

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Bioquímica e Imunologia

**Clonagem, expressão e purificação de membros
da família de Proteína de Superfície Associada a
Mucina (MASP) de *Trypanosoma cruzi* para
estudos estruturais**

Simara Semíramis de Araújo

Belo Horizonte / MG

2010

Simara Semíramis de Araújo

Clonagem, expressão e purificação de membros da família de Proteína de Superfície Associada a Mucina (MASP) de *Trypanosoma cruzi* para estudos estruturais

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica e Imunologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Bioquímica e Imunologia.

Área de concentração: Estrutura e função de proteínas e peptídeos.

Prof. Dr. Ronaldo Alves Pinto Nagem

Orientador

Prof^a. Dr^a. Daniella Castanheira Bartholomeu

Prof^a. Dr^a. Santuza Maria Ribeiro Teixeira

Co-Orientadoras

Instituto de Ciências Biológicas da UFMG

Belo Horizonte / MG

2010



Universidade Federal de Minas Gerais -
 Curso de Pós-Graduação em Bioquímica e Imunologia ICB/UFMG
 Av. Antônio Carlos, 6627 - Pampulha
 31270-901 - Belo Horizonte - MG
 e-mail: pg-biq@icb.ufmg.br (31)3499-2615



ATA DA DEFESA DA DISSERTAÇÃO DE MESTRADO DE SIMARA SEMIRAMIS DE ARAÚJO. Aos doze dias do mês de fevereiro de 2010 às 14:00 horas, reuniu-se no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais a Comissão Examinadora da dissertação de Mestrado, indicada *ad referendum* do Colegiado, para julgar, em exame final, o trabalho intitulado "Clonagem, expressão e purificação de membros da família de proteína de superfície associada a mucina (MASP) de *Trypanosoma cruzi*" requisito final para a obtenção do grau de Mestre em Bioquímica e Imunologia. Abrindo a sessão o Presidente da Comissão Prof. Ronaldo Alves Pinto Nagem da Universidade Federal de Minas Gerais, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra a candidata para apresentação de seu trabalho. Seguiu-se a argüição pelos examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: Dr. Adriano Monteiro de Castro Pimenta da Universidade federal de Minas Gerais, aprovada; Dr. Carlos Renato Machado da Universidade federal de Minas Gerais, aprovada; Dra. Daniella Castanheira Bartholomeu-co-orientadora, da Universidade federal de Minas Gerais, aprovada; Dra. Santuza Maria Ribeiro Teixeira, co-orientadora, da Universidade Federal de Minas Gerais, aprovada; Dr. Ronaldo Alves Pinto Nagem, orientador, da Universidade Federal de Minas Gerais, aprovada. Pelas indicações a candidata foi considerada APROVADA. O resultado final foi comunicado publicamente a candidata pelo Presidente da Comissão. Nada mais havendo tratar o Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 12 de fevereiro de 2010.

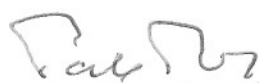

 Dr. Adriano Monteiro de Castro Pimenta – UFMG


 Dr. Carlos Renato Machado – UFMG


 Dra. Daniella Castanheira Bartholomeu – co-orientadora – UFMG


 Dra. Santuza Maria Ribeiro Teixeira – co-orientadora – UFMG


 Dr. Ronaldo Alves Pinto Nagem – orientador - UFMG


 Prof. Carlos Renato Machado
 Coordenador "protempore" do Curso
 de Pós-Graduação em
 Bioquímica - Imunologia
 ICB - UFMG

Meus sinceros agradecimentos...

Ao prof. Dr. Ronaldo Nagem, pelo entusiasmo em ensinar, pela acessibilidade sempre presente entre aluno e orientador, pelo empenho em realizar conosco os experimentos e pela vontade diária em querer aprender.

À prof^a. Dr^a. Daniella Bartholomeu, por acolher-me tão bem em seu laboratório quando o grupo de Biologia Estrutural começava a se formar. Obrigada pelos ensinamentos durante as reuniões, pela disponibilidade, paciência e confiança!

À prof^a. Dr^a. Santuza Teixeira, por receber-me também em seu laboratório e, logo, por conceder-me livre acesso a reagentes e equipamentos. Obrigada por permitir-me participar de suas reuniões de grupo e discussões de artigos, sem esquecer os incentivos às consultas semanais no PUBMED!

Às agências de fomento e instituições que permitiram a realização deste trabalho: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Organização Mundial de Saúde (OMS).

Aos membros da banca examinadora, titulares e suplente.

Ao prof. Dr. Frederic Frezard e à sua aluna de doutorado Flaviana Fernandes, ambos do Departamento de Fisiologia e Biofísica, por nos concederem o uso do equipamento de dicroísmo circular.

Ao Dr. Agenor Valadares, pelas análises no espectrômetro de massas.

Ao Juliano Leal, pelas inúmeras sequências de DNA.

À Celise Maria Ferreira Costa, secretária do Departamento de Bioquímica e Imunologia, por sempre atender-me com muita gentileza.

Ao Jamil, ao Elimar, e a todos os funcionários do Departamento.

Ao Márcio, pelo apoio na reforma do Laboratório de Biologia Estrutural.

Aos amigos do Laboratório de Biologia Estrutural e do Laboratório de Genômica de Parasitas, com quem muito aprendi sobre o respeito, a disponibilidade em ajudar e o trabalho em equipe. Obrigada pelos risos, pela convivência sadia e pelas horas de descontração.

Aos professores.

À minha família, por me ensinarem sobre o mais importante, o que permanece por toda a vida: o amor de Deus (I João 4, 9-16).

RESUMO

A família gênica *masp* (proteína de superfície associada a mucina), identificada durante a anotação do genoma do *Trypanosoma cruzi*, é a segunda maior família multigênica neste parasito, consistindo de aproximadamente 1400 membros. Membros da família MASP contêm domínios N- e C-terminais conservados que codificam para um peptídeo sinal e um sítio de adição de âncora de GPI, respectivamente. A região central é variável em comprimento e em sequência de aminoácidos, e contém um grande repertório de motivos repetitivos, os quais sugerem que MASP esteja envolvida em interações parasito-hospedeiro. Uma vez que nenhum membro da família MASP foi caracterizado até o momento, decidiu-se estudar as estruturas secundária e terciária de diferentes MASPs. Para este objetivo, optou-se por clonar e expressar em *Escherichia coli* a região central das sequências de quatro membros MASP. Para isso, algumas características das proteínas alvo foram consideradas, tais como curto comprimento da sequência de aminoácidos, em torno de 170 resíduos, baixo conteúdo de sítios potenciais de glicosilação e baixo grau de desordem estrutural. Uma vez que regiões de baixa complexidade normalmente correspondem a segmentos desenovelados ou estruturas extendidas, foi utilizado o servidor PONDR VL-XT para prever regiões desordenadas nos quatro membros MASP. Proteínas nativamente desenoveladas têm sido descritas por possuírem elevada taxa de resíduos carregados e baixa taxa de resíduos hidrofóbicos. Gráficos de carga líquida média versus taxa de hidrofobicidade foram obtidos para as regiões centrais das quatro MASPs. Estas análises revelaram a presença das quatro regiões centrais no espaço designado para proteínas nativamente desenoveladas. As sequências foram amplificadas por PCR, clonadas nos sistemas de clonagem TOPO e/ou pGEM-T e submetidas ao sequenciamento de DNA. Posteriormente, os produtos de PCR foram transferidos para dois vetores de expressão derivados do vetor pET-28a (Novagen), que foram modificados no Centro de Biologia Molecular e Estrutural (CeBiME, Campinas / SP). Em ambos os vetores, o sítio de reconhecimento por trombina foi substituído pelo sítio de reconhecimento da protease TEV (tobacco etch vírus). As quatro proteínas MASP em fusão a cauda de histidina N-terminal foram expressas na forma solúvel, enquanto que aquelas fusionadas a GST N-terminal foram expressas como corpos de inclusão insolúveis, exceto a recombinante MASP1:GST. A purificação da MASP 3 fusionada a cauda de histidina foi realizada por cromatografia de afinidade. Após a clivagem da etiqueta por TEV, a proteína foi purificada por cromatografia de troca iônica. O estudo da estrutura tridimensional de proteínas MASP permitirá uma compreensão mais detalhada destas moléculas, tal como a localização de seus epítomos e seu enovelamento, o que pode contribuir para a identificação de sua função biológica no *T. cruzi*.

Palavras-chave: *Trypanosoma cruzi*. Proteína de Superfície Associada a Mucina. Cristalografia de proteínas.

ABSTRACT

The *Trypanosoma cruzi* *masp* (mucin associated surface protein) gene family was identified during the annotation of the genome as the second largest gene family in this human pathogen, consisting of approximately 1400 members. MASP members contain N- and C- terminal conserved domains that encode a putative signal peptide and a GPI-anchor addition site, respectively. The central region is variable both in length and in amino acid sequence and contains a large repertoire of repetitive motifs, which suggests MASP may be involved in parasite-host cell interaction. Since no member of the MASP family has been characterized to date, we have decided to study the secondary and tertiary structures of distinct MASPs. For this purpose, we have decided to clone and express in *E. coli* the central region sequences of four MASP members. For this, some features of the target proteins were considered such as short amino acid sequence length, around 170 residues, low content of potential glycosylation sites and low intrinsic disorder degree. Knowing that sequence regions with low complexity nearly always correspond to nonfolding segments or extended structures, we have used the web server PONDR VL-XT to predict disordered regions in MASPs. Natively unfolded proteins have been described to possess a high ratio of charged residues and a low ratio of hydrophobic residues. The mean net charge versus mean hydrophobicity ratio was plotted for the four MASPs central region. This analysis showed that the four central regions falls in the phase space of natively unfolded proteins. After that, the sequences were PCR amplified and cloned into TOPO and/or pGEM-T cloning systems and submitted to DNA sequencing. Subsequently, the PCR products were transferred to two expression vectors derived from pET-28a vector (Novagen), which were modified in the Center of Molecular and Structural Biology (CeBiME, Campinas/SP). In both vectors, the thrombin recognition site was replaced by a TEV (*tobacco etch virus*) protease recognition site. The four N-terminal fusion His-tag MASP proteins were expressed in the soluble forms, whereas the N-terminal GST fusion MASPs were expressed as insoluble inclusion bodies, except the recombinant GST:MASP1. Purification of the Histidine-tagged MASP3 was carried out by affinity chromatography. After His-tag cleavage by TEV, the protein was purified by ion exchange chromatography. The study of the three-dimensional structure of MASP proteins will enable a more detailed understanding of these molecules, such as the location of their epitopes and its fold, which may contribute to the identification of its biological function on *T. cruzi*.

Key-words: *Trypanosoma cruzi*. Mucin Associated Surface Protein. Protein crystallography

Lista de Ilustrações

Figura 1: Distribuição geográfica da infecção humana pelo <i>Trypanosoma cruzi</i> nas Américas, segundo a OMS.....	1
Figura 2: Ciclo de vida esquemático do parasito <i>T. cruzi</i>	2
Figura 3: Distribuição de proteínas de superfície em <i>L. major</i> , <i>T. brucei</i> e <i>T. cruzi</i>	6
Figura 4: Representação em fita das duas estruturas conhecidas de VSG, ILTat 1.24 e MITat 1.2.....	8
Figura 5: Domínios estruturais e representação da superfície de Leishmanolisina.....	9
Figura 6: <i>trans</i> -Sialidase de <i>Trypanosoma cruzi</i>	11
Figura 7: Características das sequências polipeptídicas de MASP	13
Figura 8: Representação esquemática da variabilidade de proteínas MASP.....	14
Figura 9: Ensaio de imunofluorescência de tripomastigotas fixados com paraformaldeído e incubados com anticorpo anti-MASP.	15
Figura 10: Desenho esquemático dos vetores de expressão pET28a-TEV e pET28a-GST-TEV.....	30
Figura 11: Análise <i>in silico</i> para identificação de sítios de N- e O-glicosilação nas MASPs 1, 2, 3 e 4.....	40
Figura 12: Predições de desordem estrutural nas MASPs 1, 2, 3 e 4.....	42
Figura 13: Amplificação das regiões codificadoras centrais dos genes <i>masp</i> utilizando iniciadores específicos para posterior clonagem nos vetores pET28a-TEV e pET28a-GST-TEV.....	43
Figura 14: Amplificação do sítio de clonagem dos vetores TOPO e pGEM utilizando iniciadores M13 direto e reverso, após subclonagem dos insertos de MASP 1, 2, 3 e 4, para posterior clonagem em pET28a-TEV e pET28a-GST-TEV	44
Figura 15: Alinhamento entre as extremidades da região codificadora central de Tc00.1047053510359.550 e os iniciadores 7084.t00005_F:pET.GST.TEV e 7084.t00005_R, desenhados especificamente para amplificar a região codificadora central de Tc00.1047053506757.50.....	45
Figura 16: Dupla digestão enzimática dos vetores TOPO e pGEM contendo insertos de MASP 1, 2, 3 e 4	46
Figura 17: Digestão dos vetores pET28a-TEV e pET28a-GST-TEV com enzimas de restrição indicadas na própria figura.....	47
Figura 18: Amplificação dos insertos de MASP 1, 2, 3 e 4 utilizando iniciadores específicos, após clonagem em pET28a-TEV e pET28a-GST-TEV	48

Figura 19: Alinhamento entre a sequência nucleotídica codificadora central de Tc00.1047053507957.300 (MASP 3) e a sequência obtida para a construção de MASP 3 em pET28a-TEV	49
Figura 20: Perfil eletroforético de proteínas totais de bactérias <i>E. coli</i> BL21 Star, transformadas com os vetores de expressão pET28a-TEV e pET28a-GST-TEV, com ou sem inserto, antes e após indução da expressão por IPTG.	51
Figura 21: <i>Western blotting</i> das frações correspondentes à indução por 3 horas e ao sobrenadante de MASPs 1, 2, 3 e 4 fusionadas a cauda de seis resíduos de histidina (_{6x} His) N-terminal.	52
Figura 22: Purificação da proteína _{6x} His:MASP3 por cromatografia de afinidade a níquel.	54
Figura 23: Perfis de massas de MASP 3.	55
Figura 24: Purificação da proteína MASP3 por cromatografia de troca iônica.	56
Quadro 1: Oligonucleotídeos sintéticos utilizados como iniciadores nas reações de amplificação das regiões codificadoras centrais dos genes <i>masp</i>	24
Quadro 2: Enzimas e tampões utilizados nas reações de dupla digestão dos vetores de sequenciamento recombinantes.....	29
Quadro 3: Enzimas e tampões utilizados nas reações de dupla digestão dos vetores de expressão.	31
Quadro 4: Relação entre o comprimento de sequências nucleotídicas e de resíduos de aminoácidos das proteínas MASP selecionadas.....	39

Lista de Abreviaturas e Siglas

ACN	acetonitrila
<i>apud</i>	citado por
BZ	Benzonidazol
CD	dicroísmo circular
Da	Dalton
DNA	ácido desoxirribonucléico
DO	densidade óptica
DTT	ditiotreitól
GPI	glicosil-fosfatidil-inositol
GP63	Glicoproteína de 63 kDa
GST	Glutathiona-S-transferase
IPTG	isopropil- β -D-tiogalactopiranosídeo
kDa	kiloDalton
M	concentração molar (moles/L)
mM	concentração milimolar (10^{-3} moles/L)
μ M	concentração micromolar (10^{-6} moles/L)
MALDI	ionização a laser assistida por matriz
MASP	Proteína de superfície associada a mucina
MSP	Protease principal de superfície
NFX	Nifurtimox
PBS	tampão fosfato salina
PCR	reação em cadeia da polimerase
PDB	Banco de dados de proteína
PM	peso molecular
q.s.p.	quantidade suficiente para
RMN	ressonância magnética nuclear
SDS	dodecil-sulfato de sódio
TFA	ácido trifluoracético
TEV	vírus <i>etch</i> do tabaco (protease)
rpm	rotações por minuto

V	Volts
vc	volume de coluna
VSG	Glicoproteína variante de superfície

As demais abreviaturas e siglas serão explicadas quando forem citadas pela primeira vez no texto.

Abreviações (em uma e três letras) dos resíduos de aminoácidos

A	Ala	Alanina
C	Cys	Cisteína
D	Asp	Aspartato
E	Glu	Glutamato
F	Phe	Fenilalanina
G	Gly	Glicina
H	His	Histidina
I	Ile	Isoleucina
K	Lys	Lisina
L	Leu	Leucina
M	Met	Metionina
N	Asn	Asparagina
P	Pro	Prolina
Q	Gln	Glutamina
R	Arg	Arginina
S	Ser	Serina
T	Thr	Treonina
V	Val	Valina
W	Trp	Triptofano
Y	Tyr	Tirosina

Sumário

1. INTRODUÇÃO	1
1.1. O parasito <i>Trypanosoma cruzi</i> e a doença de Chagas	1
1.2. Proteínas de superfície em tripanosomatídeos	5
1.2.1. VSG	7
1.2.2. MSP/GP63	8
1.2.3. Trans-Sialidase	10
1.2.4. Mucinas	12
1.2.5. MASP	12
1.3. Caracterização estrutural de proteínas.....	16
1.3.1. Cristalografia de proteínas por difração de raios-X.....	17
1.3.2. Espectroscopia de ressonância magnética nuclear.....	19
1.3.3. Espectroscopia de dicroísmo circular.....	20
1.4. Justificativa.....	21
1.5. Objetivos	22
1.5.1. Geral	22
1.5.2. Específicos	22
2. MATERIAIS E MÉTODOS	23
2.1. Seleção das proteínas MASP.....	23
2.2. Oligonucleotídeos usados para amplificação das regiões codificadoras centrais dos genes <i>masp</i>	24
2.3. Reação em Cadeia da Polimerase (PCR)	25
2.4. Eletroforese em gel de agarose.....	25
2.5. Reação de ligação de fragmentos a vetores.....	26
2.6. Preparo de bactérias	26
2.6.1. Eletrocompetentes	26
2.6.2. Termocompetentes.....	27
2.7. Transformação de bactérias.....	27
2.7.1. Por eletroporação	27
2.7.2. Por choque térmico	28
2.8. Extração de plasmídeos recombinantes.....	28
2.9. Digestão enzimática dos vetores de sequenciamento	29
2.10. Digestão enzimática dos vetores de expressão.....	30
2.11. Sequenciamento dos fragmentos inseridos nos plasmídeos recombinantes.....	31

2.12. Indução da expressão de proteínas recombinantes	32
2.13. Teste de solubilidade	32
2.14. Eletroforese em gel de poliacrilamida-SDS	32
2.15. <i>Western blotting</i>	33
2.16. Lise celular	34
2.17. Cromatografia de afinidade	34
2.18. Diálise e determinação da concentração de proteínas	35
2.19. Remoção da cauda de histidina	35
2.20. Espectrometria de massas.....	36
2.21. Cromatografia de troca iônica	36
2.22. Espectroscopia de dicroísmo circular.....	37
3. RESULTADOS	38
3.1. Seleção das proteínas MASP.....	38
3.2. Amplificação e subclonagem das regiões codificadoras centrais dos genes <i>masp</i> em vetores de sequenciamento.....	43
3.3. Clonagem das regiões codificadoras centrais dos genes <i>masp</i> em vetores de expressão. 45	
3.4. Expressão das proteínas MASP recombinantes	50
3.5. Purificação da proteína MASP 3 recombinante e remoção da etiqueta de afinidade.....	53
4. DISCUSSÃO	57
5. CONCLUSÕES	64
REFERÊNCIAS BIBLIOGRÁFICAS	65

1. INTRODUÇÃO

1.1. O parasito *Trypanosoma cruzi* e a doença de Chagas

A doença de Chagas, também conhecida como Tripanosomíase Americana, causada pelo protozoário flagelado *Trypanosoma cruzi*, pertencente à ordem Kinetoplastida, família Tripanosomatidae, ocorre no continente americano, com distribuição geográfica desde a Argentina até o sul dos Estados Unidos da América (Fig. 1) (COURA & DIAS, 2009).

Estimativas do início da década de 90 indicaram que entre 16 e 18 milhões de pessoas estavam infectadas e entre 90 e 100 milhões estavam sob risco da infecção (WHO, 1991). Na mesma década, países do Cone Sul (Argentina, Bolívia, Brasil, Chile, Paraguai e Uruguai) iniciaram programas de controle da doença de Chagas, que visavam à interrupção da transmissão do *T. cruzi* ao eliminar insetos vetores domésticos, além da melhoria da fiscalização de bancos de sangue para reduzir o risco de transmissão transfusional. Como resultado de tais iniciativas, estimativas mais recentes indicam uma prevalência de 9 a 12 milhões de pessoas infectadas (REMME *et al.*, 2006).



Figura 1: Distribuição geográfica da infecção humana pelo *Trypanosoma cruzi* nas Américas, segundo a OMS. Extraído e adaptado de Coura e Dias, 2009.

O parasito, seu vetor e a doença foram descritos por Carlos Chagas em 1909 (CHAGAS, 1909). O parasito é digenético e infecta alternadamente um hospedeiro vertebrado mamífero e um hospedeiro invertebrado, inseto hematófago da ordem Hemiptera, família Reduviidae, subfamília Triatominae. O ciclo biológico do *T. cruzi* é complexo e envolve estágios que apresentam diferenças morfológicas, ultra-estruturais, funcionais e bioquímicas (BRENER, 1973). Classicamente, seu ciclo de vida é descrito apresentando três formas: a forma epimastigota, a qual é replicativa; a amastigota, replicativa e infectiva (GASPAR *et al.*, 2009); e a tripomastigota, não replicativa e infectiva (Fig. 2).

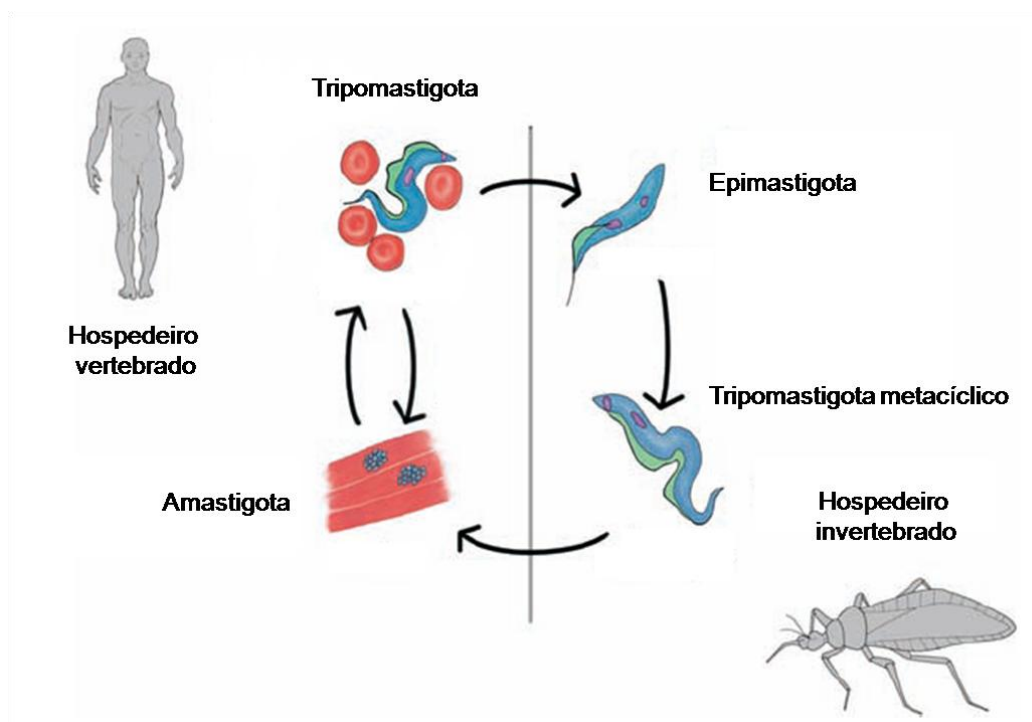


Figura 2: Ciclo de vida esquemático do parasito *T. cruzi*. Extraído e adaptado de Atwood e colaboradores, 2005.

No hospedeiro invertebrado, o ciclo biológico inicia-se após a ingestão de sangue de animais infectados durante o repasto sanguíneo pelo inseto vetor. Ao chegar ao estômago do triatomíneo, formas tripomastigotas sanguíneas iniciam sua transformação para formas epimastigotas. Estas migram para o intestino onde se multiplicam por divisão binária. Na porção posterior do intestino ocorre diferenciação celular dos epimastigotas em tripomastigotas metacíclicos, formas infectantes para o hospedeiro vertebrado que são liberadas nas fezes e na urina do triatomíneo durante o repasto sanguíneo (GARCIA & AZAMBUJA, 2000; DE SOUZA, 2000).

O ciclo de vida do *T. cruzi* no hospedeiro vertebrado inicia-se após a penetração dos tripomastigotas metacíclicos através de feridas na pele ou nas mucosas do hospedeiro. Ao atingirem a corrente sanguínea, os parasitos invadem vários tipos celulares, dentro dos quais se diferenciam em amastigotas. Após sucessivas divisões binárias no citoplasma celular, os amastigotas se diferenciam em tripomastigotas sanguíneos culminando com a ruptura da célula e a liberação na corrente sanguínea de parasitos que podem infectar novas células do hospedeiro vertebrado ou serem ingeridos pelo inseto vetor (DE SOUZA, 2000).

A transmissão do *T. cruzi* ao homem pode ainda ser efetuada através de transfusão sanguínea, via transmissão congênita e, com menor frequência, por acidentes de laboratório, transplante de órgãos e via transmissão oral (COURA, 2007; DIAS, 2009).

Atualmente, a transmissão da doença de Chagas está praticamente interrompida em extensas regiões do Brasil e de outros países latino-americanos graças à erradicação do vetor *Triatoma infestans*, mediante a luta racional com inseticidas de longa ação residual e a melhoria da habitação rural. Em contrapartida, outras espécies de triatomíneos consideradas como vetores em potencial (*Triatoma brasiliensis*, *Triatoma pseudomaculata*, *Triatoma sordida*, *Panstrongylus megistus*) tendem a persistir, esporadicamente invadindo o ambiente domiciliar. Diante disso, merecem particular atenção áreas susceptíveis à expansão de fronteiras para a agricultura, as quais se constituem em entrada no ambiente selvagem, tais como a região amazônica. Nos estados do Pará, Amapá e Amazonas, mais de 400 casos agudos foram reportados até o final da década de 90, a maioria deles causados por transmissão oral (COURA & DIAS, 2009).

Clinicamente, a doença é caracterizada por apresentar duas fases distintas: a aguda e a crônica. A fase aguda é caracterizada por elevada parasitemia sanguínea e tecidual. A sintomatologia é bastante variável, podendo ser assintomática ou apresentar sintomas brandos, tais como conjuntivite e edema ocular indolor e unilateral conhecido como sinal de Romana. Manifestações gerais e alterações sistêmicas incluem febre, mal-estar geral, cefaléia, astenia, hiporexia, aumento de volume dos linfonodos e hepatoesplenomegalia (RASSI *et al.*, 2000). A fase aguda é detectada entre 10 e 40 dias após a infecção e sua duração varia entre 4 e 12 semanas, ao fim das quais o quadro febril e a parasitemia elevada tendem a regredir. Tais elementos praticamente definem a transição da doença de Chagas para a fase crônica, caracterizada por baixa parasitemia,

lenta evolução e difícil detecção (DIAS & COURA, 1997). A maioria dos pacientes na fase crônica permanece assintomática por décadas ou por toda a vida. Em alguns casos, tais pacientes podem evoluir para uma forma sintomática da doença. As manifestações mais comuns são o grave comprometimento dos sistemas cardíaco (forma cardíaca), digestivo (forma digestiva) ou de ambos (forma cardio-digestiva) (PRATA, 2001; COURA, 2007). A principal causa de mortalidade e morbidade da doença de Chagas é a cardiopatia inflamatória crônica.

