e o erro padrão da média de um experimento representativo, realizado em triplicata, são mostradas.

### A Concanavalina A



# B domínio LamG T. rangeli

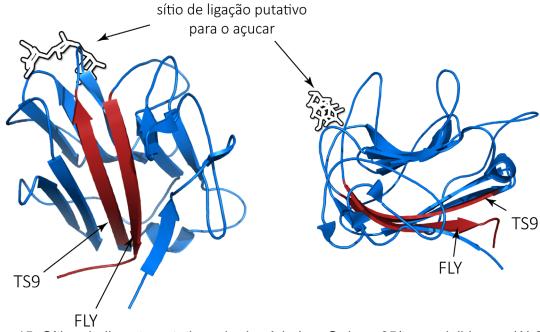


Figura 15. Sítios de ligação putativos do domínio LamG de gp85/trans-sialidases. (A) Estrutura tridimensional da concanavalina-A (PDB ID 1CVN cadeia A) com uma molécula de trimanosídeo ligada. (B) Estrutura terciária do domínio LamG da sialidase de *T. rangeli* (1WCS, aminoácidos 426-624). As setas indicam os peptídeos conservados TS9 e FLY que constituem o sítio de ligação identificação da citoqueratina (vermelho) e um sítio putativo de ligação de carboidratos identificado por analogia estrutural com a molécula de concanavalina-A.

## Discussão

As proteínas pertencentes ao grupo II da família gp85/trans-sialidases têm sido implicadas há muito tempo na adesão a células e componentes de matriz extracelular, mas os detalhes estruturais de tais interações são ainda, em muitos aspectos, desconhecidos. Uma das dificuldades para desvendar os detalhes estruturais de tais interações é o grande número de proteínas contidas nestas famílias multigênicas. Tendo centenas de membros individuais, é desafiador associar cada membro individual ao seu ligante. Além disso, a caracterização de cada membro separadamente não é prática. Essa tarefa é particularmente desafiadora no caso do *T. cruzi*, onde ferramentas genéticas como RNAi não podem ser usadas (ULLU; TSCHUDI; CHAKRABORTY, 2004). Aqui utilizamos uma abordagem sistemática para identificar motivos peptídicos conservados compartilhados por vários membros da família e testá-los em ensaios funcionais utilizando *phage display*. Esta abordagem revelou-se bem-sucedida e levou à identificação do peptídeo TS9, um novo motivo envolvido na adesão e invasão de células pelo parasita. Vale ressaltar o fato de que a identificação do peptídeo TS9 confirma ainda resultados anteriores do nosso grupo e reforça o conhecimento de que as citoqueratinas são importantes ligantes para a adesão e invasão do *T. cruzi* (MAGDESIAN et al., 2001; MATTOS et al., 2014; TONELLI et al., 2010). Também destaca a importância na biologia parasitária do domínio LamG presente nas proteínas da família das gp85/trans-sialidases.

As citoqueratinas são proteínas essenciais para a estrutura celular e principais constituintes dos filamentos intermediários. Por conseguinte, são frequentemente vistos apenas como proteínas intracelulares. No entanto, é de notar que vários estudos têm relatado a expressão ectópica de citoqueratinas, como no meio extracelular na membrana plasmática (GODFROID et al., 1991; GONIAS; HEMBROUGH; SANKOVIC, 2001; MAGDESIAN et al., 2001; TONELLI et al., 2010; WELLS et al., 1997). Ainda estudos de nosso grupo e outros enfatizaram o papel importante da citoqueratina extracelular na relação entre hospedeiro-patógeno (MAGDESIAN et al.,

2001; SAJJAN; ACKERLEY; FORSTNER, 2002; SAJJAN; SYLVESTER; FORSTNER, 2000; TONELLI et al., 2010). Há também descrito na literatura o envolvimento destas proteínas em outros processos patológicos: a expressão de citoqueratina 8 na superfície de células cancerosas tem sido associada com maior aderência à fibronectina e à matriz extracelular (KARANTZA, 2011; LIU et al., 2008). Também, anticorpos contra citoqueratina 8 inibem a ligação do receptor do ativador de plasminogênio do tipo uroquinase e a produção de plasmina pelas células de câncer da mama, aumentando o potencial invasor do tumor (OBERMAJER; DOLJAK; KOS, 2009).

Outros microrganismos causadores de doenças em humanos também são capazes de aderir em citoqueratinas durante a infecção. A proteína ClfB de *Staphylococcus aureus* é capaz de se ligar em citoqueratina 10 e fibrinogênio contribuindo para a colonização do epitélio nasal (GANESH et al., 2011). O patógeno *Streptococcus pneumoniae* usa a proteína PsrP para se ligar também à citoqueratina 10 (SCHULTE et al., 2014). Já o *Streptococcus agalactiae* se liga a citoqueratina 4 utilizando a proteína Srr-1 (SUNDARESAN; SAMEN; PONNURAJ, 2015). É curioso notar que todas essas proteínas têm estrutura semelhante à do domínio LamG: um sanduíche β, sugerindo um padrão estrutural para ligantes de citoqueratinas em

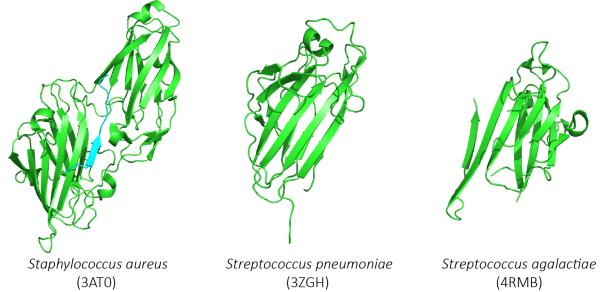


Figura 16. Estrutura cristalográfica de três proteínas de microrganismos, capazes de infectar seres humanos, que se ligam a citoqueratinas. As estruturas têm um tipo de dobramento chamado de sanduíche  $\beta$  que também caracteriza do domínio LamG das gp85/trans-sialidases. O código PDB e a espécie a que proteína pertence é indicada abaixo da estrutura.

microrganismos (Figura 16).

Vale ressaltar que as proteínas pertencentes a outros grupos de filamentos (microtúbulos e microfilamentos) também foram identificadas por cromatografia de afinidade no peptídeo TS9, porém se essas são receptores para as proteínas gp85/TS ainda deve ser investigado. Sabe-se que as formas amastigotas expressam proteínas da família gp85/TS e, portanto, podem interagir intracelularmente com citoqueratinas (LOW et al., 1998; SANTOS; GARG; TARLETON, 1997) e outras proteínas citoesqueléticas. Esta interação pode ser importante para a proliferação de parasitas no citoplasma. De fato, o silenciamento da citoqueratina-18 demonstrou afetar a capacidade do parasita de se replicar dentro das células hospedeiras (CLASER et al., 2008). Mas porque as células expressam várias citoqueratinas e dada a interação promíscua de gp85 com estas diferentes proteínas de citoqueratina, ainda é difícil avaliar os efeitos do silenciamento de um único gene de citoqueratina na invasão e proliferação de células parasitas.

Os resultados do ensaio de inibição no presente estudo sugerem que o peptídeo TS9 afeta a infecção por *T. cruzi* ao interferir com a capacidade do parasita de aderir à superfície celular, sendo acitoqueratina um dos possíveis receptores celulares envolvidos neste processo. Estes dados corroboram conclusões anteriores de nosso grupo que têm implicado o peptídeo conservado FLY e, portanto, o motivo VTVxNVxLYNRPLN, na adesão de células de *T. cruzi* a citoqueratinas/vimentina e invasão (MAGDESIAN et al., 2001). No entanto, é curioso o fato de os peptídeos mostrarem efeitos opostos em ensaios celulares: o peptídeo TS9 inibe consistentemente a infecção parasitária e bloqueia a interação do domínio LamG com citoqueratinas; O peptídeo FLY, por outro lado, também inibe a ligação do domínio LamG às citoqueratinas, mas aumenta a infecção por células parasitas (MAGDESIAN et al., 2001). Uma possível explicação é o fato de que o peptídeo FLY tem um efeito direto sobre as células hospedeiras, induzindo remodelação de citoqueratina e ativação de cascatas de sinalização ERK (MAGDESIAN et al., 2007), que por sua vez pode desencadear caminhos alternativos

utilizados pelo parasita para entrar nas células. Não sabemos se o peptídeo TS9 também afeta as células hospedeiras da mesma maneira. Não obstante, o efeito sobre a infecção celular e os outros resultados apontam para um papel importante para estes dois motivos peptídicos conservados, o domínio LamG e citoqueratinas na biologia do parasita. Consequentemente, é improvável que o mapeamento de dois peptídeos altamente conservados de ligação a citoqueratina (TS9 e FLY) ao mesmo local de ligação dentro do domínio LamG seja uma coincidência, mas serão necessários mais estudos para revelar os detalhes moleculares das suas interações e como eles operam no processo.

É notável o fato de o peptídeo TS9 se sobrepor com um epítopo imunodominante previamente descrito. O peptídeo TS9 (FTLVASVTI) partilha a sequência FTLV encontrada em dois peptídeos imunodominantes (ANYK<u>FTLV</u> e VNYD<u>FTLV</u>) derivados de membros da família gp85 / TS. Estes peptídeos são encontrados dentro da mesma região do peptídeo TS9 (MARTIN et al., 2006; ROSENBERG; MARTIN; TARLETON, 2010). Os peptídeos ANYKFTLV e VNYDFTLV são epítopos imunodominantes responsáveis pela ativação de células T CD8+ em camundongos e humanos, com até 40% de todas as células T CD8+ respondendo a eles. A resposta imune do hospedeiro representa uma importante pressão seletiva contra populações de parasitas com estas sequências. Apesar da pressão evolutiva estes peptídeos continuam a ser expressos pelo parasita. Nossos resultados sugerem que a ligação à citoqueratina pode explicar por que esses peptídeos não mudaram ao longo do tempo, uma vez que são necessários para a interação parasita célula-hospedeiro.

O domínio LamG, presente nas gp85qTS, também se encontra em outras proteínas com funções diversas: lamininas, colágenos, agrina, neurexinas, pentraxinas, proteína ligantes de hormônios sexuais (SBP/SHBG), concanavalina A, entre outros. São notáveis as diferentes funções e propriedades de ligação exibidas pelos domínios LamG nestas proteínas. Quando presentes em moléculas de adesão eles podem mediar ligação a receptores, adesão celular e migração (TISI et al., 2000).

Os domínios LamG encontrados na proteína de ligação a hormônios sexuais (SBP/SHBG) ligam moléculas pequenas como o estradiol (AVVAKUMOV et al., 2010). Por outro lado, os domínios LamG presentes nas pentraxinas interagem com uma multiplicidade de ligantes derivados de patógenos microbianos e de detritos celulares produzidos por processos inflamatórios (LU et al., 2012). Finalmente, na molécula de concanavalina-A, o domínio LamG é uma lectina que se liga a resíduos de α-D-manosil e α-D-glucosil. É possível que o domínio LamG nas proteínas gp85/TS ligue carboidratos facilitando a transferência de ácido siálico pela trans-sialidases, porém essa questão permanece em aberto. No entanto, como sugerido pela comparação com a estrutura de concanavalina-A, provavelmente o sítio de ligação seria um local distinto do sítio de citoqueratinas, formando um possível e interessante alvo para o desenvolvimento de fármacos.

# Conclusão

Este trabalho descreve uma nova abordagem sistemática para estudar famílias de genes encontradas em *T. cruzi* e outros parasitas. Combinando a análise *in silico*, para redução da complexidade de uma grande família de proteínas, com a tecnologia de *phage display*, estabelecemos um ensaio funcional para pesquisar os motivos peptídicos compartilhados pelos membros da família das gp85/TS. Esta abordagem mostrou-se bem-sucedida e levou à identificação de um peptídeo de adesão celular, importante para a invasão de células parasitas. Certamente, o estudo pode ser estendido a outras famílias de genes presentes no *T. cruzi* (ou outros organismos) tais como mucinas, *mucin-associated surface proteins* (MASP), *dispersed gene family-1* (DGF-1) e GP63. Dessa maneira pode-se identificar novos motivos peptídicos relevantes para a interação hospedeiro-parasita, tais como o caracterizado neste estudo, e ajudar a compreender os detalhes moleculares da infecção celular.

# Capítulo 2.

Mapeamento da resposta imune humoral na doença de Chagas por *phage display* de *shotgun* genômico

# Introdução

Sistema imune e tecnologias para mapeamento. O sistema imune é o conjunto de processos, moléculas, células e tecidos responsável por identificar e defender o organismo de patógenos, que podem ser simples moléculas, como no caso de toxinas, até seres macroscópicos como helmintos. Em vertebrados podemos separar o sistema imune em dois grandes componentes: sistema inato e sistema adaptativo. O sistema inato confere imunidade imediata já que não é específico para um patógeno, independendo de contato prévio. Barreiras anatômicas como a pele e células fagocíticas como macrófagos, neutrófilos e células dendríticas são parte do sistema inato. Já o sistema adaptativo, também chamado de adquirido, é específico para cada patógeno e suas moléculas. Esta resposta é desenvolvida quando o organismo invadido entra em contato com o patógeno pela primeira vez, e leva à criação da memória imunológica que permite uma rápida resposta em caso de reinfecção. As células que compõem o sistema adaptativo são os linfócitos T e B. Os primeiros são responsáveis por atividades regulatórias (células T *helper*) e de imunidade celular (células T citotóxicas), enquanto os segundos são responsáveis pela imunidade humoral (anticorpos). Ambos os tipos de linfócitos possuem receptores em sua superfície (receptor de células T e receptor de células B) que são os responsáveis por reconhecer moléculas com alta especificidade, os antígenos, desencadeando assim a resposta imune. Esses receptores passam por um complexo sistema de maturação, levando a uma enorme diversidade e possibilitando o reconhecimento das mais diversas moléculas (ABBAS; LICHTMAN; PILLAI, 2011).

O repertório imune é definido como a diversidade de receptores de antígenos presentes nos linfócitos T e B. Avanços nas técnicas de amplificação e sequenciamento destes genes permitiram uma melhor compreensão dessas coleções e da maturação do sistema imunológico (ROBINS, 2013; SHERWOOD et al., 2011). No entanto, a sequência primária dos receptores de antígenos não fornece informações sobre os antígenos e epítopos que a eles se ligam. No caso de receptores de células B

e anticorpos, pode-se usar técnicas de pequena escala como teste da reatividade do soro de indivíduos contra determinadas proteínas, mesmo aquelas oriundas de pequenas bibliotecas de cDNA por ELISA, possibilitando o estudo de dezenas ou até centenas de moléculas diferentes (CUNHA-NETO; CHEVILLARD, 2014). Técnicas mais modernas como *microarray*s de proteínas e peptídeos permitem que milhares de potenciais antígenos sejam testados simultaneamente (CARMONA et al., 2015; DRAGHICI; CHATTERJEE; TAINSKY, 2005), podendo até mesmo chegar a centenas de milhares no caso de *microarray*s de peptídeos de alta densidade. No entanto existem limitações importantes: grandes coleções de clones são necessárias para se construir um arranjo de proteínas que dependem de bactérias para sua produção e enovelamento e devem ainda ser posteriormente isoladas e imobilizadas. Também, ao se utilizar proteínas inteiras é possível obter informações relativas aos antígenos reconhecidos, porém sem determinação dos epítopos. Já microarrays de peptídeos, apesar de permitirem que uma grande quantidade de peptídeos seja testada, chegando até a algumas centenas de milhares, não são grandes o suficiente para que proteomas completos sejam cobertos. Ademais, os peptídeos estão restritos a sequências relativamente pequenas de no máximo 15 aminoácidos, o que pode ser suficiente para detecção de muitos epítopos lineares, mas falham na detecção da ligação de anticorpos que necessitam de sequências mais longas ou de conformações específicas para haver ligação.

Outro método utilizado com sucesso para mapeamento de epítopos é o *phage display*, que permite expressar na superfície de partículas virais (bacteriófagos) peptídeos e proteínas (JACOBSSON; FRYKBERG, 2001; SMITH, 1985). Essa metodologia é bastante versátil, permitindo a seleção de peptídeos ou proteínas contra diferentes alvos em diferentes situações; por exemplo, ensaios de seleção de peptídeos (chamados *biopanning*) podem ser realizados *in vitro* contra moléculas alvo imobilizadas, em células (que podem ou não ser estimuladas para alterar a expressão de receptores) (GIORDANO et al., 2001), em tecidos dissecados, ou ainda *in vivo*, injetando os fagos na circulação do animal, retirando os órgãos de interesse

e selecionando aqueles fagos que expressam peptídeos que interagem com o leito vascular de diferentes órgãos (PASQUALINI; RUOSLAHTI, 1996). Esta abordagem já permitiu identificar moléculas que têm sido utilizadas com sucesso para o desenvolvimento de fármacos e vacinas, para métodos diagnósticos e outros insumos (BARNHART et al., 2011; GIORDANO et al., 2010; HAJITOU et al., 2006).

É interessante observar que a primeira aplicação do *phage display* foi para a identificação de epítopos de anticorpos (SCOTT; SMITH, 1990). Uma biblioteca de *phage display* pode conter mais de 10º diferentes peptídeos/proteínas, o que resulta em ligantes para virtualmente qualquer alvo biológico. Mas, enquanto o *phage display* de peptídeos comumente utiliza DNA exógeno formado de oligonucleotídeos sintéticos aleatórios, no *shotgun* genômico, o DNA nuclear de organismos selecionados é fragmentado e utilizado para a construção das bibliotecas de *phage display* (JACOBSSON; FRYKBERG, 1995, 2001). Como os genes de procariotos não apresentam íntrons e seus genomas são, de forma geral, bastante compactos, as bibliotecas construídas têm se mostrado bastante eficientes na identificação e caracterização de novas proteínas bacterianas envolvidas, principalmente, em adesão celular. Por exemplo, novas adesinas de *Streptococcus* grupo B (BECKMANN et al., 2002), *Staphylococcus* (ZHANG et al., 1998) e *Leptospira* (LIMA et al., 2013) foram identificadas utilizando-se o *shotgun* genômico.

Bibliotecas de *phage display* construídas com cDNA também já foram utilizadas para identificação de autoepítopos: Minenkova et al., (2003) identificaram 16 antígenos diferentes em pacientes com câncer de mama e Chatterjee et al., (2006) analisaram 480 clones diferentes após a seleção da uma biblioteca contra amostras de pacientes com câncer de ovário, identificando 12 antígenos. Mais recentemente, a imunoprecipitação de uma biblioteca sintética contendo o proteoma humano exposto no fago T7, seguida de análise do sequenciamento de larga escala, foi utilizada com sucesso (LARMAN et al., 2011). No entanto, esta última abordagem envolve o uso de extensos *microarray*s de DNA para a construção, tornando seu custo muito alto.

Apesar de ser um organismo eucarioto, os genes de *T. cruzi*, com raras exceções, não apresentam íntrons e seu genoma é relativamente pequeno. Por isso, a metodologia de *shotgun* genômico display se torna bastante atrativa e é, portanto, uma possibilidade interessante para estudar os mais de 11.000 genes hipotéticos ou ainda não validados no genoma deste parasita (considerando-se o genoma diploide) (EL-SAYED, 2005; EL-SAYED et al., 2005). Mesmo assim um dos desafios de se fazer o *shotgun* de eucariotos ainda é o tamanho do genoma e de seu exoma em relação aos genomas de procariotos. Entretanto, uma comparação do genoma do *T. cruzi* com o genoma do *Streptococcus mutans*, que foi utilizado com sucesso para *shotgun* genômico (AJDIC; CHEN, 2013), mostra que ambos apresentam densidades gênicas aproximadamente similares (980 genes/Mb no Streptococcus contra 386 genes/Mb no genoma do *T. cruzi*). E apesar do genoma (haploide) do *T. cruzi* ser significativamente maior (15 vezes) e com um número maior de genes, isto não é uma limitação, pois o tamanho das bibliotecas ainda está dentro de um número satisfatório para manipulação (108-109 clones).