As únicas drogas utilizadas no tratamento clínico da doença de Chagas, desde o início dos anos 70, são os compostos nitroheterocíclicos Nifurtimox (Lampit[®], Bayer), um derivado 5-nitrofurano, e Benzonidazol (Rochagan[®] no Brasil, Radanil[®] na Argentina, Roche), um derivado 2-nitroimidazólico (COURA & DE CASTRO, 2002; revisado por URBINA, 2009).

No parasito, ambas as drogas devem passar por ativação mediada por enzimas nitroredutases para apresentarem efeitos citotóxicos (MURTA *et al.*, 2006; WILKINSON *et al.*, 2008). Tem sido experimentalmente demonstrado que a ação tripanosomicida do Nifurtimox (NFX) deve-se à redução metabólica do grupo nitro gerando radicais nitroânion muito reativos que, por sua vez, causam a produção de espécies reativas de oxigênio altamente tóxicas, como por exemplo, o ânion superóxido e o peróxido de hidrogênio. Quanto ao Benzonidazol (BZ), seu mecanismo de ação está envolvido com a ligação covalente de seus intermediários nitroreduzidos a vários componentes celulares como DNA, lipídeos e proteínas (COURA & DE CASTRO, 2002; URBINA, 2009).

Ambos os compostos são efetivos na fase aguda da doença conduzindo aproximadamente 80% dos pacientes tratados à cura parasitológica (URBINA & DO CAMPO, 2003). Em geral, os melhores resultados são obtidos em casos agudos e infecções recentes, especialmente entre crianças, uma vez que apresentam maior tolerância ao tratamento. Em estudos realizados no Brasil por Andrade e colaboradores (1996) e na Argentina por Sosa Estani e colaboradores (1998, 1999), foi demonstrado em crianças com infecção crônica recente cerca de 60% de cura após o tratamento com BZ (revisado por URBINA, 2009). Entretanto, BZ e NFX apresentam limitada eficácia na fase crônica tardia da doença curando somente 6 a 10% dos pacientes tratados (CANÇADO, 1999), além de significativos efeitos colaterais que podem conduzir à interrupção do tratamento pelo paciente (URBINA, 2009).

A eficácia terapêutica de ambas as drogas também varia de acordo com a área geográfica, provavelmente devido a diferenças de susceptibilidade às drogas por diferentes cepas de *T. cruzi*. (URBINA & DO CAMPO, 2003).

Durante as duas últimas décadas não se realizaram progressos satisfatórios com relação ao desenvolvimento de fármacos eficazes para o tratamento da doença de Chagas. Estudos recentes, contudo, têm se baseado no conhecimento crescente da biologia do *T. cruzi*. Dentre as abordagens mais promissoras estão os inibidores da biosíntese de ergosterol, como Pozaconazol e Ravuconazol; inibidores da cruzipaína, principal cisteína protease de *T. cruzi*, essencial para sobrevivência e proliferação do parasito; inibidores da síntese de tripanotona e do resgate de purinas (revisado por URBINA, 2009).

1.2. Proteínas de superfície em tripanosomatídeos

Em 2005 foram publicadas as sequências dos genomas dos três tripanosomatídeos (*Trytrips*) causadores de doenças tropicais negligenciadas: *Leishmania major* (IVENS *et al.*, 2005), *Trypanosoma brucei* (BERRIMAN *et al.*, 2005) e *Trypanosoma cruzi* (EL-SAYED *et al.*, 2005a).

Análises comparativas do conteúdo gênico dos três parasitos identificaram um conjunto conservado de 6.158 proteínas e numerosos membros espécie-específicos, dos quais *T. cruzi* (32%) e *T. brucei* (26%) apresentam uma proporção maior que *L. major* (12%) (EL-SAYED *et al.*, 2005b). Muitos dos membros espécie-específicos são proteínas de superfície (Fig. 3). A análise dos genomas suporta a hipótese de que as três espécies evoluíram de modo a adquirirem distintas superfícies celulares e, logo, diferentes estratégias de sobrevivência e de evasão do sistema imune como consequência da adaptabilidade a diferentes nichos em seus respectivos hospedeiros (ACOSTA-SERRANO *et al.*, 2007).

Mesmo que a identidade molecular dos componentes de superfície seja diferente para cada uma das três espécies, algumas características são compartilhadas: as moléculas de superfície são superabundantes e formam uma cobertura protetora polianiónica por toda a superfície celular do parasito, inclusive o flagelo, além de serem aderidas à face externa da membrana plasmática por ligação covalente a âncoras de GPI

(ACOSTA-SERRANO *et al.*, 2007). Nakayasu e colaboradores (2009) apresentaram a primeira análise em larga escala do “GPIoma” de *T. cruzi*. Além da identificação estrutural das âncoras de GPI, análises de predição revelaram que aproximadamente 12% dos genes em *T. cruzi* codificam para proteínas ancoradas por GPI, um número muito maior quando comparado com *T. brucei* (1,5%) e *L. major* (1,02%) (NAKAYASU *et al.*, 2009).

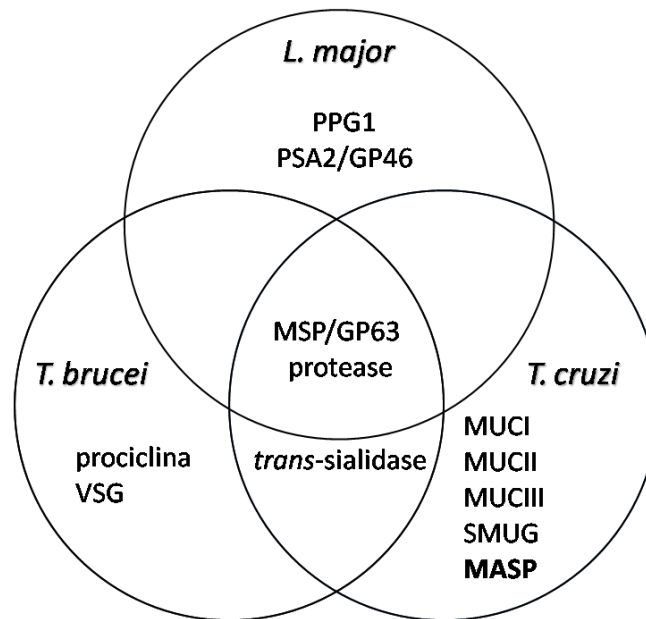


Figura 3: Distribuição de proteínas de superfície em *L. major*, *T. brucei* e *T. cruzi*. MUC I, II e III, os três grupos de mucina de *T. cruzi* pertencentes à família gênica TcMUC; MASP, proteína de superfície associada a mucina; SMUG (*small mucin-like gene*), mucinas expressas nas formas que se desenvolvem no inseto; MSP/GP63, proteína principal de superfície, também conhecida como GP63; VSG, glicoproteína variante de superfície; PSA2/GP46, antígeno 2 de superfície de promastigota, também conhecido como GP46; PPG1, proteofosfoliglicano 1. Adaptado de Acosta-Serrano e colaboradores, 2007, cap.10, p. 327.

Uma característica marcante do genoma de *T. cruzi*, quando comparado aos genomas de *T. brucei* e *L. major*, é a dramática expansão de famílias multigênicas que codificam para proteínas de superfície que incluem *trans*-sialidasas, mucinas, GP63 e uma grande e nova família nomeada MASP, ou proteína de superfície associada a mucina, somente identificada durante a anotação do projeto genoma deste parasito (EL-SAYED *et al.*, 2005a). Juntas, as três grandes famílias gênicas de *trans*-sialidasas, mucinas e MASP contêm em torno de 2.500 genes aparentemente funcionais, além de mais de 1.000 pseudogenes. Isto representa uma expansão que ocupa aproximadamente 20% da capacidade codificadora do genoma de *T. cruzi* (revisado por ACOSTA-

SERRANO *et al.*, 2007), o que evidencia a importância de *trans*-sialidases, mucinas e MASP para a biologia do parasito.

Com exceção de MASP, as proteínas de superfície mais abundantes de *T. cruzi* têm sido extensamente caracterizadas, e suas funções pelo menos parcialmente elucidadas. Estudos estruturais destas proteínas revelam mecanismos de ação particulares e propõem estratégias de interação entre parasito e hospedeiro. Nas próximas subseções serão descritas algumas das proteínas de superfície presentes nos *TriTryps*, com maior destaque para aquelas que possuem estrutura tridimensional depositada no *Protein Data Bank*, o PDB (BERMAN *et al.*, 2000) (<http://www.rcsb.org/pdb/home/home.do>). Mesmo que ainda sem estrutura tridimensional conhecida, cuja elucidação constitui o objetivo deste trabalho, ênfase maior será dada para a família MASP.

1.2.1. VSG

Há no genoma de *T. brucei* aproximadamente 1250 a 1500 genes *VSG* (em inglês “*variant surface glycoprotein*”), expressos um por vez, cada qual codificando para uma *VSG* com propriedades antigênicas distintas (BERRIMAN *et al.*, 2005; TAYLOR & RUDENKO, 2006).

A superfície de *T. brucei* é inteiramente coberta por um denso e homogêneo empacotamento de aproximadamente $5,5 \times 10^6$ dímeros de uma única glicoproteína variante de superfície (*VSG*), a qual elicitava uma resposta imune humoral do hospedeiro vertebrado rapidamente eliminando as células do parasito. Por sua vez, a população do parasito presente na corrente sanguínea do hospedeiro evade o sistema imune ao trocar, por meio de rearranjos gênicos, a glicoproteína expressa na superfície por outra *VSG* com propriedades imunogênicas distintas. Dessa forma, os poucos tripanosomas que expressam a *VSG* alternativa escapam à eliminação mediada pelos anticorpos específicos e se replicam conduzindo a um novo pico de parasitemia (PAYS *et al.*, 2004; 2007).

Tal fenômeno, muito bem descrito em *T. brucei* (revisado em BORST, 2002; PAYS *et al.*, 2004), é conhecido como variação antigênica e se traduz na habilidade do parasito em utilizar um repertório de *VSGs* com diferentes estruturas primárias.

As estruturas tridimensionais dos domínios N-terminal das VSGs MITat 1.2 e ILTat 1.24 foram resolvidas por cristalografia de raios-X a 2,9 e 2,7 Å de resolução, respectivamente (FREYMANN *et al.*, 1990; BLUM *et al.*, 1993). Apesar da baixa similaridade entre as sequências destas duas proteínas, apenas 20% de identidade, esta não é acompanhada por grandes alterações na estrutura tridimensional. O domínio variável N-terminal, que contém os epítomos antigênicos expostos na superfície do parasito, é estruturado em um enovelamento semelhante entre os membros de VSGs analisados, único em *T. brucei* e não relacionado a proteínas de outros organismos (Fig. 4) (BLUM *et al.*, 1993).

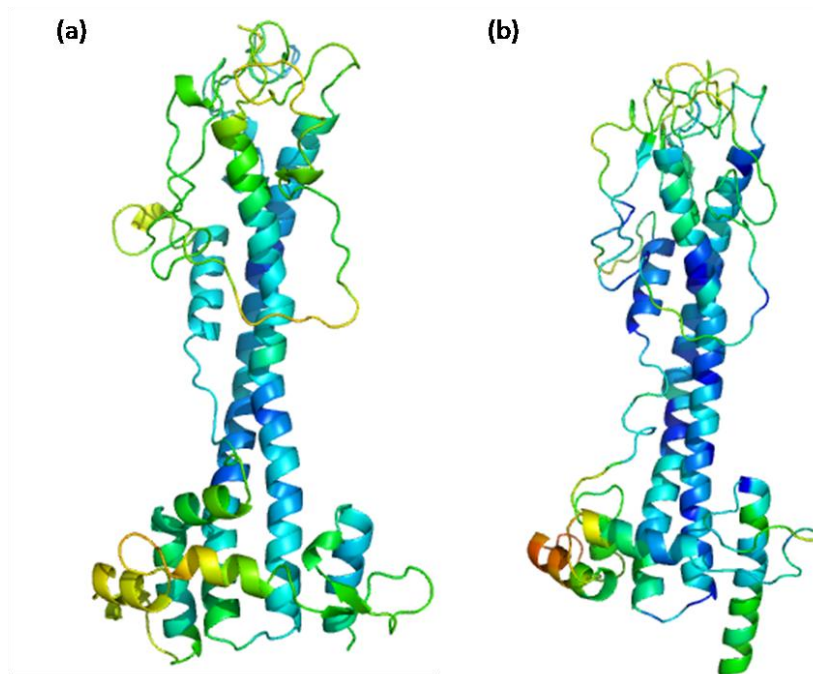


Figura 4: Representação em fita das duas estruturas conhecidas de VSG, ILTat 1.24 (a) e MITat 1.2 (b). Os elementos centrais dos dois domínios variáveis de VSG são duas α -hélices antiparalelas, em torno das quais se posicionam vários outros pequenos elementos de estrutura secundária. Para a construção das duas estruturas, foram utilizadas as cadeias A dos arquivos PDB de 1VSG (MITat 1.2) e 2VSG (ILTat 1.24) e o programa PyMOL (DELANO, 2002). As estruturas foram coloridas de acordo com o B-fator, de forma que o azul indica regiões mais bem definidas enquanto o vermelho indica regiões com grande flexibilidade. A figura original pode ser visualizada no artigo de Blum e colaboradores, 1993.

1.2.2. MSP/GP63

Metaloproteases da família MSP (protease principal de superfície), em inglês “*major surface protease*”, ou GP63 são as únicas glicoproteínas de superfície

abundantes compartilhadas entre as três espécies (Fig. 3). Enquanto a família gênica *gp63* de *L. major* possui 4 genes, em *T. brucei* há 13, contra 425 membros em *T. cruzi*, destes sendo 251 pseudogenes. (EL-SAYED *et al.*, 2005a). Em *Leishmania* spp., membros *gp63* desempenham amplo papel na patogênese da doença, enquanto que diferentes funções também podem ser delineadas para as outras espécies (YAO, 2009). Sua habilidade em hidrolisar uma diversidade de substratos fez da GP63 uma proteína convenientemente utilizada por diferentes parasitos na interação com moléculas de seus hospedeiros (YAO *et al.*, 2003).

Um dos membros de GP63 de *L. major* foi a primeira proteína de superfície deste parasito a ter a estrutura tridimensional determinada, resolvida por cristalografia de raios-X, a 2,5 e 1,8 Å de resolução (Fig. 5) (SCHLAGENHAUF *et al.*, 1998). Também conhecida como Leishmanolisina (EC 3.4.24.36, de acordo com *IUBMB Enzyme Nomenclature*), trata-se de uma molécula compacta contendo predominantemente estrutura secundária de fitas β . A proteína consiste de três domínios: N-terminal, central e C-terminal, tendo o primeiro um átomo coordenado de zinco e um enovelamento similar ao de módulos catalíticos encontrados em proteases ligadoras deste metal (SCHLAGENHAUF *et al.*, 1998).

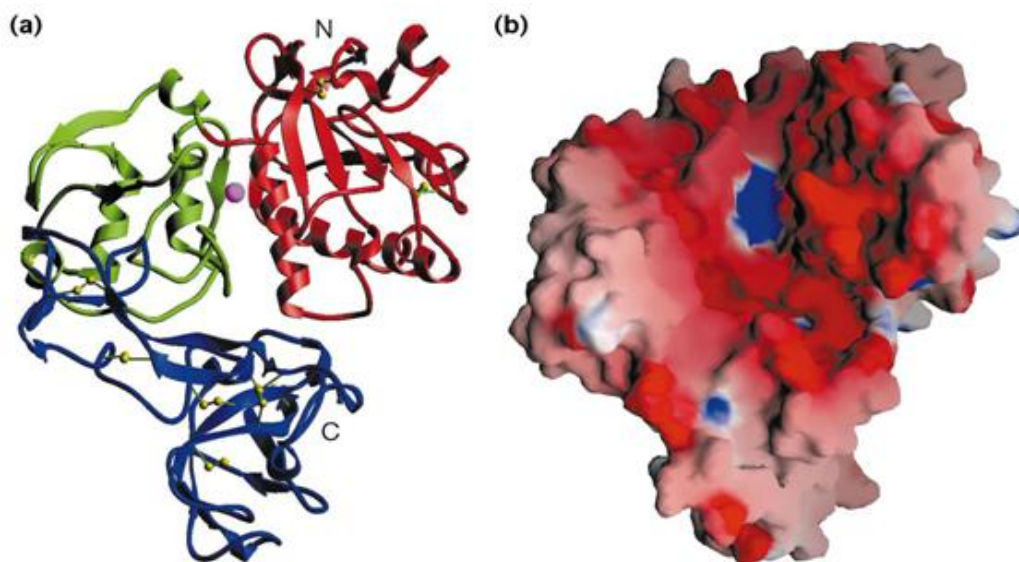


Figura 5: Domínios estruturais e representação da superfície de Leishmanolisina. (a) Representação em fita da estrutura da Leishmanolisina com o domínio N-terminal em vermelho, o domínio central em verde e o domínio C-terminal em azul. Pontes dissulfeto são mostradas em amarelo, e o átomo de zinco presente no sítio ativo é representado pela esfera em cor magenta. A molécula está posicionada de modo a evidenciar a fenda do sítio ativo. (b) Representação da superfície molecular com regiões carregadas negativamente mostradas em vermelho e porções positivas em azul. Extraído de Schlagenhaut e colaboradores, 1998.

Resíduos de aminoácidos dispostos na superfície da molécula são menos conservados que aqueles presentes no interior. Além disso, a variabilidade de sequência entre diferentes membros de leishmanolisina é correlacionada com flexibilidade estrutural. O domínio C-terminal, com três *loops* desordenados, é o menos conservado, apresentando a maior variabilidade de sequência (SCHLAGENHAUF *et al.*, 1998).

1.2.3. *Trans-Sialidase*

T. cruzi e *T. brucei* são incapazes de sintetizar ácido siálico, o qual, todavia, está presente em vários glicoconjugados nas superfícies dos dois parasitos. A incorporação do ácido siálico presente em glicoconjugados em células do hospedeiro é possível pela ação da enzima *trans*-sialidase. A *trans*-sialidase de *T. cruzi*, ancorada por glicosil-fosfatidil-inositol (GPI) à membrana de formas tripomastigotas do parasito, catalisa a transferência do monossacarídeo a partir de sialoglicoconjugados do hospedeiro para aceptores β -galactosil em moléculas de mucinas O-glicosiladas presentes na superfície do parasito (SCHENKMAN *et al.*, 1994; FRASCH, 2000).

A grande família das *trans*-sialidasas em *T. cruzi*, altamente polimórfica, possui 1430 genes (EL-SAYED *et al.*, 2005a), dos quais apenas um pequeno número codifica para proteínas que contêm os resíduos necessários para a atividade catalítica (ACOSTA-SERRANO *et al.*, 2007). As *trans*-sialidasas, ativas e inativas, consistem de dois domínios independentes sendo o N-terminal catalítico e o C-terminal antigênico, o qual possui um motivo repetitivo de doze resíduos de aminoácidos altamente imunogênico conhecido como epítipo SAPA (em inglês “*shed acute phase antigen*”) (CAZZULO & FRASCH, 1992; CREMONA *et al.*, 1999).

Durante a fase inicial da infecção pelo *T. cruzi*, uma forte resposta humoral contra o epítipo SAPA é observada, sugerindo um papel distrativo para o domínio imunodominante ao retardar a indução de uma resposta humoral capaz de neutralizar o domínio catalítico. Mais tarde, a indução de anticorpos direcionados para a região catalítica, alguns dos quais com características inibitórias, coincide com o controle da parasitemia e a sobrevivência do hospedeiro (LEGUIZAMÓN *et al.*, 1994). A presença de uma rede de epítipos com reatividade cruzada na região catalítica da molécula sugere também um mecanismo de evasão, desenvolvido para evitar a elicitação de uma

forte resposta imune humoral ao apresentar alvos indefinidos para o sistema imune do hospedeiro (PITCOVSKY *et al.*, 2001, 2002).

No que concerne à atividade catalítica da enzima, segundo Cremona e colaboradores (1995), a comparação de seqüências codificadoras para *trans*-sialidases ativas e inativas revelou que, dos 20 aminoácidos divergentes, de um total de 642, um único resíduo seria estritamente essencial para a atividade enzimática. A substituição de um resíduo de tirosina na posição 342 por uma histidina seria responsável por abolir a atividade catalítica das *trans*-sialidases. Em uma análise proteômica de *T. cruzi*, dentre as 223 proteínas *trans*-sialidases detectadas estão os produtos preditos de 15 genes codificadores de *trans*-sialidases ativas. Sugere-se que a expressão de um grande número de membros não enzimáticos da família das *trans*-sialidases possa desviar a resposta imune dos alvos enzimaticamente ativos (ATWOOD *et al.*, 2005).

Por meio da elucidação da estrutura tridimensional de um dos membros de *trans*-sialidase de *T. cruzi*, resolvida por cristalografia de raios-X a 1,65 Å de resolução (Fig. 6) (BUSCHIAZZO *et al.*, 2002), a arquitetura do sítio ativo foi caracterizada e mecanismos de ação puderam ser minuciosamente descritos (AMAYA *et al.*, 2004).

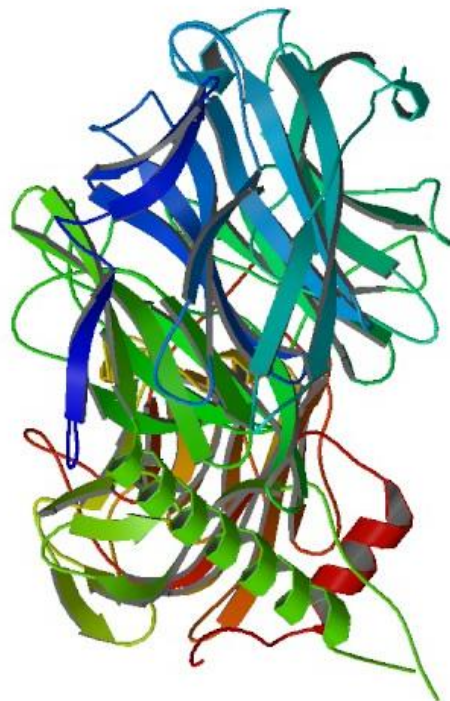


Figura 6: *Trans*-Sialidase de *Trypanosoma cruzi*. Estrutura tridimensional resolvida por cristalografia de raios-X por Buschiazzi e colaboradores, 2002. PDB: 1MS3.

1.2.4. Mucinas

As glicoproteínas de superfície mais abundantes em *T. cruzi* são as proteínas do tipo mucina. Suas sequências de aminoácidos são ricas em resíduos de serina e treonina, os quais se constituem em sítios aceptores para O-oligossacarídeos (SCHENKMAN *et al.*, 1993). Como aceptoras de ácido siálico, tem-se postulado que as mucinas possam participar de interações parasito-hospedeiro tais como adesão e invasão celular (FRASCH, 2000; ACOSTA-SERRANO *et al.*, 2001).

A família gênica de mucinas consiste de 863 membros, a qual pode ser dividida em duas sub-famílias: TcSMUG, com 19 membros expressos na forma epimastigota presente no inseto vetor, e TcMUC, com 844 membros expressos nas formas do parasito que se desenvolvem no hospedeiro mamífero (EL-SAYED *et al.*, 2005a).

Como consequência da expressão de diferentes grupos de mucinas ao longo da transição entre as formas residentes no inseto para aquelas residentes no mamífero, o parasito troca a cobertura de mucinas, antes de composição homogênea, para uma de composição protéica altamente heterogênea. A variabilidade marcante entre as cadeias polipeptídicas de mucinas expressas nas formas de *T. cruzi* presentes no hospedeiro mamífero sugere um mecanismo de imune-evasão (BUSCAGLIA *et al.*, 2006).

Até o momento, a estrutura tridimensional de membros desta família em *T. cruzi* não foi elucidada. Sabe-se, contudo, que moléculas do tipo mucinas possuem conformação estendida imposta pela extensa glicosilação ao longo da cadeia polipeptídica (HOLLINGSWORTH *apud* BUSCAGLIA *et al.*, 2006).

1.2.5. MASP

Descoberta durante a anotação do genoma de *T. cruzi* (EL-SAYED *et al.*, 2005a), *masp* (proteína de superfície associada a mucina), em inglês “*mucin-associated surface protein*”, corresponde à segunda maior família multigênica do parasito (6% do genoma diplóide), depois da família das *trans*-sialidases. A família recebe este nome porque seus genes estão, na grande maioria das vezes, fisicamente associados aos genes codificadores para mucinas (EL-SAYED *et al.*, 2005a; BARTHOLOMEU *et al.*, 2009).

Dos 1377 membros identificados, 771 codificam para proteínas caracterizadas por região central altamente variável e domínios N- e C- terminais conservados que correspondem, respectivamente, a uma sequência de peptídeo sinal e a um sítio de adição de âncora de GPI, sugerindo possível localização na superfície do parasito (Fig. 7). A região central, variável em comprimento e conteúdo de sequência, usualmente contém motivos repetitivos de resíduos de aminoácidos (EL-SAYED *et al.*, 2005a; BARTHOLOMEU *et al.*, 2009).

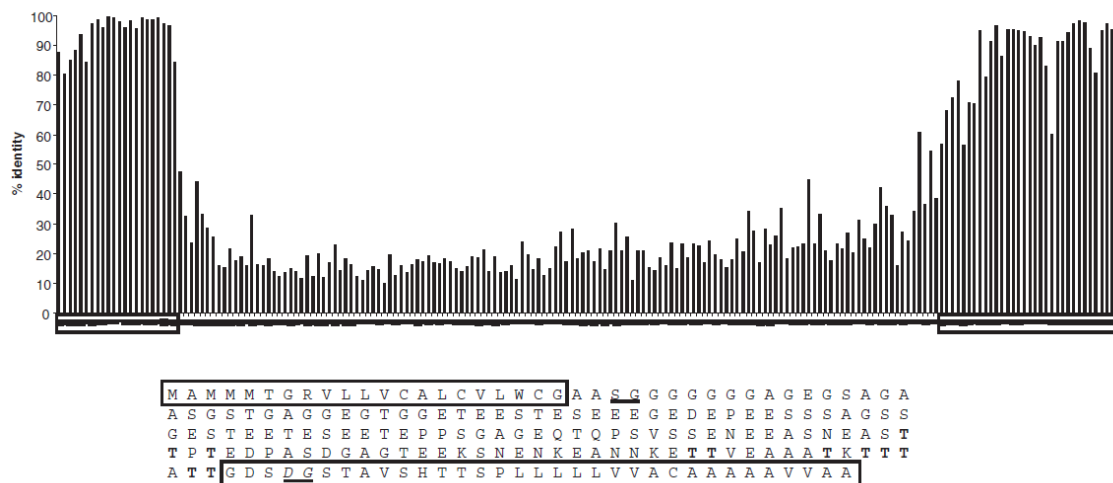


Figura 7: Características das sequências polipeptídicas de MASP. Todas as 771 proteínas MASP contendo extremidades N- e C-terminal conservadas foram alinhadas por meio do algoritmo ClustalW. Uma sequência consenso foi gerada e valores de identidade em porcentagem foram atribuídos para cada posição de resíduo de aminoácido ao longo da cadeia polipeptídica disposta no eixo x. Os domínios conservados N- e C-terminal que correspondem, respectivamente, a uma sequência de peptídeo sinal e a um sítio de adição de âncora de GPI estão dentro das caixas. Extraído e adaptado de Bartholomeu e colaboradores, 2009.