Neste projeto, construímos e utilizamos uma biblioteca para determinar o perfil antigênico humoral de pacientes com a doença de Chagas em diferentes formas clínicas da doença: assintomáticos (ou indeterminados) e portadores de cardiomiopatia crônica, leve e grave. Para isto, combinamos o *phage display* de *shotgun* genômico (Figura 17) com a plataforma de sequenciamento em larga escala Illumina MiSeq, sendo possível determinar de forma extensiva o repertório imunológico humoral de indivíduos infectados, identificando mais de 30.000 peptídeos distintos que se ligam às imunoglobulinas, sendo estes agrupados em mais de 2.000 epítopos diferentes reconhecidos especificamente por indivíduos infectados. As sequências identificadas são potencialmente um rico material para estudos da progressão da doença assim como desenvolvimento de método diagnóstico e também prognóstico.

A cardiomiopatia chagásica. Na fase aguda da infecção humana por *T. cruzi*, há envolvimento cardíaco em 90% dos indivíduos infectados, com predominância de

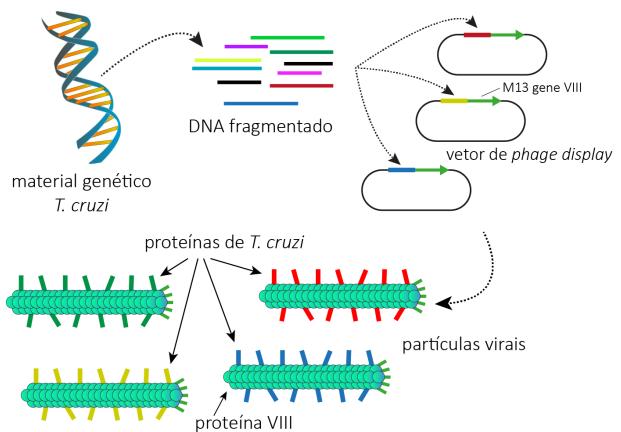


Figura 17. Representação esquemática da construção da biblioteca de phage display usando a estratégia de shotgun genômico. O material genético do parasita é purificado e fragmentado. Estes fragmentos são ligados de forma aleatória no vetor de phage display. Quando dentro de bactérias do tipo E. coli e na presença do helper phage, esta biblioteca dá origem a partículas virais que expressam fragmentos de proteínas de T. cruzi em sua superfície.

miocardite. Porém, após 6 ou 8 semanas, a parasitemia diminui, os sintomas da fase aguda desparecem e os pacientes entram na chamada fase indeterminada da doença. Como já mencionado, a grande maioria dos indivíduos não apresenta mais sintomas físicos da doença (p. ex., alterações no eletrocardiograma), mas permanecem positivos para exames sorológicos ou parasitários (aproximadamente 60-70%). Os demais, num período de 5-10 anos, irão manifestar problemas cardíacos (30%) ou digestivos (5-10%). Entres os primeiros, um terço (ou seja, 10% dos infectados), irão manifestar uma forma cardíaca particular, fatal, com disfunção ventricular, parada cardíaca e arritmia (CUNHA-NETO; CHEVILLARD, 2014). É interessante notar, que a gravidade do comprometimento cardíaco depende da presença de miocardite. Pacientes assintomáticos apresentam miocardite menos importante, enquanto os pacientes com disfunções cardíacas severas apresentam miocardite frequente com elevado grau de inflamação (HIGUCHI et al., 1987). A cardiomiopatia chagásica crônica (CCC) é a forma

mais importante da doença. São variadas suas manifestações: arritmias, bloqueios de condução do impulso nervoso, tromboembolismo, infarto, podendo chegar a paradas cardíacas e morte súbita.

Assim, a patogênese da CCC é dependente de uma intrincada relação parasita hospedeiro, onde a baixíssima carga parasitária, que é crônica, constante e sistêmica, dispara reações autoimunes e inflamatórias, levando à gradual destruição do tecido cardíaco (BIOLO; RIBEIRO; CLAUSELL, 2010; CUNHA-NETO; CHEVILLARD, 2014). O processo inflamatório cardíaco é predominante na fase aguda da doença, mas persiste na fase crônica, apesar de subclínico. Como mencionado para a miocardite, a perda celular e consequente fibrose associada a esta inflamação se correlaciona com a gravidade da cardiomiopatia (HIGUCHI et al., 1987). Entretanto, reações autoimunes também têm sido implicadas neste processo: a exposição de antígenos (lise celular) induz o mimetismo molecular entre antígenos do parasita e moléculas do tecido hospedeiro, além de uma ativação policional, resultando na produção de autoanticorpos (CUNHA-NETO et al., 2006; MARIN-NETO et al., 2007). De fato, autoanticorpos específicos contra antígenos presentes no coração, sistema nervoso e outros tecidos, foram identificados em modelos animais da doença, assim como em pacientes. Entre os antígenos mapeados, podemos citar o receptor β-adrenérgico e a miosina cardíaca (CUNHA-NETO et al., 1995; KIERSZENBAUM, 1999), porém a lista de antígenos com reação cruzada é bem mais extensa (BONNEY; ENGMAN, 2015). A resposta citotóxica por linfócitos CD8+ também é importante no processo (ROSENBERG; MARTIN; TARLETON, 2010). Estas células parecem reconhecer um epítopo dominante presente em todos os membros da superfamília de proteínas de superfície do parasita conhecidas como gp85/trans-sialidases (COLLI, 1993; ROSENBERG; MARTIN; TARLETON, 2010; TEIXEIRA et al., 2015), além de cruzipaína e proteína FL-160 (FONSECA et al., 2005). Porém, ainda não se sabe qual é a exata contribuição de todos estes fatores no desenvolvimento e progressão da doença e os mecanismos moleculares e imunológicos na CCC. Desta forma, é importante conhecer com mais detalhes os componentes imunológicos desta doença, para compreender melhor seus mecanismos e desenvolver alternativas terapêuticas que possam impedir a progressão da cardiomiopatia nestes pacientes.

### **Material e Métodos**

**Vetor** *phagemid.* O vetor pG8SAET foi gentilmente cedido pelo Dr. Paulo Lee Ho (Instituto Butantan) que os recebeu do laboratório de Dr. Lars Frykberg (Department of Microbiology, Swedish University of Agricultural Sciences, Uppsala).

**Cultura de parasitas.** A forma epimastigota de *T. cruzi* foi cultivada em meio LIT em incubadora a 28°C (cepa Silvio X10/1, gentilmente cedidas pela Dr<sup>a</sup>. Bianca Zingales).

Extração de DNA genômico de *T. cruzi*. DNA genômico foi obtido a partir da forma epimastigota do parasita. As células foram centrifugadas e lavadas três vezes com PBS, ressuspendidas em tampão de lise contendo 1% SDS e 100 μg/mL de proteinase K. O material foi incubado por duas horas a 50°C, centrifugado e o sobrenadante recuperado. Foram realizadas duas extrações com fenol:clorofórmio seguida da precipitação do DNA com etanol. O DNA foi ressuspendido em tampão TE (10 mM Tris pH 8, 1 mM EDTA).

Fragmentação do DNA genômico de *T. cruzi*. O DNA genômico previamente extraído foi submetido a fragmentação no aparelho COVARIS S2 em alíquotas de 30 μg seguindo as seguintes condições: *Duty cycle* – 10%, *Intensity* – 4, *Cycles/burst*: 200, *Time*: 80 segundos. Os fragmentos foram submetidos a eletroforese em gel de agarose com baixo ponto de fusão e a região entre 100 e 500 pb foi cortada e purificada realizando-se extrações com fenol e clorofórmio. O DNA obtido foi tratado com T4 DNA polimerase (Thermo Scientific) para o reparo das extremidades seguindo-se o protocolo do fabricante.

**Produção do vetor de clonagem pG8SAET.** O plasmídeo pG8SAET foi transformado em bactérias *E. coli* (cepa DH10B) e cultivado em meio LB por 16 horas. O DNA plasmideal foi isolado utilizando-se kit QIAGEN Plasmid Maxi Kit, seguindo as instruções

do fabricante. O material foi submetido a purificação por gradiente de CsCl/brometo de etídeo (HEILIG; ELBING; BRENT, 2001), seguida de extrações com n-butanol para remoção do brometo de etídio e diálise contra tampão TE para remoção do sal.

Produção da biblioteca. O vetor pG8SAET puro foi digerido com a enzima Eco105I (Thermo Fisher Scientific) seguindo-se as instruções do fabricante e purificado por extração fenol:clorofórmio e precipitação por etanol para remoção da enzima após o processo de digestão. O material foi incubado em reação junto com os fragmentos de DNA de *T. cruzi* gerados e T4 DNA ligase (Thermo Scientific) seguindo-se as instruções do fabricante da enzima. Várias relações molares de vetor:inserto foram utilizadas para determinação da melhor condição. No processo final de ligação, utilizamos 100 µg do vetor digerido e uma proporção de vetor:inserto de 1:30. A reação foi incubada por 16h a 16°C e depois purificada por extrações com fenol e clorofórmio e precipitação com etanol. O DNA resultante foi transformado em bactéria E. coli (cepa DH10B) eletrocompetente. As bactérias transformadas foram recuperadas e selecionadas durante a noite em meio LB contendo 50 ug/ml de carbenicilina. Diluições seriadas foram plaqueadas em meio LB ágar para determinação do número de transformantes e análise dos clones. No dia seguinte, o DNA plasmideal da biblioteca foi purificado, ressuspendido em TE e congelado. Para a produção dos fagos, a biblioteca na forma de plasmídeo é transformada em bactéria E. coli (cepa TG1) e infectada com o helper phage M13KO7 que possibilita a criação das partículas virais.

**Soro dos pacientes.** Os soros utilizados neste estudo fazem parte da soroteca do Dr. Edécio Cunha Neto, e foram coletados ao longo de vários anos, com o consentimento dos doadores, tendo estes estudos sido aprovados pelo comitê de ética em pesquisa com seres humanos da Faculdade de Medicina da USP. Todos os pacientes foram considerados chagásicos após pelo menos dois ou três resultados positivos para a presença de anticorpos anti-*T. cruzi*. Os pacientes positivos foram submetidos a eletrocardiografia e ecocardiografia e aqueles com alterações no ECG (bloqueio do ramo direito do feixe de His e/ou bloqueio fascicular anterior esquerdo) foram classificados

como sendo cardiopatas leves (fração ejetada do ventrículo esquerdo, FEVE > 40%) ou cardiopatas graves (FEVE ≤ 40%). Todos os demais, sem alterações no ECG foram considerados chagásicos assintomáticos (Tabela 2). Os soros dos pacientes foram agrupados em conjuntos de 10 (2 grupos para cada condição) e as imunoglobulinas purificadas por cromatografia de afinidade utilizando resina de sepharose-proteína G, seguindo-se as recomendações do fabricante (Thermo Fischer).

Tabela 2. Coorte dos soros utilizados neste estudo.

Estágio da Doença	Número de pacientes
Doadores soronegativos para Chagas	20
Doadores soropositivos para Chagas, sem manifestação da doença (assintomáticos)	20
Chagásicos crônicos com cardiopatia leve	20
Chagásicos crônicos com cardiopatia grave	20

Purificação de IgG do soro dos pacientes. Os soros dos pacientes de cada grupo (controle, assintomático, CCC leve e CCC grave) foram combinados, totalizando 1 ml de soro para cada grupo e purificados por cromatografia de afinidade. As amostras foram adsorvidas por gravidade em coluna contendo resina de sepharose-proteína G (GE Healthsciences). A resina foi lavada com 10 volumes de PBS para remover contaminantes e finalmente as imunoglobulinas foram eluídas com tampão 100 mM glicina pH 2.5, seguido de neutralização com tampão 1 M tris pH 8. A concentração de imunoglobulinas foi medida pelo ensaio de BCA (Thermo Scientific) e a pureza verificada por eletroforese em gel de SDS-poliacrilamida.

**Seleção em imunoglobulinas.** As IgGs purificadas de cada grupo foram imobilizadas em poços individuais de placas de poliestireno (96-poços) durante a noite a 4°C (1 ug em 50 ul). Os poços foram lavados três vezes com 200 ul de PBS, bloqueados com 20 ul de PBS 1% BSA por uma hora e depois incubados com 50 ul da biblioteca de *phage display* por 2h. Para minimizar a seleção de clones que se liguem a imunoglobulinas

de forma inespecífica, assim como a seleção de antígenos naturais reconhecidos por anticorpos presentes na população em geral, durante a seleção foram adicionados na solução 5 µg de IgG do grupo controle (soros negativos para Chagas). Ao final da incubação, os poços foram lavados 10 vezes com 200 ul de PBS e os fagos ligados às imunoglobulinas recuperados por infecção com 200 ul de cultura saturada de bactéria *E. coli* (cepa TG1). As bactérias foram então infectadas com o *helper phage* M13KO7 e cultivadas durante a noite na presença de carbenicilina e kanamicina para a amplificação dos fagos. No outro dia os fagos foram precipitados do sobrenadante e utilizados para um novo ciclo de seleção.

Sequenciamento em larga escala. Foram utilizados para o sequenciamento os DNAs plasmidiais das bibliotecas *naïve* e do terceiro ciclo de seleção das amostras. Foi realizada uma primeira reação de PCR de 20 ciclos utilizando *primers* específicos que amplificam a região codificadora do peptídeo exógeno e também adiciona sequências necessárias para a plataforma Illumina (Tabela 3). O produto de PCR foi purificado e submetido a uma nova etapa de PCR de 20 ciclos, agora utilizando a enzima KAPA HiFi HotStart (Kapa Biosystems) e *primers* fornecidos no kit Nextera XT (Illumina). As amostras foram purificadas utilizando AMPures *beads* (Beckman Coulter) de acordo com as instruções do fabricante e quantificadas por fluorescência utilizando o reagente picogreen (ThermoFischer). Foi realizado PCR em tempo real utilizando o kit de quantificação de bibliotecas da Kapa Biosystems. Ao final, as amostras foram diluídas para a concentração de 4 nM e desnaturadas por calor e 0.1M NaOH e adicionadas na célula de sequenciamento. Foi utilizado cartucho MiSeq Reagent Kit v2 2x250bp.

Processamento dos dados de sequenciamento. As sequências pareadas obtidas no MiSeq foram montadas utilizando-se o software PEAR 0.9.6 (ZHANG et al., 2014). Para todas as outras análises foram utilizados softwares desenvolvidos no laboratório utilizando a linguagem Python. As sequências dos insertos foram extraídas e contadas para se estimar a frequência de cada sequência única. Sequências de DNA que foram observadas apenas uma vez (*singletons*) foram descartadas. As sequências foram

Tabela 3. Sequência dos *primers* utilizados para a amplificação da região codificando os insertos na biblioteca construída. A região sublinhada corresponde à específica para o vetor, sendo seguida a montante (*upstream*) de degeneração para adicionar variabilidade ao sequenciamento e região específica para o kit Nextera XT.

Nome	Sequência
Fw-0N	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG <u>CTGCGCAACACGATGACC</u>
Fw-1N	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNCTGCGCAACACGATGACC
Fw-2N	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNCTGCGCAACACGATGACC
Fw-3N	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNCTGCGCAACACGATGACC
Rv-0N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG <u>GTTCCAGTGGGTCCGGATA</u>
Rv-1N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNGTTCCAGTGGGTCCGGATA
Rv-2N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNGTTCCAGTGGGTCCGGATA
Rv-3N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNN <u>GTTCCAGTGGGTCCGGATA</u>

alinhadas ao genoma do *T. cruzi* (cepas CL Brener, Sylvio X10, DM28c e Marinkellei) e aquelas que não apresentavam pelo menos 90% de identidade foram descartadas. Foram identificadas as sequências que mantinham o passo correto de leitura e que não continham códons de parada. Estas sequências que formavam peptídeos viáveis foram então comparadas ao proteoma do *T. cruzi* (cepas CL Brener, Sylvio X10, DM28c e Marinkellei) utilizando BlastP (CAMACHO et al., 2009) em uma instância local. Peptídeos com menos de 60% de similaridade com o proteoma foram descartados. Os dados foram então tabulados e analisados.

Agrupamento dos peptídeos em epítopos. Os peptídeos foram organizados por tamanho crescente. O primeiro é retirado da lista de sequências e forma um novo epítopo que é comparado contra todos os outros peptídeos pela distância parcial de Levenshtein (LEVENSHTEIN, 1966) como implementada no pacote FuzzyWuzzy ("FuzzyWuzzy: Fuzzy String Matching in Python - ChairNerd", [s.d.]). Se a correspondência entre as sequências for igual ou maior que 80%, a nova sequência é retirada da lista e adicionada ao epítopo. O processo é feito até que todas as sequências tenham sido adicionadas a um epítopo. Os peptídeos pertencentes a cada epítopo são alinhados entre si utilizando o software MAFFT (KATOH; STANDLEY, 2013) e a sequência consenso é determinada utilizando o software HMMER (MISTRY et al., 2013). Essas sequências são utilizadas para representar cada epítopo e um novo ciclo de agrupamento se inicia. O processo continua até que novos agrupamentos não

possam mais ser realizados.

**Análise de hidrofilicidade.** Utilizamos o índice criado por PARKER; GUO; HODGES (1986) para o cálculo da hidrofilicidade dos peptídeos identificados tanto na biblioteca *naïve* quanto após as seleções.

Análise de sequências repetitivas. As sequências repetitivas presentes no proteoma do *T. cruzi* (cepas CL Brener, Sylvio X10, DM28c e Marinkellei) foram identificadas utilizando o software XSTREAM (NEWMAN; COOPER, 2007). Foi gerado um perfil HMM para cada região repetitiva encontrada e os peptídeos obtidos durante as seleções foram comparados utilizando o software HMMER (MISTRY et al., 2013). Foram considerados como positivos aqueles peptídeos que obtiveram *E-value* menor ou igual a 10-5.

### Resultados

### Produção e caracterização da biblioteca.

Para produzir a biblioteca genômica do parasita *T. cruzi*, separamos o processo em etapas distintas: produção em larga escala do vetor de clonagem, produção dos insertos, otimização das condições de ligação e a ligação/transformação/produção da biblioteca. Dois vetores para a produção das bibliotecas estavam disponíveis em nosso laboratório, um que permite a fusão na proteína VIII e outra na proteína III, ambas de superfície no bacteriófago. Enquanto esta última está presente com apenas cinco cópias por partícula viral, a proteína VIII poder chegar a três mil cópias em uma única partícula; isso tem consequências importantes como aumento significativo da probabilidade de expressão das proteínas fusionadas e aumento da avidez na interação, permitindo identificar pares de interação com menor afinidade e aumento da pressão seletiva para fagos que expressem proteínas no passo de leitura correto (JACOBSSON; FRYKBERG, 1996). Bibliotecas foram produzidas utilizando ambos os vetores, sendo que os resultados com a biblioteca em pVIII foram superiores, assim, estes são os dados que iremos apresentar nesta tese.

Para a geração dos insertos, duas cepas se destacavam como candidatas para servir de fonte do DNA genômico: CL Brener e Sylvio X10/1. Isto porque, ambas tiveram seu genoma sequenciado e anotado (EL-SAYED et al., 2005; FRANZÉN et al., 2011). Porém, pelo fato da cepa CL Brener ser híbrida, possivelmente oriunda da fusão de duas outras cepas de *T. cruzi*, optamos por utilizar a cepa Sylvio X10/1 como fonte de material genético para construção da biblioteca. Em teoria, precisaríamos do dobro de insertos nas nossas bibliotecas para cobrir o genoma da cepa CL Brener, em comparação com a outra cepa. Há também a vantagem de esta cepa apresentar um genoma mais compacto e com menos regiões repetitivas quando comparada à cepa CL Brener, a primeira a ter seu genoma sequenciado (FRANZÉN et al., 2011).

Para a geração da biblioteca é necessário que os cromossomos que compõem o material genético do parasita sejam quebrados em fragmentos menores. É possível fazer isso utilizando diferentes técnicas como digestão enzimática ou fragmentação mecânica. Neste trabalho, após testarmos algumas opções disponíveis, optamos por utilizar equipamento COVARIS S2 para este fim. Neste equipamento a amostra fica dentro de um tubo imerso em banho de água enquanto ondas ultrassônicas são geradas pelo aparelho causando a quebra do DNA. O uso deste equipamento nos dá algumas vantagens sobre outros métodos, como menor dano por calor ao DNA pois o banho possui temperatura controlada, também evita contaminação cruzada como é frequente no uso de nebulizador e aparelhos ultrassônicos que utilizam sondas que entram em contato com a amostra, e principalmente, os resultados obtidos no aparelho são reprodutíveis.