El-Sayed e colaboradores (2005a), na tentativa de identificar motivos de sequência conservados na família, conseguiram agrupar as 771 proteínas preditas MASP em 135 subgrupos que compartilham a mesma combinação de motivos, evidenciando uma organização em mosaico entre os membros, reflexo da variabilidade da região central (Fig. 8).

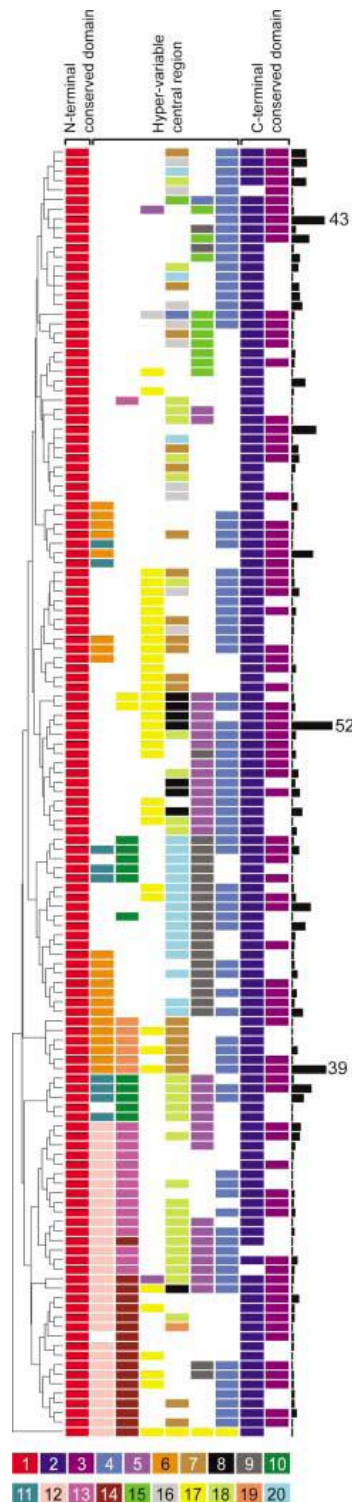


Figura 8: Representação esquemática da variabilidade de proteínas MASP. O algoritmo MEME (versão 3.0) foi utilizado para identificar motivos compartilhados por membros da família MASP. Cada linha corresponde a uma combinação de motivos dispostos ao longo de uma dada sequência de MASP, a partir do domínio N-terminal (bloco vermelho) ao domínio C-terminal (blocos azul e lilás). Os motivos, codificados por cores, estão numerados em ordem decrescente de significância estatística. O motivo em vermelho, por exemplo, que representa o domínio N-terminal, é o mais conservado e apresenta maior número de ocorrência nas MASPs e, logo, é o de maior significância estatística. Os números relativos em cada um dos subgrupos definidos estão representados como um histograma à direita. Os domínios conservados N- e C-terminal e a região central variável estão indicados no topo da figura. Extraído de El-Sayed e colaboradores, 2005a.

A fim de obter anticorpos para verificar a expressão protéica de MASP nas diferentes formas do parasito, Bartholomeu e colaboradores (2009) selecionaram um peptídeo, encontrado em 109 membros, com base nos critérios de conservação, imunogenicidade e especificidade única para MASP. Ensaio de *western blotting* revelaram a expressão preferencial de MASP nas formas tripomastigotas infectivas circulantes. Em estudo proteômico dos quatro estágios morfológicos do parasito, foram identificados 9 de 5.792 peptídeos únicos de MASP provenientes das formas tripomastigotas (ATWOOD *et al.*, 2005). Em ensaios de imunofluorescência de tripomastigotas incubados com os anticorpos anti-MASP descritos acima, aproximadamente 5% dos parasitos foram visualizados, indicando que a expressão das proteínas MASP representadas por aquele peptídeo selecionado é restrita a uma parcela da população (Fig. 9) (BARTHOLOMEU *et al.*, 2009). Até o momento, não se sabe se na população analisada há uma expressão preferencial de um subgrupo específico não representado pelo peptídeo utilizado, se há uma expressão heterogênea entre os diferentes tripomastigotas, se cada tripomastigota expressa um único gene *masp* ou co-expressa diferentes membros da família, e se há parasitos que não expressam MASP (BARTHOLOMEU *et al.*, 2009).

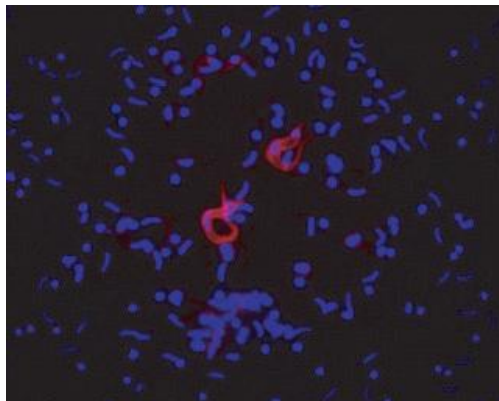


Figura 9: Ensaio de imunofluorescência de tripomastigotas fixados com paraformaldeído e incubados com anticorpo anti-MASP. DNA do parasito corado com DAPI, em azul. Extraído de Bartholomeu e colaboradores, 2009.

A expressão de MASP nas formas tripomastigotas infectivas circulantes, a extensa variabilidade de sequência, a natureza repetitiva e a localização na superfície do parasito são características que sugerem a participação de MASP nos processos de adesão – invasão celular e/ou mecanismos de escape do sistema imune. Assim, a compreensão da função biológica desta família é primordial para um maior entendimento da sua participação na interação entre parasito e hospedeiro.

1.3. Caracterização estrutural de proteínas

Os projetos genoma têm oferecido um suprimento ilimitado de sequências para análise e exploração de genes e de seus produtos protéicos correspondentes. Sequências biológicas podem ser comparadas, por métodos de alinhamentos, com todos os membros de um banco de dados de sequências previamente caracterizadas, o que pode servir como uma fonte para inferências evolutivas, estruturais e funcionais (OTTO *et al.*, 2010).

Entretanto, muitos avanços em nossa compreensão de sistemas biológicos em nível molecular têm sido possíveis através do conhecimento de estruturas detalhadas de proteínas e ácidos nucleicos (DURBIN & FEHER, 1996). Modelos tridimensionais das macromoléculas são amplamente utilizados como ferramentas para desvendar detalhes moleculares de processos biológicos. Cientistas utilizam estes modelos para aprender como as moléculas atuam: como enzimas catalisam reações metabólicas, como proteínas transportadoras carregam suas moléculas-cargo, como anticorpos se ligam a antígenos, e como proteínas interagem com o DNA, podendo “ligar” e “desligar” genes (RHODES, 2000).

Um objetivo em comum dos grandes consórcios e centros de genômica estrutural desenvolvidos em alguns países é decifrar, o mais rápido possível, as estruturas tridimensionais de uma vasta gama de proteínas codificadas pelo genoma de um dado organismo, utilizando as técnicas de cristalografia de raios-X e espectroscopia de ressonância magnética nuclear (RMN) (JOACHIMIAK, 2009). Ambas as técnicas são ferramentas indispensáveis na elucidação de estruturas tridimensionais. De um total de 71.415 estruturas de macromoléculas biológicas depositadas no PDB (<http://www.rcsb.org/pdb/home/home.do>) até a data de 23/02/2011, 62.095 foram resolvidas por cristalografia de raios-X e 8.740 por espectroscopia de RMN. Destas, 115 referem-se a macromoléculas biológicas em *T. cruzi*, sendo 112 estruturas resolvidas por cristalografia de raios-X e apenas 3 por espectroscopia de RMN.

1.3.1. Cristalografia de proteínas por difração de raios-X

Para se obter informações estruturais com resolução atômica, o principal método tem sido a cristalografia de raios-X. Entretanto, um dos passos-chave limitantes para se determinar a estrutura tridimensional de uma proteína por difração de raios-X é a sua cristalização (PUSEY *et al.*, 2005). A cristalização é considerada por muitos como um processo de tentativa e erro no qual as proteínas são precipitadas, lentamente, a partir de uma solução que as contém (NAGEM, 2003).

A técnica de cristalização mais conhecida é o método da gota suspensa. Uma única gota, formada pela mistura de 1 a 5 μL de uma solução de proteína com o mesmo volume de uma solução precipitante, é colocada sobre uma lamínula de vidro silanizada. A lamínula é invertida sobre um pequeno reservatório preenchido pela mesma solução precipitante, vedando-o. Neste sistema fechado, a difusão de vapor do solvente da gota para a solução precipitante contida no reservatório promoverá lentamente o estado de supersaturação na gota, permitindo a formação dos primeiros núcleos de cristalização (RHODES, 2000; NAGEM, 2003).

Uma vez obtidos cristais, estes são submetidos a uma fonte monocromática de raios-X, diante da qual são rotacionados por certo intervalo de tempo. A radiação espalhada pelo cristal é medida por um detector bidimensional de raios-X e o resultado da medida é chamado de padrão de difração (NAGEM, 2003)

Simplificadamente, a difração é consequência do espalhamento de várias ondas de raios-X pelas nuvens eletrônicas dos átomos presentes em cada molécula constituinte do cristal. Cada onda resultante, presente no padrão de difração, possui frequência, amplitude e fase. A frequência é a mesma da fonte de raios-X utilizada. A amplitude pode ser obtida através da intensidade de cada uma das ondas medidas nos padrões de difração. Contudo, a informação sobre a fase é perdida no processo de difração, constituindo assim o conhecido “problema das fases” (RHODES, 2000).

Existem alguns métodos para a solução do problema das fases. O método do Átomo Pesado (*Heavy-Atom method*) requer a presença de átomos de elevado peso atômico (zinco, mercúrio, urânio, iodo, cério, ferro, por exemplo) na estrutura cristalina utilizada. Estes átomos atuam como difratores fortes, uma vez que apresentam grande número de elétrons e, logo, contribuem mais para o espalhamento total da radiação. As perturbações medidas no padrão de difração podem ser utilizadas para se obter

estimativas das fases. A obtenção das fases por meio de cristais derivados de átomos pesados se beneficia também da capacidade de alguns átomos em absorver raios-X de comprimentos de onda específicos. Sabe-se que, quando a frequência da radiação incidente usada para a coleta de dados de difração se aproxima de uma das frequências naturais de oscilação de um sistema elétron-núcleo presente no cristal, um efeito de ressonância passa a ocorrer, contribuindo para o espalhamento anômalo da radiação pela matéria. Métodos como “difração anômala em vários comprimentos de onda” (*Multiwavelength Anomalous Diffraction – MAD*) fazem uso deste efeito para a resolução do problema das fases. Pelo método de Substituição Molecular (*Molecular Replacement*), o cristalógrafo pode optar por utilizar um modelo tridimensional já conhecido para resolver a estrutura de interesse, supondo, obviamente, que exista uma grande semelhança entre as estruturas. Assim, as fases de um dado modelo servem como estimativas iniciais das fases para o modelo de interesse (RHODES, 2000; NAGEM, 2003).

A maioria das estruturas cristalinas aparenta ser idêntica às estruturas em solução. Provavelmente, a evidência mais convincente de que estruturas cristalinas podem, seguramente, ser utilizadas para inferir função molecular é o fato de muitas macromoléculas ainda serem funcionais no estado cristalino. Substratos adicionados às suspensões de enzimas cristalizadas são convertidos em produtos, mesmo que a taxas reduzidas, sugerindo que os sítios de ligação e de catálise enzimática permanecem intactos. Outra evidência para a similaridade de estruturas cristalinas e em solução é a compatibilidade de modelos cristalográficos com aqueles obtidos por RMN. As maiores discrepâncias restringem-se a porções flexíveis dos modelos, tais como *loops* na superfície. Em outros casos, a estrutura da proteína foi obtida a partir de mais de um tipo de cristal, e os modelos obtidos foram bastante similares, sugerindo que a estrutura molecular não foi alterada pela cristalização. Além disso, análises cuidadosas de mapas de densidade eletrônica revelam muitas moléculas de água, ordenadas ou não, na superfície de proteínas cristalizadas. Logo, proteínas em um cristal ainda estão em um ambiente aquoso, sujeitas aos mesmos efeitos do solvente que estabilizam a estrutura em solução (RHODES, 2000).

1.3.2. Espectroscopia de ressonância magnética nuclear

Modelos tridimensionais de moléculas em solução podem ser obtidos por espectroscopia de RMN. Certos núcleos atômicos, como ^1H , ^{13}C , ^{15}N e ^{31}P , possuem *spins* nucleares como resultado de momentos magnéticos de seus componentes prótons e nêutrons. Os *spins* fazem com que os núcleos se comportem como pequenos magnetos e, como resultado, adotem orientações em um campo magnético. Quando as moléculas protéicas são submetidas a um campo magnético forte, o *spin* de ^1H , por exemplo, se alinha com o campo. Este alinhamento de equilíbrio pode ser mudado para um estado excitado ao serem aplicados sobre a amostra pulsos de radiação eletromagnética na faixa de rádio-frequência (RF). Núcleos idênticos em ambientes químicos diferentes interagem de forma ligeiramente diferente com a radiação incidente, absorvendo diferentes frequências de energia. Quando os núcleos retornam para o estado de equilíbrio, há emissão de radiação de RF que pode ser medida. A frequência da radiação emitida por cada núcleo, calculada em relação a um sinal de referência e então chamada de deslocamento químico, depende do tipo de núcleo e do ambiente molecular em que este se encontra. Logo, núcleos específicos podem ser detectados por sua característica energia de absorção. (BRANDEN & TOOZE, 1999; RHODES, 2000).

Além de apresentarem deslocamentos químicos particulares, os *spins* nucleares interagem magneticamente uns com os outros por um processo chamado de acoplamento *spin-spin*. O efeito resultante do acoplamento *spin-spin* é desdobrar uma determinada absorção em várias absorções individuais, normalmente em um padrão que depende do número de núcleos equivalentes acoplados. Tendo em vista que os núcleos precisam estar próximos uns dos outros para que ocorra o acoplamento, este efeito pode ser usado para determinar quais núcleos são vizinhos na molécula em estudo (RHODES, 2000).

Na maioria dos casos tem-se trabalhado com as propriedades magnéticas de ^1H , uma vez que átomos de hidrogênio são mais abundantes em proteínas. Todavia, para o estudo de proteínas maiores, cuja massa molecular se aproxima de 30 kDa, a combinação de experimentos de tripla ressonância e amostras marcadas isotopicamente com ^2H , ^{13}C , ^{15}N é crítica para a determinação de tais estruturas em solução (GARDNER & KAY, 1998).

As aplicações do RMN vão além da determinação de estruturas tridimensionais em resolução atômica e abrangem, por exemplo, a dinâmica molecular, mudanças conformacionais e orientações de domínios e subunidades entre moléculas que se interagem, a caracterização de polipeptídeos desestruturados e a sua transição para estruturas funcionais quando em complexos ativos (FOSTER *et al.*, 2007; BILLETTER *et al.*, 2008).

1.3.3. Espectroscopia de dicroísmo circular

Em sua grande maioria, polipeptídeos e proteínas apresentam estruturas estáveis e funcionais; arranjos altamente ordenados, como α -hélices, folhas β paralelas e antiparalelas e voltas β . Estes arranjos constituem os elementos de estrutura secundária. Todavia, essas estruturas não são estáticas e, durante a atividade biológica, podem ocorrer mudanças conformacionais nas proteínas.

As formas regulares dos arranjos atômicos das estruturas secundárias interagem de modo particular com a radiação polarizada, fazendo, por exemplo, com que uma luz com polarização circular no sentido horário seja absorvida diferentemente do que uma luz com polarização no sentido anti-horário. A espectroscopia de dicroísmo circular, ou simplesmente CD (em inglês “*circular dichroism*”), mede justamente essa diferença de absorção de radiação com polarização circular e permite verificar assim o conteúdo e o tipo de estrutura secundária presente na forma ativa da proteína e conseqüentemente seu estado enovelado (SREERAMA & WOODY, 2004).

Os espectros de CD para proteínas são geralmente divididos em três faixas de comprimentos de onda, baseados na energia das transições eletrônicas que predominam em determinada faixa. As três faixas são: (1) UV distante, abaixo de 250 nm, onde predominam as contribuições das ligações peptídicas, mais precisamente do arranjo espacial das amidas; (2) UV próximo, entre 250 e 300 nm, no qual contribuem as cadeias laterais aromáticas e, (3) UV visível, entre 300 e 700 nm, no qual grupos cromóforos externos são os principais contribuintes, permitindo estudo de ligantes (SREERAMA & WOODY, 2004).

1.4. Justificativa

A doença de Chagas, causada pelo protozoário flagelado *Trypanosoma cruzi*, constitui um grave problema de saúde pública no continente Americano, com distribuição geográfica desde a Argentina até o sul dos Estados Unidos da América (COURA & DIAS, 2009). Não há vacinas contra a doença de Chagas e o tratamento terapêutico disponível é insatisfatório, pois apresenta limitada eficácia na fase crônica da doença (CANÇADO, 1999), significativos efeitos colaterais que podem conduzir à interrupção do tratamento pelo paciente (COURA & DE CASTRO, 2002), além da existência de diferenças de susceptibilidade a drogas por cepas distintas de *T. cruzi*. (URBINA & DO CAMPO, 2003).

Uma característica marcante do genoma de *T. cruzi* é a dramática expansão de famílias multigênicas que codificam para proteínas de superfície. As três grandes famílias gênicas de *trans*-sialidases, mucinas e MASP representam uma expansão que ocupa aproximadamente 20% da capacidade codificadora do genoma de *T. cruzi* (revisado por ACOSTA-SERRANO *et al.*, 2007), o que evidencia a importância destas moléculas para a biologia do parasito.

Estudos realizados por Bartholomeu e colaboradores (2009) demonstraram a expressão de MASP nas formas tripomastigotas infectivas circulantes, extensa variabilidade de sequência, natureza repetitiva e localização na superfície do parasito, características que sugerem a participação de MASP nos processos de adesão - invasão celular e/ou mecanismos de escape do sistema imune.

A elucidação das estruturas tridimensionais de membros da família MASP constitui uma informação valiosa para o entendimento mais detalhado destas moléculas. Através desta caracterização estrutural, pode-se determinar a localização dos seus epítomos e o tipo de enovelamento protéico, deduzir domínios funcionais e talvez obter uma indicação sobre sua função no parasito, fato este desconhecido até o momento. A relação entre estrutura e função de MASP é um fato importante a ser explorado no projeto, visto que até o momento nada se sabe a este respeito. Proteínas de superfície em *T. cruzi* como as *trans*-sialidases, por exemplo, tiveram seu sítio catalítico detalhadamente descrito e mecanismos de ação propostos por meio de análises das estruturas cristalográficas da proteína sozinha e complexada com açúcares ligantes (BUSCHIAZZO *et al.*, 2002; AMAYA *et al.*, 2004).

É possível supor ainda que as proteínas MASP possuam, entre si, estruturas similares, a despeito da extensa variabilidade de sequência entre os membros, como é observado com as VSGs em *T. brucei*, as quais se estruturam em um enovelamento semelhante apesar da baixa similaridade entre as sequências (BLUM *et al.*, 1993).

1.5. Objetivos

1.5.1. Geral

Caracterizar as estruturas secundárias de proteínas da família MASP de *Trypanosoma cruzi*.

1.5.2. Específicos

- Selecionar as sequências de interesse por meio de análises de predição *in silico*.
- Clonar os genes codificadores para as proteínas MASP em sistemas de expressão com cauda de histidina e Glutathione-S-Transferase (GST).
- Expressar e purificar quantidades suficientes das proteínas MASP recombinantes para estudos estruturais.
- Obter informações sobre a estrutura secundária das proteínas e seu enovelamento.

2. MATERIAIS E MÉTODOS

2.1. Seleção das proteínas MASP

Os dados originados do projeto genoma da cepa CL Brener de *T. cruzi*, estruturados em um banco relacional MySQL v5.0, encontram-se em um servidor do Laboratório de Imunologia e Genômica de Parasitos no Departamento de Parasitologia (ICB/UFMG).

Neste trabalho utilizou-se sequências nucleotídicas, e suas respectivas sequências protéicas preditas, de genes anotados como MASP, excluindo-se pseudogenes e genes parciais. Optou-se por trabalhar apenas com a região central de MASP, uma vez que proteínas desta família são caracterizadas por apresentarem uma região central altamente variável e domínios N- e C-terminal conservados que codificam para um peptídeo sinal e um sítio de adição de âncora de GPI, respectivamente. Tendo em vista vantagens e limitações das técnicas de cristalografia de raios-X e espectroscopia de RMN, preferência foi dada para cadeias polipeptídicas de massa molecular inferior a 20 kDa. As sequências referentes ao peptídeo sinal e ao sítio de adição de âncora de GPI foram respectivamente identificadas pelos algoritmos SignalIP (BENDTSEN *et al.*, 2004) (<http://www.cbs.dtu.dk/services/SignalP/>) e GPI-SOM (FANKHAUSER & MASER, 2005) (<http://gpi.unibe.ch/>).

Análises *in silico* de um determinado conjunto de MASPs foi realizada com a finalidade de investigar algumas características das sequências de resíduos de aminoácidos, tais como:

1. Sítios de glicosilação. Predições de N- e O-glicosilação foram realizadas, respectivamente, pelos algoritmos NetNGlyc 1.0 (GUPTA *et al.*, 2004) (<http://www.cbs.dtu.dk/services/NetNGlyc/>), o qual identifica a sequência consenso Asn-Xaa-Ser/Thr, e NetOGlyc 3.1 (JULENIUS *et al.*, 2005) (<http://www.cbs.dtu.dk/services/NetOGlyc/>), o qual identifica resíduos de serina e treonina passíveis de O-glicosilação.
2. Regiões de desordem estrutural, identificadas pelo algoritmo PONDR[®] VL-XT (LI *et al.*, 1999; ROMERO *et al.*, 2001) (<http://www.pondr.com/>), o qual

combina um preditor de regiões desordenadas internas e dois preditores de regiões desordenadas nas extremidades N- e C-terminal. Para cada região central, gráficos de carga líquida média *versus* hidrofobicidade média foram gerados conforme descrito por Uversky e colaboradores, 2000.

Foram selecionadas as sequências gênicas depositadas no banco de dados TriTrypDB (ASLETT *et al.*, 2010) (<http://tritrypdb.org/tritrypdb/>) sob os códigos de acesso Tc00.1047053510013.170, Tc00.1047053507955.20, Tc00.1047053507957.300, Tc00.1047053506757.50. Ao longo deste trabalho, estas sequências foram denominadas MASP1, MASP2, MASP3 e MASP4, respectivamente.

2.2. Oligonucleotídeos usados para amplificação das regiões codificadoras centrais dos genes *masp*

Para amplificar as regiões codificadoras centrais dos genes *masp* a partir do DNA genômico de *T. cruzi*, cepa híbrida CL Brener, foram utilizados os oligonucleotídeos descritos no Quadro 1. Em cada iniciador foi adicionada uma sequência específica reconhecida por uma enzima de restrição facilitar a etapa de clonagem.

ID do gene <i>masp</i> (TriTrypDB)	Nome do oligonucleotídeo	Sequência	Sítio de enzima de restrição	T _m (°C)
Tc00.1047053510013.170	8273.m00017_F:pET.TEV	5' <u>GCCGCTAGCGCTGTTGGTGTGGTT</u> 3'	<i>NheI</i>	74
Tc00.1047053510013.170	8273.m00017_F:pET.GST.TEV	5' <u>GCCGAATTCGCTGTTGGTGTGGTT</u> 3'	<i>EcoRI</i>	72
Tc00.1047053507955.20	7543.m00002_F:pET.TEV	5' <u>GCCGCTAGCGATGTTGGTGTGGTT</u> 3'	<i>NheI</i>	72
Tc00.1047053507955.20	7543.m00002_F:pET.GST.TEV	5' <u>GCCGAATTCGATGTTGGTGTGGTT</u> 3'	<i>EcoRI</i>	70
Tc00.1047053510013.170 Tc00.1047053507955.20	8273.m00017_7543.m00002_R	5' <u>CGCAAGCTTTCAGTCACTGTCGTC</u> CGT 3'	<i>HindIII</i>	74
Tc00.1047053507957.300	7544.t00030_F:pET.TEV	5' <u>TCGGCTAGCAAGGAGGCGGAAGC</u> 3'	<i>NheI</i>	73
Tc00.1047053507957.300	7544.t00030_F:pET.GST.TEV	5' <u>TGCGGATCCAAGGAGGCGGAAGC</u> 3'	<i>BamHI</i>	76
Tc00.1047053507957.300	7544.t00030_R	5' <u>AGCGAGCTCTCAGTCACTGTCGGCAGTC</u> 3'	<i>SacI</i>	74
Tc00.1047053506757.50	7084.t00005_F:pET.TEV	5' <u>CAGGCTAGCGACAAGGTGGTGGAAAG</u> 3'	<i>NheI</i>	71
Tc00.1047053506757.50	7084.t00005_F:pET.GST.TEV	5' <u>ACTGGATCCGACAAGGTGGTGGAAAG</u> 3'	<i>BamHI</i>	70
Tc00.1047053506757.50	7084.t00005_R	5' <u>TCTGAGCTCTCAGTCGCTGTCGGTAGTC</u> 3'	<i>SacI</i>	71

Quadro 1: Oligonucleotídeos sintéticos utilizados como iniciadores nas reações de amplificação das regiões codificadoras centrais dos genes *masp*. O código identificador (ID) é válido para o gene e a respectiva proteína. Os nucleotídeos sublinhados correspondem às sequências de reconhecimento pelas enzimas de restrição. F: *forward* (iniciador direto). R: *reverse* (iniciador indireto). TEV: *tobacco etch virus*. GST: Glutathione-S-transferase. T_m: Temperatura de anelamento.

2.3. Reação em Cadeia da Polimerase (PCR)

A amplificação das sequências nucleotídicas codificadoras centrais de *masp* foi realizada por PCR em um volume final de 30 μL /reação, sendo: 3 μL de tampão de reação 10x (*Tris-HCl* 10 mM pH 8,4; *NaCl* 40 mM; *MgCl₂* 1,5 mM; *Triton-X* 100 0,1%), 1 μL de dNTPs (200 μM de cada um dos desoxirribonucleotídeos), 1 μL (10 pmoles/ μL) de cada oligonucleotídeo iniciador (direto, específico da região 5', e reverso, complementar a região 3'), 1 μL (1 ng/ μL) de DNA genômico de formas epimastigotas de *T. cruzi*, cepa híbrida CL Brener, como fita molde, 0,25 μL da enzima Taq DNA polimerase (5 U/ μL) (Phoneutria, Belo Horizonte, Brasil) e água deionizada bidestilada q.s.p. 30 μL .

Cada reação foi realizada nas seguintes condições: desnaturação inicial a 94°C por 2 minutos, seguida de 30 ciclos de três etapas (desnaturação a 94°C, hibridização a 65°C e extensão a 72°C, por 1 minuto cada). Após os 30 ciclos, fez-se uma extensão final por 10 minutos a 72°C.

Os amplicons foram visualizados em gel de agarose 1% corado com brometo de etídio (0,3 $\mu\text{g}/\text{mL}$).

2.4. Eletroforese em gel de agarose

Para o preparo do gel, 1% de agarose foi dissolvida por aquecimento em tampão TAE 1x (4,8 g/L *Tris-base* pH 8,0; 1,14 mL ácido acético glacial; 2 mL EDTA 0,5 M). A esta solução adicionou-se brometo de etídio (0,3 $\mu\text{g}/\mu\text{L}$) ou SYBR[®] Safe (1x) (Invitrogen, Carlsbad, USA), fluoróforo menos mutagênico e mais sensível que o anterior.