O DNA genômico da cepa Sylvio X10/1 foi extraído, fragmentado e. insertos na faixa de 100 e 500 pb foram eluídos de gel (Figura 18A), assim temos preferencialmente insertos de aproximadamente 33 a 130 aminoácidos. Fragmentos muito pequenos (codificando peptídeos menores que 15 aminoácidos) não são desejados pois geram dificuldades no momento de identificar a qual gene o inserto pertence. Já fragmentos muito grandes têm maior probabilidade de conter códon de parada ou incluírem regiões

não transcritas ou/e intergênicas.

Após testar diferentes razões molares (Figura 18B), os insertos obtidos foram ligados ao vetor pG8SAET obtendo-se 4.4x10<sup>8</sup> transformantes. Para uma análise profunda da biblioteca produzida, realizamos o sequenciamento em larga escala, o

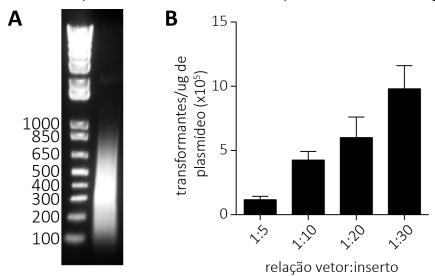


Figura 18. Análise por eletroforese em gel de agarose do material genético do *T. cruzi* (cepa Sylvio X10) fragmentado utilizando Covaris S2. (B) Testes de eficiência de ligação o vetor pG8SAET digerido com a enzima Eco105I aos fragmentos de DNA de *T. cruzi* em diferentes razões molares entre vetor e inserto.

que possibilita a análise de centenas de milhares de sequências simultaneamente.

Neste experimento, foram obtidas um total de 261.178 sequências e, tendo em vista a grande quantidade de dados, um protocolo de bioinformática é essencial para o processamento dos resultados. Dessa maneira, desenvolvemos programas específicos para as nossas necessidades assim como utilizamos alguns já descritos na literatura. Como nossos dados foram obtidos sequenciando-se ambas as fitas dos insertos (paired-end read), o primeiro passo é a montagem dessas duas sequências curtas em apenas uma mais longa. Utilizamos o software PEAR (PEAR - Paired-End reAd mergeR, v. 0.9.6) (ZHANG et al., 2014) que foi capaz de montar 98.2% das sequências (Figura 19A). As sequências que flanqueiam os insertos foram então identificadas e removidas. Nos casos onde essas sequências não puderam ser identificadas, a leitura de DNA foi desprezada. Após estes passos, 85,2% das sequências continuaram sendo analisadas. Para avaliar a profundidade do sequenciamento e diversidade da

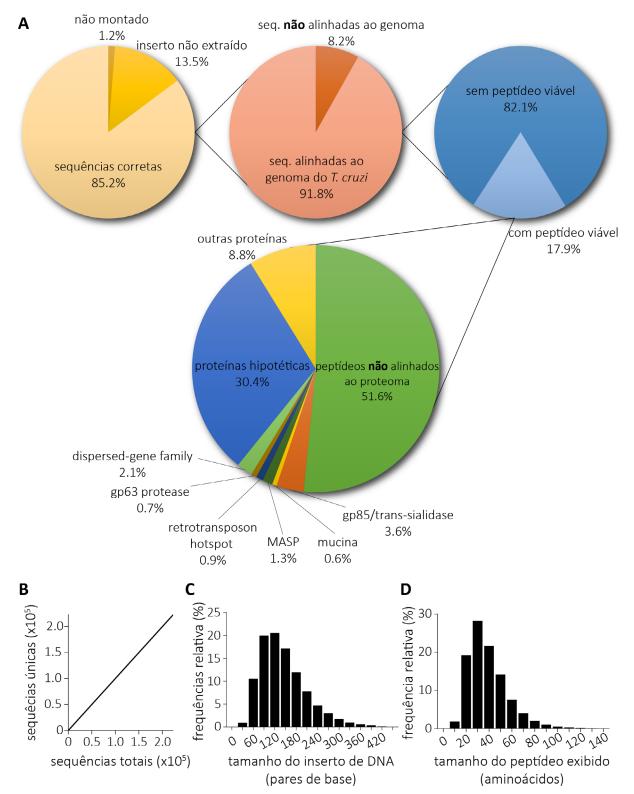


Figura 19. Análise do sequenciamento da biblioteca naïve. (A) Após o sequenciamento, as leituras de DNA que puderam ser montadas e ter seus insertos extraídos foram alinhadas ao genoma do parasita. Estas sequências foram examinadas para a presença de passos de leitura abertos e depois comparadas ao proteoma do parasita para identificação da proteína correspondente. (B) Curva de saturação do sequenciamento da biblioteca mostrando o número de sequências únicas e sequências totais. (C) Distribuição de frequência do tamanho dos insertos de DNA identificados na biblioteca. (D) Distribuição de frequência do tamanho dos peptídeos identificados na biblioteca.

biblioteca, montamos a curva de saturação para correlacionar o número de sequências únicas com o número de sequências totais (Figura 19B). O sequenciamento se mostrou longe da saturação, pois observamos que a quantidade de sequências únicas (eixo Y) é praticamente a mesma da de sequências totais (eixo X). Esse resultado é esperado, já que a diversidade da biblioteca é estimada na ordem de 10<sup>8</sup> e o sequenciamento cobriu apenas 10<sup>5</sup> sequências. Também avaliamos o tamanho dos insertos de DNA na biblioteca (Figura 19C). Em média os insertos têm extensão de 143 pb (mediana = 132 e máximo = 423) e 50% dos clones têm tamanho entre 96 e 178 pb.

A seguir, realizamos o alinhamento no genoma do *T. cruzi*, para descartar sequências contaminantes (p.ex., da bactéria vetor *Escherichia coli*). Devido à grande fragmentação dos dados de genômica do parasita, em grande parte decorrente da natureza repetitiva do genoma, combinamos os dados das quatro cepas já sequenciadas e disponíveis para essa análise: CL Brener (EL-SAYED et al., 2005), Sylvio X10 (FRANZÉN et al., 2011), DM28c (GRISARD et al., 2014) e Marinkellei (FRANZÉN et al., 2012). No total, 91.8% dos insertos (que passaram pelos filtros anteriores) puderam ser alinhados utilizando-se um limite inferior de 90% de identidade. Muitas das sequências não alinhadas correspondem a DNA derivado do cinetoplasto do parasita, assim como sequências curtas onde o software de alinhamento (BLAT) não foi capaz de realizar o alinhamento.

Aqui vale uma observação inerente do sistema utilizado: as inserções ocorrem entre a sequência do peptídeo sinal e da proteína VIII do capsídeo viral. Dessa maneira, os insertos devem ter uma quantidade específica de nucleotídeos para que o passo de leitura não se altere e a proteína VIII seja produzida corretamente. Há também a possibilidade da presença de códons de parada, impedindo a produção correta da proteína e consequentemente não havendo apresentação na superfície do fago. Apenas 17,9% dos insertos resultam na produção de proteína VIII viável, com peptídeos com tamanho médio de 38 aminoácidos (mediana = 35; mínimo = 3; máximo = 139). (Figura 19D). Nota-se uma diminuição no tamanho dos insertos quando comparados

ao que era esperado pelo padrão de fragmentação, onde observamos uma média de tamanho de aproximadamente 300 pb (100 aminoácidos). É possível que fragmentos menores de DNA tenham maior eficiência de ligação quando comparados aos que fragmentos maiores, levando a essa alteração.

Os peptídeos contidos na biblioteca podem recriar regiões de proteínas existentes no *T. cruzi*, porém é também possível que sejam decorrentes de passos de leitura formados artificialmente, que não existem nos genes do parasita, dessa maneira, não tendo qualquer relação com a sua biologia (até onde conhecemos). Dessa forma, todos os peptídeos identificados foram comparados com os proteomas disponíveis. Novamente, utilizamos informações de quatro cepas diferentes para evitar que algumas sequências não fossem identificadas por falta de anotação. Consideramos que peptídeos com pelo menos 60% de similaridade com sequências do banco de dados poderiam ser representantes fiéis das proteínas correspondentes. Utilizamos um valor de corte relativamente baixo, por causa do grande número de famílias multigênicas, com membros que compartilham similaridade nesta faixa, ou ainda mais baixa. Quase metade dos peptídeos tinha correspondência com proteínas presentes no banco de dados, sendo mais da metade destas anotadas como "hipotéticas", ou seja, moléculas cuja função não é sabida e nem pode ser inferida através de similaridade com outras proteínas conhecidas (Figura 19A).

Por fim, podemos concluir que a biblioteca tem grande cobertura: o tamanho do genoma do *T. cruzi* (cepa Sylvio X10) foi determinado em 4,4 x 10<sup>7</sup> pb; nossa biblioteca tem 4,4 x 10<sup>8</sup> clones com uma média de tamanho de 143 pb de extensão, podendo-se estimar então um tamanho total de 4,4 x 10<sup>8</sup> x 143 = 6,31 x 10<sup>10</sup> pb, correspondendo a uma cobertura de mais de 1.400 vezes o genoma. Notadamente, nem todos os clones formam peptídeos viáveis e quando formam nem sempre tem correspondência com as proteínas do parasita. Porém, mesmo assumindo-se que apenas ~8,5% dos transformantes contêm inserto que produzem proteínas de *T. cruzi*, ainda assim a cobertura seria de pelo menos 120 vezes. Desta forma é seguro afirmar

que a biblioteca possivelmente contém todos (ou ao menos a maioria) dos fragmentos proteicos existentes no *T. cruzi*.

### Seleção da biblioteca contra soro de pacientes

Uma vantagem da tecnologia de *phage display* é o acoplamento entre fenótipo e genótipo: ou seja, a parte funcional (proteína - fenótipo) fica presente na mesma partícula onde está a sequência codificadora (DNA – genótipo). Desta forma, podemos

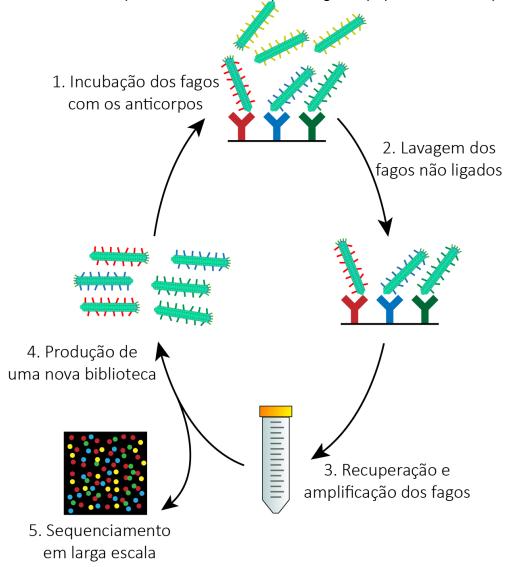


Figura 20. Esquema de seleção *in vitro* dos fagos presentes na biblioteca. As partículas virais são incubadas com as imunoglobulinas previamente imobilizadas. Os clones que não tiverem afinidades por estas são removidos após extensas lavagens. Os fagos remanescentes são recuperados por infecção de bactérias do tipo *E. coli* e depois amplificados para gerar uma nova biblioteca que será submetida a mais um ciclo de seleção. Após três ciclos, a biblioteca é submetida a sequenciamento de larga escala para identificação dos clones e suas abundâncias relativas na biblioteca.

utilizar a biblioteca de *phage display* para experimentos de evolução *in vitro* onde selecionamos os clones com as características de interesse.

Neste projeto, utilizamos um protocolo para selecionar clones que se ligam às imunoglobulinas (IgG) purificadas de soro de pacientes portadores da doença de Chagas (Figura 20). As IgGs são imobilizadas em uma placa de poliestireno de 96 poços e depois incubadas com a biblioteca. Em solução está presente IgG derivada de indivíduos soronegativos para *T. cruzi*, competindo com as imunoglobulinas imobilizadas e, dessa maneira, impedindo que peptídeos que mimetizem antígenos de outros patógenos ou sequências que se ligam de forma inespecífica aos anticorpos sejam selecionadas. Os poços são lavados e os fagos remanescentes são amplificados e utilizados em um novo ciclo de seleção. O processo é então repetido mais duas vezes para que haja enriquecimento de sequências com as características desejadas.

Para a seleção, utilizamos amostras de soro de 80 indivíduos divididos em quatro grupos: controle, assintomático, CCC leve e CCC grave. O grupo controle consiste de indivíduos doadores de sangue que são soronegativos para o *T. cruzi*; os assintomáticos são indivíduos que ao doarem sangue são identificados como soropositivos, no entanto não apresentam sintomas da doença; CCC leve são soropositivos e apresentam alterações no eletrocardiograma, já CCC graves, além destas alterações também apresentam insuficiência cardíaca definida pela baixa fração de ejeção do ventrículo esquerdo (menor que 40%). Cada grupo foi ainda subdividido em dois subgrupos com 10 pacientes cada (Figura 21A), de forma que as seleções foram realizadas independentemente nestes grupos (replicatas biológicas). A mediana das idades dos pacientes variou de 48 a 60 anos nos grupos, no entanto não há diferença estatística significativa entre eles (Figura 21B, p-valor > 0.05 teste Kruskal-Wallis).

O soro dos pacientes de cada um dos oito grupos foi combinado, as imunoglobulinas presentes foram purificadas e posteriormente analisadas por eletroforese em gel de poliacrilamida para verificação da pureza (Figura 21C). O

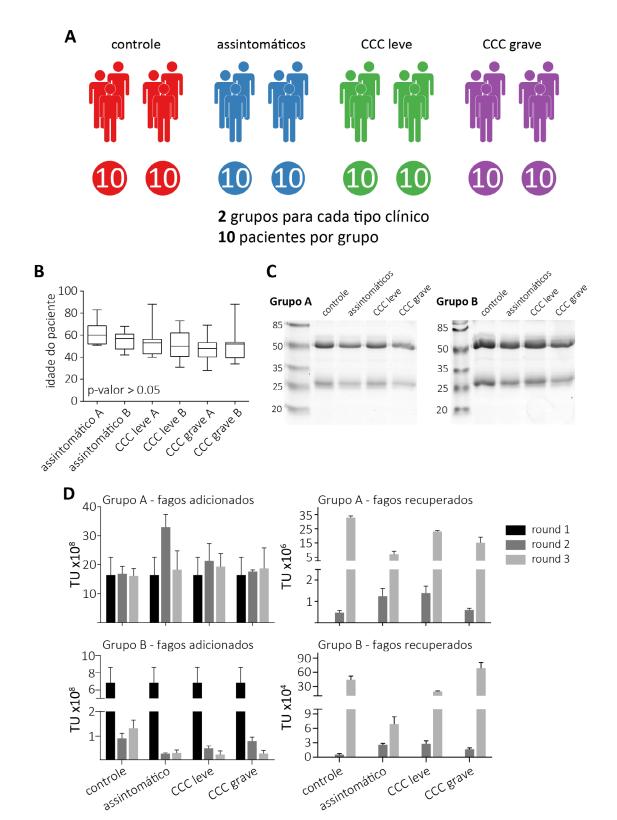


Figura 21. (A) Esquema representativo dos grupos de pacientes que tiveram as imunoglobulinas do soro analisadas neste estudo (B) Diagrama de caixas representando a idade dos pacientes de cada grupo. As caixas representam os percentis 25 e 75 com um traço na mediana, os whiskers representam o valor máximo e mínimo. Não há diferença estatística significativa entre os grupos (p-valor > 0,05 - ANOVA) (C) Eletroforese em gel de poliacrilamida e SDS das imunoglobulinas purificadas de cada grupo de pacientes. O padrão de bandas é característico de IgG, com a banda da cadeia pesada com aproximadamente 50 kDa e da cadeia leve com aproximadamente 25 kDa. (D) Número de fagos adicionados e recuperados em cada ciclo de seleção.

material purificado foi então utilizado para realização das seleções como descrito anteriormente. Três ciclos foram realizados e ao fim do segundo e terceiro ciclos, a quantidade de fagos retida nos poços após as lavagens foi quantificada (Figura 21D). No primeiro ciclo a quantificação não foi realizada para evitar perda de antígenos únicos, já que estes poderiam estar presentes em poucas cópias (ou mesmo, uma única cópia) e a remoção para quantificação poderia impedir sua propagação para os ciclos subsequentes. Na seleção, espera-se que o número de fagos recuperados aumente ao longo dos ciclos, já que estamos selecionando a população de partículas virais pela capacidade de aderir à proteína imobilizada. Também pode haver enriquecimento de clones que apresentem alguma vantagem de crescimento e, dessa maneira, também observamos um aumento na recuperação no grupo controle. No segundo experimento de seleção, realizado com o grupo B de cada tipo de paciente, observamos que a recuperação de fagos foi em média 50 vezes menor do que no experimento com o grupo A. Essa diferença pode ocorrer devido a variações na infectividade e rendimento da produção de fagos da bactéria em diferentes experimentos.

O enriquecimento de fagos que observamos sugere que a seleção foi bem sucedida. Por isso, decidimos analisar por sequenciamento de larga escala os insertos selecionados. Da mesma maneira como foi realizada com a biblioteca *naïve*, após o terceiro ciclo de seleção as sub-bibliotecas geradas foram sequenciadas e submetidas às mesmas formas de análise (Figura 22), com adição de uma nova etapa de agrupamento de peptídeos por epítopo. Após montagem das leituras de ambas as extremidades de cada clone e extração da sequência do inserto, ficamos com média de 672,566 leituras de DNA por amostras (mediana = 611.202; máximo = 1.130.870; mínimo = 449.598 - Tabela 4). Dada a grande cobertura, optamos, como feito em trabalho anterior (DIAS-NETO et al., 2009), por excluir sequências de inserto que apareciam apenas uma vez em nossa amostra (*singletons*). Muitas destas sequências com baixíssima abundância podem ser frutos de erros de sequenciamento ou mutações introduzidas pela etapa de amplificação por PCR. Estas sequências representam menos de um terço das leituras (entre de 4 a 30%).

Após este primeiro filtro, as sequências foram alinhadas ao genoma do parasita, novamente utilizando quatro cepas diferentes e o critério de 90% de identidade. Em média, 97% das sequencias puderam se alinhadas (Tabela 5) e são, portanto, de *T. cruzi*, com exceção do grupo assintomático B que apresentou apenas 70% das sequências selecionadas mapeando no genoma do parasita, devido a um clone abundante que não pode ser alinhado.

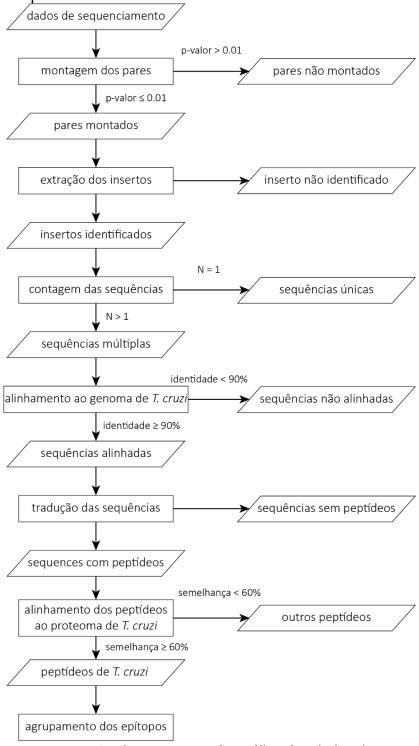


Figura 22. Fluxograma representando o processo de análise dos dados de sequenciamento.

Tabela 4. Estatísticas do número de sequências analisadas de cada amostra após seleção e sequenciamento.

grupo	# de seq. com inserto identificado	# de seq. após remoção de singletons	# de insertos diferentes	# de insertos diferentes após remoção de singletons
controle A	449.598	323.430 (71,9%)	154.058	27.890 (18,1%)
controle B	573.203	477.715 (83,3%)	169.027	31.051 (18,4%)
assintomático A	454.477	316.501 (69,6%)	260.077	48.003 (18,5%)
assintomático B	742.131	653.373 (88,0%)	243.561	42.580 (17,5%)
CCC leve A	865.965	653.891 (75,5%)	112.168	16.680 (14,9%)
CCC leve B	649.201	525.747 (81,0%)	105.324	16.566 (15,7%)
CCC grave A	1.130.870	929.889 (82,2%)	145.086	21.632 (14,9%)
CCC grave B	515.085	494.513 (96,0%)	24.012	3.440 (14,3%)

Tabela 5. Número de sequências totais e sequências diferentes alinhadas ao genoma do *T. cruzi* (identidade ≥ 90%).

grupo	# de seq.	# de seq. alinhadas	# de seq. diferentes	# de seq.
				diferentes
				alinhadas
controle A	323.430	312.947 (96,8%)	27.890	26.578 (95,3%)
controle B	477.715	473.002 (99,0%)	16.680	15.666 (93,9%)
assintomático A	316.501	309.327 (97,7%)	31.051	29.893 (96,3%)
assintomático B	653.373	459.321 (70,3%)	16.566	14.418 (87,0%)
CCC leve A	653.891	639.351 (97,8%)	48.003	45.819 (95,5%)
CCC leve B	525.747	509.561 (96,9%)	21.632	20.268 (93,7%)
CCC grave A	929.889	910.692 (97,9%)	42.580	40.199 (94,4%)
CCC grave B	494.513	482.396 (97,5%)	3.440	2.427 (70,6%)

As sequências alinhadas ao genoma foram então analisadas quanto à presença de passos de leitura abertos (ORFs), ou seja, se tinham a capacidade de codificar um peptídeo que seria expresso em fusão com a proteína VIII do capsídeo viral (Tabela 6). Na biblioteca *naïve* observamos que menos de 18% dos clones possuíam peptídeos viáveis, no entanto após três ciclos de seleção houve um aumento significativo deste valor, chegando a até 99,2% no caso da seleção contra o grupo CCC grave. Esse resultado evidencia a pressão seletiva por clones que apresentem algum peptídeo (epítopo) em sua superfície reconhecido pelas IgGs utilizadas, em detrimento de bacteriófagos sem peptídeos exógenos. Portanto, a presença de um grande número de clones que não produzem proteínas de *T. cruzi* na nossa biblioteca foi uma vantagem durante a seleção, pois serviu de indicador para mensurar o sucesso de um

experimento. Nota-se que os grupos controle A e B, tiveram o menor enriquecimento para clones com algum peptídeo, já que a princípio não devem reconhecer peptídeos derivados de *T. cruzi* e também porque em solução estavam presentes IgGs do próprio grupo, evitando assim a ligação dos fagos.

### Peptídeos selecionados têm uma larga faixa de tamanho

Epítopos de anticorpos podem variar desde pequenas sequências lineares de alguns aminoácidos até grandes sequências dependentes de conformação. Em nossa Tabela 6. Análise do passo-de-leitura dos clones alinhados ao genoma do parasita.

grupo	# de seq. alinhadas contendo ORFs	# de seq. diferentes alinhadas contendo ORFs	# de peptídeos diferentes
controle A	145.470 (46.5%)	15,695 (56.3%)	13,096
controle B	60,696 (12.8%)	4,079 (24.5%)	1,971
assintomático A	192,233 (62.1%)	20,418 (65.8%)	16,123
assintomático B	350,887 (76.4%)	9,861 (59.5%)	6,239
CCC leve A	419,853 (65.7%)	30,819 (64.2%)	24,833
CCC leve B	245,662 (48.2%)	13,452 (62.2%)	10,252
CCC grave A	668,137 (73.4%)	25,842 (60.7%)	20,027
CCC grave B	478,512 (99.2%)	2,150 (62.5%)	1,641

biblioteca possuímos clones que podem potencialmente satisfazer ambos os requisitos, no entanto, em casos de epítopos conformacionais em que regiões determinantes para a ligação se encontrem separadas por algumas centenas de aminoácidos, não seria possível mapear a interação. Analisando a distribuição de tamanho dos peptídeos selecionados (Figura 23) podemos notar que o perfil de tamanho difere da biblioteca naïve, havendo uma maior presença de peptídeos grandes (>50 aminoácidos) do que anteriormente, mas a faixa predominante continua entre 30 e 40 aminoácidos. Isso sugere que não há um viés com tendência à seleção de sequências curtas, que podem ser mais facilmente produzidas e expressas em relação a sequencias maiores. Esse padrão também pode significar que peptídeos maiores podem estar levando a identificação de epítopos conformacionais ou que a presença de um segmento maior da proteína estabiliza a sequência linear na conformação correta para o reconhecimento pelo anticorpo.

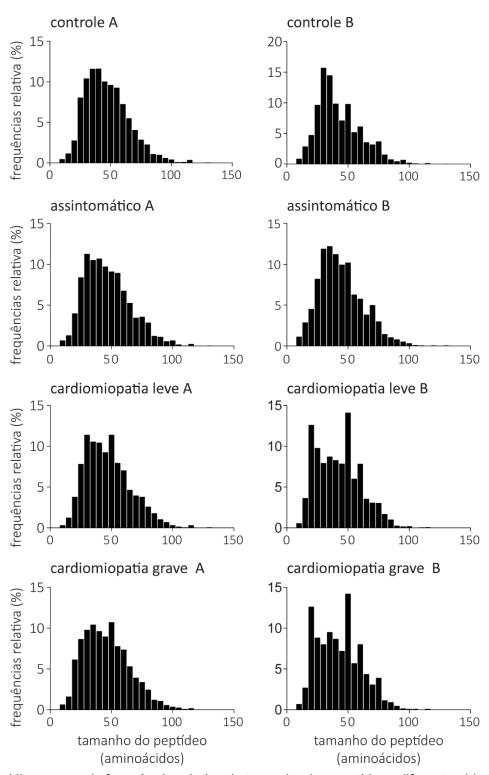


Figura 23. Histograma de frequência relativa do tamanho dos peptídeos diferentes identificados em cada amostra após seleção e sequenciamento.

### Epítopos identificados são hidrofílicos

Diferente dos linfócitos T que necessitam que antígenos sejam processados e apresentados na forma de pequenos peptídeos, os linfócitos B e seus receptores, durante o processo de geração de anticorpos, reconhecem os antígenos na sua forma nativa, fazendo com que frequentemente reconheçam sequências presentes na superfície da proteína. Como as regiões das proteínas que ficam em contato com a água tendem a possuir características hidrofílicas, essa característica pode ser um bom indicativo da possibilidade de um peptídeo ser um epítopo.

Utilizando uma escala de hidrofilicidade desenvolvida para avaliação de regiões antigênicas (Figura 24A) (PARKER; GUO; HODGES, 1986), avaliamos essa característica físico-química na biblioteca *naïve* e também após os experimentos de seleção. Nota-se um aumento significativo da hidrofilicidade quando selecionamos a biblioteca contra as imunoglobulinas (Figura 24B). Enquanto na biblioteca *naïve* o índice segue uma distribuição normal com um pico em 60 (Figura 24C) e com a presença de peptídeos hidrofóbicos (índice > 0), após as seleções há um favorecimento de sequências mais hidrofílicas, alterando a distribuição de frequência (Figura 24D). Esse resultado indica que os peptídeos identificados seguem o padrão esperado.

### Enriquecimento de peptídeos de *T. cruzi* durante as seleções

Apesar dos peptídeos identificados até o momento serem derivados do DNA do *T. cruzi*, é necessário verificar se estas sequências ocorrem no proteoma do parasita ou são resultado da formação artificial de um passo de leitura aberto. Utilizamos a ferramenta BlastP (CAMACHO et al., 2009) para comparar os peptídeos às proteínas anotadas do parasita (cepas CL Brener, Sylvio X10, DM28c e Marinkellei). Novamente consideramos os peptídeos com mais de 60% de similaridade como sendo correspondentes às proteínas encontradas. A distribuição dos peptídeos com relação à sua similaridade com proteínas de *T. cruzi* (Figura 25) é bimodal (exceção ao grupo controle B), onde há um primeiro pico na região de aproximadamente 40% de similaridade e um segundo pico em 100% de similaridade. Isso sugere que correspondências aleatórias surgem

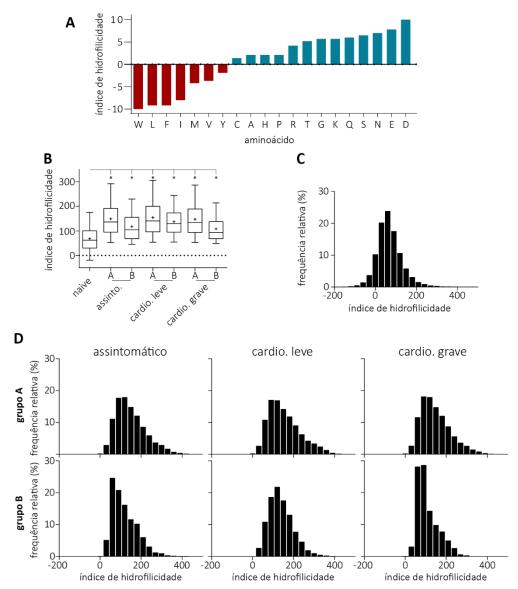


Figura 24. (A) Escala utilizada para o cálculo da hidrofilicidade dos peptídeos (B) Diagrama de caixas representando a hidrofilicidade dos peptídeos presentes na biblioteca não selecionada (naïve) e após as seleções. As caixas representam o percentil 25 e 75, com um traço na mediana e uma cruz na média, e os whiskers representam os percentis 5 e 95. (C) Histograma de distribuição do índice de hidrofilicidade dos peptídeos da biblioteca naïve (D) Histograma de distribuição do índice de hidrofilicidade dos peptídeos identificados após a seleção.

em torno desse primeiro pico da distribuição.

Notadamente, as seleções dos pacientes infectados dos grupos B tiveram uma proporção maior de peptídeos próximos a 100% de correspondência e muito poucos na região de 40%, enquanto o controle B teve uma proporção muito baixa de peptídeos próximos a 100%. Esse resultado indica que a seleção nesse grupo foi possivelmente mais estringente, favorecendo sequências naturais do parasita e impedindo que quaisquer destes peptídeos se ligasse no grupo controle. Durante os

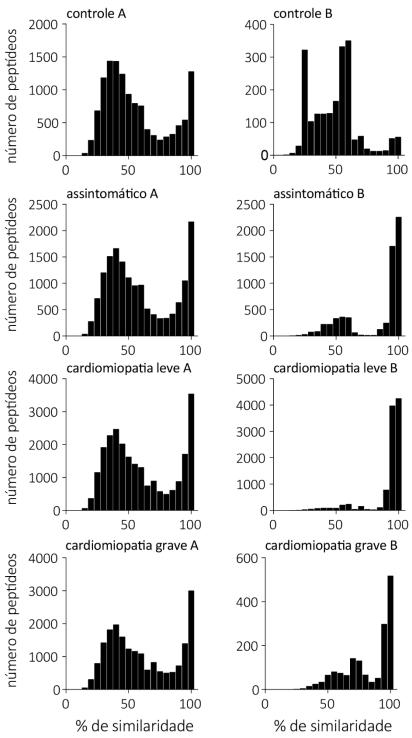


Figura 25. Histogramas representando a frequência de peptídeos com determinada similaridade a proteínas anotadas no proteoma de *T. cruzi*. A relação entre os peptídeos e as proteínas foi determinada utilizando o *software* BlastP (CAMACHO et al., 2009).

experimentos do grupo B notamos também uma quantidade bem menor de fagos recuperados ao final de cada ciclo (Figura 21D), sugerindo o mesmo efeito.

Quando analisamos todos os peptídeos diferentes encontrados, sem levar em consideração a abundância de cada um (Tabela 7), notamos que os grupos controle

têm uma baixa proporção de sequências similares a proteínas do parasita (31,8% e 20,8%). Já nos grupos infectados essa proporção é maior, chegando a 92.3% no grupo B dos pacientes que possuem CCC leve. Quando se leva em consideração a abundância de cada peptídeo após o sequenciamento (Tabela 8) o efeito fica mais pronunciado, com os controles agora com valores de 27,4% e 16,6% de peptídeos com correspondência positiva e um aumento desse valor nos grupos de pacientes infectados. O grupo compreendendo os pacientes assintomáticos apresenta consistentemente uma menor proporção de peptídeos com correspondência positiva quando comparado aos grupos formados por cardiopatas.

Em seguida, analisamos a distribuição dos peptídeos selecionados e seu mapeamento nas grandes famílias gênicas do parasita. Curiosamente, poucos peptídeos mapearam com as mucinas, DGF1, RHS, gp63 e MASP. Assim, das grandes famílias gênicas analisadas, observamos que a família gp85/trans-sialidase aparenta ter uma mais alta imunogenicidade (Tabela 7 e Tabela 8). É importante ressaltar que mucinas, por exemplo, têm extensas cadeias de carboidratos adicionadas a sua estrutura após a tradução, e esta característica não é representada na biblioteca. Outro fato relevante, é o fato das proteínas hipotéticas estarem bastante representadas após as seleções. Esse resultado não chega a ser surpreendente visto que estas proteínas não possuem semelhança com qualquer proteína humana, assim sendo potencialmente imunogênicas. É de se ressaltar também o aumento na proporção de leituras de peptídeos de gp85/trans-sialidases nos pacientes com CCC grave - 63,5% e 88% - enquanto os valores para os outros pacientes são mais baixos.

Tabela 7. Peptídeos diferentes alinhados ao proteoma do 7. cruzi utilizando BlastP e estatísticas relacionadas às grandes famílias gênicas. Apenas peptídeos derivados de sequências de DNA alinhadas ao genoma foram considerados.

Surface protease (gp63) <sup>2</sup>	17 (0.4%)	0 (0.0%)	20 (0.3%)	23 (0.5%)	37 (0.4%)	23 (0.2%)	25 (0.3%)	(%0:0)
Mucin- associated protein (MASP) <sup>2</sup>	29 (0.7%)	6 (1.5%)	37 (0.6%)	2 (0.0%)	185 (1.8%)	2 (0.0%)	45 (0.5%)	0.0%)
Proteína Hipotética²	3,527 (84.6%)	318 (77.8%)	4,509 (71.8%)	966 (21.1%)	6,968 (69.6%)	5,438 (57.5%)	6,202 (72.5%)	388 (29.0%)
gp85/trans- sialidase²	98 (2.4%)	52 (12.7%)	155 (2.5%)	318 (7.0%)	380 (3.8%)	1,173 (12.4%)	381 (4.5%)	212 (15.8%)
RHS protein <sup>2</sup>	31 (0.7%)	1 (0.2%)	37 (0.6%)	2 (0.0%)	54 (0.5%)	1 (0.0%)	40 (0.5%)	(0.1%)
$Mucina^2$	10 (0.2%)	0.0%)	5 (0.1%)	1 (0.0%)	31 (0.3%)	3 (0.0%)	14 (0.2%)	2 (0.1%)
Dispersed gene family 1 (DGF1) <sup>2</sup>	66 (1.6%)	2 (0.5%)	80 (1.3%)	2 (0.0%)	118 (1.2%)	1 (0.0%)	95 (1.1%)	(0.1%)
# peptídeos alinhados	4,170 (31.8%)	409 (20.8%)	6,282 (39.0%)	4,570 (73.2%)	10,011 (40.3%)	9,462 (92.3%)	8,557 (42.7%)	1,340 (81.7%)
# peptídeos diferentes	13,096	1,971	16,123	6,239	24,833	10,252	20,027	1,641
grupo	controle A	controle B	assintomático A	assintomático B	CCC leve A	CCC leve B	CCC grave A	CCC grave B

Porcentagem relativa ao número de peptídeos diferentes derivados de sequências de DNA alinhadas ao genoma.
Porcentagem relativa ao número de peptídeos diferentes alinhados ao proteoma.

Nesta análise é levada em consideração a frequência com que aparece cada peptídeo. Apenas peptídeos derivados de sequências de DNA Tabela 8. Leituras de peptídeos alinhados ao proteoma do T. cruzi utilizando BlastP e estatísticas relacionadas às grandes famílias gênicas. alinhadas ao genoma foram considerados

Surface protease (gp63)²	82 (0.2%)	0 (%0.0)	141 (0.2%)	1,090 (0.6%)	233 (0.1%)	1,502 (0.6%)	285 (0.1%)	0 (0.0%)
Mucin- associated protein (MASP) <sup>2</sup>	145 (0.4%)	60 (0.6%)	215 (0.3%)	10 (0.0%)	11,736 (5.9%)	17 (0.0%)	314 (0.1%)	0 (0.0%)
Proteína Hipotética²	34,508 (86.6%)	6,600 (65.5%)	49,840 (66.9%)	61,292 (31.2%)	139,274 (70.5%)	112,845 (48.8%)	154,655 (30.9%)	35,211 (8.4%)
gp85/trans- sialidase²	725 (1.8%)	3,204 (31.8%)	1,393 (1.9%)	9,569 (4.9%)	6,044 (3.1%)	25,686 (11.1%)	318,176 (63.5%)	368,513 (88.0%)
RHS protein²	224 (0.6%)	3 (0.0%)	269 (0.4%)	57 (0.0%)	693 (0.4%)	7 (0.0%)	510 (0.1%)	3 (0.0%)
Mucina²	34 (0.1%)	0.0%)	18 (0.0%)	3 (0.0%)	176 (0.1%)	(%0.0)	56 (0.0%)	7 (0.0%)
Dispersed gene family 1 <sup>2</sup>	297 (0.7%)	4 (0.0%)	368 (0.5%)	6 (0.0%)	763 (0.4%)	2 (0.0%)	606 (0.1%)	5 (0.0%)
# leituras de peptídeo alinhadas	39,868 (27.4%)	10,073 (16.6%)	74,457 (38.7%)	196,409 (56.0%)	197,653 (47.1%)	231,240 (94.1%)	500,753 (74.9%)	418,701 (87.5%)
# leituras de peptídeo	145,470	969'09	192,233	350,887	419,853	245,662	668,137	478,512
grupo	controle A	controle B	assintomático A	assintomático B	CCC leve A	CCC leve B	CCC grave A	CCC grave B

1. Porcentagem relativa ao número de leituras de DNA contendo peptídeos que foram alinhadas ao genoma.

# Agrupamento dos peptídeos em epítopos

Quando construímos uma biblioteca utilizando a técnica de *shotgun*, ou seja, por fragmentação e clonagem aleatória, geramos sequências de tamanhos variáveis. Ou seja, cada segmento de uma proteína pode estar representado diversas vezes por peptídeos diferentes, mas sobrepostos. Estes peptídeos apesar de serem provenientes da mesma região, não possuem exatamente a mesma sequência, pois podem ser mais longos ou mais curtos, também podem estar deslocados ligeiramente a montante (*upstream*) ou a jusante (*downstream*) (Figura 26), mas por compreenderem o mesmo epítopo, todos são amplificados durante o processo de seleção. No caso específico do *T. cruzi*, há também outras duas particularidades: a existência de muitas repetições *in tandem* em regiões codificadoras pode gerar uma série de clones que expressam peptídeos muito parecidos por possuírem DNA proveniente dessas regiões repetitivas; ademais, a presença de grandes famílias multigênicas leva à existência de peptídeos semelhantes provenientes de diferentes regiões do genoma. Dessa maneira, para

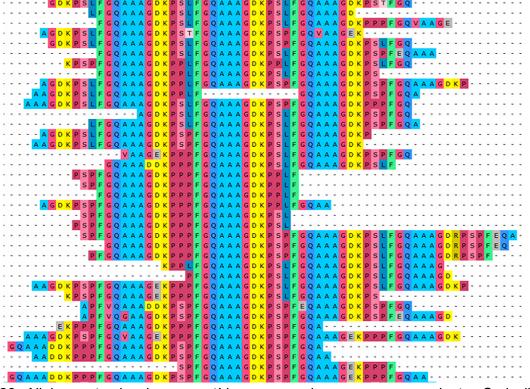


Figura 26. Alinhamento de alguns peptídeos agrupados no mesmo conjunto. Cada linha representa um peptídeo diferente identificado durante a seleção. Apesar de terem sequências e tamanhos diferentes, é possível inferir que todos representam o mesmo epítopo. Cores foram adicionadas para facilitar a visualização dos aminoácidos. As sequências mostradas correspondem ao epítopo B13, previamente identificado (CUNHA-NETO et al., 1995).

identificarmos os epítopos selecionados, é necessário que esta redundância seja removida, agregando os semelhantes em um único grupo, pois estes potencialmente são reconhecidos pelos mesmos anticorpos. Esse passo facilita inferir quais antígenos e epítopos têm a maior probabilidade de serem específicos para cada tipo de caso clínico. Ainda, a verificação da existência destas diversas sequências em nossas seleções comprova alta cobertura da biblioteca, mostrando ser um reagente capaz de representar todo o genoma do parasita.