Após polimerização da agarose, as amostras foram aplicadas no gel em tampão de amostra de DNA 1x (*Ficoll* 5%; azul de Bromofenol 0,04%; xileno cianol 0,04%). O sistema foi então submetido à voltagem constante de 100 V. A visualização das bandas foi realizada por radiação ultravioleta.

2.5. Reação de ligação de fragmentos a vetores

Os amplicons, após purificação a partir do gel de agarose utilizando *QIAquick*[®] *Gel Extraction Kit* (Qiagen, Hilden, Alemanha), foram subclonados nos vetores pCR[®]2.1 TOPO (Invitrogen), utilizando a enzima topoisomerase ligada covalentemente ao vetor, e pGEM[®]-T (Promega, Madison, USA), utilizando a enzima T4DNA ligase (Promega), segundo instruções dos fabricantes. Vetores de expressão e insertos também foram submetidos à reação de ligação com a enzima T4 DNA Ligase (Promega) a 4°C *overnight*, conforme recomendações do fabricante

2.6. Preparo de bactérias

2.6.1. Eletrocompetentes

Bactérias *Escherichia coli* cepa XL-1 Blue foram preparadas para transformação por choque elétrico. Uma colônia de bactérias foi inoculada em 100 mL de meio 2xYT (*Bacto-triptona* 1,6%; *extrato de levedura* 1%; *NaCl* 0,5%; *pH* 7,0) sem antibiótico e incubada *overnight* a 37°C sob agitação constante de 200 rpm. Inoculou-se 10 mL da cultura crescida em 1000 mL de meio 2xYT fresco sem antibiótico e cultivou-se a 37°C sob agitação de 200 rpm até ser atingida a DO_{600} de 0,5-0,7. Após 20 minutos no gelo, as células foram coletadas por centrifugação (4000 x g, por 15 minutos a 4°C) e o *pellet* ressuspenso em 500 mL de glicerol 10% gelado. Foram realizados outros dois ciclos de centrifugação e ressusensão do *pellet* respectivamente em 250 mL e 20 mL de glicerol 10%, conforme descrito anteriormente, ao fim das quais o *pellet* foi ressuspenso em 1-2 mL de glicerol 10% gelado e aliquoteado. As alíquotas de 50 µL foram imediatamente transferidas para o gelo seco e armazenadas em freezer -80 °C (AUSUBEL *et al.*, 1987).

2.6.2. Termocompetentes

Bactérias *Escherichia coli* cepa BL-21 (DE3) Star (Invitrogen) foram preparadas para transformação por choque térmico. Uma colônia de bactérias foi inoculada em 100 mL de meio 2xYT (*Bacto-triptona 1,6%; extrato de levedura 1%; NaCl 0,5%; pH 7,0*) sem antibiótico e incubada *overnight* a 37°C sob agitação constante de 200 rpm. Inoculou-se 10 mL da cultura crescida em 1000 mL de meio 2xYT fresco sem antibiótico e cultivou-se a 37°C sob agitação de 200 rpm, até ser atingida a DO₆₀₀ de 0,6-0,8. As células foram transferidas para tubos autoclavados e mantidas a 0°C por 10 minutos sendo posteriormente coletadas por centrifugação (8000 rpm, por 10 minutos a 4°C, em centrífuga HITACHI Himac CR 21GII, rotor R14A). O *pellet* foi ressuscitado em 100 mL de CaCl₂ 0,1M gelado, e uma segunda etapa de centrifugação foi realizada nas mesmas condições descritas acima, ao fim da qual o *pellet* foi ressuscitado em 2 mL de CaCl₂ 0,1M gelado e aliquoteado. As alíquotas de 50 µL foram imediatamente transferidas para o gelo seco e armazenadas em freezer -80 °C. (SAMBROOK *et al.*, 1989).

2.7. Transformação de bactérias

2.7.1. Por eletroporação

Produtos de ligação de fragmentos a vetores foram incubados com 100 µL de bactérias *E. coli* XL-1 Blue eletrocompetentes e transferidos para uma cubeta apropriada, tamanho de 0,2 cm (Bio-Rad, Berkeley, USA). Após a eletroporação (2,50kV, 1 pulso, em eletroporador *MicroPulser*, Bio-Rad), foram adicionados 300 µL de meio 2xYT líquido, e incubou-se a 37°C por 1 hora em agitação a 200 rpm para expressão do gene de resistência. Posteriormente, as bactérias foram plaqueadas em meio seletivo 2xYT-ágar 1,5% com ampicilina (100 µg/mL), caso a transformação tenha sido com os vetores TOPO ou pGEM, e kanamicina (50 µg/mL), caso tenham sido utilizados os vetores de expressão (item 2.10). Após incubação em estufa

bacteriológica a 37°C por 16 horas, os clones obtidos foram analisados através de PCR de colônias utilizando os iniciadores específicos descritos no quadro 1 e o programa apresentado no item 2.3, seguido de ensaio eletroforético em gel de agarose para visualização do inserto. Também foram utilizados os iniciadores M13 direto (5' CGC CAG GGT TTT CCC AGT CAC GAC 3') e M13 reverso (5' TCA CAC AGG AAA CAG CTA TAG C 3', T_{anelamento}: 50°C), que se anelam nos vetores TOPO e pGEM.

2.7.2. Por choque térmico

Para a transformação de bactérias *E. coli* BL-21(DE3) Star (Invitrogen), plasmídeos recombinantes de interesse foram incubados com 50 µL de bactérias termocompetentes competentes durante 30 minutos em gelo. Em seguida, incubou-se a 42°C por 30 segundos e resfriou-se em gelo para o choque térmico. Posteriormente, em 300 µL de meio 2xYT sem antibiótico adicionou-se o produto da transformação e incubou-se a 37°C por 1 hora em agitação. Finalmente, as bactérias foram plaqueadas em meio seletivo 2xYT/ágar 1,5% com antibiótico específico (50 µg/mL de Kanamicina) e incubadas a 37°C durante 16 horas.

2.8. Extração de plasmídeos recombinantes

Clones positivos foram selecionados e cultivados por 16 horas em 5-10 mL de meio 2xYT líquido com antibiótico específico (100 µg/mL de ampicilina ou 50 µg/mL de kanamicina) para posterior extração dos plasmídeos recombinantes conforme recomendações do fabricante do *QIAprep*[®] *Spin Miniprep Kit* (Qiagen).

2.9. Digestão enzimática dos vetores de sequenciamento

Como mencionado no item 2.2, para cada iniciador foi adicionada uma sequência reconhecida por uma enzima de restrição. As sequências de cada iniciador de um par direto e reverso reconhecidas pelas enzimas de restrição foram compatíveis com sequências presentes no sítio de clonagem dos vetores de expressão utilizados (item 2.10), de modo a gerar extremidades coesivas no inserto e no vetor durante a digestão destes pelas enzimas de restrição.

Para a obtenção de cada um dos insertos referentes às regiões centrais codificadoras dos genes *masp* selecionados, os vetores TOPO e pGEM subclonados foram submetidos à dupla digestão com as devidas enzimas de restrição (Quadro 2), a uma temperatura de 37°C por 16 horas, conforme recomendações dos fabricantes (Promega; Fermentas, Burlington, Canadá).

Vetor de expressão de destino: pET28a-TEV				Vetor de expressão de destino: pET28a-GST-TEV			
Construção	Tampão	Enzimas de restrição (Fermentas): unidades (U) e estimativa da atividade enzimática (%)		Construção	Tampão	Enzimas de restrição (Promega): unidades (U) e estimativa da atividade enzimática (%).	
pGEM:MASP1	Tango® (Fermentas)	<i>NheI</i> - 15U (100%)	<i>HindIII</i> - 15U (75%)	pGEM:MASP1	E (Promega)	<i>EcoRI</i> - 18U (50-75%)	<i>HindIII</i> - 15U (100%)
pGEM:MASP2	Tango®	<i>NheI</i> - 15U (100%)	<i>HindIII</i> - 15U (75%)	pGEM:MASP2	E	<i>EcoRI</i> - 18U (50-75%)	<i>HindIII</i> - 15U (100%)
TOPO:MASP3	Tango®	<i>NheI</i> - 15U (100%)	<i>SacI</i> - 15U (75%)	TOPO:MASP3	E	<i>BamHI</i> - 15U (100%)	<i>SacI</i> - 15U (50-75%)
TOPO:MASP4	Tango®	<i>NheI</i> - 15U (100%)	<i>SacI</i> - 15U (75%)	pGEM:MASP4	E	<i>BamHI</i> - 15U (100%)	<i>SacI</i> - 15U (50-75%)

Quadro 2: Enzimas e tampões utilizados nas reações de dupla digestão dos vetores de sequenciamento recombinantes. Aproximadamente 250-350 ng de plasmídeo recombinante de cada construção foi duplamente digerido. Tampão Tango® 1x (Fermentas): *Tris-acetato* 33 mM (pH 7,9 a 37°C); *acetato de Magnésio* 10 mM; *acetato de Potássio* 66 mM; *BSA* 0,1 mg/mL. Tampão E 1x (Promega): *Tris-HCl* 6 mM (pH 7,5 a 37°C); *cloreto de Magnésio* 6 mM; *cloreto de Sódio* 100 mM; *DTT* 1 mM. Os vetores de expressão pET28a-TEV e pET28a-GST-TEV são descritos no item 2.10.

Os produtos das reações de digestão enzimática foram submetidos à eletroforese em gel de agarose para visualizar os insertos liberados. As bandas correspondentes aos insertos foram excisadas do gel, e o DNA foi purificado com *QIAquick® Gel Extraction Kit* (Qiagen) conforme recomendações do fabricante.

2.10. Digestão enzimática dos vetores de expressão

Modificados no Centro de Biologia Molecular e Estrutural (CeBiME, Campinas/SP) (CARNEIRO *et al.*, 2006) a partir do vetor pET28a (Novagen, Darmstadt, Alemanha), os vetores de expressão utilizados são o pET28a-TEV, que apresenta uma cauda de 6 resíduos de histidina N-terminal como fusão, e o pET28a-GST-TEV, cuja proteína de fusão é uma Glutathione-S-Transferase (GST) N-terminal. Em ambos, o sítio de clivagem por trombina foi substituído pelo da protease do vírus *etch* do tabaco (*tobacco etch virus* - TEV) (Fig. 10).

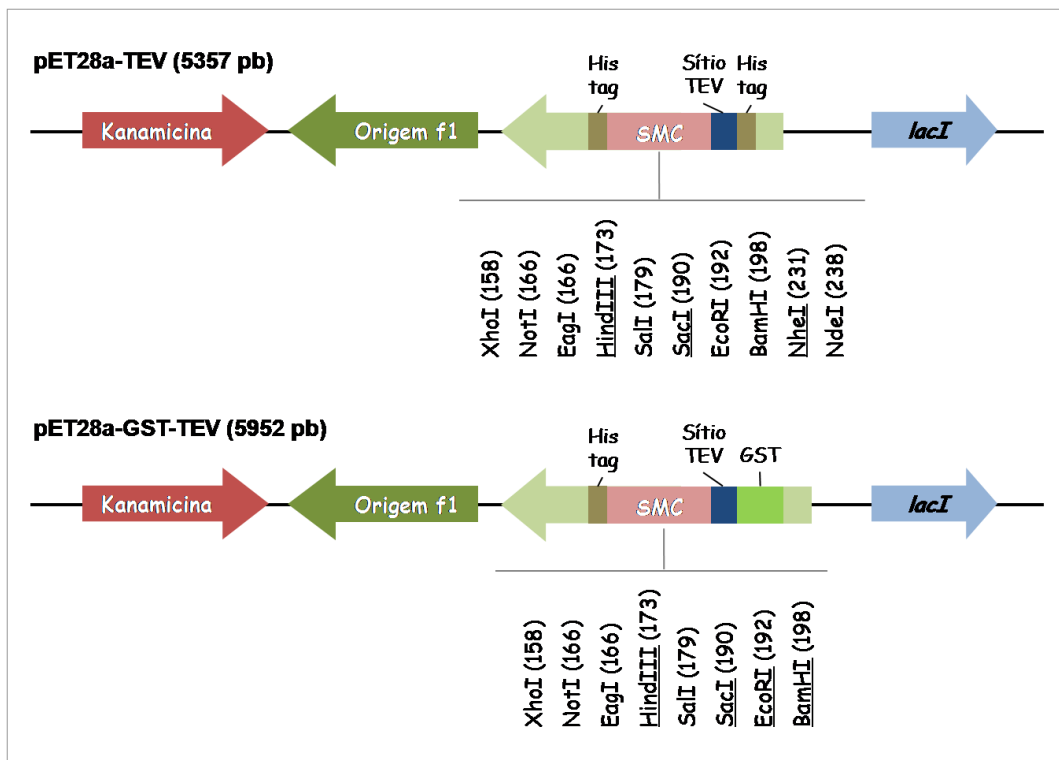


Figura 10: Desenho esquemático dos vetores de expressão pET28a-TEV e pET28a-GST-TEV. As enzimas selecionadas para a etapa de clonagem estão sublinhadas. SMC: sítio múltiplo de clonagem.

A fim de linearizar os vetores pET28a-TEV e pET28a-GST-TEV e gerar extremidades coesivas, fez-se dupla digestão enzimática a 37°C por 16 horas (Quadro 3), conforme recomendações dos fabricantes (Fermentas; Promega). Para visualização e obtenção dos fragmentos, fez-se conforme descrito no item 2.9.

Vetor de expressão	Tampão	Enzimas de restrição: unidades (U) e estimativa da atividade enzimática (%)	
pET28a-TEV (3µg)	Tango®	<i>NheI</i> - 10U (100%)	<i>HindIII</i> - 10U (75%)
	Tango®	<i>NheI</i> - 10U (100%)	<i>SacI</i> - 10U (75%)
pET28a-GST-TEV (3µg)	E	<i>EcoRI</i> - 12U (50-75%)	<i>HindIII</i> - 10U (100%)
	E	<i>BamHI</i> - 10U (100%)	<i>SacI</i> - 10U (50-75%)

Quadro 3: Enzimas e tampões utilizados nas reações de dupla digestão dos vetores de expressão. Aproximadamente 3 µg de cada vetor foi duplamente digerido. Tampão Tango® 1x (Fermentas): *Tris-acetato* 33 mM (pH 7,9 a 37°C); *acetato de Magnésio* 10 mM; *acetato de Potássio* 66 mM; *BSA* 0,1 mg/mL. Tampão E 1x (Promega): *Tris-HCl* 6 mM (pH 7,5 a 37°C); *cloreto de Magnésio* 6 mM; *cloreto de Sódio* 100 mM; *DTT* 1 mM.

2.11. Sequenciamento dos fragmentos inseridos nos plasmídeos recombinantes

Os sequenciamentos dos plasmídeos recombinantes foram realizados pelo método didesoxiterminal (SANGER *et al.*, 1977) no núcleo de Análise do Genoma e Expressão Gênica (NAGE/ICB/UFMG) com o auxílio do *DYEnamic ET Dye Terminator Cycle Sequencing Kit For MegaBACE DNA Analysis* (Amersham Biosciences, Buckinghamshire, Inglaterra). Para confirmação das sequências alvo selecionadas, no sequenciamento dos plasmídeos TOPO e pGEM recombinantes foram utilizados os iniciadores M13 direto e reverso descritos no item 2.7.1.

Para confirmar a integridade das sequências codificadoras para as etiquetas de afinidade e o sítio de clivagem para a protease TEV, bem como a janela de leitura para cada proteína recombinante, o sequenciamento dos vetores de expressão contendo os fragmentos gênicos de interesse foi realizado utilizando os iniciadores T7 direto (5' TAA TAC GAC TCA CTA TAG GG 3') e T7 reverso (5' GCT AGT TAT TGC TCA GCG G 3', $T_{anelamento}$: 50°C). As construções de MASP 1, 2 e 4 em pET28a-TEV também foram sequenciadas pela Macrogen, companhia designada para fornecer serviço de sequenciamento de DNA (Seul, Coréia do Sul), em sequenciador automático *ABI Prism® 3730xl DNA Analyser* (Applied Biosystems, Foster City, USA).

2.12. Indução da expressão de proteínas recombinantes

Colônias únicas de *E. coli* BL-21 (DE3) Star (Invitrogen) previamente transformadas e plaqueadas (item 2.7.2) foram inoculadas separadamente em 3 mL de meio de cultura 2xYT na presença do antibiótico Kanamicina 50 µg/mL e incubadas *overnight* a 37°C sob agitação de 200 rpm. Após 16 horas, as culturas iniciais foram diluídas 1:20 em 50 mL de meio 2xYT/Kanamicina (50 µg/mL) e cultivadas até DO₆₀₀ de 0,6-0,8, quando fez-se a indução com 1 mM de IPTG (Invitrogen) por 3 horas a 37°C. Alíquotas de 1 mL antes e depois de 3 horas de indução foram coletadas e submetidas à eletroforese em gel de poliacrilamida-SDS.

2.13. Teste de solubilidade

Após 3 horas de indução, as células foram recuperadas por centrifugação (3000 x g, por 25 minutos a 4°C, em centrífuga Eppendorf 5804R) e ressuspendidas em 5 mL de PBS (*NaCl* 137 mM; *KCl* 2,7 mM; *Na₂HPO₄* 10 mM; *KH₂PO₄* 1,8 mM; pH 7,4) com adição de lisozima (100 µg/mL). O lisado foi submetido a 5 ciclos de congelamento e descongelamento em nitrogênio líquido e banho-maria a 37°C, e passado várias vezes por seringa (1 mL/cc, 26^{1/2} G, 13x0,45) até reduzir completamente a viscosidade. O sobrenadante foi separado da fração insolúvel (*pellet*) através de centrifugação (14.000 rpm, por 10 minutos a 4°C, em centrífuga Eppendorf 5402), e ambas as amostras foram submetidas à eletroforese em gel de poliacrilamida-SDS para verificar a solubilidade das proteínas recombinantes.

2.14. Eletroforese em gel de poliacrilamida-SDS

Para a análise eletroforética de proteínas, os géis foram preparados a partir da mistura de N,N'-metileno-bis-acrilamida 0,8% e acrilamida 30% (p/v). O gel de

separação (12%) foi feito adicionando-se a esta mistura Tris-HCl 1,5 M pH 8,8 e SDS 0,01%. O gel foi polimerizado pela adição de persulfato de amônio a 10% em água destilada (0,5% v/v) e TEMED 0,05% (v/v). O gel de concentração (4% de acrilamida) foi preparado como descrito anteriormente para o gel de separação, porém utilizando-se tampão Tris-HCl 0,5 M pH 6,8 (LAEMMLI, 1970).

As amostras de extratos protéicos foram ressuspensas em tampão de amostra (*SDS 10%; Tris-HCl 0,5 mM pH 6,8; azul de Bromofenol 1%; 2-β-mercaptoetanol 5%; glicerol 10%*), fervidas em banho-maria por 5 minutos e aplicadas às canaletas do gel de concentração. A eletroforese foi realizada em tampão de corrida 1x (*Tris-HCl 25 mM; glicina 192 mM; SDS 0,1%; pH 8,3*) em voltagem constante de 100 V. Após a corrida, os géis foram corados por duas horas em azul de Coomassie (*Coomassie Brilliant Blue G-250 0,25%; metanol 50%; ácido acético 10%*) e descorados em solução descorante (*etanol 30%; ácido acético 10%*), ou então submetidos a ensaios de *western blotting*.

2.15. Western blotting

Segundo a técnica adaptada de Towbin e colaboradores (1979), após a corrida eletroforética das proteínas de interesse, o conteúdo protéico foi transferido do gel de poli-acrilamida-SDS 12% para membranas de nitrocelulose (BioAgency, São Paulo, Brasil) em tampão de transferência (*Tris 25 mM; glicina 192 mM; metanol 20%, pH 8,3*). O sistema foi colocado em cuba específica para eletrotransferência (BioRad, Hercules, USA) e submetido a voltagem constante de 100 V durante 1 hora em banho de gelo.

Após ser retirada do sistema, a membrana foi embebida em corante Ponceau (*Ponceau S 1%; ácido acético 10%*) para certificação da transferência das proteínas. Posteriormente, foi lavada com água, bloqueada com leite em pó desnatado 5% em TBS (*Tris 50 mM; NaCl 150 mM, pH 7,6*) contendo Tween 0,05% por 1 hora ou *overnight* e então lavada com TBS-Tween por 3 vezes de 5 minutos cada. Feito isto, a membrana foi incubada por 1 hora com o anticorpo primário anti-histidina Penta-His (Qiagen) diluído a 1:2000, em TBS-Tween contendo 3% de leite em pó desnatado. Após ser lavada com TBS-Tween (3 vezes de 5 minutos), a membrana foi incubada por mais 1 hora, agora com o anticorpo secundário anti-IgG de camundongo conjugado com

peroxidase diluído a 1:3000. A revelação da membrana foi feita por método colorimétrico com 3,3'-diaminobenzidina (DAB) 1 mg/mL em TBS, 4-cloronaftol 0,5 mg/mL em metanol e H₂O₂ 0,03% (v/v). A reação foi interrompida com água deionizada.

A partir deste ponto, o trabalho foi conduzido com a MASP 3 clonada no vetor pET28a-TEV, tendo em vista os resultados obtidos e a necessidade de otimização do tempo para a condução dos experimentos.

2.16. Lise celular

Uma vez otimizadas as condições de indução, a expressão da MASP 3 em fusão com cauda de histidina foi feita de modo semelhante ao descrito no item 2.12, seguindo porém a proporção para volumes maiores de cultura, como 1L, a fim de obter maior rendimento.

Após a indução, as células foram coletadas por centrifugação (3000 x g, por 25 minutos a 4°C), ressuspendidas em tampão A (*fosfato de Sódio 20 mM; NaCl 500 mM; Imidazol 30 mM; pH 7,5*) e incubadas com lisozima (100 µg/mL) em gelo por 15 minutos. O tratamento do lisado foi feito de modo semelhante ao descrito no item 2.13. O sobrenadante foi submetido a métodos cromatográficos para purificação da proteína de interesse.

2.17. Cromatografia de afinidade

A proteína MASP 3 fusionada à cauda de histidina foi purificada inicialmente através de cromatografia de afinidade a níquel. O extrato foi injetado em coluna HisTrap HP de 5 mL (GE Healthcare Life Sciences, Buckinghamshire, Inglaterra) previamente equilibrada com 5 volumes de coluna (vc) de tampão A. Após a aplicação da amostra, a coluna foi lavada com 5 vc de tampão A, ou até que nenhuma leitura de

absorbância fosse observada a 280 nm. A eluição foi feita em concentração crescente de imidazol utilizando 25 volumes de tampão B (*fosfato de Sódio 20 mM; NaCl 500 mM; Imidazol 500 mM; pH 7,5*). Amostras recolhidas ao longo de todo o processo foram submetidas à eletroforese em gel de poliacrilamida-SDS.

2.18. Diálise e determinação da concentração de proteínas

Para remoção do sal e do imidazol das frações provenientes da coluna de afinidade, as amostras foram dialisadas contra tampão Tris-HCl 10 mM pH 7,5 em membrana de celulose (6000 a 8000 Da; SIGMA, St. Louis, USA). Para as amostras provenientes da coluna de troca iônica (item 2.21), a diálise foi feita contra tampão fosfato de Sódio 10 mM pH 7,5 novamente em membrana de celulose.

Fez-se a dosagem das proteínas através da medida da absorbância a 280 nm. Como é conhecida a sequência de resíduos de aminoácidos das proteínas em estudo, é possível prever o coeficiente de extinção molar de cada uma e aplicar a seguinte fórmula:

$$A = \epsilon * l * c \quad (1)$$

onde **A** é a absorbância a 280 nm; **ϵ** , o coeficiente de extinção molar; **l**, o caminho óptico da cubeta em cm, e **c**, a concentração molar da proteína. Para a MASP 3 em fusão com cauda de histidina, $\epsilon = 11460 \text{ M}^{-1} \text{ cm}^{-1}$, e após clivagem por TEV, $\epsilon = 9970 \text{ M}^{-1} \text{ cm}^{-1}$. Estes dados foram obtidos através do programa ProtParam (GASTEIGER *et al.*, 2005) (<http://www.expasy.ch/tools/protparam.html>) hospedado no servidor proteômico ExPASy – *Expert Protein Analysis System* (GASTEIGER *et al.*, 2003) (<http://www.expasy.ch>).

2.19. Remoção da cauda de histidina

A cauda de histidina presente na extremidade N-terminal da proteína recombinante foi clivada através de proteólise com TEV. Testes de clivagem foram

feitos a 30°C por 3 horas, com a protease TEV a 5%, 10% e 20% (p/p) em tampão Tris-HCl 50 mM pH 8,0; EDTA 0,05 mM com adição de DTT 1 mM. A remoção da cauda de histidina foi confirmada por análises de espectrometria de massa.

2.20. Espectrometria de massas

As amostras, antes e após a clivagem por TEV, foram analisadas em espectrômetro de massas, tipo MALDI-TOF, Bruker Daltonics autoflex III smartbeam. Para isso, foram aplicados em duplicata sobre placa MTP Anchor Chip 600/384 (Bruker Daltonics, Billerica, USA) 0,5 µL de amostra e 0,5 µL da matriz ácido sinapínico (10 mg/mL em ACN 30%, 0,3% TFA). Para minimizar a quantidade de sal nas amostras, os spots, após secagem à temperatura ambiente, foram lavados por duas vezes com 10 µL de TFA 0,1%. As amostras foram analisadas no espectrômetro de massa utilizando o software Flexcontrol 3.0. O equipamento foi calibrado utilizando padrões externos, com proteínas de peso molecular conhecido (Protein calibration standart I). Os espectros obtidos foram analisados utilizando o software Flex analysis 3.0 para identificação das respectivas razões massa/carga medidas e inferência das massas moleculares das espécies ionizadas.

2.21. Cromatografia de troca iônica

O produto da clivagem por TEV foi injetado em coluna RESOURCE Q de 1 mL (GE Healthcare Life Sciences) previamente equilibrada com 10 vc de tampão Tris-HCl 20 mM pH 7,5. Após aplicação da amostra, a coluna foi lavada com 10 vc do mesmo tampão, ou até que nenhuma leitura de absorbância fosse observada a 280 nm. A eluição foi feita em gradiente linear com concentração crescente de NaCl utilizando 25 vc de tampão Tris-HCl 20 mM pH 7,5; NaCl 1 M. Amostras recolhidas ao longo de todo o processo foram submetidas à eletroforese em gel de poliacrilamida-SDS.

2.22. Espectroscopia de dicroísmo circular

Amostras protéicas da MASP 3 provenientes de cada etapa de cromatografia, dialisadas contra tampão fosfato de sódio 10 mM pH 7,4, foram centrifugadas (13.000 rpm, por 15 minutos a 4°C, em centrífuga Thermo Scientific HERAEUS Fresco 17) e submetidas a análises de dicroísmo circular. Os espectros de CD foram registrados pelo espectrômetro Chirascan (Applied Photophysics, Leatherhead, Inglaterra), software ProDataChirascan, em intervalos de 1,0 nm entre 190 a 260 nm de comprimento de onda, nas temperaturas de 20, 37, 60 e 80°C. As amostras da proteína com e sem etiqueta de afinidade foram utilizadas nas concentrações de 109 e 72 µM, respectivamente, e dispostas em cubetas de quartzo de 1 cm de caminho óptico.

Os dados, obtidos em termos de elipcidade (θ) expressa em millidegrees (mdeg), foram convertidos em elipcidade molar residual $[\theta]_{MR}$ segundo a fórmula:

$$[\theta]_{MR} = 100 * \theta / (C_{MR} * l) \quad (2)$$

$[\theta]_{MR}$ é a elipcidade molar residual a ser expressa em degrees.cm².dmol⁻¹;

θ é a elipcidade obtida pelo espectrômetro em mdeg;

C_{MR} é a concentração molar residual da proteína; para a qual $C_{MR} = C * N$, onde C é a concentração molar da proteína e N é o número de resíduos de aminoácidos;

l é o caminho óptico da cubeta em cm.

Para predição da estrutura secundária em porcentagem, a deconvolução dos dados foi feita pelo software CDNN Versão 2, desenvolvido pelo Dr. Gerald Böhm, Institut für Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Alemanha (<http://bioinformatik.biochemtech.uni-halle.de/cdnn>).

3. RESULTADOS

3.1. Seleção das proteínas MASP

No banco contendo os dados do genoma diplóide de *T. cruzi*, cepa CI Brener, são encontradas 1377 sequências anotadas como MASP, a partir do qual foram obtidas 810 sequências excluindo-se pseudogenes e genes parciais. Destas, 34 sequências protéicas apresentam região central inferior a 20 kDa. A partir deste conjunto, foram selecionados quatro membros, cujas sequências gênicas estão depositadas no banco de dados TriTrypDB (<http://tritrypdb.org/tritrypdb/>). Os identificadores e o alinhamento das sequências de resíduos de aminoácidos são apresentados. O alinhamento foi feito utilizando o algoritmo MAFFT (KATO *et al.*, 2002) (<http://mafft.cbrc.jp/alignment/server/>).

```

1          11          21          31          41          50
|          |          |          |          |          |
MAMMMTGRVL LVCALCVLWS VAADG----- ---AVGVVSG --GDDNSLKE 40 Tc00_1047053510013_170
MAMMMTGRVL LVCALCVLWS VVADG----- ---DVGVVSG --GDDNSLKE 40 Tc00_1047053507955_20
MAMMMTGRVL LVCALCVLWC GAGGGFA-KE AEASANGVSS KTTPADRIIL 49 Tc00_1047053507957_300
MAMMMTGRVL LVCALCVLWC GVGGGFADKV VEAPAGVVAS KRTDEENLIL 50 Tc00_1047053506757_50

LFIPVARLQE RPEQRAAEAT ADAKAAAAAK ASAEAAERAK IAT---EATK 87 Tc00_1047053510013_170
LFIPVARLQE RPEQRAAGAT ADAKAAAAAK ASAEAAERAK IATAKASAEK 90 Tc00_1047053507955_20
NWHVLMK--- --EECATENT KNGTVNVPAM KR--CIQVAM KGICDTFYNK 92 Tc00_1047053507957_300
NWFYFLIR--- --EECANEST TGGKLNVTAE RI--CIHKVM KGVCDAFYDK 93 Tc00_1047053506757_50

TA----- -TAKAEAEA AAEALRGTTV REQEVKAAIH DQDNSVEHHS 128 Tc00_1047053510013_170
AAAAATEADA KTTAAKTAEA AAEALRGTTI REQEVKTAIH DQDNSVEHHS 140 Tc00_1047053507955_20
TPSEIHDPEV KGMCTYYAAI PDEPDEPPTP QGPPPSAV- ---SNTTSNE 138 Tc00_1047053507957_300
TPGVTHDRDV EFICTYYATI PEEPVEPPTP QGPPSDSAV- ---SSTTSNE 139 Tc00_1047053506757_50

EEKQELLQE- EPERQEKEQH EKQHQHQHEH SAGNGDEPAK DNTANGTNAS 177 Tc00_1047053510013_170
GEKQELLQEQ EPERQEKEQH EKQHQHQREH SAGNGEESPK EKTANGTNAS 190 Tc00_1047053507955_20
GTPKNAPESD DAGRGEGKED EKEH----- --GNTKQKAV ESAAMKHITK 180 Tc00_1047053507957_300
GTPKNAPESD DAGRGEGKED EKEH----- --GNTKQTPV ETAAVENDTK 181 Tc00_1047053506757_50

AITDSDGST ALSHTSLLL LLLVVVACAA AAVVAA 214 Tc00_1047053510013_170
AITDSDGST AVSHNTSLLL LLLVVVACAA AAVVAA 227 Tc00_1047053507955_20
--TADSDGRT AVTHITFPLL LLLL---AA AAALVAA 212 Tc00_1047053507957_300
--TTSDSGST AVSHATSPLL LLLLVV-CAA AAVVAA 215 Tc00_1047053506757_50

```

Há dois pares de sequências muito parecidas, entre Tc00.1047053510013.170 e Tc00.1047053507955.20 (MASPs 1 e 2) e entre Tc00.1047053507957.300 e Tc00.1047053506757.50 (MASPs 3 e 4). A divergência entre os pares é grande, entretanto, observa-se que a conservação se restringe às regiões N e C-terminal.

Para MASPs 1 e 2, o sítio de clivagem de peptídeo sinal predito pelo algoritmo SignalIP situa-se entre ADG-AV e ADG-DV, respectivamente. Para MASPs 3 e 4, a clivagem ocorre respectivamente entre GFA-KE e GFA-DK. Quanto aos sítios de adição de âncora de GPI preditos pelo algoritmo GPI SOM, a clivagem ocorre entre os resíduos de ácido aspártico e glicina sublinhados.

O Quadro 4 relaciona o comprimento de sequências nucleotídicas e de resíduos de aminoácidos das quatro proteínas MASP recombinantes.

Identificador (TriTrypDB)	Número de nucleotídeos		Número de resíduos de aminoácidos e (peso molecular em Daltons), antes e após remoção da etiqueta de afinidade da proteína recombinante expressa em:	
	Região central codificadora	Amplicon	pET28a-TEV	pET28a-GST-TEV
MASP 1 (Tc00.1047053510013.170)	477	498	178 (19256.5) 164 (17450.6)	385 (43397.6) 164 (17444.5)
MASP 2 (Tc00.1047053507955.20)	516	537	191 (20404.9) 177 (18598.9)	398 (44546.0) 177 (18592.9)
MASP 3 (Tc00.1047053507957.300)	474	495	177 (19262.2) 163 (17456.2)	382 (43127.0) 161 (17173.8)
MASP 4 (Tc00.1047053506757.50)	477	498	178 (19522.2) 164 (17716.3)	383 (43387.0) 162 (17433.9)

Quadro 4: Relação entre o comprimento de sequências nucleotídicas e de resíduos de aminoácidos das proteínas MASP selecionadas. Cada amplicon consiste da região central codificadora somada aos nucleotídeos adicionados pelos iniciadores. Os valores correspondentes ao número de resíduos de aminoácidos e ao peso molecular em Daltons foram obtidos pelo programa ProtParam hospedado no servidor ExPASy (<http://www.expasy.ch/tools/protparam.html>) (GASTEIGER *et al.*, 2005).

Para a seleção destas quatro MASPs, foi investigada a presença de sítios de glicosilação na cadeia polipeptídica, a fim de serem escolhidas aquelas com menores probabilidades para esta modificação pós-traducional. A N-glicosilação, que ocorre no nitrogênio da amida da cadeia lateral do resíduo de asparagina presente na sequência Asn-Xaa-Ser/Trn, sendo Xaa qualquer aminoácido exceto prolina, foi identificada pelo algoritmo NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>). O resultado desta predição é mostrado na figura 11. Os gráficos ilustram os sítios de glicosilação preditos ao longo da cadeia polipeptídica disposta no eixo x. Enquanto que para MASP 1 não foi

predito este tipo de glicosilação, há poucos sítios em potencial (linhas verticais) nas outras MASPs, sendo que alguns não ultrapassam o limiar de 0.5, uma vez que a sequência tripeptídica consenso é um requerimento, mas nem sempre suficiente para ditar a glicosilação da asparagina. A O-glicosilação, por sua vez, ocorre no hidróxi-oxigênio de resíduos de serina e treonina e foi identificada pelo algoritmo NetOGlyc 3.1 (<http://www.cbs.dtu.dk/services/NetOGlyc/>), cujos resultados também são apresentados na figura 11. Há vários sítios preditos presentes nas quatro MASPs, entretanto nem todos configuram-se em sítios prováveis de glicosilação, os quais ultrapassam o limiar de 0.5.

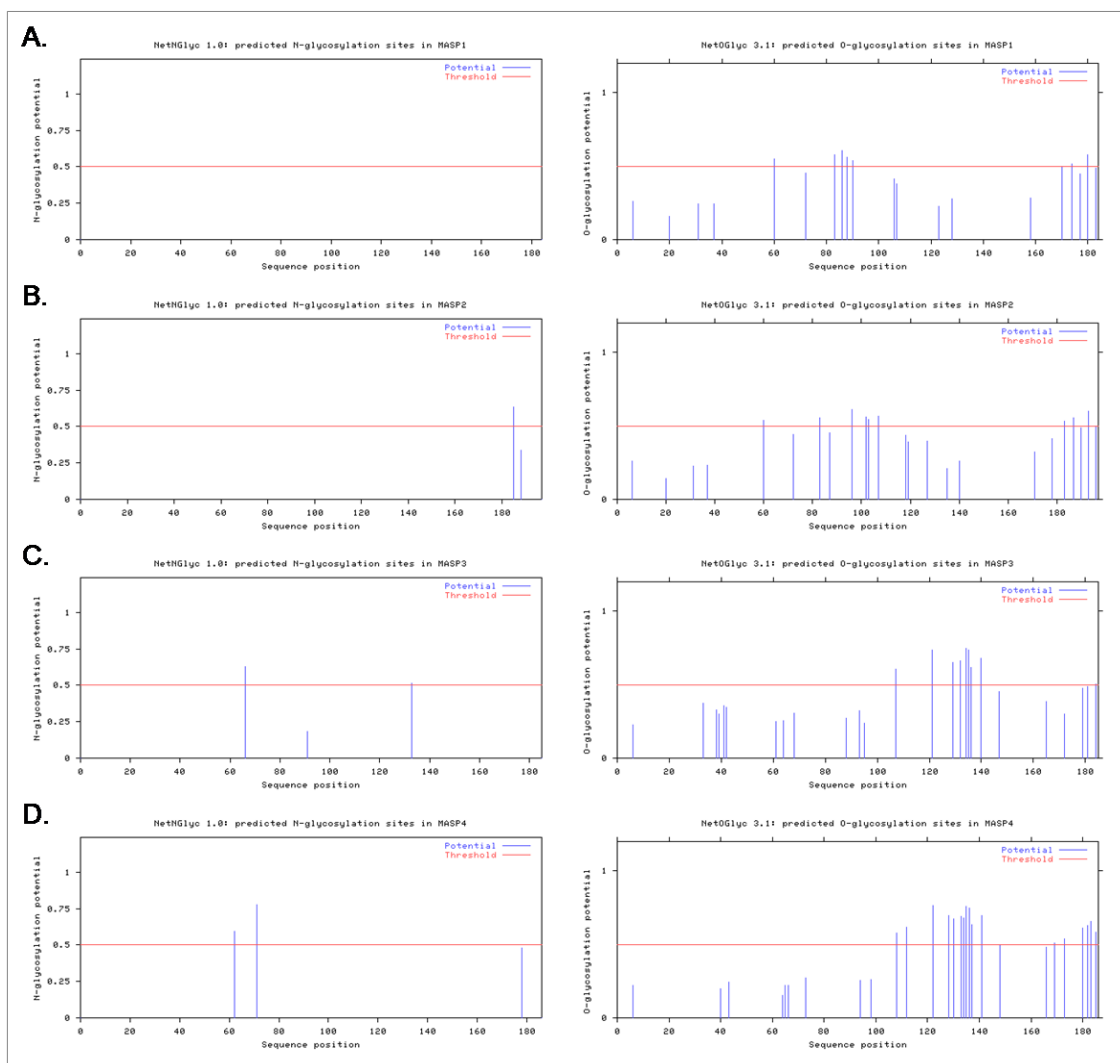


Figura 11: Análise *in silico* para identificação de sítios de N- (gráficos à esquerda) e O-glicosilação (gráficos à direita) nas MASPs 1 (A), 2 (B), 3 (C) e 4 (D). Para esta análise, foi excluída a sequência correspondente ao sítio de adição de âncora de GPI, mas utilizada a região central acrescida do peptídeo sinal, uma vez que este último é identificado pelos preditores, pois constitui uma exigência para o reconhecimento de proteínas de membrana. Os gráficos foram gerados pelos algoritmos NetNGlyc 1.0 (à esquerda) e NetOGlyc 3.1 (à direita). Sítios com grande probabilidade de glicosilação ultrapassam o limiar (linha horizontal em 0.5).

Outra característica investigada foi a presença de regiões de desordem estrutural nas cadeias polipeptídicas de MASP. Regiões desordenadas abrangem desde pequenas porções até proteínas inteiras, as quais não possuem estrutura terciária fixa e, logo, são parcial ou completamente desenoveladas em condições fisiológicas.

As análises de desordem estrutural foram realizadas com a sequência completa das quatro MASPs, pois o algoritmo PONDR[®] VL-XT (<http://www.pondr.com/>) consiste de um preditor de regiões desordenadas internas e dois outros preditores apenas para as extremidades N- e C-terminal (LI *et al.*, 1999; ROMERO *et al.*, 2001). O resultado desta predição é mostrado na figura 12. Os gráficos ilustram as regiões de ordem e desordem (abaixo e acima do limiar de 0.5, respectivamente) preditas ao longo da cadeia polipeptídica disposta no eixo x. MASPs 1 e 2 apresentam longas regiões centrais desordenadas de aproximadamente 150 e 100 resíduos, respectivamente. Porções menos extensas são encontradas nas MASPs 3 e 4.

Diante disso, gráficos de carga líquida *versus* hidrofobicidade foram obtidos apenas para as regiões centrais das quatro MASPs (Fig. 12). Estas análises revelam que as porções centrais localizam-se no espaço designado para proteínas desenoveladas, porém estão próximas do limite (linha cheia) que representa uma estimativa da vizinhança entre proteínas enoveladas e desenoveladas.

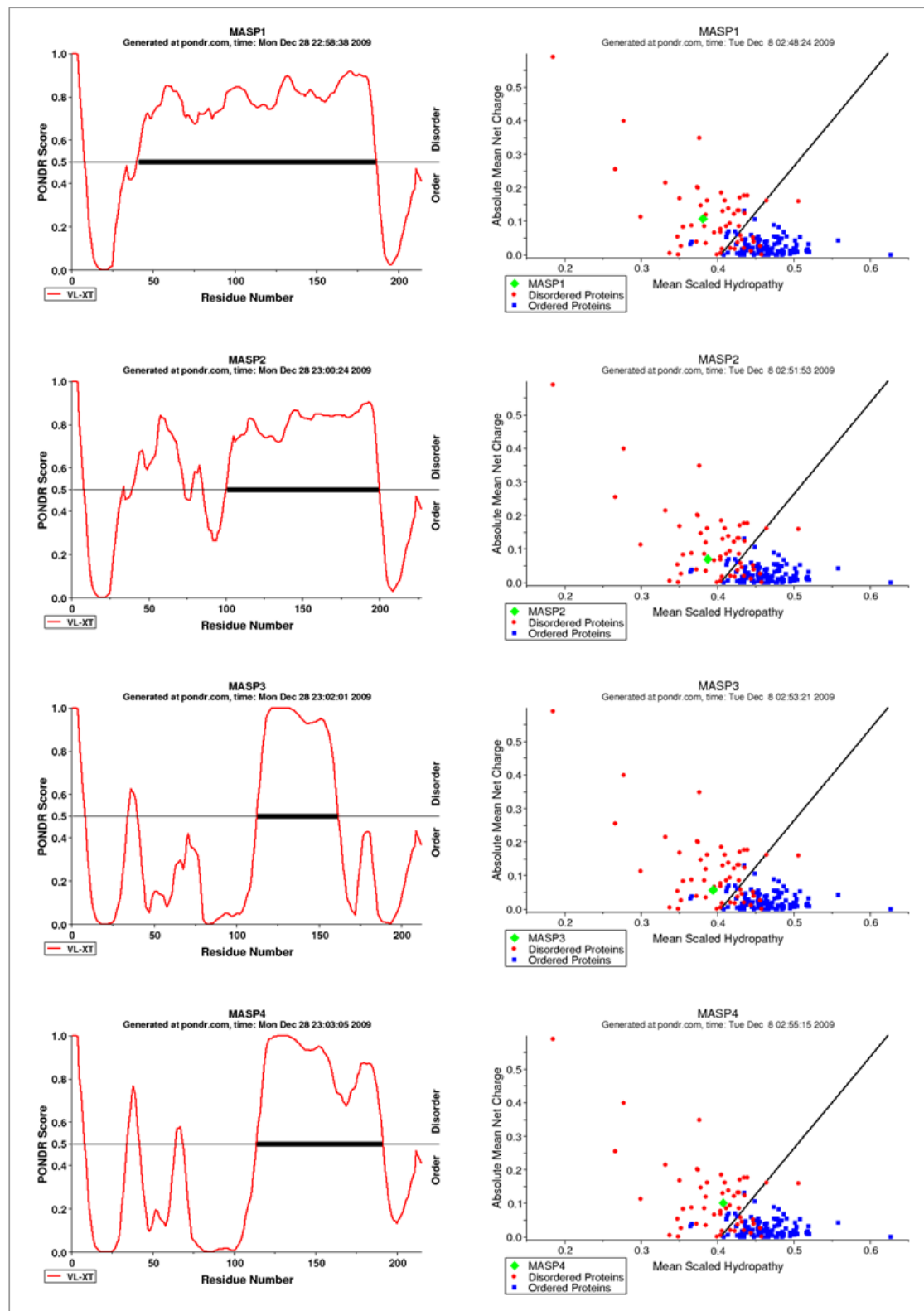


Figura 12: Predições de desordem estrutural nas MASP1, 2, 3 e 4 (indicadas na figura). Os gráficos (à esquerda) obtidos pelo algoritmo PONDR VL-XT revelam regiões desordenadas previstas ao longo da cadeia polipeptídica completa de cada MASP. Os gráficos correspondentes (à direita) revelam as médias de carga líquida (R) e hidrofobicidade (H). As taxas de R/H foram analisadas apenas para as porções centrais de cada MASP. A linha cheia ($R = 2.785H - 1.151$) representa uma estimativa do limite entre proteínas desenoveladas (esferas vermelhas) e enoveladas (quadrados azuis) determinada pelo método descrito por Uversky e colaboradores, 2000.

3.2. Amplificação e subclonagem das regiões codificadoras centrais dos genes *masp* em vetores de sequenciamento

Como apenas a porção central hipervariável das proteínas MASP é expressa na superfície preferencialmente de formas tripomastigotas do parasito *T. cruzi*, foram sintetizados oligonucleotídeos iniciadores específicos para as regiões codificadoras centrais dos quatro genes *masp* previamente selecionados. Em cada iniciador foi adicionada uma sequência específica reconhecida por uma enzima de restrição para posterior utilização na etapa de clonagem, conforme descrito no Quadro 1 (Materiais e Métodos, item 2.2).

Tendo em vista, para cada região codificadora central, a possibilidade de clonagem em sistemas de expressão com cauda de histidina ou GST na porção N-terminal, foram obtidos oito amplicons, cada qual correspondendo ao peso esperado de aproximadamente 500 pares de bases (pb) conforme mencionado no Quadro 4 (item 3.1) (Fig. 13).

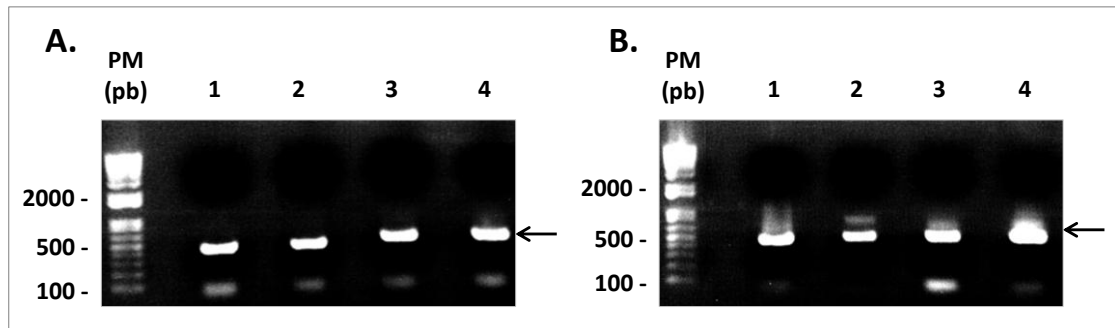


Figura 13: Amplificação das regiões codificadoras centrais dos genes *masp* (1: Tc00.1047053510013.170; 2: Tc00.1047053507955.20; 3: Tc00.1047053507957.300 e 4: Tc00.1047053506757.50) utilizando iniciadores específicos para posterior clonagem nos vetores pET28a-TEV (A) e pET28a-GST-TEV (B). Os fragmentos, amplificados a partir do DNA genômico purificado do clone CL Brener, correspondem aproximadamente a 500 pb indicados pelas setas. Géis de agarose 1% corados com brometo de etídeo. PM (pb): peso molecular em pares de bases (1 kb Plus DNA Ladder - Invitrogen).

Produtos de PCR apresentam na extremidade 3' uma desoxiadenosina (A) livre adicionada pela enzima *Taq* polimerase durante a reação de amplificação. Os vetores linearizados dos sistemas TOPO[®] (Invitrogen) e pGEM[®] (Promega), por sua vez, possuem na extremidade 3' uma desoxitimidina (T) livre, característica que favorece a ligação eficiente dos produtos de PCR (INVITROGEN, 2006; PROMEGA, 2009).

Dessa forma, após amplificação e subclonagem das regiões centrais codificadoras em vetores TOPO e/ou pGEM, vários transformantes foram obtidos. Alguns clones analisados por PCR de colônias utilizando iniciadores M13 direto e reverso mostraram-se positivos (dado não mostrado). Os produtos da amplificação dos vetores TOPO e pGEM contendo inserto, quando visualizados em gel de agarose, apresentaram peso molecular esperado de aproximadamente 700 pb, sendo 200 pb correspondentes ao sítio de clonagem do vetor (Fig. 14).

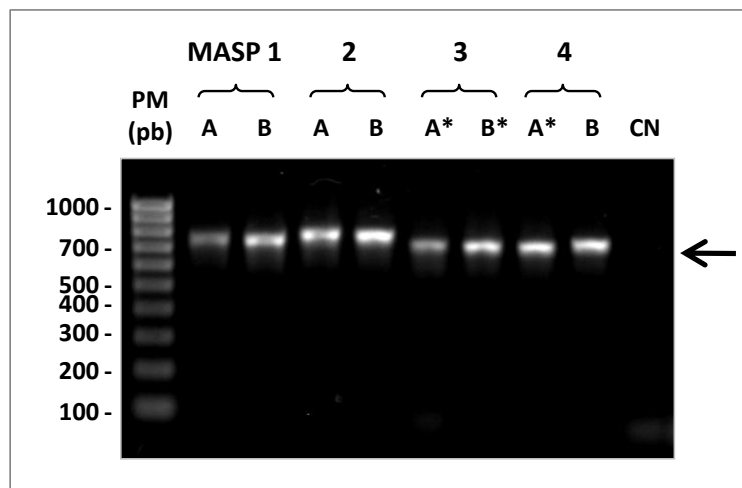


Figura 14: Amplificação do sítio de clonagem dos vetores TOPO (*) e pGEM utilizando iniciadores M13 direto e reverso, após subclonagem dos insertos de MASP 1, 2, 3 e 4, para posterior clonagem em pET28a-TEV (A) e pET28a-GST-TEV (B). Os fragmentos amplificados correspondem aproximadamente a 700 pb indicados pela seta. O DNA plasmidial utilizado nesta reação foi purificado dos clones positivos previamente selecionados por PCR de colônias (dado não mostrado). Gel de agarose 1% corado com SYBR Safe® (Invitrogen). PM (pb): peso molecular em pares de bases (100 bp ladder - Bio-Rad). CN: controle negativo da reação de PCR.

Análises de sequenciamento dos clones positivos utilizando primers M13 direto e reverso, apesar da baixa qualidade das sequências, confirmaram que sete construções trabalhadas correspondem às sequências gênicas selecionadas, exceto a construção do gene Tc00.1047053506757.50 (MASP 4) em vetor pGEM (para posterior clonagem no vetor pET28a-GST-TEV). Nesta, foi amplificada e subclonada a região codificadora central do gene Tc00.1047053510359.550, tendo em vista o alinhamento quase perfeito entre os iniciadores desenhados para a MASP 4 e as extremidades da sequência nucleotídica codificadora central de Tc00.1047053510359.550 (Fig. 15).

Tc00.1047053510359.550	1	ATGGCGATGATGATGACTGGCCGTGTGCTGCTGGTGTGTGCCCTCTGCGT	50
Tc00.1047053510359.550	51	GCTGTGGTGC GGTTGCGGCTGTGGTTTTGCGG GACAAGGTGGTGGAAAGCTG	100
		
7084.t00005_F:pET.GST.TEV	1	ACTGGATCC GACAAGGTGGTGGAAAG	25
Tc00.1047053510359.550	101	CTGCCGGTGTGGTTGCCAGTAACAGAACGGATGAGGAGAATTTGATCTTG	150
Tc00.1047053510359.550	151	AATGGTATGTTCTGACGCAGGAAGAATGCGGAATGAAAGCACGCACAGG	200
Tc00.1047053510359.550	201	AGGGAAGTTGAATGTGACAGCTGAGAGGATCTGCATGCACAAAAGTGATGA	250
Tc00.1047053510359.550	251	AGGGGGTTTGTGACGCTTTTTATAATAAAACGTCGGGTGTAACTCATGAC	300
Tc00.1047053510359.550	301	CGTGACGTTGAGTTTATTTGTACGTATTACGCTACCATTCCAGAGGAGCC	350
Tc00.1047053510359.550	351	CGTCGAGCCCCAACACCGCAAGGCCACCATCACATTCAGCTGTCAGCA	400
Tc00.1047053510359.550	401	ATGCCACCAGTGATGAGGGGACGCCGAAGAATGCACCGGAATCCGATGTT	450
Tc00.1047053510359.550	451	GCAGGGAGAGGCCAAGGAAAACAAGATGACAAGGAGCACGGCAATACGAA	500
Tc00.1047053510359.550	501	GCAAAAAGCAGTCGAAACC GCCGTGTGGAAAGTATCACAGGACTACCG	550
7084.t00005_R	1	GACTACCG	8
Tc00.1047053510359.550	551	ACAGTGACGGCAGCACCGCGGCCTCCCACACCACCTCCCATCTTTTGCTT	600
		
7084.t00005_R	9	ACAGCGACTGAGAGCTCAGA	28
Tc00.1047053510359.550	601	CTTCTTCTTGTGTGTGCGTGTGCGGCTGCTGCTGCGGTGGTGGCCGCGTG	650
Tc00.1047053510359.550	651	A	651

Figura 15: Alinhamento entre as extremidades da região codificadora central de Tc00.1047053510359.550 (destacada em negrito) e os iniciadores 7084.t00005_F:pET.GST.TEV (direto) e 7084.t00005_R (reverso), desenhados especificamente para amplificar a região codificadora central de Tc00.1047053510359.550. As porções dos iniciadores que são complementares às extremidades da sequência codificadora central de Tc00.1047053510359.550 estão em azul. O alinhamento entre estas é praticamente perfeito, exceto pela citosina na posição 16 (5'→3') do iniciador reverso. Os sítios das enzimas de restrição estão sublinhados. O alinhamento foi realizado através do algoritmo EMBOS *Pairwise Alignment Algorithms* (<http://www.ebi.ac.uk/Tools/emboss/align/index.html>).

3.3. Clonagem das regiões codificadoras centrais dos genes *masp* em vetores de expressão

Com a finalidade de obtermos as regiões centrais codificadoras com as extremidades coesivas, correspondentes aos sítios de restrição adicionados durante a amplificação, os vetores TOPO e pGEM contendo inserto foram submetidos à dupla digestão com as devidas enzimas de restrição (Materiais e Métodos, item 2.9, Quadro 2).