Metodologias de agrupamento descritas na literatura como Hammock (KREJCI et al., 2016), MMseqs (HAUSER; STEINEGGER; SÖDING, 2016) e OrthoMCL (LI; STOECKERT; ROOS, 2003) foram experimentadas, mas nenhuma delas proporcionou resultados satisfatórios. Problemas como peptídeos semelhantes não serem incluídos em um certo conjunto enquanto peptídeos divergentes o são, são comuns a estes métodos. Métodos de alinhamento múltiplo, como o implementado pelos softwares ClustalW (THOMPSON; HIGGINS; GIBSON, 1994) e MAFFT (KATOH; STANDLEY, 2013), são eficientes para comparar dados onde todas as sequências apresentam alguma semelhança entre si. No entanto, neste caso, estamos lidando com uma grande quantidade de sequências onde a maior parte delas não apresenta nenhuma relação umas com as outras. Por isso, criamos o nosso método para que ele fosse mais adequado ao problema em questão. Optamos por realizar alinhamentos em pares, ou seja, apenas duas sequências são alinhadas por vez. Um limitante desta abordagem é que se formos comparar todas as sequências entre si, a quantidade de comparações a ser feita é igual ao quadrado da quantidade de sequências dividido por dois, ou seja, para um conjunto de 10 peptídeos, são necessárias 50 comparações; se o número de peptídeos aumenta 100 vezes para 1.000, o número de comparações sobe 10.000 vezes para 500,000. Em nosso conjunto, considerando apenas os peptídeos alinhados ao proteoma (similaridade ≥ 60%), temos 30.475 sequências diferentes, o que resultaria na necessidade de realizar quase 500 milhões de comparações, tornando o custo computacional proibitivo, dado a capacidade de processamento disponível. Para contornar este problema, implementamos uma abordagem, utilizando um algoritmo conhecido como "ganancioso" (do inglês *greedy*).

No programa desenvolvido, todos os peptídeos foram organizados em ordem crescente de tamanho. O primeiro da lista é então removido e adicionado a um novo conjunto. Todos os outros peptídeos são então comparados com essa sequência fundadora do conjunto; caso as sequências sejam pelo menos 80% semelhantes, o peptídeo é removido da lista e adicionado ao conjunto. Dessa maneira, peptídeos que já foram incluídos em algum conjunto não precisam mais ser comparados com os outros, reduzindo drasticamente o número de comparações necessárias. Terminados os alinhamentos desta sequência, temos a formação do primeiro conjunto, e os peptídeos restantes são novamente organizados por ordem crescente de tamanho e o processo é repetido até que não restem mais sequências fora de conjuntos - que podem ser formados por apenas uma sequência em casos onde essa não apresente similaridade com nenhuma outra.

Findada esta primeira etapa ainda temos que lidar com conjuntos de peptídeos que são semelhantes entre si e potencialmente representam o mesmo epítopo, apesar de não terem sido agrupados ainda. Para isso, criamos uma sequência consenso para representar cada conjunto: um alinhamento múltiplo é gerado para cada conjunto de peptídeos utilizando o *software* MAFFT e depois é gerada uma sequência consenso para este alinhamento utilizando-se o pacote HMMER3 (MISTRY et al., 2013). Conjuntos que possuem apenas um peptídeo são representados por ele mesmo. Tendo uma sequência para representar cada conjunto, o algoritmo de agrupamento é repetido como descrito anteriormente. Todo o processo é realizado repetidas vezes até que nenhum novo agrupamento seja possível.

Ao final, os 30.475 peptídeos diferentes formaram 3.741 conjuntos diferentes, ou 3.741 epítopos. Destes, 1.471 contém mais de um peptídeo diferente identificado. É interessante observar, que o maior conjunto contém 4.862 peptídeos diferentes, correspondentes a uma região repetitiva presente em uma proteína anotada como hipotética no proteoma do *T. cruzi* (RefSeq ID XP\_815993). O segundo maior conjunto

contem 3.568 peptídeos diferentes e é representado pelo epítopo repetitivo, pertencente a proteína *surface antigen 2* (RefSeq ID XP\_813516), já identificado anteriormente como antígeno B13 (CUNHA-NETO; CHEVILLARD, 2014).

O antígeno surface antigen 2, citado anteriormente, aparece na lista dos 10 mais abundantes nas três formas clínicas diferentes. Outros três antígenos também estão presentes entre os 10 mais abundantes das três formas: ribosomal protein L19, RNA-binding protein, calpain cysteine peptidase. Ainda, os antígenos microtubule-associated protein e trans-sialidase (EKG04410.1) são comuns nas listas dos 10 mais abundantes nos grupos assintomáticos e CCC leve. Já a proteína hipotética EAN88361.1 aparece em ambos os grupos de cardiopatas.

Para verificar se mais peptídeos identificados eram semelhantes a outros já descritos na literatura, comparamos nosso conjunto de dados aos epítopos da doença de Chagas presentes no banco de dados IEDB, que reúne estas informações.(VITA et al., 2015). Dos 2.189 epítopos peptídicos presentes nos banco de dados, 2.029 (93%) apresentam pelo menos 60% de identidade com peptídeos identificados em nossos experimentos, sendo que 469 (21,4%) tem 100% de identidade. Indiretamente, isto valida nossa metodologia, uma vez que uma grande proporção de epítopos já descritos na literatura foi também identificada em nossos experimentos (Figura 27).

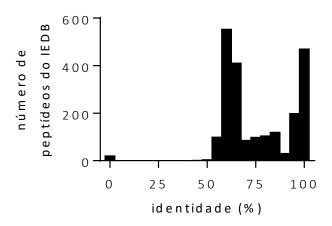


Figura 27. Histograma de frequência representando a identidade dos peptídeos do IEDB comparados aos identificados neste trabalho. Os epítopos relacionados à doença de Chagas foram obtidos no banco de dados e comparados a todos os presentes em nosso conjunto de dados pelo mesmo algoritmo utilizado no passo de agrupamento de epítopos.

É possível observar, também, o enriquecimento de alguns epítopos no grupo controle. Uma hipótese é que tais sequências podem estar mimetizando antígenos presentes em outros patógenos com os quais os pacientes possam já ter entrado em contato, ou ainda é possível que sejam peptídeos que se ligam de forma inespecífica a IgG e não puderam ser neutralizados pelos anticorpos mantidos em solução durante o processo de seleção.

Tabela 9. Os 10 antígenos mais abundantes na seleção contra os grupos assintomáticos. Abundância é dada pela média da frequência relativa das leituras de DNA observadas nos grupos A e B. Cada epítopo é representado pelo peptídeo mais abundante do conjunto e pelo antígeno correspondente (RefSeq ID e nome). São ignorados na lista epítopos que estejam entre os 10 mais abundantes do grupo controle.

Epítopo	Antígeno	Abundância
FGQAAAGDKPPLFGQAAAGDKPSL	EAN97076.1 surface antigen 2 (CA-2), putative	12.329%
GQGSSDTPCRGLNYRFPVECGDVLFLGSDGVF DNLF	EKG03462.1 hypothetical protein	2.938%
YKRALPQEEEEDVGPRHVDPDHFRSTTQDAY	EAN87716.1 microtubule-associated protein, putative	2.072%
IEEVRGAAPLGKYALVNTLEGSDGHFSAG	EKG01211.1 hypothetical protein, partial	1.742%
AAPAKAAAAPAKAAAAPAKAAAAPA	EAN99144.1 60S ribosomal protein L19, putative	0.922%
PAAKPAAKPAAKAPAPKA	EKG00673.1 RNA- binding protein, putative	0.815%
NAGSHELLGTEMPVSGEHFPPNIDSPLMGQV	EKG04410.1 trans- sialidase, putative	0.738%
SSLAGSDGVGLAGGASSIESFEGLPPGIDLSSIKD MPGAS	EAN83090.1 hypothetical protein, conserved	0.710%
VRNGECTAYVDGQPLGSLVEKHPFTLGTPPGAG VPEQEKIPEMVSLIFFGSDGVNLEGT	ESS62287.1 hypothetical protein	0.483%
FLDQKPEGVPLRELPLDDDSDFVAMEQERRQLL ER	EAN83138.1 calpain cysteine peptidase, putative	0.331%

Tabela 10. Os 10 antígenos mais abundantes na seleção contra os grupos com CCC leve. Abundância é dada pela média da frequência relativa das leituras de DNA observadas nos grupos A e B. Cada epítopo é representado pelo peptídeo mais abundante do conjunto e pelo antígeno correspondente (RefSeq ID e nome). São ignorados na lista epítopos que estejam entre os 10 mais abundantes do grupo controle.

Epítopo	Antígeno	Abundância
GSAAHTSTPAAGGFGSATTTSTPA AGGFGSAAHTSTPAVGGFGSA	EAN88361.1 hypothetical protein, conserved	7.292%
FGQAAAGDKPPLFGQAAAGDKPSL	EAN97076.1 surface antigen 2 (CA-2), putative	4.879%
YKRALPQEEEEDVGPRHVDPDHFRSTTQDAY	EAN87716.1 microtubule- associated protein, putative	4.414%
NAGSHELLGTEMPVSGEHFPPNIDSPLMGQV	EKG04410.1 trans- sialidase, putative	2.478%
ASQQAAVAAAKTAMGEAGG RSWSNVVKSPHSPRDAT	EKG07867.1 hypothetical protein	2.250%
PAAKPAAKPAAKAPAPKA	EKG00673.1 RNA- binding protein, putative	1.026%
AATEKRQSVNNYTTPGDSDGSTAVS	EKF99474.1 mucin- associated surface protein (MASP), putative	0.819%
FLDQKPEGVPLRELPLDDDSDFVAMEQERRQL LERDPRRNAREIAALEESMNARAQELAREKKL ADRAFLDQKPE	EAN83138.1 calpain cysteine peptidase, putative	0.562%
NKASVYIDGKSLGEEEVPLTGEKPLELF	EKG06014.1 trans- sialidase, putative	0.239%
AAPAKAAAAPAKAAAAPAKAAAAPA	EAN99144.1 60S ribosomal protein L19, putative	0.221%

### Os antígenos e epítopos da doença de Chagas

Do total de 3.741 epítopos identificados, consideramos 2.655 como sendo específicos para doença de Chagas. Estes epítopos são resultantes de todos os filtros colocados até agora (aparecimento da sequência de DNA pelo menos duas vezes durante o sequenciamento, alinhamento positivo com o genoma e com o proteoma) e ainda mais dois filtros novos: (1) a abundância do epítopo no grupo controle não pode ultrapassar 0.1%; (2) A abundância do epítopo em um dos grupos infectados deve ser pelo menos duas vezes maior que em qualquer grupo controle. Dessa maneira,

excluímos dois epítopos pelo primeiro critério e mais 1.092 pelo segundo critério.

Para relacionar cada epítopo ao seus antígeno, primeiro agrupamos as proteínas do *T. cruzi* utilizando o software CD-HIT (LI; GODZIK, 2006). Como estamos utilizando quatro proteomas de cepas diferentes e ainda sabemos que existem genes muito semelhantes nas diversas famílias multigênicas, removemos essa redundância pelo agrupamento de sequências com pelo 70% de identidade. As 51.899 sequências proteicas das quatro cepas foram reduzidas a 12.628 grupos. O peptídeo mais abundante de cada epítopo foi escolhido para representa-lo, e a proteína encontrada

Tabela 11. Os 10 epítopos mais abundantes na seleção contra os grupos com CCC grave. Abundância é dada pela média da frequência relativa das leituras de DNA observadas nos grupos A e B. Cada epítopo é representado pelo peptídeo mais abundante do conjunto e pelo antígeno correspondente (RefSeq ID e nome). São ignorados na lista epítopos que estejam entre os 10 mais abundantes do grupo controle.

Epítopo	Antígeno	Abundância
PPHTRRVTVRCGPPSCADE	ESS62170.1 trans- sialidase	61.093%
FGQAAAGDKPPLFGQAAAGDKPSL	EAN97076.1 surface antigen 2 (CA-2), putative	1.861%
AAPAKAAAAPAKAAAAPAKAAAAPA	EAN99144.1 60S ribosomal protein L19, putative	0.283%
YGPLRPTGMWNVEEVVDVKNSTVD FRRIDDVESEVIEALSQPDDAVVP	EKG04799.1 hypothetical protein	0.263%
PHGTQRRVATRVEAV	EAN92931.1 hypothetical protein, conserved	0.226%
ELQEERRAVARAEVELKKRLQ	EKG01449.1 hypothetical protein	0.223%
DAKPPAESPFKNAFGAPSSTAAKTPA ESPFKNVFGAPSSTAAKPPAESPFKN	EAN97077.1 hypothetical protein, conserved	0.207%
PAAKPAAKPAAKAPAPKA	EKG00673.1 RNA-binding protein, putative	0.199%
FLDQKPEGVPLRELPLDDDSDFVAME QERRQLLERDPRRNAREIAALEESMNA RAQELAREKKLADRAFLDQKPE	EAN83138.1 calpain cysteine peptidase, putative	0.189%
GSAAHTSTPAAGGFGSATTTSTP AAGGFGSAAHTSTPAVGGFGSA	EAN88361.1 hypothetical protein, conserved	0.144%

pelo software BlastP para este peptídeo foi considerada como sendo o antígeno.

Os 2.655 se espalharam por 1.830 antígenos diferentes. Não houve grande diferença na composição antigênica entre os casos clínicos (Tabela 12 e Figura 28), com proteínas hipotéticas chegando a formar quase 80% dos antígenos identificados. As gp85/trans-sialidases são a família gênica com o maior número de epítopos, representando cerca de 4% de todos os identificados.

Quando comparamos a lista dos 100 antígenos mais abundantes em nosso experimento para cada forma clínica, verificamos que 51 são comuns aos três tipos estudados. Ainda, existem 32 antígenos entre os 100 mais abundantes na forma assintomática que não compõem esta lista em nenhuma das outras formas. É possível que tais antígenos possam ser importantes para uma resposta efetiva contra o parasita, impedindo a progressão da doença. Os 100 antígenos mais abundantes no experimento e seus respectivos antígenos são listados na Tabela 13.

Tabela 12. Contagem do número de epítopos pertencentes a antígenos de grandes famílias genicas, proteínas hipotéticas e outras proteínas.

Família gênica	Todos	Assintomático	CCC leve	CCC grave
gp85/trans- sialidase	109 (4.1%)	45 (3.8%)	72 (4.5%)	56 (4.2%)
DGF1	49 (1.8%)	22 (1.9%)	28 (1.7%)	22 (1.6%)
gp63	25 (0.9%)	11 (0.9%)	13 (0.8%)	12 (0.9%)
RHS protein	25 (0.9%)	9 (0.8%)	14 (0.9%)	10 (0.7%)
MASP	21 (0.8%)	6 (0.5%)	12 (0.7%)	11 (0.8%)
mucin	13 (0.5%)	3 (0.3%)	12 (0.7%)	4 (0.3%)
Proteínas hipotéticas	2023 (76.2%)	930 (79.6%)	1248 (77.6%)	1025 (76.5%)
outros	390 (14.7%)	143 (12.2%)	210 (13.1%)	199 (14.9%)
Total	2655 (100.0%)	1169 (100.0%)	1609 (100.0%)	1339 (100.0%)

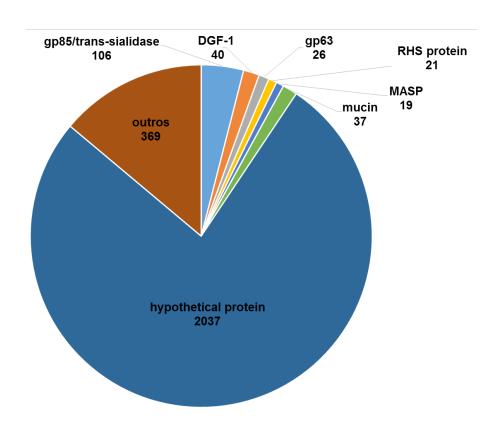


Figura 28. Distribuição dos 2.655 epítopos e seus antígenos nas grandes famílias gênicas do *T. cruzi*, proteínas hipotéticas e outras proteínas.

Tabela 13. Os 100 antígenos mais abundantes identificados considerando-se todos os grupos. É indicado se este contém epítopos já descritos no IEDB (mais de 60% de identidade com sequências depositadas), e se faz parte do conjunto dos 100 antígenos mais representados nas diferentes formas clínicas analisadas.

		100	100 mais abundantes			
Antígeno	IEDB	Assinto.	CCC leve	CCC grave		
EAN90772.1 trans-sialidase, putative	-	SIM	SIM	SIM		
EAN97076.1 surface antigen 2 (CA-2), putative	SIM	SIM	SIM	SIM		
EAN94142.1 hypothetical protein, conserved	SIM	SIM	SIM	SIM		
EAN87716.1 microtubule-associated protein, putative	SIM	SIM	SIM	SIM		
EAN85898.1 hypothetical protein, conserved	-	SIM	-	-		
EAN95247.1 trans-sialidase, putative	-	SIM	SIM	SIM		
EAN96157.1 hypothetical protein, conserved	-	-	SIM	-		
EKG08156.1 hypothetical protein TCSYLVIO_000703, partial	-	SIM	-	-		
EAN97450.1 RNA-binding protein, putative	SIM	SIM	SIM	SIM		
EAN89881.1 mucin-associated surface protein (MASP), putative	SIM	-	SIM	-		

EAN83090.1 hypothetical protein, conserved	-	SIM	-	SIM
EAN99145.1 60S ribosomal protein L19, putative	SIM	SIM	SIM	SIM
ESS62287.1 hypothetical protein TCDM 10060	-	SIM	-	-
EAN83138.1 calpain cysteine peptidase, putative	SIM	SIM	SIM	SIM
EAN87967.1 hypothetical protein, conserved	SIM	SIM	-	-
EAN94706.1 trans-sialidase, putative	-	-	SIM	-
EAN96486.1 hypothetical protein, conserved	-	-	-	SIM
EAN83685.1 hypothetical protein Tc00.1047053510863.20	SIM	SIM	SIM	SIM
EAN97077.1 hypothetical protein, conserved	-	SIM	SIM	SIM
EAN95422.1 hypothetical protein, conserved	-	-	-	SIM
EAN96361.1 R27-2 protein, putative	SIM	SIM	SIM	SIM
EAN92808.1 hypothetical protein, conserved	-	-	SIM	SIM
EAN99467.1 hypothetical protein Tc00.1047053510629.60	-	-	SIM	SIM
EAN96852.1 surface protease GP63, putative	-	-	SIM	-
ESS55066.1 surface protease GP63	-	SIM	-	-
EAN96067.1 hypothetical protein, conserved	-	-	SIM	SIM
EAN92318.1 69 kDa paraflagellar rod protein, putative	-	-	SIM	-
ESS61308.1 trans-sialidase	SIM	SIM	-	-
EAN88590.1 kinesin, putative	SIM	SIM	-	-
ESS62057.1 hypothetical protein TCDM_10300	-	SIM	SIM	SIM
EAN89037.1 hypothetical protein, conserved	SIM	SIM	SIM	SIM
EAN83090.1 hypothetical protein, conserved	-	SIM	-	-
EAN81678.1 trans-sialidase, putative	SIM	SIM	SIM	SIM
ESS64107.1 putative trans-sialidase	SIM	-	SIM	-
ESS55733.1 hypothetical protein TCDM_12778	-	SIM	SIM	SIM
EAN91689.1 hypothetical protein, conserved	-	-	-	SIM
ESS61042.1 hypothetical protein TCDM_11387	-	SIM	SIM	SIM
EAO00075.1 cytochrome C oxidase subunit IV, putative	-	-	SIM	-