Quando visualizados em gel de agarose, os produtos confirmaram o sucesso da digestão enzimática devido à presença de bandas de aproximadamente 500 pb, correspondentes ao peso esperado para os insertos liberados (Fig. 16). Estas bandas foram excisadas do gel para posterior purificação dos insertos.

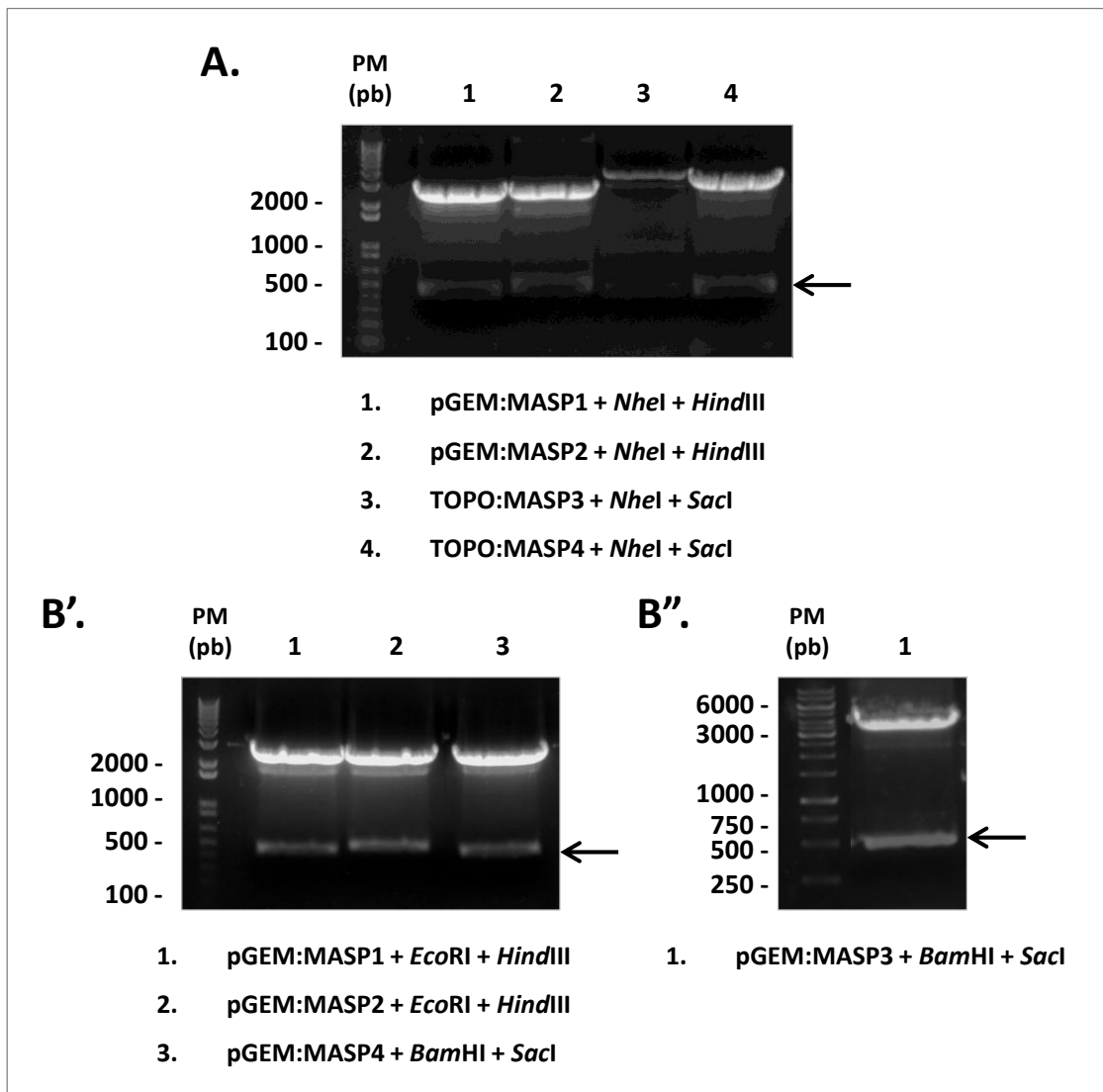


Figura 16: Dupla digestão enzimática dos vetores TOPO e pGEM contendo insertos de MASP 1, 2, 3 e 4. Os fragmentos liberados indicados pelas setas correspondem aproximadamente a 500 pb, os quais serão posteriormente clonados em pET28a-TEV (A) e pET28a-GST-TEV (B' e B''). As enzimas de restrição utilizadas estão indicadas na figura. Géis de agarose 1% corados com brometo de etídeo. PM (pb): peso molecular em pares de bases (A e B': 1 kb Plus DNA Ladder - Invitrogen. B'': GeneRuler® 1 kb DNA Ladder - Fermentas).

Os vetores de expressão pET28a-TEV e pET28a-GST-TEV foram submetidos à dupla digestão com as devidas enzimas de restrição a fim de gerar também extremidades coesivas compatíveis com aquelas presentes nos insertos. Após eletroforese em gel de agarose, os vetores de expressão apresentaram padrão de migração semelhante ao observado para as digestões controle, nas quais os vetores foram linearizados com as mesmas enzimas separadamente (Fig. 17). As bandas correspondentes a aproximadamente 6000 pb foram excisadas do gel para posterior purificação dos vetores.

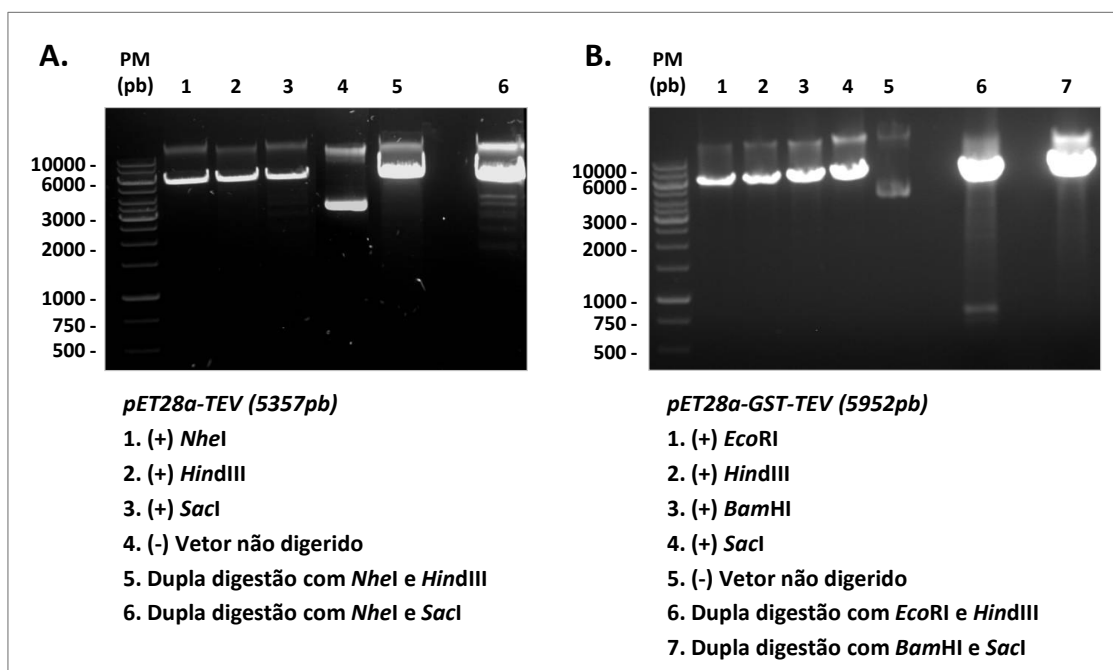


Figura 17: Digestão dos vetores pET28a-TEV (A) e pET28a-GST-TEV (B) com enzimas de restrição indicadas na própria figura. (+) corresponde ao controle positivo, no qual o vetor foi submetido à digestão por cada uma das enzimas separadamente. (-) representa o controle negativo, no qual o vetor não foi submetido à digestão. Géis de agarose 1% corados com SYBR Safe® (Invitrogen) (A) e brometo de etídeo (B). PM (pb): peso molecular em pares de bases (GeneRuler® 1 kb DNA Ladder - Fermentas).

Insertos e vetores de expressão descritos acima foram utilizados em reações de ligação. Bactérias *E. coli* XL-1 Blue foram transformadas, cada qual, com um produto das reações de ligação. Alguns transformantes obtidos, resistentes à kanamicina presente no meio 2xYT-ágar, foram analisados por PCR utilizando os iniciadores específicos descritos no Quadro 1 (Materiais e Métodos, item 2.2). Das oito construções inicialmente planejadas, foram obtidos clones positivos para sete, exceto para MASP 2 clonada em pET28a-GST-TEV (Fig 18).

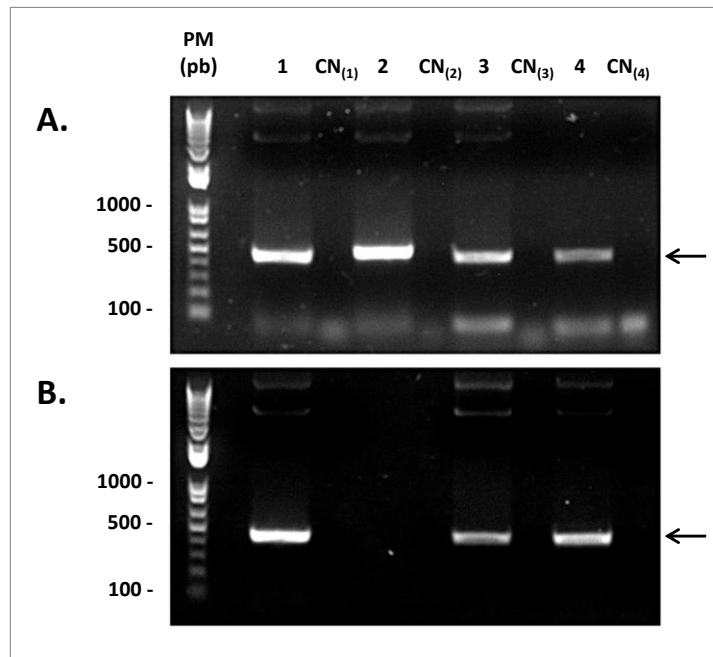


Figura 18: Amplificação dos insertos de MASP 1, 2, 3 e 4 utilizando iniciadores específicos (Quadro 1), após clonagem em pET28a-TEV (A) e pET28a-GST-TEV (B). Os fragmentos amplificados correspondem aproximadamente a 500 pb indicados pelas setas. O DNA plasmidial utilizado nesta reação foi purificado dos clones positivos previamente selecionados por PCR de colônias (dado não mostrado). Géis de agarose 1% corados com SYBR Safe® (Invitrogen). PM (pb): peso molecular em pares de bases (1 kb Plus DNA Ladder - Invitrogen). CN_(n): controle negativo da reação de PCR.

A comparação de uma das sequências obtidas da construção de MASP3 em pET28a-TEV com a sequência nucleotídica codificadora central de Tc00.1047053507957.300, depositada no banco de dados TriTrypDB (<http://tritrypdb.org/>), foi realizada através do programa EMBOSS *Pairwise Alignment Algorithms* (<http://www.ebi.ac.uk/Tools/emboss/align/index.html>) (Fig. 19). Observa-se um alinhamento praticamente perfeito entre as duas sequências nucleotídicas, com uma mutação silenciosa no nucleotídeo 453 da sequência obtida, correspondente à terceira base do códon que codifica para o mesmo aminoácido (ácido glutâmico). No início da sequência, observa-se a cauda de seis resíduos de histidina e o sítio de clivagem por TEV inseridos pelo plasmídeo.

As outras construções foram novamente submetidas ao sequenciamento utilizando primers T7 direto e reverso.

Tc00.1047053507957.300	1		0
6xHis:MASP3	1	ATGGGCCATCATCATCATCACGAAAACCTGTATTTCAGGGCCATAT	50
		M G H H H H H H E N L Y F Q G H M	
Tc00.1047053507957.300	1	AAGGAGGCGGAAAGCTTCTGCCAATGGTGTTCAGTAAGACAA	43
6xHis:MASP3	51	GGCTAGCAAGGAGGCGGAAAGCTTCTGCCAATGGTGTTCAGTAAGACAA	100
		A S K E A E A S A N G V S S K T T	
Tc00.1047053507957.300	44	CACCTGCTGACAGAATCATCTTGAATTGGCATGTCTGATGAAGGAAGAA	93
6xHis:MASP3	101	CACCTGCTGACAGAATCATCTTGAATTGGCATGTCTGATGAAGGAAGAA	150
		P A D R I I L N W H V L M K E E	
Tc00.1047053507957.300	94	TGCGCGACGGAAAATACGAAGAATGGGACGGTAAATGTGCCAGCGATGAA	143
6xHis:MASP3	151	TGCGCGACGGAAAATACGAAGAATGGGACGGTAAATGTGCCAGCGATGAA	200
		C A T E N T K N G T V N V P A M K	
Tc00.1047053507957.300	144	GAGATGCATACAAGTTGCGATGAAGGGGATTTGTGACACTTTTATAATA	193
6xHis:MASP3	201	GAGATGCATACAAGTTGCGATGAAGGGGATTTGTGACACTTTTATAATA	250
		R C I Q V A M K G I C D T F Y N K	
Tc00.1047053507957.300	194	AGACGCCGAGTGAATTCATGACCCCTGAGGTTAAGGGTATGTGTACGTAT	243
6xHis:MASP3	251	AGACGCCGAGTGAATTCATGACCCCTGAGGTTAAGGGTATGTGTACGTAT	300
		T P S E I H D P E V K G M C T Y	
Tc00.1047053507957.300	244	TACGCTGCCAATCCAGATGAGCCGACGAGCCTCCAAACCCGCAAGGGCC	293
6xHis:MASP3	301	TACGCTGCCAATCCAGATGAGCCGACGAGCCTCCAAACCCGCAAGGGCC	350
		Y A A I P D E P D E P P T P Q G P	
Tc00.1047053507957.300	294	ACCACCAAGATTCAGCTGTCAGCAATACCAAGTAAATGAGGGGACGCCGA	343
6xHis:MASP3	351	ACCACCAAGATTCAGCTGTCAGCAATACCAAGTAAATGAGGGGACGCCGA	400
		P P D S A V S N T T S N E G T P K	
Tc00.1047053507957.300	344	AGAATGCGCCGGAAATCCGATGACGAGGGAGAGGAGAAGAAAAGAGAT	393
6xHis:MASP3	401	AGAATGCGCCGGAAATCCGATGACGAGGGAGAGGAGAAGAAAAGAGAT	450
		N A P E S D D A G R G E G K E D	
Tc00.1047053507957.300	394	GAAAAGGAGCACGGCAATACGAAGCAAAAAGCAGTCGAAAGCGCAGCTAT	443
6xHis:MASP3	451	GAAAAGGAGCACGGCAATACGAAGCAAAAAGCAGTCGAAAGCGCAGCTAT	500
		E K E H G N T K Q K A V E S A A M	
Tc00.1047053507957.300	444	GAAACATATCACGAAGACTGCCGACAGTGAC	474
6xHis:MASP3	501	GAAACATATCACGAAGACTGCCGACAGTGACTGA	534
		K H I T K T A D S D *	

Figura 19: Alinhamento entre a sequência nucleotídica codificadora central de Tc00.1047053507957.300 (MASP 3) e a sequência obtida para a construção de MASP 3 em pET28a-TEV. As bases idênticas estão unidas por traços e a mutação, por ponto. A sequência de resíduos de aminoácidos da tradução dos códons correspondentes foi alinhada abaixo das sequências nucleotídicas. Destacam-se os códons correspondentes à metionina inicial, em verde; à cauda de polihistidina, em amarelo; e ao sítio de clivagem por TEV, em cinza.

3.4. Expressão das proteínas MASP recombinantes

Para a expressão das proteínas de interesse foi utilizada a cepa BL-21(DE3) Star[®] (Invitrogen) de bactérias *E. coli*. Esta cepa possui uma mutação no gene *rne*, que codifica para uma RNase E truncada, sem a extremidade C-terminal necessária para a degradação de RNAs mensageiros, o que permite um nível maior de expressão de genes heterólogos.

Das oito construções planejadas inicialmente, foi realizada a expressão de sete MASPs recombinantes (Fig. 20).

Para as quatro construções em fusão com cauda de histidina, observa-se em gel de poliacrilamida-SDS, após indução da expressão por 3 horas, o aparecimento de uma banda com peso molecular aproximado de 30 kDa, ausente no controle negativo e presente na fração solúvel (Fig. 20: 1A, 2A, 3A, 4A). Para o controle negativo, bactérias BL-21 Star foram transformadas com o vetor pET28a-TEV sem inserto, cujo resultado da indução por 3 horas, correspondente a expressão da cauda de histidina seguida pelo sítio múltiplo de clonagem, não pôde ser visualizado em gel de poliacrilamida-SDS devido ao baixo peso molecular deste peptídeo (6076 Da).

Para as três construções em fusão com GST, observou-se no tempo de indução por 3 horas a expressão de proteínas com peso molecular entre 45-50 kDa, sendo 26 kDa correspondentes à GST. Destas, apenas a GST:MASP1 mostrou-se presente na fração solúvel (Fig. 20: 1B), enquanto GST:MASP3 e GST:MASP4 foram encontradas abundantemente na fração insolúvel (Fig. 20: 3B e 4B, respectivamente). Para o controle negativo, bactérias BL-21 Star foram transformadas com o vetor pET28a-GST-TEV sem inserto, cujo resultado da indução por 3 horas, correspondente à expressão da GST seguida pelo sítio múltiplo de clonagem (28849.4 Da), pôde ser visualizado em gel de poliacrilamida-SDS (Fig. 20: 1B, 3B e 4B).

Vale ressaltar que o peso molecular observado para as MASPs em fusão à cauda de histidina apresentou um acréscimo de aproximadamente 10 kDa, quando comparado ao peso molecular esperado (aproximadamente 20 kDa). Mesmo tendo sido visualizada tal discrepância com relação à mobilidade eletroforética destas proteínas, as análises de espectrometria de massa confirmaram o peso molecular predito para a _{6x}His:MASP 3, conforme está apresentado no item 3.5.

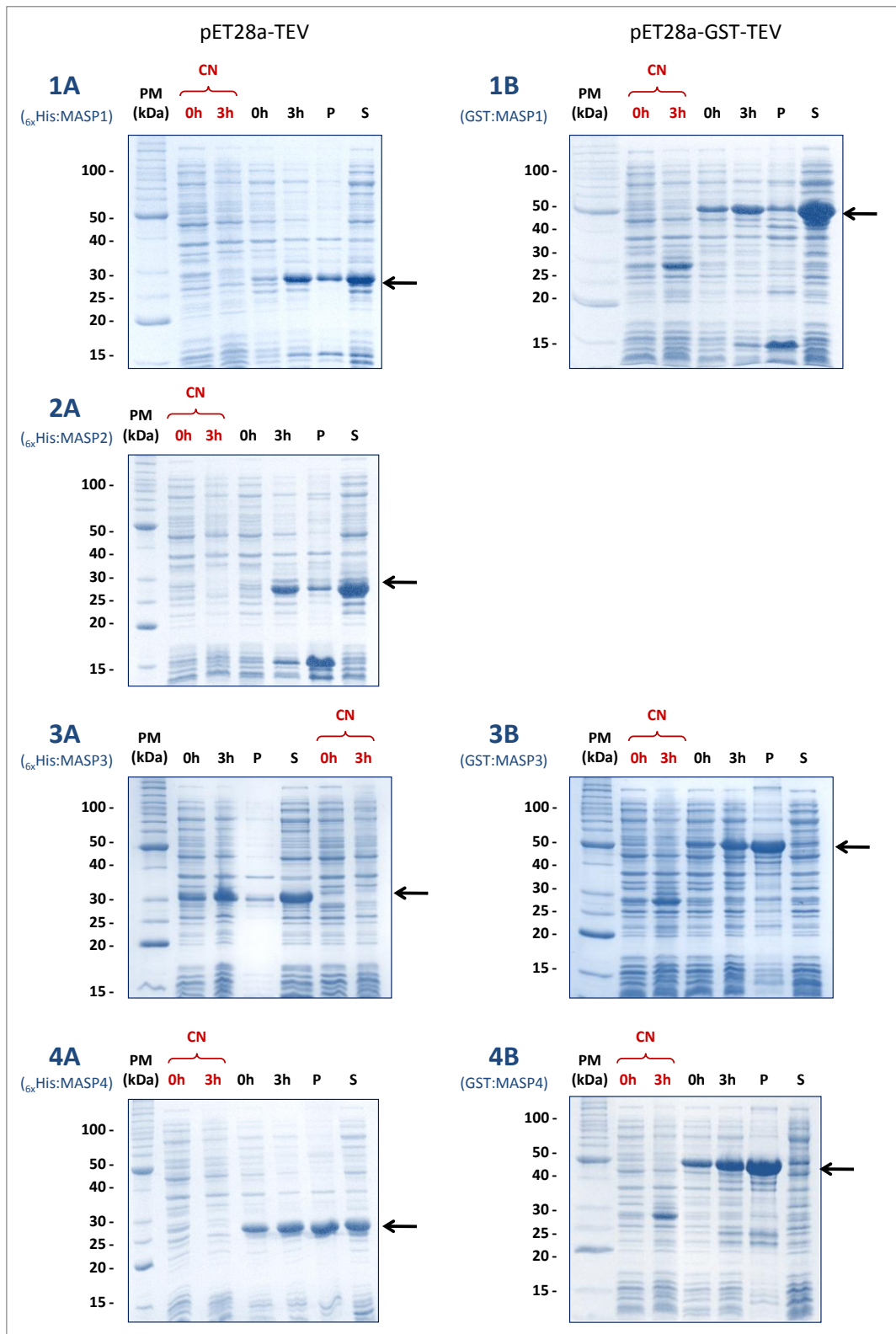


Figura 20: Perfil eletroforético de proteínas totais de bactérias *E. coli* BL-21 Star, transformadas com os vetores de expressão pET28a-TEV (A) e pET28a-GST-TEV (B), com ou sem inserto, antes (0h) e após (3h) indução da expressão por IPTG. P: *pellet* (fração insolúvel do teste de solubilidade). S: sobrenadante (fração solúvel do teste de solubilidade). As bandas referentes às MASPs 1, 2, 3 e 4 recombinantes estão indicadas pelas setas. As canaletas indicadas por 0h e 3h destacados em vermelho correspondem ao controle negativo (CN) da expressão, no qual bactérias *E. coli* BL-21 Star foram transformadas com vetor de expressão sem inserto. Géis de poliacrilamida-SDS 12% corados com azul de Coomassie. PM (kDa): peso molecular em kiloDaltons (Benchmark[®] Protein Ladder - Invitrogen).

Quanto mais solúvel for a proteína alvo presente no lisado celular, mais fácil espera-se obtê-la pura e, como resultado, melhores serão as análises estruturais. Tendo isto em mente, optou-se por prosseguir com as quatro MASPs em fusão com cauda de histidina.

Para confirmar a expressão destas quatro MASPs recombinantes, foi realizado ensaio de *western blotting* das frações correspondentes à indução por 3 horas (3h) e ao sobrenadante (S). As bandas de expressão de aproximadamente 30 kDa previamente identificadas em gel de poliacrilamida-SDS foram reconhecidas por anticorpo específico para a sequência de polihistidina (Fig. 21).

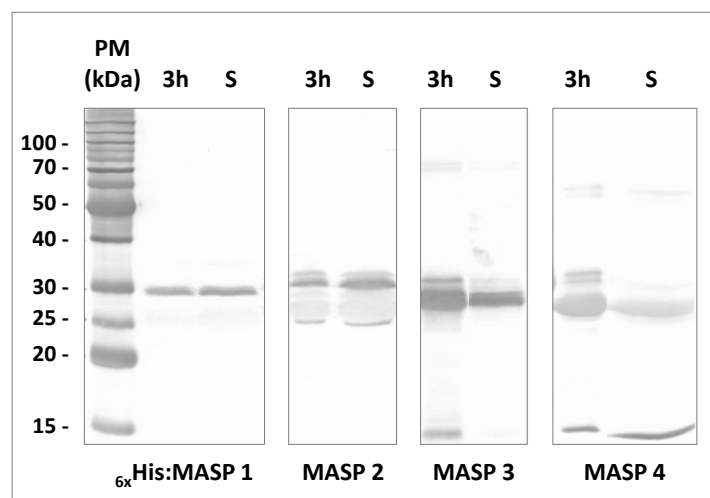


Figura 21: *Western blotting* das frações correspondentes à indução por 3 horas (3h) e ao sobrenadante (S) de MASPs 1, 2, 3 e 4 fusionadas a cauda de seis resíduos de histidina ($_{6x}$ His) N-terminal. A membrana de nitrocelulose foi incubada com anticorpo primário Penta-His (Qiagen), diluição 1:2000, e anticorpo secundário anti-IgG de camundongo conjugado a peroxidase, diluição 1:3000, seguida de revelação por DAB. PM (kDa): peso molecular em kiloDaltons (Benchmark[®] Protein Ladder - Invitrogen).

Tendo o projeto avançado até este ponto, novas sequências das construções de MASP 1, 2 e 4 em pET28a-TEV foram obtidas pelo NAGE (ICB/UFMG) e pela Macrogen (Seul, Coréia do Sul) utilizando os iniciadores T7 direto e reverso. As análises confirmaram a integridade das sequências codificadoras para a etiqueta de seis resíduos de histidina e o sítio de clivagem por TEV, bem como a janela de leitura correta para as três MASPs recombinantes. Entretanto, quanto às sequências gênicas clonadas, foi observada a ocorrência de substituições em alguns resíduos de aminoácidos em $_{6x}$ His:MASP1 e $_{6x}$ His:MASP2. Mesmo tendo sido sequenciada pela Macrogen uma única vez com um oligonucleotídeo iniciador, $_{6x}$ His:MASP1 apresentou a substituição N30Y na cadeia polipeptídica. Para $_{6x}$ His:MASP2, tanto as sequências obtidas pelo NAGE quanto pela Macrogen indicaram as três substituições E34G, A78T

e Q161R. No que se refere à $_{6x}$ His:MASP4, os dados de sequenciamento são ainda inconclusivos.

Frente a isto, os estudos com a construção $_{6x}$ His:MASP3 foram priorizados, já que a mesma foi a única que apresentou 100% de identidade da sequência de aminoácidos com aquela depositada em bancos de dados (item 3.3, Fig. 19).

3.5. Purificação da proteína MASP 3 recombinante e remoção da etiqueta de afinidade

A proteína MASP 3 fusionada a cauda de histidina N-terminal foi inicialmente submetida à cromatografia de afinidade em coluna de níquel-sepharose no sistema ÄKTAprime plus. A coluna utilizada His Trap HP consiste de uma matriz sólida de agarose com grupos quelantes imobilizados carregados com íons Ni^{2+} , os quais ligarão às histidinas expostas das proteínas e, com maior afinidade, às caudas de quatro ou mais resíduos de histidina.

Após equilíbrio da coluna com tampão de ligação contendo 30 mM de imidazol, o sobrenadante do lisado de 1L de cultura de *E. coli* (amostra em volume final de 15 mL), também no mesmo tampão, foi injetado na coluna. A eluição da proteína $_{6x}$ His:MASP3 foi feita em gradiente linear, o qual consistiu no aumento gradativo da concentração de imidazol de 30 mM para 500 mM (Fig. 22 A). O imidazol, composto que contém radical semelhante ao da histidina, atua como competidor pelos íons Ni^{2+} da resina favorecendo a liberação da proteína ligada. Submetidas à eletroforese em gel de poliacrilamida-SDS, as frações recolhidas confirmaram a presença da proteína $_{6x}$ His:MASP3 pela banda de aproximadamente 30 kDa (Fig. 22 B).

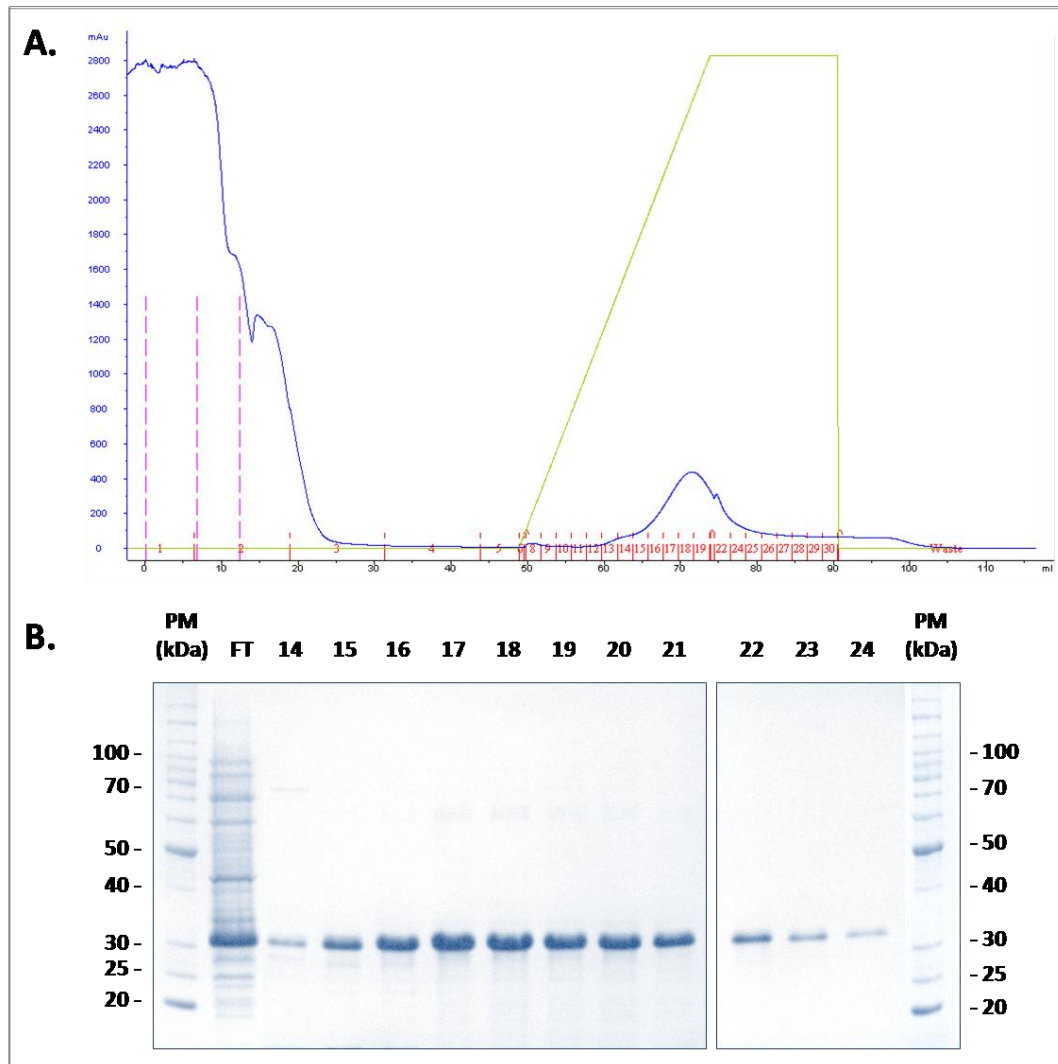


Figura 22: Purificação da proteína $_{6x}$ His:MASP3 por cromatografia de afinidade a níquel. (A) Cromatograma obtido pelo software ÄKTA PrimeView[®] (GE Healthcare Life Sciences). O lisado de 1L de cultura de *E. coli* foi carregado em coluna HisTrap HP 5mL (linha pontilhada rosa) utilizando o sistema ÄKTAprime plus. A eluição foi feita em gradiente linear de imidazol até 500 mM (em verde). (B). Gel de poliácridamida-SDS 12% corado com azul de Coomassie contendo as frações de 14 a 24 eluídas por gradiente de imidazol. FT: (*Flow-through*) Material recolhido após passagem pela coluna. PM (kDa): peso molecular em kiloDaltons (Benchmark[®] Protein Ladder - Invitrogen).

As frações de 15 a 21 foram reunidas e dialisadas contra tampão Tris-HCl 10mM pH 7,5. A leitura da absorbância a 280 nm revelou um total de aproximadamente 35 mg de amostra protéica, um rendimento equivalente a 35 μ g de proteína recombinante para cada 1 mL de meio de cultura.

Parte da amostra foi submetida à proteólise por TEV a 5%, 10% e 20% p/p durante 3 horas a 30°C para remoção da cauda de histidina. Os testes de clivagem para as três concentrações de TEV foram confirmados por espectrometria de massas.

Pelas análises dos perfis de massas obtidos, as três concentrações de TEV testadas apresentaram resultados semelhantes para a clivagem, uma vez que MASP 3

possui massa molecular esperada de 19262,2 Da quando fusionada a cauda de histidina e 17456,2 Da após a clivagem por TEV (Fig. 23). A concentração de 5% de TEV foi selecionada para a proteólise em maior quantidade.

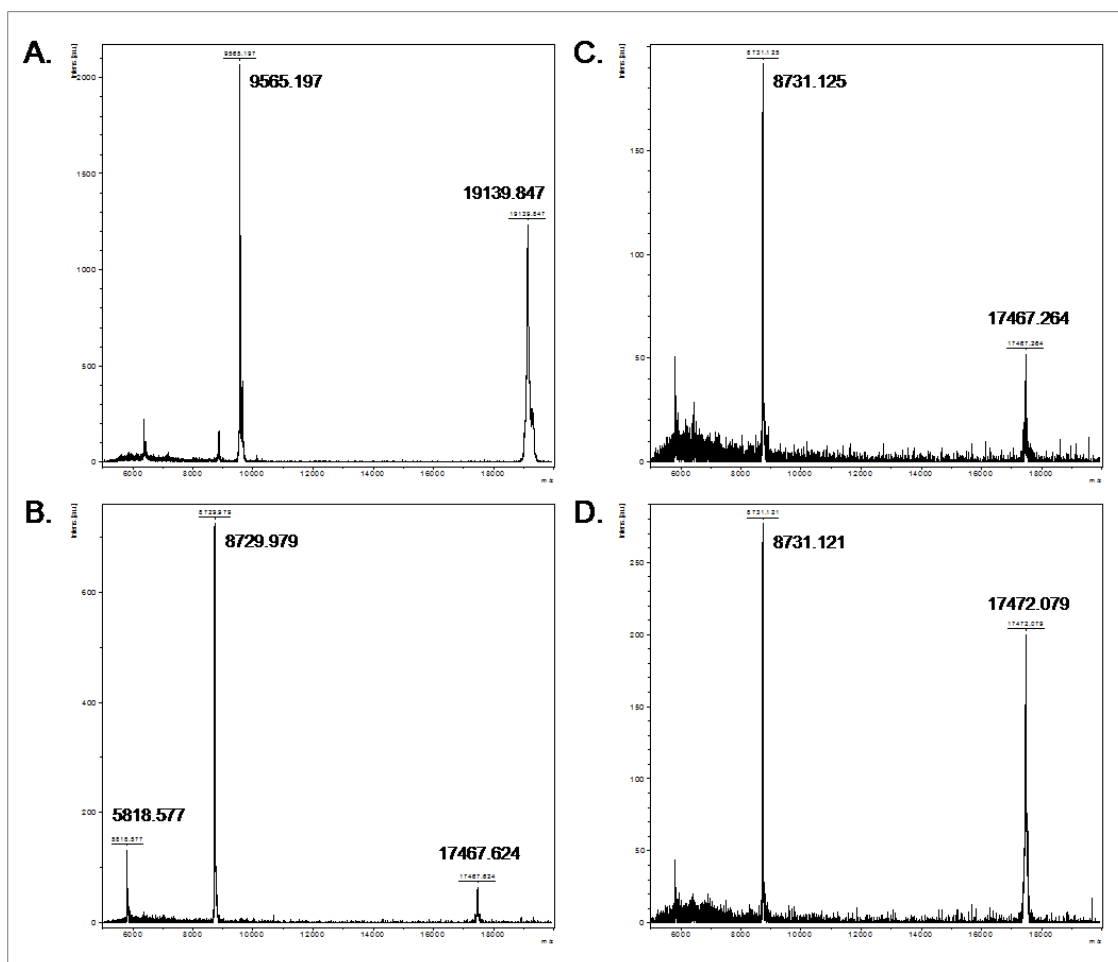


Figura 23: Perfis de massas de MASP 3, ainda em fusão a cauda de histidina (A) e após clivagem por TEV (B) 5%, (C) 10% e (D) 20% p/p durante 3 horas a 30°C.

Com o objetivo de separar a protease TEV e a MASP 3 recombinante, agora sem cauda de histidina, uma segunda etapa de purificação foi realizada por cromatografia de troca iônica em coluna RESOURCE Q de 1 mL utilizando o sistema ÄKTAprime plus, com tampão A contendo Tris-HCl 20 mM pH 7,5 e tampão B de mesma constituição acrescido de NaCl 1 M (Fig. 24).

A coluna utilizada, RESOURCE Q, consiste de uma matriz de poliestireno e divinilbenzeno com grupos ligantes, carregados positivamente, de amônio quaternário (Q), um trocador aniônico forte, pois mantém a sua carga em um largo intervalo de pH.

No pH escolhido para a cromatografia de troca iônica (pH 7,5), a carga líquida de MASP 3 recombinante clivada é negativa e de TEV, positiva, tendo em vista que o

ponto isoelétrico teórico de cada proteína corresponde, respectivamente, a 5,04 e 8,39. Nesse caso, MASP 3 é capaz de se ligar ao grupo trocador aniônico da coluna, podendo ser eluída com o aumento da força iônica por meio de NaCl presente no tampão de eluição B. De um total de aproximadamente 10 mg de proteína recombinante utilizada na proteólise por TEV, o rendimento obtido após esta etapa de cromatografia foi em torno de 15%. A proteína foi obtida em dois picos principais (Fig. 24 A), sendo que parte foi eliminada juntamente com a protease TEV (Fig. 24 B).

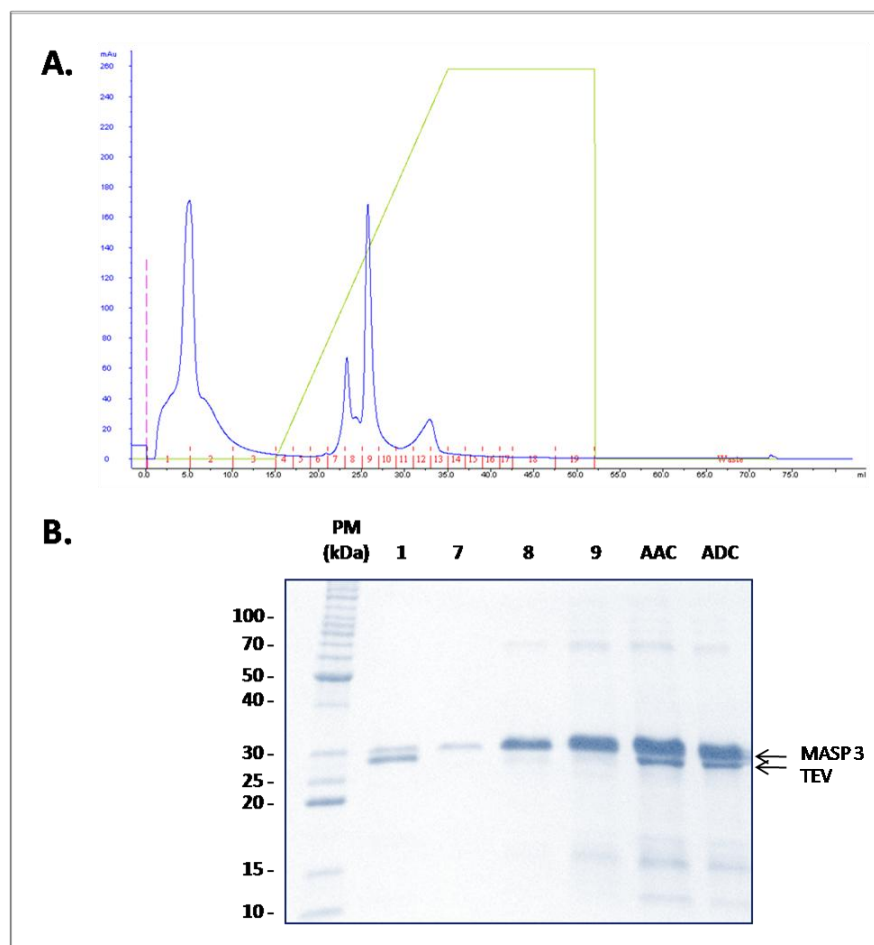


Figura 24: Purificação da proteína MASP3 por cromatografia de troca iônica. (A) Cromatograma obtido pelo software ÄKTA PrimeView® (GE Healthcare Life Sciences). A amostra de 10 mg de proteína recombinante, proveniente da proteólise da cauda de histidina por TEV 5%, foi carregada em coluna RESOURCE Q 1mL utilizando o sistema ÄKTAprime plus. A eluição foi feita em gradiente linear de 1 M de NaCl. (B). Gel de poliacrilamida-SDS 12% corado com azul de Coomassie contendo as frações: 1 (coletada durante lavagem da coluna) e 7 a 9 (coletadas durante a eluição). A protease TEV (28,3 kDa) não se liga à coluna e é liberada logo após a injeção da amostra (fração 1). A MASP 3 sem cauda de histidina foi eluída nas frações 7, 8 e 9. AAC – amostra antes da clivagem. ADC – amostra depois da clivagem por TEV. PM (kDa): peso molecular em kiloDaltons (Benchmark® Protein Ladder - Invitrogen).

4. DISCUSSÃO

A elucidação de estruturas tridimensionais de macromoléculas biológicas pelas técnicas de cristalografia de raios-X e espectroscopia de ressonância magnética nuclear tem contribuído expressivamente para a melhor compreensão de sistemas biológicos em nível molecular. Análises precisas quanto aos mecanismos de ação e interação entre proteínas e seus ligantes tornam-se possíveis, favorecendo estudos mais avançados como o desenho racional de fármacos.

Somado a isto, com o advento dos projetos genoma, uma gama de informações tem sido disponibilizada, bem como um suprimento ilimitado de sequências para análise e exploração de genes e de seus produtos protéicos correspondentes.

Análises comparativas dos genomas dos três tripanosomatídeos *Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major* representam uma oportunidade única para a identificação de novos alvos moleculares e o desenvolvimento de métodos terapêuticos ainda não explorados. Drogas desenhadas especificamente para conjuntos moleculares conservados nos três organismos, porém ausentes nos hospedeiros mamíferos, têm a vantagem de serem potencialmente úteis no combate aos três parasitos, ao mesmo tempo em que não apresentam, aparentemente, riscos de toxicidade ao homem (EL-SAYED *et al.*, 2005b).

Além disso, as informações disponibilizadas pelos projetos genoma dos *Trityps* têm contribuído para o conhecimento da arquitetura molecular das superfícies celulares dos três parasitos e para as bases genéticas e evolutivas que os conduziram a explorarem diferentes estratégias de sobrevivência e de evasão do sistema imune como consequência da adaptabilidade a diferentes nichos em seus respectivos hospedeiros.

Em *T. cruzi*, a grande expansão de famílias multigênicas que codificam para proteínas de superfície constitui uma diversidade molecular que pode muito bem ser explorada pelo parasito em mecanismos de escape do sistema imune e interação com células do hospedeiro. Além do extenso repertório de proteínas de superfície altamente polimórficas, o *T. cruzi* combina sequências de diferentes famílias gênicas para aumentar a diversidade. Bartholomeu e colaboradores (2009) sugerem, por exemplo, que a organização genômica dos genes *masp* distribuídos em *clusters* com genes de outras famílias de proteínas de superfície seja um fator contribuinte para a manutenção da diversidade da família MASP ao evitar a homogeneização das sequências. Assim,

eventos de recombinação podem gerar um repertório ilimitado de proteínas do parasito expostas ao hospedeiro (BARTHOLOMEU *et al.*, 2009), conduzindo o sistema imune a uma série de respostas espúrias e ineficientes e, portanto, ao atraso de uma resposta protetora efetiva na fase inicial da infecção (TARLETON, 2007).

A massiva expansão de genes *masp* no genoma de *T. cruzi* sugere que esta família desempenhe um papel crítico para a sobrevivência do parasito. Sua expressão nas formas tripomastigotas infectivas circulantes, a extensa variabilidade de sequência, a natureza repetitiva e a localização na superfície do parasito sugerem a participação de MASP em interações parasito-hospedeiro (BARTHOLOMEU *et al.*, 2009). Embora não haja uma associação clara entre um perfil específico de moléculas de superfície e a habilidade do parasito de infectar e se replicar numa determinada célula hospedeira, Macedo e colaboradores (2004) indicam o polimorfismo genético das glicoproteínas de superfície de *T. cruzi* como um fator importante que contribua para tal fenômeno (BARTHOLOMEU *et al.*, 2009). É possível que a extensa variabilidade de MASP possa contribuir com um grande repertório de peptídeos capazes de interagir com diferentes receptores de vários tipos de células hospedeiras (BARTHOLOMEU *et al.*, 2009).

Estudos funcionais da família MASP têm sido conduzidos no Laboratório de Genômica de Parasitos, situado no Departamento de Parasitologia (ICB/UFMG) e coordenado pela prof^a Daniella Castanheira Bartholomeu. Acredita-se que, aliados a estes, estudos estruturais de alguns membros da família constituem uma informação valiosa para o entendimento mais detalhado destas moléculas. Através da caracterização estrutural, pode-se determinar a localização de epítomos e o tipo de enovelamento protéico, o que pode permitir a dedução de domínios funcionais e, possivelmente, sugerir algo sobre sua função.

Em vista disso, quatro membros da família MASP, identificados como MASP 1, 2, 3 e 4, foram selecionados para caracterização estrutural.

Análises *in silico* quanto à presença de regiões de desordem estrutural nas quatro cadeias polipeptídicas completas revelaram porções desordenadas nas quatro sequências, mais extensas, porém, nas MASPs 1 e 2. Gráficos de carga líquida *versus* hidrofobicidade localizaram a região central das quatro MASPs dentro da região designada para proteínas desenoveladas. A combinação de baixa hidrofobicidade e carga líquida relativamente alta representa um importante pré-requisito para a ausência de estrutura compacta em proteínas sob condições fisiológicas, conduzindo a proteínas

nativamente desenoveladas (UVERSKY *et al.*, 2000). Além disso, a maioria das proteínas intrinsecamente desordenadas é enriquecida pelos resíduos de aminoácidos Glu, Lys, Arg, Gly, Gln, Ser, Pro e Ala (UVERSKY, 2002). Como se sabe, a região central de MASP é altamente variável e repetitiva. Repetições de resíduos únicos são mais comuns, sendo as repetições de resíduos de ácido glutâmico mais frequentes, seguidas em ordem decrescente pelos motivos ricos em Gly, Ala, Pro e Ser, principalmente (BARTHOLOMEU *et al.*, 2009).

A expressão de proteínas com sequências repetitivas é um fenômeno comum observado em parasitos como forma de escape do ataque do sistema imune do hospedeiro (BARTHOLOMEU *et al.*, 2009). A infecção por *T. cruzi* induz uma forte produção de anticorpos contra várias proteínas contendo sequências repetitivas (HOFT; ALVAREZ; BUSCAGLIA; DAROCHA; PAIS *apud* BARTHOLOMEU *et al.*, 2009).

Tem sido relatado que os genomas de alguns protozoários patogênicos codificam para proteínas com regiões de baixa complexidade de sequência, as quais são tipicamente um indicativo da presença de desordem intrínseca (FENG *et al.*, 2006; MOHAN *et al.*, 2008). É interessante observar que tal característica não é exclusiva de *T. cruzi*, mas, sim, compartilhada por diversos parasitos talvez como estratégia de sobrevivência em seus hospedeiros.

No que diz respeito à possibilidade de MASPs apresentarem regiões desordenadas, pode-se sugerir, por exemplo, que a falta de estrutura globular rígida para os membros desta família represente uma vantagem funcional considerável, à medida que sua plasticidade lhes permita interagir eficientemente com diferentes moléculas do hospedeiro. Diante disso, interações promíscuas poderiam favorecer a sobrevivência do parasito ao facilitar interações com moléculas necessárias para adesão e invasão das células hospedeiras, além de poderem inibir a geração de uma resposta efetiva de anticorpos de alta afinidade. Também, uma resposta humoral baseada na especificidade a muitas das conformações alternativas da molécula poderia ter sua efetividade reduzida.

Vale ressaltar, contudo, que as análises para identificação de regiões desordenadas nas MASPs correspondem a predições e, portanto, necessitam de comprovação experimental.

Dessa forma, para que os quatro membros selecionados da família MASP sejam caracterizados estruturalmente, grandes quantidades de proteína purificada, mais que 5mg, são necessárias. Para isso, o uso de proteínas recombinantes torna-se imperativo.

As vantagens de utilizar proteínas recombinantes tipicamente incluem, entre outras, maior rendimento na obtenção da proteína, facilidade em purificar por meio de etiquetas de afinidade, possibilidade de modificações na cadeia polipeptídica ou uso de domínios isolados, e facilidade de marcação com átomos pesados ou isótopos para ajudar na caracterização das estruturas tridimensionais destas macromoléculas em resolução atômica.

O primeiro passo para a produção de proteínas recombinantes consiste na obtenção de um clone de expressão do gene de interesse. Bactérias *Escherichia coli* são as células hospedeiras mais largamente utilizadas para a produção de proteínas heterólogas (GOULDING & PERRY, 2003). Além disso, a expressão de proteínas recombinantes em bactérias pode ser fortemente regulada, uma vez que vetores do sistema pET (Novagen) contém o promotor T7 *lac*. Estes plasmídeos também possuem o promotor e a sequência codificadora para o repressor *lac* (*lacI*) que, em lisógenos DE3, atua no cromossomo bacteriano de modo a reprimir a transcrição do gene da enzima T7 RNA polimerase pela célula hospedeira, além de bloquear a transcrição do gene alvo ao agir sobre o promotor T7 *lac* presente no vetor. A expressão basal em níveis muito baixos é alcançada, enquanto níveis elevados de expressão são obtidos após indução (DUBENDORFF & STUDIER, 1991; NOVAGEN, 2005).

Diante disso, os produtos de PCR das regiões centrais codificadoras das quatro MASPs foram inicialmente subclonados em vetores de sequenciamento para posterior clonagem nos vetores de expressão pET28a-TEV e pET28a-GST-TEV, que possibilitam a expressão de proteínas recombinantes em fusão, respectivamente, a cauda de histidina e a GST na extremidade N-terminal (CARNEIRO *et al.*, 2006). Algumas análises de sequenciamento permaneceram inconclusivas e outras revelaram mutações, além da identificação de um quinto membro MASP clonado. Uma provável causa para as mutações pode ter sido a utilização de uma DNA polimerase (Taq DNA polimerase, Phoneutria, Belo Horizonte, Brasil) sem atividade revisora de erros. Quanto à clonagem de um quinto membro MASP (Tc00.1047053510359.550), a sua amplificação foi possível mediante a utilização de iniciadores sintetizados para Tc00.1047053506757.50 devido ao grau de similaridade entre as extremidades das regiões centrais codificadoras.

A expressão da $_{6x}\text{His}:\text{MASP3}$, abundantemente presente na fração solúvel, foi confirmada por ensaio de *western blotting* utilizando anticorpo específico para a cauda de polihistidina. A discrepância de mobilidade eletroforética entre o peso molecular predito de $_{6x}\text{His}:\text{MASP3}$ (19,3 kDa), o qual foi posteriormente confirmado por

espectrometria de massas, e o visualizado em gel de poliacrilamida-SDS (aproximadamente 30 kDa) também foi observada para uma proteína repetitiva de *T. cruzi* (MACÊDO *et al.*, 2006). Segundo Banker e Cotman (1972), algumas proteínas não migram em géis de poliacrilamida-SDS de acordo com seu peso molecular real. Tais proteínas apresentam valores incomuns de mobilidade eletroforética e de coeficiente de retardamento que podem ser explicados com base na interação anômala com o SDS, conformação atípica dos complexos proteína-SDS, ou propriedades incomuns da proteína nativa que são mantidas em solução com SDS. Cada um desses fatores pode afetar o comportamento de proteínas em géis de poliacrilamida-SDS (BANKER & COTMAN, 1972).

A mini-expressão das outras construções revelou que as MASPs 1, 2 e 4 em fusão às histidinas também apresentaram-se nas frações solúveis, sendo posteriormente reconhecidas por anticorpo específico para a cauda de histidina em ensaio de *western blotting*. Quanto às construções das MASPs 1, 3 e 4 em fusão com GST, apenas MASP 1 esteve solúvel.

Normalmente, a expressão de proteínas solúveis corresponde ao primeiro gargalo em todo o processo de obtenção de proteínas recombinantes para estudos estruturais. Várias estratégias são abordadas com o objetivo de superar este obstáculo (SØRENSEN & MORTENSEN, 2005). Mesmo que em alguns casos seja suficiente alterar as condições do protocolo de expressão, as etiquetas de afinidade têm sido atualmente as melhores ferramentas para aumentar a solubilidade de proteínas recombinantes (ESPOSITO & CHATTERJEE, 2006). Dentre as proteínas de fusão que mais favorecem a solubilidade de suas parceiras está a MBP, proteína de ligação a maltose, até muito mais efetiva que a própria GST (KAPUST & WAUGH, 1999). Dessa forma, uma boa opção para trabalhos posteriores a serem desenvolvidos no Grupo de Biologia Estrutural (BioEst) é o uso de vetores de expressão em fusão com a proteína MBP na tentativa de obter as recombinantes solúveis.

Frente aos resultados obtidos, optou-se por seguir adiante apenas com a construção $_{6x}$ His:MASP3, devido ao sequenciamento inconclusivo e à presença de mutações nas outras sequências polipeptídicas, bem como às limitações de tempo impostas para a realização deste trabalho.

Inicialmente, foi aplicada a cromatografia de afinidade para a purificação da $_{6x}$ His:MASP3. Proteínas fusionadas a cauda de histidina permitem que sua purificação por cromatografia de afinidade seja relativamente fácil (DEREWENDA, 2004). Esta

técnica é frequentemente utilizada como uma primeira etapa em protocolos de purificação, mais precisamente como uma etapa de captura da proteína de interesse, a qual é obtida de forma purificada e concentrada. Após esta etapa, foi obtido um total de aproximadamente 35 mg de proteína para 1L de expressão em *E. coli* BL-21. Tal rendimento é suficiente, por exemplo, para estudos de cristalografia de raios-X, tendo em vista que 3 mg de proteína já permite testar até 600 condições de cristalização.