EKF98095.1 hypothetical protein TCSYLVIO_011012, partial	-	SIM	SIM	SIM
EAN93557.1 mucin-associated surface protein (MASP), putative	SIM	-	SIM	-
EAN85153.1 hypothetical protein, conserved	-	SIM	SIM	SIM
EKG04779.1 hypothetical protein TCSYLVIO 004165	-	SIM	SIM	SIM
EAN96081.1 hypothetical protein, conserved	-	-	SIM	SIM
EAN98223.1 trans-sialidase, putative	SIM	SIM	-	-
EAN92093.1 hypothetical protein, conserved	SIM	SIM	SIM	SIM
EAN85891.1 hypothetical protein Tc00.1047053504031.40	SIM	SIM	SIM	SIM
EAN94163.1 DNA topoisomerase III, putative	-	SIM	SIM	SIM
EAN99273.1 arginyl-tRNA synthetase, putative	SIM	-	-	SIM
ESS55057.1 hypothetical protein TCDM_13494	SIM	SIM	SIM	SIM
EAN95247.1 trans-sialidase, putative	SIM	SIM	-	-
EAN95247.1 trans-sialidase, putative	SIM	SIM	SIM	-
EAN94320.1 serine/threonine protein kinase, putative	-	-	-	SIM
EAN99112.1 hypothetical protein, conserved	SIM	SIM	-	-
EAN89612.1 hypothetical protein, conserved	-	-	SIM	-
EAN98533.1 hypothetical protein, conserved	-	SIM	SIM	-
EAN97543.1 hypothetical protein Tc00.1047053506321.180	-	SIM	SIM	SIM
EAN99205.1 hypothetical protein Tc00.1047053508741.250	SIM	SIM	SIM	SIM
EAN81053.1 alpha tubulin, putative	SIM	SIM	SIM	-
EAO00077.1 hypothetical protein Tc00.1047053506529.370	SIM	SIM	SIM	SIM
EAN95944.1 protein kinase, putative	-	-	-	SIM
EAN99196.1 imidazolonepropionase, putative	-	-	-	SIM
ESS70005.1 hypothetical protein TCDM_01098	SIM	SIM	SIM	SIM
ESS55205.1 trans-sialidase	SIM	SIM	-	-
EAN94473.1 dynein heavy chain, putative	-	-	-	SIM
EAN87816.1 DNA-directed RNA polymerase I largest subunit, putative	-	-	SIM	-
ESS56901.1 hypothetical protein TCDM_12455	-	SIM	SIM	SIM

ESS63065.1 mucin-associated surface protein (MASP)	SIM	-	SIM	SIM
ESS60614.1 hypothetical protein TCDM_11848	-	SIM	SIM	SIM
EAN98364.1 hypothetical protein, conserved	-	SIM	-	SIM
ESS63070.1 hypothetical protein TCDM 09194	SIM	SIM	SIM	SIM
EAN96187.1 hypothetical protein, conserved	-	SIM	SIM	SIM
EKG00226.1 trans-sialidase, putative	SIM	SIM	SIM	-
ESS55663.1 trans-sialidase		-		SIM
EKG08826.1 hypothetical protein TCSYLVIO_000017, partial	-	SIM	SIM	SIM
ESS54998.1 hypothetical protein TCDM_13559	SIM	SIM	SIM	SIM
ESS55263.1 hypothetical protein TCDM_13279	SIM	SIM	SIM	SIM
EAN98520.1 hypothetical protein, conserved	SIM	SIM	SIM	SIM
EAN87641.1 vesicular transport- associated repeat protein, putative	-	SIM	SIM	-
ESS67917.1 hypothetical protein TCDM_03431	SIM	SIM	SIM	-
ESS55520.1 hypothetical protein TCDM_13000	SIM	SIM	SIM	SIM
EAN99422.1 hypothetical protein, conserved	-	SIM	-	-
ESS60717.1 hypothetical protein TCDM_11743	SIM	SIM	SIM	SIM
EKG02232.1 hypothetical protein TCSYLVIO_006749, partial	SIM	SIM	SIM	SIM
EKG01114.1 hypothetical protein TCSYLVIO_007904	SIM	-	SIM	SIM
ESS55751.1 hypothetical protein TCDM_12757	SIM	SIM	SIM	SIM
ESS62779.1 hypothetical protein TCDM_09519	SIM	-	SIM	-
EAN95652.1 hypothetical protein, conserved	SIM	-	SIM	-
ESS55020.1 hypothetical protein TCDM_13535	-	SIM	SIM	SIM
EAN99684.1 hypothetical protein, conserved	SIM	SIM	-	-
EAN83482.1 hypothetical protein Tc00.1047053510983.10	SIM	-	SIM	SIM
ESS60714.1 hypothetical protein TCDM_11740	SIM	SIM	-	SIM
ESS63171.1 hypothetical protein TCDM_09088	SIM	SIM	SIM	SIM

ESS55954.1 hypothetical protein TCDM_12545	SIM	SIM	SIM	SIM
EAN90304.1 protein kinase, putative	-	SIM	-	SIM
EAN82145.1 surface protease GP63, putative	-	SIM	-	SIM
ESS55888.1 hypothetical protein TCDM_12620	SIM	-	SIM	SIM
ESS62988.1 hypothetical protein TCDM_09254	SIM	SIM	SIM	-
EAN93642.1 trans-sialidase, putative	-	SIM	-	-
ESS63187.1 hypothetical protein TCDM_09055	SIM	SIM	SIM	SIM
ESS61925.1 hypothetical protein TCDM_10454	-	SIM	SIM	SIM

# Sequencias repetitivas formam epítopos

É de se notar que muitos antígenos identificados contêm sequências com repetições em *tandem*. Utilizando o software XSTREAM (NEWMAN; COOPER, 2007) para identificar essas unidades repetitivas em todo os proteoma do parasita, encontramos 227 sequências com unidades repetitivas variando de 106 a 2 aminoácidos, que aparecem pelo menos cinco vezes lado-a-lado. Algumas destas sequências podem ser repetidas até 177 vezes lado-a-lado, como no caso da sequência PKPAE, presente em gp85/trans-sialidases. Já o motivo repetitivo da *calpain cysteine peptidase*, apesar de aparecer menos vezes – seis – tem 68 aminoácidos de extensão.

Os epítopos identificados foram então comparados com essas sequências repetitivas para determinar correspondência. Um total de 43 epítopos provenientes de sequências repetitivas foram encontrados (Tabela 14). Ainda, muitos destes estão entre os epítopos mais abundantes em nossos resultados como: PFGQAAAGDKPS (surface antigen 2), AAGGFGSATTTSAP (EAN94142.1 hypothetical protein), GPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDV (microtubule-associated protein), AKPPAESPFKNVFGAPSSTA (EAN97077.1 hypothetical protein) e KQRAAEATKVAEAE (R27-2 protein).

Tabela 14. Unidades repetitivas que compõem epítopos de antígenos identificados. O número de repetições mostrado corresponde ao máximo identificado em uma proteína.

Unidade repetitiva	#	 Antígeno
·	repetições	·
PKPAE	177	EAN91735.1 trans-sialidase
EEAEARRLA	106	EAN83595.1 Tb-291 membrane- associated protein-like
EAEERAQR	103	EAN81788.1 membrane associated protein, putative
EDVQELVAPA	79	EAN89929.1 hypothetical protein, conserved
KQRAAEATKVAEAE	76	EAN93943.1 hypothetical protein
PFGQAAAGDKPS	72	EAN91665.1 surface antigen 2 (CA-2), putative
GSSMSEGSSHERSEG	65	EAN86927.1 hypothetical protein
PAEPKSAEPK	59	EAN87106.1 trans-sialidase
LAQREADNEKLAED	49	EAN86336.1 myosin heavy chain, putative
AAGGFGSATTTSAP	44	EAN88361.1 hypothetical protein
DSSAHSTPSTPV	40	EAN84377.1 trans-sialidase
ENERLADELEQKAA	40	EAN81003.1 hypothetical protein, conserved
ERQLQERDDALAALRDRLEEY GREKSALESRSSESVDALAAL	38	EAN88119.1 hypothetical protein
LQRQNEELQSQLKESRRGEEKLDA	34	EAN99975.1 antigenic protein, putative
GFRG	31	EAN86592.1 nucleolar RNA- binding protein
AKPPAESPFKNVFGAPSSTA	28	EAN97077.1 hypothetical protein,
GFGDRGGR	24	EAN95753.1 hypothetical protein
TITTKVPTTTTTTMAPT	23	EAN82230.1 hypothetical protein Tc00.1047053510433.10
ENEELRAEDEHKTRGLQ EVSEQAEDLQRQLEELRA	21	EAN94898.1 hypothetical protein
LNDGGQRYSLSISVCVFTH WTPNNPELSAAWRNFASSCPHG CCTRLIVFERLLHHTLWSKFLFPI	21	EAN84488.1 hypothetical protein
QAARQQEDVQELVA PAENVRELVAPTEDV	20	EAN98020.1 hypothetical protein, conserved

GPRHVDPDHFRSTTQDAYR PVDPSAYKRALPQEEEEDV	19	EAN87716.1 microtubule- associated protein, putative
VTAEREELAENLRATEDAKAEVERNLES	17	EAN84693.1 kinesin-like protein
MEPTTITTKAPTTITT	17	EAN82230.1 hypothetical protein
GGQRGGGRRDYGDQGNQGDY	16	EAN87685.1 mitochondrial oligo_U binding protein TBRGG1, putative
MATDELTTEARPLDVTVGS ALLSDSAAVTAPADRDATAL	14	EAN94669.1 hypothetical protein Tc00.1047053508119.200
TNDLQERAAAAEDAARRRC AAAREKEEAAKRLEAELEVR	13	EAN84189.1 hypothetical protein
KPEEK	11	EKF26748.1 hypothetical protein
SAKKRPAVLSSQEVQAVVNA RRATPRNDEYDDDYESETSA	10	EAN85405.1 hypothetical protein
GRGGDRGGFRG	10	EAN86596.1 nucleolar RNA- binding protein
KAAAPK	6	EAN99142.1 60S ribosomal protein L23a, putative
TTTTTTTKAP	10	EAN83649.1 mucin TcMUCI
ASASEEGSRGDSSMP	9	EAN98211.1 trans-sialidase, putative
AAVPDSSSDDDDAPVR KPAQKAKASPAMRPAPKK	8	EAN84654.1 hypothetical protein, conserved
STSSTPVGSGAK	8	EAN85648.1 trans-sialidase
NADDARSPLRRQPA	8	EAN85710.1 hypothetical protein, conserved
ARGQLVGEEREGTCGLHRDAV DSEERAVRRCLERGEAAAVDELGEEYG SATHERVVEALAAEEDA	8	EAN91803.1 hypothetical protein, conserved
KPEDAKPEEK	8	EAN83891.1 hypothetical protein
TKPVSSPP	8	EAN91857.1 hypothetical protein, conserved
LPKKKGRPKKLNEAEDAVAEM TLMEEQEEMSAAVTEAEEAP	7	EAN91857.1 hypothetical protein, conserved
LTQEVQRLERELEEARREQERA VGATEAMQRDAARERDEAVAA	7	EAN99422.1 hypothetical protein, conserved
LLEKDPRRNAKEIAALEESMNAR AQELAREKKLADRAFLDQKPEGVSLR ELPLDDDSDFVAMEQERRQ	6	EAN82258.1 calpain cysteine peptidase
RHSENADNVRSPLK	6	EAN85710.1 hypothetical protein, conserved

# Grupos com cardiomiopatia grave apresentam alta correlação

As amostras foram comparadas em pares para avaliar a abundância dos epítopos compartilhados através da correlação de Pearson (Figura 29). As amostras de pacientes CCC grave apresentaram alta correlação entre si, porém não se assemelham às outras amostras. Isso sugere que a resposta humoral nessa forma clínica é mais homogênea, podendo estar relacionada com a progressão da doença. Já as amostras de CCC leve não apresentaram alta correlação entre si, mas uma correlação mais alta com assintomáticos do mesmo grupo (assintomático A vs. CCC leve A e assintomático B vs. CCC leve B). Entre assintomáticos, a correlação também foi significativa: 0,6.

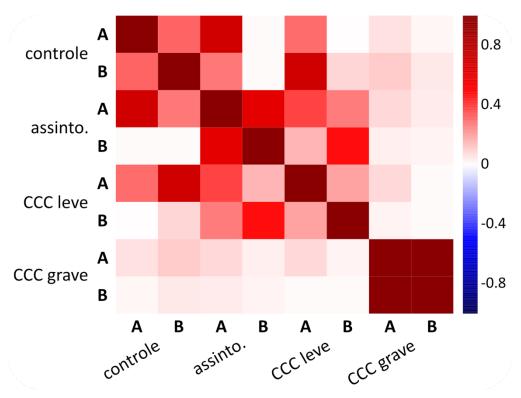


Figura 29. Correlação de Pearson entre as amostras. A abundância de cada epítopo encontrado foi comparada entre os diferentes grupos. A cores representam o valor de correlação como indicado pela escala à esquerda.

Chama atenção o fato de existir uma correlação tão grande entre as amostras de CCC grave, especialmente se levarmos em consideração a hipótese de que essa forma da doença pode ter um componente autoimune. O epítopo do antígeno

EAN90772.1 trans-sialidase representa 40,3% e 81,8% das leituras nas amostras CCC grave do grupo A e B, respectivamente. Ao mesmo tempo, tem uma abundância média de 0,06% considerando todos os outros grupos. Um experimento independente de seleção foi realizado apenas na amostra CCC grave do grupo A. O resultado novamente revelou o mesmo epítopo como mais abundante, compondo 15.806% das leituras de DNA. No entanto, estudos adicionais são necessários para que se entenda melhor o reconhecimento deste antígeno pelos pacientes e como isso pode estar relacionado com a progressão da doença.

Em resumo, nossos resultados indicam que a resposta imune humoral na doença de Chagas é mais ampla do que inicialmente antecipada. Antígenos previamente descritos foram observados. Também é de se destacar o grande número de antígenos associados a proteínas hipotéticas, o que mostra que ainda existe uma lacuna muito grande no nosso conhecimento não só da resposta imune na doença de Chagas como também da biologia do parasita.

# Discussão

A resposta imune é um componente de papel central na história natural de doenças infeciosas. Neste trabalho propomos uma nova forma de estudar o repertório imune humoral, aliando a técnica de *phage display* de *shotgun* genômico, com sequenciamento de nova geração, seguido de análises computacionais. Isto é possível porque o genoma do *T. cruzi* não contem íntrons, e nos leva a crer que esta tecnologia pode ser expandida para outras doenças causadas por tripanosomatídeos como leishmaniose (*Leishmania spp.*) e doença do sono (*T. brucei*), por exemplo.

A caracterização detalhada de componentes da reposta imune é importante para o desenvolvimento racional de métodos sorológico de diagnóstico e prognóstico e também de novas vacinas. Avanços já foram feitos na área utilizando estudos de larga escala, inclusive para a doença de Chagas (CARMONA et al., 2015; MENDES et al., 2013; REIS-CUNHA et al., 2014). Estes métodos geralmente utilizam *microarray*s

de peptídeos de 15 aminoácidos em média que, no caso das plataformas mais modernas, pode chegar a uma escala de centenas de milhares de peptídeos em um *microarray*, representando algumas centenas de genes. Em nosso trabalho testamos simultaneamente dezenas de milhões de peptídeos derivados de *T. cruzi*, que possuíam tamanhos variados e, em geral, mais longos que os utilizados em *microarray*s. Dessa maneira, a abordagem utilizada se mostra abrangente, levando a uma mais profunda caracterização de epítopos relacionados a uma doença infeciosa.

Apesar das vantagens descritas, alguns detalhes devem ser ressaltados: a construção uma biblioteca de alta cobertura não é trivial. Ainda que seja possível trabalhar com bibliotecas com menor cobertura, isso pode comprometer os resultados. O tamanho dos peptídeos estudados pode impedir a identificação de alguns epítopos conformacionais que tenham extensão na ordem de algumas centenas de aminoácidos e mesmo que se construa uma biblioteca com sequências mais longa, não há como garantir que as proteínas se enovelarão corretamente para que o reconhecimento seja feito pelos anticorpos. Para o estudo de eucariotos superiores, é necessário utilizar cDNA produzido a partir do RNA mensageiro para a construção da biblioteca devido a existência de íntrons no genoma. Nessa abordagem, a biblioteca fica limitada ao grupo de genes que está sendo expresso no momento da extração do material genético, limitando o potencial de descoberta. Ainda, é possível se construir bibliotecas sintéticas onde clones são desenhados *in silico* e depos sintetizados (LARMAN et al., 2011).

Hoje, um dos maiores impeditivos para que se possa estudar e desenvolver drogas efetivas para a doença de Chagas é a falta de um método diagnóstico que ateste cura (CANÇADO, 1999; JANNIN; VILLA, 2007; PICADO et al., 2017; PINAZO et al., 2014). Mesmo técnicas sensíveis como PCR podem gerar falsos negativos, dada a diminuta quantidade de parasitas presentes no sangue, dessa maneira, métodos sorológicos continuam sendo a melhor opção. No entanto, muitos destes testes utilizam antígenos provenientes de extratos de epimastigotos (DA SILVEIRA;

UMEZAWA; LUQUETTI, 2001), e mesmo em casos onde há a completa eliminação do parasita, a redução do títulos de anticorpos pode demorar anos. Dessa forma, o desenvolvimento de novos métodos diagnósticos através da identificação de antígenos específicos para a doença de Chagas, cuja reatividade acompanhe a dinâmica da doença, é uma ferramenta fundamental para o desenvolvimento de novas terapias. O conjunto de antígeno/epítopos identificados neste estudo poderá, se devidamente validado em grandes coortes de pacientes, ser utilizado como ponto de partida para testes em pacientes curados *vs.* não-curados, possivelmente revelando um grupo de moléculas capaz de indicar a cura da doença.

Um indicativo do sucesso da seleção foi a identificação de diversos epítopos já descritos na literatura. Podemos destacar o epítopo B13 (CUNHA-NETO et al., 1995), a sequência repetitiva SAPA (shed-acute phase antigen) (AFFRANCHINO et al., 1989), Tc40 (LESÉNÉCHAL et al., 1997), flagellar-repetitive antigen (FRA) (LAFAILLE et al., 1989), cytoplasmic repetitive antigen (CRA) (LAFAILLE et al., 1989), entre outros. No caso do epítopo B13, diferentes variações contendo substituições já foram testadas quanto à reatividade de soros de pacientes, visto que este é proveniente de uma região repetitiva com degenerações. Curiosamente, o peptídeo mais abundante em nosso estudo representando este epítopo contém a forma mais reativa (FGQAAAGDKPSL) (THOMAS et al., 2012), indicando mais uma vez a validade dos resultados. Um aspecto ainda a ser elucidado é se a abundância de leituras de DNA observada para uma sequência ao final dos nossos experimentos se correlaciona com título do soro para este epítopo. Também, se a ausência de um epítopo no mapeamento implica em baixa ou nenhuma reatividade para este.

Epítopos em regiões de proteínas onde se observam repetições em tandem já foram reconhecidas anteriormente. Mucinas associadas a câncer de mama possuem a região repetitiva PDTRPAPGSTAPPAHGVTSA altamente imunogênica, fato explorado para o diagnóstico da doença (BURCHELL et al., 1989; FRIEDMAN; HAYES; KUFE, 1986; GENDLER et al., 1988; GIMMI et al., 1996). Muitos dos epítopos conhecidos

para a infecção por *Plasmodium falciparum*, causador da malária, também são provenientes de regiões repetitivas de proteínas (ANDERS, 1986; PALACPAC et al., 2011; YAGI et al., 2014). Essa característica também é conhecida no *T. cruzi* (FRASCH, 1994; VALIENTE-GABIOUD et al., 2011). Nota-se que as sequências repetitivas estão conservadas em diferentes cepas de parasita, ainda não se tendo uma explicação definitiva para sua existência. A princípio, conservar e expandir sequencias que geram uma forte resposta imune por parte do hospedeiro parece contraproducente para o ciclo de vida do parasita. De certa maneira, estas sequências devem também conferir vantagens: o direcionamento da resposta imune a certos epítopos pouco relevantes para a biologia do parasita pode evitar que outros mais importantes sejam alvo de anticorpos. Estas sequências também podem estar relacionadas à ativação policional de células B, levando à expansão de uma diversa população de clones diferentes, enfraquecendo a resposta imune.

A progressão da doença de Chagas para a forma cardíaca é frequentemente associada a autoimunidade (CUNHA-NETO et al., 2006; MARIN-NETO et al., 2007). Em nosso trabalho, utilizamos uma biblioteca com fragmentos proteicos derivados de T. cruzi para mapear epítopos relacionados à doença, os quais podem eventualmente mimetizar autoantígenos, como já sugerido. Embora possamos realizar buscas em bancos de dados procurando por proteínas humanas que compartilhem similaridade de sequências com nossos antígenos, essa abordagem é limitada pelos próprios algoritmos de busca. No caso de busca com longos trechos das sequências das proteínas de T. cruzi, é improvável encontrar correspondências, a não ser em proteínas conservadas evolutivamente. Se a busca se limitar a pequenos trechos (aproximadamente 8-10 aminoácidos) correspondentes ao epítopo, muitas proteínas candidatas serão encontradas, sendo difícil filtrar os falsos positivos. Uma alternativa interessante seria utilizar uma biblioteca de phage display construída com fragmentos de cDNA humano (p.ex., cardiomióticos) e testá-las contra o soro dos pacientes com CCC grave, fortes candidatos a apresentar autoimunidade. Dessa maneira, a descoberta de autoantígenos se tornaria mais direta e precisa.

Outra possibilidade é a de se adaptar esta metodologia para a identificação em larga escala de peptídeos ligantes de receptores de células T. Essa é uma perspectiva interessante no caso da doença de Chagas, visto que a imunidade celular é importante no combate a agentes infecciosos intracelulares, identificando e destruindo células infectadas (DUTRA; ROCHA; TEIXEIRA, 2005). O desafio neste caso, é que diferente das imunoglobulinas, que se ligam diretamente ao antígeno, a interação do TCR com o peptídeo ocorre no contexto do complexo de histocompatibilidade e de diversas moléculas assessórias. A determinação dos epítopos de células T, especialmente aqueles específicos de pacientes que se mantém assintomáticos, seria um importante passo para o desenvolvimento racional de vacinas para a doença. Embora os pacientes assintomáticos não atinjam esterilidade, sua atividade imunológica contra o *T. cruzi* é suficiente para mantê-los livres dos sintomas da doença.

# Conclusão

A resposta humoral é um componente essencial do sistema imunológico e do combate a infecções. No caso da doença de Chagas, há ainda o agravante da produção de autoanticorpos como possíveis moduladores da cardiomiopatia e progressão da doença. Neste estudo, construímos um dos mapas mais completos da resposta imune humoral da doença de Chagas. Milhares de epítopos foram identificados sendo que grande parte deles até então eram desconhecidos. Uma grande porção desses epítopos se concentra em proteínas anotadas como hipotéticas, reforçando a necessidade de se entender melhor a biologia do parasita para se criar novas estratégias de combate à doença. Esses novos dados podem contribuir para o desenvolvimento de métodos diagnóstico que melhor acompanham a progressão da doença e também ajudar a explicar porque indivíduos apresentam quadros clínicos diferentes.

# Conclusão da tese

O reconhecimento entre biomoléculas é central em todos os processos biológicos, incluindo a interação entre diferentes organismos, como no caso de doenças infecciosas. Neste trabalho dividido em duas partes, estudamos diferentes aspectos de como o *Trypanosoma cruzi*, protozoário causador da doença de Chagas, interage com o organismo hospedeiro e também, como os seres humanos reconhecem o parasita dentro de seu corpo. Em comum a ambas as partes está a delineação de regiões específicas de macromoléculas responsáveis por estas interações. Utilizando metodologias combinatórias, em alguns casos inovadoras, quando comparadas à literatura atual, geramos um conjunto de resultados que poderá auxiliar no desenvolvimento de novas terapias e diagnósticos para uma das doenças mais negligenciadas do mundo.

# Referências Bibliográficas

ABBAS, A. K.; LICHTMAN, A. H. H.; PILLAI, S. **Cellular and Molecular Immunology**. [s.l.] Elsevier Health Sciences, 2011.

ABUIN, G. et al. A surface antigen of Trypanosoma cruzi involved in cell invasion (Tc-85) is heterogeneous in expression and molecular constitution. **Molecular and Biochemical Parasitology**, v. 35, n. 3, p. 229–237, jul. 1989.

AFFRANCHINO, J. L. et al. Identification of a Trypanosoma cruzi antigen that is shed during the acute phase of Chagas' disease. **Molecular and Biochemical Parasitology**, v. 34, n. 3, p. 221–228, 15 maio 1989.

AJDIC, D.; CHEN, Z. A novel phosphotransferase system of Streptococcus mutans is responsible for transport of carbohydrates with  $\alpha$ -1,3 linkage. **Molecular oral microbiology**, v. 28, n. 2, p. 114–128, abr. 2013.

ALARCÓN DE NOYA, B. et al. Large Urban Outbreak of Orally Acquired Acute Chagas Disease at a School in Caracas, Venezuela. **The Journal of Infectious Diseases**, v. 201, n. 9, p. 1308–1315, 1 maio 2010.

ALVES, M. J. et al. Partial inhibition of trypomastigote entry into cultured mammalian cells by monoclonal antibodies against a surface glycoprotein of Trypanosoma cruzi. **Molecular and Biochemical Parasitology**, v. 21, n. 1, p. 75–82, out. 1986.

ALVES, M. J. M.; COLLI, W. Trypanosoma cruzi: adhesion to the host cell and intracellular survival. **IUBMB Life**, v. 59, n. 4–5, p. 274–279, maio 2007.

ANDERS, R. F. Multiple cross-reactivities amongst antigens of Plasmodium falciparum impair the development of protective immunity against malaria. **Parasite Immunology**, v. 8, n. 6, p. 529–539, 1 nov. 1986.

ANDRADE, L. O.; ANDREWS, N. W. The Trypanosoma cruzi-host-cell interplay: location, invasion, retention. **Nature reviews. Microbiology**, v. 3, n. 10, p. 819–823, out. 2005.

ANDREWS, N. W.; COLLI, W. Adhesion and interiorization of Trypanosoma cruzi in mammalian cells. **The Journal of Protozoology**, v. 29, n. 2, p. 264–269, maio 1982.

ARAYA, J. E. et al. Cloning and characterization of a gene for the stage-specific 82-kDa surface antigen of metacyclic trypomastigotes of Trypanosoma cruzi. **Molecular and Biochemical Parasitology**, v. 65, n. 1, p. 161–169, maio 1994.

ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic Acids Research**, v. 38, n. suppl 1, p. D457–D462, 1 jan. 2010.

AUFDERHEIDE, A. C. et al. A 9,000-year record of Chagas' disease. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 7, p. 2034–2039, 17 fev. 2004.

AVVAKUMOV, G. V. et al. Structural analyses of sex hormone-binding globulin reveal

novel ligands and function. **Molecular and Cellular Endocrinology**, v. 316, n. 1, p. 13–23, 5 mar. 2010.

BARNHART, K. F. et al. A Peptidomimetic Targeting White Fat Causes Weight Loss and Improved Insulin Resistance in Obese Monkeys. **Science translational medicine**, v. 3, n. 108, p. 108ra112, 9 nov. 2011.

BASTOS, C. J. C. et al. Clinical Outcomes of Thirteen Patients with Acute Chagas Disease Acquired through Oral Transmission from Two Urban Outbreaks in Northeastern Brazil. **PLOS Neglected Tropical Diseases**, v. 4, n. 6, p. e711, 15 jun. 2010.

BECKMANN, C. et al. Identification of Novel Adhesins from Group B Streptococci by Use of Phage Display Reveals that C5a Peptidase Mediates Fibronectin Binding. **Infection and Immunity**, v. 70, n. 6, p. 2869–2876, 1 jun. 2002.

BERMAN, H. M. et al. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 1 jan. 2000.

BESTETTI, R. B. et al. Predictors of Sudden Cardiac Death for Patients with Chagas' Disease: A Hospital-Derived Cohort Study. **Cardiology**, v. 87, n. 6, p. 481–487, 1996.

BIOLO, A.; RIBEIRO, A. L.; CLAUSELL, N. Chagas Cardiomyopathy—Where Do We Stand After a Hundred Years? **Progress in Cardiovascular Diseases**, Unusual Cardiomyopathies. v. 52, n. 4, p. 300–316, jan. 2010.

BONNEY, K. M.; ENGMAN, D. M. Autoimmune pathogenesis of Chagas heart disease: looking back, looking ahead. **The American Journal of Pathology**, v. 185, n. 6, p. 1537–1547, jun. 2015.

BURCHELL, J. et al. A short sequence, within the amino acid tandem repeat of a cancer-associated mucin, contains immunodominant epitopes. **International Journal of Cancer**, v. 44, n. 4, p. 691–696, 15 out. 1989.

CAMACHO, C. et al. BLAST+: architecture and applications. **BMC bioinformatics**, v. 10, p. 421, 2009.

CANÇADO, J. R. Criteria of Chagas disease cure. **Memórias do Instituto Oswaldo Cruz**, v. 94, p. 331–335, set. 1999.

CARMONA, S. J. et al. Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants. **Molecular & Cellular Proteomics: MCP**, v. 14, n. 7, p. 1871–1884, jul. 2015.

CHAGAS, C. Uber eine neue trypanosomiasis des menschen. **Archiv für Schiffsund Tropen-Hygiene**, v. 13, p. 351–353, 1909a.

CHAGAS, C. Nouvelle espèce de trypanosomiase humaine. **Bulletin de la Société de Pathologie Exotique**, v. 2, n. 6, p. 304–307, 1909b.

CHAGAS, C. Neue Trypanosomen. Archiv für Schiffs-und Tropenhygiene, n. 13, p.

120-122, fev. 1909c.

CHAGAS, C. Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem. **Memórias do Instituto Oswaldo Cruz**, v. 1, n. 2, p. 159–218, ago. 1909d.

CHATTERJEE, M. et al. Diagnostic Markers of Ovarian Cancer by High-Throughput Antigen Cloning and Detection on Arrays. **Cancer Research**, v. 66, n. 2, p. 1181–1190, 15 jan. 2006.

CLASER, C. et al. Silencing cytokeratin 18 gene inhibits intracellular replication of Trypanosoma cruzi in HeLa cells but not binding and invasion of trypanosomes. **BMC Cell Biology**, v. 9, p. 68, 17 dez. 2008.

COLLI, W. Trans-sialidase: a unique enzyme activity discovered in the protozoan Trypanosoma cruzi. **FASEB journal: official publication of the Federation of American Societies for Experimental Biology**, v. 7, n. 13, p. 1257–1264, out. 1993.

CÓRDOVA, E. et al. Neurological manifestations of Chagas' disease. **Neurological Research**, v. 32, n. 3, p. 238–244, abr. 2010.

CORTEZ, C. et al. The gp82 surface molecule of Trypanosoma cruzi metacyclic forms. **Sub-Cellular Biochemistry**, v. 74, p. 137–150, 2014.

COURA, J. R.; VIÑAS, P. A. Chagas disease: a new worldwide challenge. **Nature**, v. 465, n. n7301\_supp, p. S6–S7, 24 jun. 2010.

CROOKS, G. E. et al. WebLogo: a sequence logo generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, jun. 2004.

CROSS, G. A.; TAKLE, G. B. The surface trans-sialidase family of Trypanosoma cruzi. **Annual Review of Microbiology**, v. 47, p. 385–411, 1993.

CUNHA-NETO, E. et al. Autoimmunity in Chagas disease cardiopathy: biological relevance of a cardiac myosin-specific epitope crossreactive to an immunodominant Trypanosoma cruzi antigen. **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 8, p. 3541–3545, 11 abr. 1995.

CUNHA-NETO, E. et al. Induction of cardiac autoimmunity in Chagas heart disease: A case for molecular mimicry. **Autoimmunity**, v. 39, n. 1, p. 41–54, 1 jan. 2006.

CUNHA-NETO, E.; CHEVILLARD, C. Chagas Disease Cardiomyopathy: Immunopathology and Genetics. **Mediators of Inflammation**, v. 2014, p. e683230, 19 ago. 2014.

DA SILVEIRA, J. F.; UMEZAWA, E. S.; LUQUETTI, A. O. Chagas disease: recombinant Trypanosoma cruzi antigens for serological diagnosis. **Trends in Parasitology**, v. 17, n. 6, p. 286–291, 1 jun. 2001.

DIAS-NETO, E. et al. Next-Generation Phage Display: Integrating and Comparing

Available Molecular Tools to Enable Cost-Effective High-Throughput Analysis. **PLoS ONE**, v. 4, n. 12, p. e8338, 17 dez. 2009.

DIAZGRANADOS, C. A. et al. Chagasic encephalitis in HIV patients: common presentation of an evolving epidemiological and clinical association. **The Lancet Infectious Diseases**, v. 9, n. 5, p. 324–330, maio 2009.

DRAGHICI, S.; CHATTERJEE, M.; TAINSKY, M. A. Epitomics: serum screening for the early detection of cancer on microarrays using complex panels of tumor antigens. **Expert Review of Molecular Diagnostics**, v. 5, n. 5, p. 735–743, set. 2005.

DUTRA, W. O.; ROCHA, M. O. C.; TEIXEIRA, M. M. The clinical immunology of human Chagas disease. **Trends in Parasitology**, v. 21, n. 12, p. 581–587, 1 dez. 2005.

ELIZARI, M. V.; CHIALE, P. A. Cardiac arrhythmias in Chagas' heart disease. **Journal of Cardiovascular Electrophysiology**, v. 4, n. 5, p. 596–608, out. 1993.

EL-SAYED, N. M. The Genome Sequence of Trypanosoma cruzi, Etiologic Agent of Chagas Disease. **Science**, v. 309, n. 5733, p. 409–415, 15 jul. 2005.

EL-SAYED, N. M. et al. Comparative Genomics of Trypanosomatid Parasitic Protozoa. **Science**, v. 309, n. 5733, p. 404–409, 15 jul. 2005.

FILARDI, L. S.; BRENER, Z. Susceptibility and natural resistance of Trypanosoma cruzi strains to drugs used clinically in Chagas disease. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 81, n. 5, p. 755–759, 1987.

FONSECA, S. G. et al. 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 / Institut Pasteur**, v. 7, n. 4, p. 688–697, abr. 2005.

FOSTINIS, Y. et al. Heat shock protein HSP90 and its association with the cytoskeleton: a morphological study. **Biochemistry and Cell Biology**, v. 70, n. 9, p. 779–786, 1 set. 1992.

FRANZÉN, O. et al. Shotgun Sequencing Analysis of Trypanosoma cruzi I Sylvio X10/1 and Comparison with T. cruzi VI CL Brener. **PLoS Negl Trop Dis**, v. 5, n. 3, p. e984, 8 mar. 2011.

FRANZÉN, O. et al. Comparative genomic analysis of human infective Trypanosoma cruzi lineages with the bat-restricted subspecies T. cruzi marinkellei. **BMC Genomics**, v. 13, p. 531, 5 out. 2012.

FRASCH, A. C. C. Trans-sialidase, SAPA amino acid repeats and the relationship between <span class="italic">Trypanosoma cruzi</span> and the mammalian host. **Parasitology**, v. 108, n. S1, p. S37–S44, mar. 1994.

FREITAS, L. M. et al. Genomic Analyses, Gene Expression and Antigenic Profile of the Trans-Sialidase Superfamily of Trypanosoma cruzi Reveal an Undetected Level of Complexity. **PLoS ONE**, v. 6, n. 10, p. e25914, 19 out. 2011.

FRIEDMAN, E. L.; HAYES, D. F.; KUFE, D. W. Reactivity of monoclonal antibody DF3 with a high molecular weight antigen expressed in human ovarian carcinomas. **Cancer Research**, v. 46, n. 10, p. 5189–5194, out. 1986.

**FuzzyWuzzy:** Fuzzy String Matching in Python - ChairNerd. Disponível em: <a href="http://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/">http://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/</a>. Acesso em: 13 jun. 2017.

GANESH, V. K. et al. Structural and biochemical characterization of Staphylococcus aureus clumping factor B/ligand interactions. **The Journal of Biological Chemistry**, v. 286, n. 29, p. 25963–25972, 22 jul. 2011.

GENDLER, S. et al. A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. **Journal of Biological Chemistry**, v. 263, n. 26, p. 12820–12823, 15 set. 1988.

GIMMI, C. D. et al. Breast cancer–associated antigen, DF3/MUC1, induces apoptosis of activated human T cells. **Nature Medicine**, v. 2, n. 12, p. 1367–1370, dez. 1996.

GIORDANO, R. et al. An acidic component of the heterogeneous Tc-85 protein family from the surface of Trypanosoma cruzi is a laminin binding glycoprotein. **Molecular and Biochemical Parasitology**, v. 65, n. 1, p. 85–94, maio 1994.

GIORDANO, R. et al. Cloning of a surface membrane glycoprotein specific for the infective form of Trypanosoma cruzi having adhesive properties to laminin. **The Journal of Biological Chemistry**, v. 274, n. 6, p. 3461–3468, 5 fev. 1999.

GIORDANO, R. J. et al. Biopanning and rapid analysis of selective interactive ligands. **Nature Medicine**, v. 7, n. 11, p. 1249–1253, 1 nov. 2001.

GIORDANO, R. J. et al. From combinatorial peptide selection to drug prototype (I): Targeting the vascular endothelial growth factor receptor pathway. **Proceedings of the National Academy of Sciences**, 26 fev. 2010.

GIRONÈS, N. U.; FRESNO, M. Etiology of Chagas disease myocarditis: Autoimmunity, parasite persistence, or both? **Trends in Parasitology**, v. 19, n. 1, p. 19–22, 2003.

GODFROID, E. et al. Cytokeratins are exposed on the outer surface of established human mammary carcinoma cells. **Journal of Cell Science**, v. 99, n. 3, p. 595–607, 1 jul. 1991.

GONIAS, S. L.; HEMBROUGH, T. A.; SANKOVIC, M. Cytokeratin 8 functions as a major plasminogen receptor in select epithelial and carcinoma cells. **Frontiers in Bioscience: A Journal and Virtual Library**, v. 6, p. D1403-1411, 1 nov. 2001.

GRISARD, E. C. et al. Trypanosoma cruzi Clone Dm28c Draft Genome Sequence. **Genome Announcements**, v. 2, n. 1, p. e01114-13, 27 fev. 2014.

HAJITOU, A. et al. A Hybrid Vector for Ligand-Directed Tumor Targeting and Molecular Imaging. **Cell**, v. 125, n. 2, p. 385–398, 21 abr. 2006.

HAUSER, M.; STEINEGGER, M.; SÖDING, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. **Bioinformatics (Oxford, England)**, v. 32, n. 9, p. 1323–1330, 1 maio 2016.

HEILIG, J. S.; ELBING, K. L.; BRENT, R. Large-Scale Preparation of Plasmid DNA. In: **Current Protocols in Molecular Biology**. [s.l.] John Wiley & Sons, Inc., 2001.

HIGUCHI, M. L. et al. The role of active myocarditis in the development of heart failure in chronic Chagas' disease: a study based on endomyocardial biopsies. **Clinical Cardiology**, v. 10, n. 11, p. 665–670, nov. 1987.

JACOBSSON, K.; FRYKBERG, L. Cloning of ligand-binding domains of bacterial receptors by phage display. **BioTechniques**, v. 18, n. 5, p. 878–885, maio 1995.

JACOBSSON, K.; FRYKBERG, L. Phage display shot-gun cloning of ligand-binding domains of prokaryotic receptors approaches 100% correct clones. **BioTechniques**, v. 20, n. 6, p. 1070–1076, 1078, 1080–1081, jun. 1996.

JACOBSSON, K.; FRYKBERG, L. Shotgun phage display cloning. **Combinatorial chemistry & high throughput screening**, v. 4, n. 2, p. 135–143, abr. 2001.

JANNIN, J.; VILLA, L. An overview of Chagas disease treatment. **Memórias do Instituto Oswaldo Cruz**, v. 102, p. 95–98, out. 2007.

KARANTZA, V. Keratins in health and cancer: more than mere epithelial cell markers. **Oncogene**, v. 30, n. 2, p. 127–138, 13 jan. 2011.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.

KHUSAL, K. G. et al. Prokineticin receptor identified by phage display is an entry receptor for Trypanosoma cruzi into mammalian cells. **Parasitology Research**, v. 114, n. 1, p. 155–165, jan. 2015.

KIERSZENBAUM, F. Chagas' Disease and the Autoimmunity Hypothesis. **Clinical Microbiology Reviews**, v. 12, n. 2, p. 210–223, abr. 1999.

KOENIG, T. et al. Robust Prediction of the MASCOT Score for an Improved Quality Assessment in Mass Spectrometric Proteomics. **Journal of Proteome Research**, v. 7, n. 9, p. 3708–3717, 5 set. 2008.

KREJCI, A. et al. Hammock: a hidden Markov model-based peptide clustering algorithm to identify protein-interaction consensus motifs in large datasets. **Bioinformatics** (Oxford, England), v. 32, n. 1, p. 9–16, 1 jan. 2016.

KROPF, S. P.; SÁ, M. R. The discovery of Trypanosoma cruzi and Chagas disease (1908-1909): tropical medicine in Brazil. **História, Ciências, Saúde-Manguinhos**, v. 16, p. 13–34, jul. 2009.

LAFAILLE, J. J. et al. Structure and expression of two Trypanosoma cruzi genes

encoding antigenic proteins bearing repetitive epitopes. **Molecular and Biochemical Parasitology**, v. 35, n. 2, p. 127–136, 15 jun. 1989.

LARANJA, F. S. et al. Chagas' Disease. **Circulation**, v. 14, n. 6, p. 1035–1060, 1 dez. 1956.

LARMAN, H. B. et al. Autoantigen discovery with a synthetic human peptidome. **Nature Biotechnology**, v. 29, n. 6, p. 535–541, jun. 2011.

LEE, B. Y. et al. Global economic burden of Chagas disease: a computational simulation model. **The Lancet Infectious Diseases**, v. 13, n. 4, p. 342–348, abr. 2013.

LESÉNÉCHAL, M. et al. Cloning and characterization of a gene encoding a novel immunodominant antigen of Trypanosoma cruzi. **Molecular and Biochemical Parasitology**, v. 87, n. 2, p. 193–204, ago. 1997.

LEVENSHTEIN, V. Binary codes capable of correcting deletions, insertions, and reversals. **Soviet Physics Doklady**, v. 10, n. 8, p. 707–710, 1966.

LI, L.; STOECKERT, C. J.; ROOS, D. S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. **Genome Research**, v. 13, n. 9, p. 2178–2189, 1 set. 2003.

LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**, v. 22, n. 13, p. 1658–1659, 1 jul. 2006.

LIMA, S. S. et al. Adhesin activity of Leptospira interrogans lipoprotein identified by in vivo and in vitro shotgun phage display. **Biochemical and Biophysical Research Communications**, v. 431, n. 2, p. 342–347, 8 fev. 2013.

LIU, F. et al. Overexpression of Cell Surface Cytokeratin 8 in Multidrug-Resistant MCF-7/MX Cells Enhances Cell Adhesion to the Extracellular Matrix. **Neoplasia**, v. 10, n. 11, p. 1275–1284, nov. 2008.

LOW, H. P. et al. Amastigote Surface Proteins of Trypanosoma cruzi Are Targets for CD8+ CTL. **The Journal of Immunology**, v. 160, n. 4, p. 1817–1823, 15 fev. 1998.

LU, J. et al. Pentraxins and Fc receptors. **Immunological Reviews**, v. 250, n. 1, p. 230–238, nov. 2012.

MADY, C. et al. Survival and predictors of survival in patients with congestive heart failure due to Chagas' cardiomyopathy. **Circulation**, v. 90, n. 6, p. 3098–3102, 1 dez. 1994.

MAGDESIAN, M. H. et al. Infection by Trypanosoma cruzi. Identification of a parasite ligand and its host cell receptor. **The Journal of Biological Chemistry**, v. 276, n. 22, p. 19382–19389, 1 jun. 2001.

MAGDESIAN, M. H. et al. A conserved domain of the gp85/trans-sialidase family activates host cell extracellular signal-regulated kinase and facilitates Trypanosoma cruzi infection. **Experimental Cell Research**, v. 313, n. 1, p. 210–218, 1 jan. 2007.

MARCHLER-BAUER, A. et al. CDD: a Conserved Domain Database for the functional annotation of proteins. **Nucleic Acids Research**, v. 39, n. Database issue, p. D225-229, jan. 2011.

MARIN-NETO, J. A. et al. Pathogenesis of chronic Chagas heart disease. **Circulation**, v. 115, n. 9, p. 1109–1123, 6 mar. 2007.

MARROQUIN-QUELOPANA, M. et al. Modeling the Trypanosoma cruzi Tc85-11 protein and mapping the laminin-binding site. **Biochemical and Biophysical Research Communications**, v. 325, n. 2, p. 612–618, 10 dez. 2004.

MARTIN, D. L. et al. CD8+ T-Cell responses to Trypanosoma cruzi are highly focused on strain-variant trans-sialidase epitopes. **PLoS pathogens**, v. 2, n. 8, p. e77, ago. 2006.

MATTOS, E. C. et al. The Gp85 surface glycoproteins from Trypanosoma cruzi. **Sub-Cellular Biochemistry**, v. 74, p. 151–180, 2014.

MENDES, T. A. DE O. et al. Identification of Strain-Specific B-cell Epitopes in Trypanosoma cruzi Using Genome-Scale Epitope Prediction and High-Throughput Immunoscreening with Peptide Arrays. **PLOS Neglected Tropical Diseases**, v. 7, n. 10, p. e2524, 31 out. 2013.

MENEGHELLI, U. G. Chagas' disease: a model of denervation in the study of digestive tract motility. **Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas E Biologicas**, v. 18, n. 3, p. 255–264, 1985.

MILES, M. A.; FELICIANGELI, M. D.; DE ARIAS, A. R. American trypanosomiasis (Chagas' disease) and the role of molecular epidemiology in guiding control strategies. **BMJ: British Medical Journal**, v. 326, n. 7404, p. 1444–1448, 28 jun. 2003.

MINENKOVA, O. et al. Identification of tumor-associated antigens by screening phage-displayed human cDNA libraries with sera from tumor patients. **International Journal of Cancer**, v. 106, n. 4, p. 534–544, 10 set. 2003.

MISTRY, J. et al. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. **Nucleic Acids Research**, v. 41, n. 12, p. e121, jul. 2013.

MORENO, M. et al. Trypanosoma cruzi benznidazole susceptibility in vitro does not predict the therapeutic outcome of human Chagas disease. **Memórias Do Instituto Oswaldo Cruz**, v. 105, n. 7, p. 918–924, nov. 2010.

NEWMAN, A. M.; COOPER, J. B. XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. **BMC Bioinformatics**, v. 8, p. 382, 2007.

OBERMAJER, N.; DOLJAK, B.; KOS, J. Cytokeratin 8 ectoplasmic domain binds urokinase-type plasminogen activator to breast tumor cells and modulates their adhesion, growth and invasiveness. **Molecular Cancer**, v. 8, n. 1, p. 1–10, 1 dez. 2009.

OKONECHNIKOV, K.; GOLOSOVA, O.; FURSOV, M. Unipro UGENE: a unified bioinformatics toolkit. **Bioinformatics**, 24 fev. 2012.

OUAISSI, M. A. et al. Trypanosoma Cruzi Infection Inhibited by Peptides Modeled from a Fibronectin Cell Attachment Domain. **Science**, v. 234, n. 4776, p. 603–607, 31 out. 1986.

PALACPAC, N. M. Q. et al. Plasmodium falciparum serine repeat antigen 5 (SE36) as a malaria vaccine candidate. **Vaccine**, v. 29, n. 35, p. 5837–5845, 11 ago. 2011.

PARIS, G. et al. A sialidase mutant displaying trans-sialidase activity. **Journal of Molecular Biology**, v. 345, n. 4, p. 923–934, 28 jan. 2005.

PARKER, J. M.; GUO, D.; HODGES, R. S. 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. **Biochemistry**, v. 25, n. 19, p. 5425–5432, 23 set. 1986.

PASQUALINI, R.; RUOSLAHTI, E. Organ targeting in vivo using phage display peptide libraries. **Nature**, v. 380, n. 6572, p. 364–366, 28 mar. 1996.

PEREIRA, K. S. et al. Chagas' Disease as a Foodborne Illness. **Journal of Food Protection**, v. 72, n. 2, p. 441–446, 1 fev. 2009.

PEREIRA, M. E. et al. Invasive Phenotype of Trypanosoma Cruzi Restricted to a Population Expressing Trans-Sialidase. **Infection and Immunity**, v. 64, n. 9, p. 3884–3892, 1 set. 1996.

PICADO, A. et al. Development of Diagnostics for Chagas Disease: Where Should We Put Our Limited Resources? **PLOS Neglected Tropical Diseases**, v. 11, n. 1, p. e0005148, 1 maio 2017.

PINAZO, M.-J. et al. Biological markers for evaluating therapeutic efficacy in Chagas disease, a systematic review. **Expert Review of Anti-infective Therapy**, v. 12, n. 4, p. 479–496, 1 abr. 2014.

RASSI JR, A.; RASSI, A.; MARIN-NETO, J. A. Chagas disease. **The Lancet**, v. 375, n. 9723, p. 1388–1402, 17.

RASSI JR, A.; RASSI, S. G.; RASSI, A. Sudden death in Chagas' disease. **Arquivos Brasileiros de Cardiologia**, v. 76, n. 1, p. 86–96, jan. 2001.

REIS-CUNHA, J. L. et al. Genome-wide screening and identification of new Trypanosoma cruzi antigens with potential application for chronic Chagas disease diagnosis. **PloS One**, v. 9, n. 9, p. e106304, 2014.

ROBINS, H. Immunosequencing: applications of immune repertoire deep sequencing. **Current Opinion in Immunology**, Special section: Systems biology and bioinformatics / Immunogenetics and transplantation. v. 25, n. 5, p. 646–652, 1 out. 2013.

ROSENBERG, C. S.; MARTIN, D. L.; TARLETON, R. L. CD8+ T Cells Specific for

- Immunodominant Trans-sialidase Epitopes Contribute to Control of Trypanosoma cruzi Infection but are Not Required for Resistance. **Journal of immunology (Baltimore, Md.: 1950)**, v. 185, n. 1, p. 560–568, 1 jul. 2010.
- ROSSI, M. A. The pattern of myocardial fibrosis in chronic Chagas' heart disease. **International Journal of Cardiology**, v. 30, n. 3, p. 335–340, 1 mar. 1991.
- ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p. 725–738, abr. 2010.
- SAJJAN, U.; ACKERLEY, C.; FORSTNER, J. Interaction of cblA/adhesin-positive Burkholderia cepacia with squamous epithelium. **Cellular Microbiology**, v. 4, n. 2, p. 73–86, fev. 2002.
- SAJJAN, U. S.; SYLVESTER, F. A.; FORSTNER, J. F. Cable-Piliated Burkholderia cepacia Binds to Cytokeratin 13 of Epithelial Cells. **Infection and Immunity**, v. 68, n. 4, p. 1787–1795, abr. 2000.
- SAMUEL, J. et al. Cardiac thrombosis and thromboembolism in chronic Chagas' heart disease. **The American Journal of Cardiology**, v. 52, n. 1, p. 147–151, jul. 1983.
- SANTOS, M. A.; GARG, N.; TARLETON, R. L. The identification and molecular characterization of Trypanosoma cruzi amastigote surface protein-1, a member of the trans-sialidase gene super-family. **Molecular and Biochemical Parasitology**, v. 86, n. 1, p. 1–11, maio 1997.
- SCHULTE, T. et al. The basic keratin 10-binding domain of the virulence-associated pneumococcal serine-rich protein PsrP adopts a novel MSCRAMM fold. **Open Biology**, v. 4, p. 130090, 2014.
- SCOTT, J.; SMITH, G. Searching for peptide ligands with an epitope library. **Science**, v. 249, n. 4967, p. 386–390, 27 jul. 1990.
- SHERWOOD, A. M. et al. Deep Sequencing of the Human TCR $\gamma$  and TCR $\beta$  Repertoires Suggests that TCR $\beta$  Rearranges After  $\alpha\beta$  and  $\gamma\delta$  T Cell Commitment. **Science Translational Medicine**, v. 3, n. 90, p. 90ra61-90ra61, 6 jul. 2011.
- SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, p. 539, 2011.
- SMITH, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. **Science**, v. 228, n. 4705, p. 1315–1317, 14 jun. 1985.
- SOUZA-LIMA, R. DE C. DE et al. Outbreak of acute Chagas disease associated with oral transmission in the Rio Negro region, Brazilian Amazon. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 46, n. 4, p. 510–514, ago. 2013.
- STAQUICINI, D. I. et al. Role of GP82 in the Selective Binding to Gastric Mucin during Oral Infection with Trypanosoma cruzi. **PLoS Negl Trop Dis**, v. 4, n. 3, p. e613, 2 mar. 2010.

SUNDARESAN, R.; SAMEN, U.; PONNURAJ, K. Structure of KRT4 binding domain of Srr-1 from Streptococcus agalactiae reveals a novel β-sheet complementation. **International Journal of Biological Macromolecules**, v. 75, p. 97–105, abr. 2015.

TARLETON, R. L. Chagas disease: a role for autoimmunity? **Trends in Parasitology**, v. 19, n. 10, p. 447–451, out. 2003.

TEIXEIRA, A. A. R. et al. Trypanosoma cruzi Binds to Cytokeratin through Conserved Peptide Motifs Found in the Laminin-G-Like Domain of the gp85/Trans-sialidase Proteins. **PLoS neglected tropical diseases**, v. 9, n. 9, p. e0004099, set. 2015.

TEIXEIRA, A. R. L. et al. Pathogenesis of Chagas' Disease: Parasite Persistence and Autoimmunity. **Clinical Microbiology Reviews**, v. 24, n. 3, p. 592–630, jul. 2011.

THE LANCET. Chagas' disease—an epidemic that can no longer be ignored. **The Lancet**, v. 368, n. 9536, p. 619, 19 ago. 2006.

THOMAS, M. C. et al. Characterization of an Immunodominant Antigenic Epitope from Trypanosoma cruzi as a Biomarker of Chronic Chagas' Disease Pathology. **Clinical and Vaccine Immunology: CVI**, v. 19, n. 2, p. 167–173, fev. 2012.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673–4680, 11 nov. 1994.

TISI, D. et al. Structure of the C-terminal laminin G-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin. **The EMBO iournal**, v. 19, n. 7, p. 1432–1440, 3 abr. 2000.

TONELLI, R. R. et al. 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**, v. 4, n. 11, p. e864, 2 nov. 2010.

ULLU, E.; TSCHUDI, C.; CHAKRABORTY, T. RNA interference in protozoan parasites. **Cellular Microbiology**, v. 6, n. 6, p. 509–519, jun. 2004.

VALIENTE-GABIOUD, A. A. et al. Effect of repetitiveness on the immunogenicity and antigenicity of Trypanosoma cruzi FRA protein. **Experimental Parasitology**, v. 127, n. 3, p. 672–679, mar. 2011.

VIOTTI, R. et al. Side effects of benznidazole as treatment in chronic Chagas disease: fears and realities. **Expert Review of Anti-Infective Therapy**, v. 7, n. 2, p. 157–163, mar. 2009.

VITA, R. et al. The immune epitope database (IEDB) 3.0. **Nucleic Acids Research**, v. 43, n. Database issue, p. D405-412, jan. 2015.

WELLS, M. J. et al. Cytokeratin 18 Is Expressed on the Hepatocyte Plasma Membrane Surface and Interacts with Thrombin-Antithrombin Complexes. **Journal of Biological Chemistry**, v. 272, n. 45, p. 28574–28581, 7 nov. 1997.

**WHO | What is Chagas disease?** Disponível em: <a href="http://www.who.int/chagas/disease/en/">http://www.who.int/chagas/disease/en/</a>>. Acesso em: 25 abr. 2017.

YAGI, M. et al. Protective Epitopes of the Plasmodium falciparum SERA5 Malaria Vaccine Reside in Intrinsically Unstructured N-Terminal Repetitive Sequences. **PLOS ONE**, v. 9, n. 6, p. e98460, 6 fev. 2014.

ZHANG, L. et al. A second IgG-binding protein in Staphylococcus aureus. **Microbiology**, v. 144, n. 4, p. 985–991, 1 abr. 1998.

ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. **Proteins: Structure, Function, and Bioinformatics**, v. 57, n. 4, p. 702–710, 1 dez. 2004.

# Anexo - Súmula curricular

Dados pessoais

Nome: André Azevedo Reis Teixeira

Data de nascimento: 16 de setembro de 1988

Naturalidade: Ribeirão Preto, São Paulo

# Formação acadêmica

**2012-2017** Doutorado em Ciências Biológicas (Bioquímica)

Universidade de São Paulo, São Paulo, Brasil Orientador: Prof. Dr. Ricardo José Giordano

Coorientadora: Profª. Drª. Maria Júlia Manso Alves

Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo

**2016-2017** Estágio de Pesquisa no Exterior

University of New Mexico, Novo México, Estados Unidos da América

Supervisão: Dr. Renata Pasqualini

Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo

2007-2011 Graduação em Farmácia Bioquímica

Universidade de São Paulo, Ribeirão Preto, Brasil

#### Artigo publicados em periódicos

<u>TEIXEIRA, A. A.</u> et al. Going viral? Linking the etiology of human prostate cancer to the PCA3 long noncoding RNA and oncogenic viruses. **EMBO molecular medicine**, 27 jul. 2017.

<u>TEIXEIRA, A. A. R.</u> et al. Trypanosoma cruzi Binds to Cytokeratin through Conserved Peptide Motifs Found in the Laminin-G-Like Domain of the gp85/Trans-sialidase Proteins. **PLOS Neglected Tropical Diseases**, v. 9, n. 9, p. e0004099, 23 set. 2015.

<u>TEIXEIRA, A. A. R.</u>; LUND, M.; DA SILVA, F. L. B. Fast Proton Titration Scheme for Multiscale Modeling of Protein Solutions. **J. Chem. Theory Comput.**, v. 6, n. 10, p. 3259–3266, 2010.

### Capítulos de livro publicados

MAGALHAES, L.S.; <u>TEIXEIRA, A.A.R.</u>; CARNEIRO, J.L.V.; NUNES, D.N.; DIASNETO, E.; GIORDANO, R.J. Phage display: aspectos básicos e perspectivas atuais. Em **Biotecnologia Aplicada a Saúde** - Vol 3, 2015

### Apresentação de trabalhos em eventos

<u>TEIXEIRA, A. A. R.</u>; TANG, F. H. F.; FERREIRA, F. L.; LEE HO, P.; COLLI, W.; ALVES, M. J. M.; Giordano, R. J. . Exploring the Trypanosoma cruzi Genome Using Phage Display Whole Genome Shotgun. XXXI Reunião Anual da Sociedade Brasileira

de Protozoologia. 2015.

<u>TEIXEIRA, A. A. R.</u>; FERREIRA, F. L.; LEE HO, P.; COLLI, W.; ALVES, M. J. M.; Giordano, R. J.. Whole Genome Shotgun Phage-Display For The De Novo Discovery Of Trypanosoma cruzi And Host Cell Interactions. 23rd Congress of the International Union of Biochemistry and Molecular Biology. 2015.

<u>TEIXEIRA, A. A. R.</u>; ALVES, M. J. M.; COLLI, W. Investigating Trypanosoma cruzi adhesion to host cell via gp85/trans-sialidase protein. Gordon Research Conference: Biology of Host-Parasite Interactions. 2014.

<u>TEIXEIRA, A. A. R.</u>; ALVES, M. J. M.; Giordano, R. J. .Identification of a Novell Cell Binding Peptide Belonging to the gp85/transialidase Superfamily. XLIII Annual Meeting of the Brazilian Society for Biochemistry and Molecular Biology (SBBq) 2014.

<u>TEIXEIRA, A. A. R.</u>; DA SILVA, F. L. B. . Agregação Protéica: Estudos in-silico por modelos moleculares simplistas. SIICUSP 2009.

<u>TEIXEIRA, A. A. R.</u>; DA SILVA, F. L. B. . Physico chemical properties of the complex Calcium-Calbindin D9K: The mutation E60Q puzzle. IV Escola de Modelagem Molecular em SIstemas Biológicos. 2008.

### Monitoria em disciplinas

2014. QBQ0313 – Bioquímica. Instituto de Química, Universidade de São Paulo, SP.

2014. QBQ0314 – Bioquímica da Nutrição. Instituto de Química, Universidade de São Paulo, SP.

2015. QBQ0317 – Biologia Molecular. Instituto de Química, Universidade de São Paulo, SP.

2016. QBQ4025 – Bioquímica Experimental. Instituto de Química, Universidade de São Paulo, SP.

### Palestras ministradas

23/08/2017. Anticorpos como ferramentas para o estudo de patógenos. Curso de Bioquímica Experimental Avançada, Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo.

27/04/2016. Explorando a Interação do Parasita *Trypanosoma cruzi* com a célula hospedeira. Ciclo de Seminários dos Pós-Graduandos e Pós-Doutorandos do Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo.

### Participação em cursos e workshops

2015. EuPathDB Workshop. University of Georgia

2013. Il Curso Internacional de Biologia Molecular de Tripanosomatídeos. Fiocruz (PR)

2013. MiSeq System Operation and Nextera Sample Preparation. Illumina Inc.