No que diz respeito à manutenção da proteína em fusão à cauda de histidina, esta não constitui, em uma boa parcela dos casos, um obstáculo para o correto enovelamento da proteína nem para o sucesso da etapa de cristalização, visto que não aumenta substancialmente o tamanho da proteína e geralmente não apresenta efeito significativo em sua estrutura (DEREWENDA, 2004; CARSON *et al.*, 2007). Diante disso, pode-se optar por prosseguir os estudos de caracterização estrutural com a proteína recombinante ainda em fusão à cauda de histidina. Opcionalmente, esta, ou qualquer etiqueta de afinidade, pode ser removida por uma protease, sendo necessária outra etapa de purificação. Neste trabalho, os vetores de expressão utilizados adicionam nas construções recombinantes o sítio de clivagem por TEV logo após a etiqueta de afinidade N-terminal (CARNEIRO *et al.*, 2006). A protease TEV se liga à sequência Glu-Asn-Leu-Tyr-Phe-Gln-Gly. Após a clivagem, apenas o resíduo de glicina permanece na extremidade N-terminal da proteína de interesse. Além disso, a TEV utilizada por nós também é uma proteína recombinante fusionada a cauda de histidina N-terminal (BLOMMEL & FOX, 2007), característica que pode ser explorada em vias de facilitar a separação desta e da proteína de interesse, após a clivagem da etiqueta, por um segundo processo de cromatografia de afinidade em coluna de níquel.

Pelas análises de espectrometria de massas, confirmou-se a clivagem da cauda de histidina após incubação de $_{6x}$ His:MASP3 com TEV a 5%, 10% e 20% (p/p) por 3 horas a 30°C. Tendo em vista a combinação de bom rendimento e uso racional de reagentes, optou-se por realizar a clivagem em maior escala com TEV a 5% p/p, seguida de cromatografia de troca iônica. Foi observado que o rendimento final nesta etapa foi de apenas 15%.

Pode-se dizer que a combinação de rendimento e pureza da amostra protéica consiste em outro gargalo do processo. Em alguns casos, a pureza da amostra protéica só é conseguida em detrimento do rendimento. Uma amostra altamente pura, mas cujo rendimento foi inferior a 25%, por exemplo, dependendo do objetivo do trabalho é muito mais desejável que 75% de rendimento e alguns contaminantes. Como neste

trabalho não se sabe se a perda de proteína está relacionada à etapa de clivagem ou à troca iônica, uma possível alternativa consiste em realizar uma segunda cromatografia de afinidade em coluna de níquel após a proteólise por TEV, conforme mencionado anteriormente, ao invés da troca iônica.

Quanto a obtenção da MASP 3 em dois picos de eluição pela cromatografia de troca iônica, pode-se sugerir, por exemplo, que isto se deve ao estado de oligomerização da proteína. Uma análise das duas amostras em gel de poliacrilamida nativo poderia revelar quanto à presença de complexos multiméricos. Medições de espalhamento dinâmico da luz (*Dynamic Light Scattering - DLS*) poderiam ser realizadas a fim de investigar o estado de agregação protéico sem ocasionar a perda da amostra. Moléculas de proteína monodispersas em solução constituem em um bom ponto de partida para a cristalização (FERRÉ-D'AMARÉ & BURLEY, 1994), apesar de não ser o único requerimento para isto. Ao contrário, a polidispersividade pode impedir a formação de cristais.

Como continuidade deste projeto, será necessário concentrar a amostra para dar início às condições de cristalização. Para os testes, a concentração da proteína utilizada deve ser a mais alta possível (até 10 mg/mL), além de contarem também pureza e homogeneidade da amostra. Medições de espalhamento dinâmico da luz podem ser conduzidas durante o passo de concentração, não apenas para verificar a dispersividade da proteína em solução, mas também para determinar a concentração máxima limite acima da qual agregados protéicos começam a se formar.

As técnicas espectroscópicas de dicroísmo circular e H^1 -RMN podem ser realizadas para revelar se a proteína encontra-se enovelada, e, logo, se regiões desordenadas estão presentes ou ausentes, uma vez que a presença de regiões flexíveis na estrutura de uma proteína pode interferir no empacotamento ordenado que ocorre durante a cristalização.

Ensaio de proteólise limitada também podem ser conduzidos a fim de identificar regiões mais compactas e estáveis. A clivagem proteolítica limitada e específica ocorre normalmente em regiões flexíveis e acessíveis ao solvente, como N- e C- terminais desordenados, *loops* expostos ou regiões de ligação entre domínios globulares. Intermediários estáveis gerados por proteólise podem ser analisados por espectrometria de massas através das técnicas de MALDI (*matrix-assisted laser desorption ionization*) ou ESI (*electrospray ionization*) e podem ser mais fáceis de serem cristalizados (GAO *et al.*, 2005).

5. CONCLUSÕES

Em análises *in silico*, regiões de desordem estrutural foram previstas ao longo das cadeias polipeptídicas completas das quatro MASPs selecionadas. Em gráficos de carga líquida *versus* hidrofobicidade, as quatro regiões centrais foram posicionadas dentro do espaço designado para proteínas nativamente desenoveladas.

Com relação à metodologia utilizada, a fusão a cauda histidina parece ter favorecido a solubilidade das quatro proteínas MASP recombinantes quando comparada à fusão com GST. A primeira etapa cromatográfica (afinidade) foi suficiente para a obtenção de grande quantidade de proteína pura. Todavia, quanto à segunda etapa (troca iônica), a separação da proteína de interesse e da protease ocorreu em detrimento do rendimento.

REFERÊNCIAS BIBLIOGRÁFICAS

ACOSTA-SERRANO, A; ALMEIDA I. C; FREITAS-JUNIOR L. H, YOSHIDA N; SERGIO SCHENKMAN, S. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. **Molecular and Biochemical Parasitology**, v. 114, p. 143-150, 2001.

ACOSTA-SERRANO, A; HUTCHINSON, C; NAKAYASU, E. S; ALMEIDA I. C; CARRINGTON, M. Comparison and evolution of the surface architecture of trypanosomatid parasites. In: BARRY, D; MCCULLOCH, R; MOTTRAM, J; ACOSTA-SERRANO, A. (eds.). **Trypanosomes after the genome**. Horizon Bioscience, Norfolk, 2007. Cap. 10, p. 319-337.

AMAYA, M. F; WATTS, A. G; DAMEGER, I; WEHENKEL, A; NGUYEN, T; BUSCHIAZZO, A; PARIS, G; FRASCH, A. C; WITHERS, S. G; ALZARI, P. M. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. **Structure**, v. 12, p. 775-784, 2004.

ANDRADE, A. L; ZICKER, F; DE OLIVEIRA, R. M; ALMEIDA SILVA, S; LUQUETTI, A; TRAVASSOS, L. R; ALMEIDA, I. C; DE ANDRADE, S. S; DE ANDRADE, J. G; MARTELLI, C. M. Randomized trial of efficacy of benznidazole in treatment of early *Trypanosoma cruzi* infection. **The Lancet**, v. 348, p. 1407-1413, 1996.

ASLETT, M; AURRECOECHEA, C; BERRIMAN, M; BRESTELLI, J; BRUNK, B. P. *et al.*. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic Acids Research**, v. 38, p. 457-462, 2009.

ATWOOD, J. A; WEATHERLY, D. B; MINNING, T. A; BUNDY, B; CAVOLA, C; OPPERDOES, F. R; ORLANDO, R; TARLETON, R. L. The *Trypanosoma cruzi* proteome. **Science**, v. 309, p. 473-476, 2005.

AUSUBEL, F. M; BRENT R; KINGSTON R. E; MOORE, D.D; SEIDMAN J. G; SMITH J. A; STRUHL K. **Current Protocols in Molecular Biology**, Greene Publishing Associates/Wiley Interscience, New York. 1987.

BARTHOLOMEU, D. C; CERQUEIRA, G. C; LEÃO, A. C. A; DAROCHA, W. D; PAIS, F. S; MACEDO, C; DJIKENG, A; TEIXEIRA, S. M. R; EL-SAYED, N. M. Genomic organization and expression profile of the mucin-associated surface protein (*masp*) family of the human pathogen *Trypanosoma cruzi*. **Nucleic Acids Research**, v. 37, n. 10, p. 3407-3417, 2009.

BENDTSEN, J. D; NIELSEN, H; VON, H. G; BRUNAK, S. Improved prediction of signal peptides: SignalP 3.0. **Journal of Molecular Biology**, v. 340, p. 783-795, 2004.

BERMAN, H. M; WESTBROOK, J; FENG, Z; GILLILAND, G; BHAT, T. N; WEISSIG, H; SHINDYALOV, I. N. The Protein Data Bank. **Nucleic Acids Research**, v. 28, p. 235-242, 2000.

- BERRIMAN, M; GHEDIN, E; HERTZ-FOWLER, C; BLANDIN, G; RENAULD, H. *et al.*. The genome of the african trypanosome *Trypanosoma brucei*. **Science**, v. 309. n. 5733, p. 416-422, 2005.
- BILLETERA, M; WAGNERB, G; WÜTHRICH, K. Solution NMR structure determination of proteins revisited. **Journal of Biomolecular NMR**, v. 42, p. 155-158, 2008.
- BLOMMEL, P. G. & FOX, B. G. A combined approach to improving large-scale production of tobacco etch virus protease. **Protein Expression and Purification**, v. 55, p. 53-68, 2007.
- BLUM, M. L; DOWN, J. A; GURNETT, A. M; CARRINGTON, M; TURNER, M. J; WILEY, D. C. A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. **Nature**, v. 362, p. 603-609, 1993.
- BORST, P. Antigenic variation and allelic exclusion. **Cell**, v. 109, p. 5-8, 2002.
- BRANDEN, C. & TOOZE, J. Determination of protein structures. In: BRANDEN, C; TOOZE, J. **Introduction to protein structure**. 2nd ed, Garland Publishing, New York, 1999. Cap. 18, p. 373-392.
- BRENER, Z. Biology of *Trypanosoma cruzi*. **Annual Review of Microbiology**, v. 27, p. 347-382, 1973.
- BUSCAGLIA, C. A; CAMPO, V. A; FRASCH, A. C. C; DI NOIA, J. M. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. **Nature Reviews Microbiology**, v. 4, p. 229-236, 2006.
- BUSCHIAZZO, A; AMAYA, M. F; CREMONA, M. L; FRASCH, A. C; ALZARI, P. M. The crystal structure and mode of action of *trans*-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. **Molecular Cell**, v. 10, p. 757-768, 2002.
- CANÇADO, J. R. Criteria of Chagas disease cure. **Memórias do Instituto Oswaldo Cruz**, v. 94, p. 331-336, 1999.
- CARNEIRO, F. R. G; SILVA, T. C. L; ALVES, A. C; HALINE-VAZ, T; GOZZO, F. C; ZANCHIN, N. I. T. Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. **Biochemical and Biophysical Research Communications**, v. 343, n. 1, p. 260-268, 2006.
- CARSON, M; JOHNSON, D. H; MCDONALD, H; BROUILLETTE, C; DELUCAS, L. J. His-tag impact on structure. **Acta Crystallographica Section D**, v. 63, p. 295-301, 2007.
- CAZZULO, J. J. & FRASCH, A. C. C. SAPA/*trans*-sialidase and cruzipain: two antigens from *Trypanosoma cruzi* contain immunodominant but enzymatically inactive domains. **The FASEB Journal**, v. 6, p. 3259-3264, 1992

CHAGAS, C. Nova tripanosomíase humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade mórbida do homem. **Memórias do Instituto Oswaldo Cruz**, v. 1, p. 59-218, 1909.

COURA, J. R. Chagas disease: what is known and what is needed – A background article. **Memórias do Instituto Oswaldo Cruz**, v. 102, p. 113-122, 2007.

COURA, J. R. & DE CASTRO, S. L. A critical review on Chagas disease chemotherapy. **Memórias do Instituto Oswaldo Cruz**, v. 97, p. 3-24, 2002.

COURA, J. R. & DIAS, J. C. P. Epidemiology, control and surveillance of Chagas disease – 100 years after its discovery. **Memórias do Instituto Oswaldo Cruz**, v. 104, p. 31-40, 2009.

CREMONA, M. L; SÁNCHEZ, D. O; FRASCH, A. C. C; CAMPETELLA, O. A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. **Gene**, v. 160, p. 123-128, 1995.

DELANO, W. L. The PyMOL Molecular Graphics System, 2002. Disponível em: <http://www.pymol.org>

DE SOUZA, W. O parasito e sua interação com os hospedeiros. In: BRENER, Z; ANDRADE, Z. A; BARRAL-NETTO, M. (eds.). **Trypanosoma cruzi e Doença de Chagas**. 2 ed. Ed. Guanabara Koogan, Rio de Janeiro, 2000. p. 88-126.

DEREWENDA, Z. S. The use of recombinant methods and molecular engineering in protein crystallization. **Methods**, v. 34, p. 354-363, 2004.

DIAS, J. C. P. Elimination of Chagas disease transmission: perspectives. **Memórias do Instituto Oswaldo Cruz**, v. 104, p. 41-45, 2009.

DIAS, J.C. P. & COURA, J. R. Epidemiologia. In: DIAS, J. C. P. & COURA, J. R. (eds.). **Clínica e terapêutica da doença de Chagas. Uma abordagem prática para o clínico geral**. Rio de Janeiro, Fiocruz Editora, p 33-36, 1997.

DUBENDORFF, J. W. & STUDIER, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. **Journal of Molecular Biology**, v. 219, n. 1, p. 45-59, 1991.

DURBIN, S. D. & FEHER, G. Protein crystallization. **Annual Review of Physical Chemistry**, v. 47, p. 171-204, 1996.

EL-SAYED, N. M; MYLER, P. J; BARTHOLOMEU, D. C; NILSSON, D; AGGARWAL, G. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309. n. 5733, p. 409-415, 2005a.

EL-SAYED, N. M; MYLER, P. J; BLANDIN, G; BERRIMAN, M; CRABTREE, J *et al.* Comparative genomics of trypanosomatid parasitic protozoa. **Science**, v. 309. n. 5733, p. 404-409, 2005b.

- ESPOSITO, D. & CHATTERJEE, D. K. Enhancement of soluble protein expression through the use of fusion tags. **Current Opinion in Biotechnology**, v 17, p. 353-358, 2006.
- FANKHAUSER, N. & MASER, P. Identification of GPI anchor attachment signals by a Kohonen self-organization map. **Bioinformatics**, v. 21, p. 1846-1852, 2005.
- FENG, Z. P; Zhang, X; Han, P; Arora, N; Anders, R. F; Norton; R. S. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. **Molecular and Biochemical Parasitology**, v. 150, n 2, p. 256-267, 2006.
- FERRÉ-D'AMARÉ, A. R. & BURLEY, S. K. Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. **Structure**, 2: 357-359, 1994.
- FOSTER, M. P; MCELROY, A C; AMERO, C. D. Solution NMR of large molecules and assemblies. **Biochemistry**, v. 46, p. 331-340, 2007.
- FRASCH, A. C. C. Functional diversity in the *trans*-sialidase and mucin families in *Trypanosoma cruzi*. **Parasitology Today**, v. 16, p. 282-286, 2000.
- FREYMAN, D; DOWN, J; CARRINGTON, M; RODITI, I; TURNER, M; WILEY, D. 2.9 Å resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. **Journal of Molecular Biology**, v. 216, p. 141-160, 1990.
- GAO, X., BAIN, K., BONANNO, J. B., BUCHANAN, M., HENDERSON, D., LORIMER, D., MARSH, C., REYNES, J. A., SAUDER, J. M., SCHWINN, K., THAI, C., BURLEY, S. K. High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. **Journal of Structural and Functional Genomics**, 6: 129-134, 2005.
- GARCIA, E. S & AZAMBUJA, P. Fisiologia de triatomíneos: desenvolvimento, reprodução e interação com *Trypanosoma cruzi*. In: BRENER, Z; ANDRADE, Z. A; BARRAL-NETTO, M. (eds.) *Trypanosoma cruzi e Doença de Chagas*. 2 ed. Ed. Guanabara Koogan, Rio de Janeiro, 2000.
- GARDNER, K. H & KAY, L. E. The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. **Annual Review of Biophysics and Biomolecular Structure**, v. 27, p. 357-406, 1998.
- GASPAR, E. B; MORTARA, R. A; ANDRADE, L. O; DA SILVA, C. V. Lysosomal exocytosis: an important event during invasion of lamp deficient cells by extracellular amastigotes of *Trypanosoma cruzi*. **Biochemical and Biophysical Research Communications**, v. 384, p. 265-269, 2009.
- GASTEIGER, E; GATTIKER, A; HOOGLAND, C; IVANYI, I; APPEL, R. D; BAIROCH, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, v. 31, n. 13, p. 3784-3788, 2003.

GASTEIGER, E; HOOGLAND, C; GATTIKER, A; DUVAUD, S; WILKINS, M. R; APPEL, R. D; BAIROCH, A. Protein Identification and Analysis Tools on the ExPASy Server. In: WALKER, J. M. (ed.). **The Proteomics Protocols Handbook**, Humana Press p. 571-607, 2005.

GOULDING, C. W & PERRY, L. J. Protein production in *Escherichia coli* for structural studies by X-ray crystallography. **Journal of Structural Biology**, v. 142, p. 133-143, 2003.

GUPTA, R; JUNG, E; BRUNAK, S. Prediction of N-glycosylation sites in human proteins. *In prep.* 2004.

INVITROGEN. **TOPO[®] TA Cloning Kit Handbook**, 25-0184, 2006.

IVENS, A. C; PEACOCK, C. S; WORTHEY, E. A; MURPHY, L; AGGARWAL, G. *et al.*. The genome of the Kinetoplastid parasite, *Leishmania major*. **Science**, v. 309, n. 5733, p. 436-442, 2005.

JOACHIMIAK, A. High-throughput crystallography for structural genomics. **Current Opinion in Structural Biology**, v. 19, p. 573-584, 2009.

JULENIUS, K; MOLGAARD, A; GUPTA, R; BRUNAK, S. Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. **Glycobiology**, v. 15, p. 153-164, 2005.

KAPUST, R. B. & WAUGH, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. **Protein Science**, v. 8, p. 1668-1674, 1999.

KATOH, K; MISAWA, K; KUMA, K; MIYATA, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, p. 3059-3066, 2002.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of head of bacteriophage T4. **Nature**, v. 227, p. 680-683, 1970.

LEGUIZAMÓN, M. S; CAMPETELLA, O. E; GONZÁLEZ CAPPAS, S. M; FRASCH, A. C. Mice infected with *Trypanosoma cruzi* produce antibodies against the enzymatic domain of *trans*-sialidase that inhibit its activity. **Infection and Immunity**, v. 62, p. 3441-3446, 1994.

LI, X; ROMERO, P; RANI, M; DUNKER, A. K; OBRADOVIC, Z. Predicting protein disorder for N-, C-, and internal regions, **Genome Informatics**, v. 10, p. 30-40, 1999.

MACEDO, A. M; MACHADO, C. R; OLIVEIRA, R. P; PENA, S. D. *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease. **Memórias do Instituto Oswaldo Cruz**, v. 99, p. 1-12, 2004.

MACÊDO C. D; DAROCHA, W. D; MORTARA, R. A; PRADO, M. A; TEIXEIRA, S. M. Characterization of a *Trypanosoma cruzi* antigen with homology to intracellular mammalian lectins. **International Journal for Parasitology**, v. 36, p. 1473-1484, 2006.

MINNING, T. A; BUA, J; GARCIA, G. A; MCGRAW, R. A; TARLETON, R. L. Microarray profiling of gene expression during trypomastigote to amastigote transition in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 131, p. 55-64, 2003.

MOHAN, A; SULLIVAN JR, W. J; RADIVOJAC, P; DUNKER; A. K; UVERSKY, V. N. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. **Molecular BioSystems.**, v. 4, 328-340, 2008.

MURTA, S. M. F; KRIEGER, M. A; MONTENEGRO, L. R; CAMPOS, F. F. M; PROBST, C. M. *et al.* Deletion of copies of the gene encoding old yellow enzyme (TcOYE), a NAD(P)H flavin oxidoreductase, associates with in vitro-induced benzimidazole resistance in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 146, p. 151–162, 2006.

NAGEM, R. A. P. **Otimização das condições para aquisição de dados de derivados e para determinação das fases dos fatores de estrutura de cristais de proteínas por meio da difração da luz síncrotron.** 2003. Tese (Doutorado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin, Campinas, São Paulo.

NAKAYASU, E. S; YASHUNSKY, D. V; NOHARA, L. L; TORRECILHAS, A. C; NIKOLAEV, A. V; ALMEIDA, I. C. GPIomics: global analysis of glycosylphosphatidylinositol-anchored molecules of *Trypanosoma cruzi*. **Molecular Systems Biology**, v. 5, número do artigo: 261, 2009.

NOVAGEN. **pET System Manual.** 11th edition. p. 45. 2005.

OTTO, T. D; CATANHO, M; TRISTÃO, C; BEZERRA, M; FERNANDES, R. M; ELIAS, G. S; SCAGLIA, A. C; BOVERMANN, B; BERSTIS, V; LIFSCHITZ, S; DE MIRANDA, A. B; DEGRAVE, W. ProteinWorldDB: Querying radical pairwise alignments among protein sets from complete genomes. **Bioinformatics**, PMID: 20089515, 2010.

PAYS, E; VANHAMME, L; PÉREZ-MORGA, D. Antigenic variation in *Trypanosoma brucei*: facts, challenges and mysteries. **Current Opinion in Microbiology**, v. 7, p. 369-374, 2004.

PAYS, E; SALMON, D; MORRISON, L. J; MARCELLO, L; BARRY, J. D. Antigenic variation in *Trypanosoma brucei*. In: BARRY, D; MCCULLOCH, R; MOTTRAM, J; ACOSTA-SERRANO, A. (eds.). **Trypanosomes after the genome.** Horizon Bioscience, Norfolk, 2007. Cap. 11, p. 339-372.

PITCOVSKY, T. A; MUCCI, J; ALVAREZ, P; LEGUIZAMÓN, M. S; BURRONE, O; ALZARI, P. M; CAMPETELLA, O. Epitope mapping of *trans*-sialidase from

Trypanosoma cruzi reveals the presence of several cross-reactive determinants. **Infection and Immunity**, v. 69, n. 3, p. 1869-1875, 2001.

PITCOVSKY, T. A; BUSCAGLIA, C. A; MUCCI, J; CAMPETELLA, O. A functional network of intramolecular cross-reacting epitopes delays the elicitation of neutralizing antibodies to *Trypanosoma cruzi* trans-sialidase. **The Journal of Infectious Diseases**, v. 186, p. 397-404, 2002.

PRATA, A. Clinical and epidemiological aspects os Chagas disease. **The Lancet Infectious Diseases**, v. 1, n. 2, p. 92-100, 2001.

PROMEGA. **pGEM[®]-T and pGEM[®]-T Easy Vector Systems Technical Manual**, TM042, 2009.

PROTEIN DATA BANK. Disponível em: <<http://www.rcsb.org/pdb/home/home.do>>. Acesso em: jan 2010.

PUSEY, M. L; LIU, Z. J; TEMPEL, W; PRAISSMAN, J; LIN, D; WANG, B. C; GAVIRA, J. A; NG, J. D. Life in the fast lane for protein crystallization and X-ray crystallography. **Progress in Biophysics and Molecular Biology**, v. 88, p. 359-386, 2005.

RASSI, A; RASSI JR, A; RASSI, G. G. Fase aguda. In: BRENER, Z; ANDRADE, Z. A; BARRAL-NETTO, M. (eds.). **Trypanosoma cruzi e Doença de Chagas**. 2 ed. Ed. Guanabara Koogan, Rio de Janeiro, 2000.

REMME, J. H. F; PIET FEENSTRA; LEVER, P. R; MÉDICI, A; MOREL, C. *et al.* Tropical diseases targeted for elimination: Chagas disease, lymphatic filariasis, onchocerciasis and leprosy. In: JAMISON, D. T; BREMAN, J. G; MEASHAM, A. R; ALLEYNE, G; CLAESON, M; EVANS, D. B; JHA, P; MILLS, A; MUSGROVE, P. (eds.). **Disease Control Priorities in Developing Countries**. 2nd ed. World Bank/Oxford University Press, 2006. Cap. 22, p. 433-449.

RHODES, G. **Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models**. 2nd ed, Academic Press, 2000.

ROMERO, P; OBRADOVIC, Z; LI, X; GARNER, E. C; BROWN, C. J; DUNKER, A. K. Sequence complexity of disordered protein. **Proteins**, v. 42, p 38-48, 2001.

SAMBROOK, J; FRITSCH, E. F; MANIATIS, T. **Molecular Cloning - A laboratory manual**, v. 1, 2nd ed, Cold Spring Harbour Laboratory Press, New York, 1989.

SANGER, F; NICKLEN, S; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the USA**, v. 74, n. 12, p. 5463-5467, 1977.

SCHENKMAN, S; FERGUSON, M. A; HEISE, N; DE ALMEIDA, M. L; MORTARA, R. A; YOSHIDA, N. Mucin-like glycoproteins linked to the membrane by glycosylphosphatidylinositol anchor are the major acceptors of sialic acid in a reaction

catalyzed by *trans*-sialidase in metacyclic forms of *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 59, p. 293-303, 1993.

SCHENKMAN, S; EICHINGER, D; PEREIRA, M. E. A; NUSSENZWEIG, V. Structural and functional properties of *Trypanosoma trans*-sialidase. **Annual Review of Microbiology**, v. 48, p. 499-523, 1994.

SCHLAGENHAUF, E; ETGES, R; METCALF, P. The crystal structure of the *Leishmania major* surface proteinase leishmanolysin (gp63). **Structure**, v. 6, n. 8, p. 1035-1046, 1998.

SOSA ESTANI, S; SEGURA, E. L; RUIZ, A. M; VELAZQUEZ, E; PORCEL, B. M; YAMPOTIS, C. Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas disease. **American Journal of Tropical Medicine and Hygiene**, v. 59, p. 526-529, 1998.

SOSA ESTANI, S. & SEGURA, E. L. Treatment of *Trypanosoma cruzi* infection in the undetermined phase. Experience and current guidelines in Argentina. **Memórias do Instituto Oswaldo Cruz**, v. 94, p. 363-365, 1999.

SØRENSEN, H. P & MORTENSEN, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. **Microbial Cell Factories**, v. 4, 2005.

SREERAMA, N. & WOODY, R. W. Computation and analysis of protein circular dichroism spectra. **Methods in Enzymology**, n. 383, p. 318-351, 2004.

TARLETON, R. L. Immune system recognition of *Trypanosoma cruzi*. **Current Opinion in Immunology**, v. 19, p. 430-434, 2007.

TAYLOR, J. E. & RUDENKO, G. Switching trypanosome coats: what's in the wardrobe? **Trends in Genetics**, v. 22, n. 11, 614-620, 2006.

TOWBIN, H; STAHELIN, T; GORDON, J. Eletrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. **Proceedings of the National Academy of Sciences of the USA**, v. 76, p. 4350- 4354, 1979.

TriTrypDB: Disponível em: <http://tritrypdb.org/>

URBINA; J. A. & DO CAMPO, R. Specific chemotherapy of Chagas disease: controversies and advances. **Trends in Parasitology**, v. 19, n. 11, p. 495-501, 2003.

URBINA, J. A. Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches. **Acta Tropica**, 2009. doi:10.1016/j.actatropica.2009.10.023.

UVERSKY, V. N; GILLESPIE, J. R; FINK, A. L. Why Are “Natively Unfolded” Proteins Unstructured Under Physiologic Conditions? **PROTEINS: Structure, Function, and Genetics**, v. 41, p. 415-427, 2000.

UVERSKY, V. N. What does it mean to be natively unfolded? **European Journal of Biochemistry**, v. 269, p. 2-12, 2002.

WHO. Control of Chagas Disease. (**WHO Technical Report Series 811**), World Health Organization, 1991.

WILKINSON, S. R; TAYLOR, M. C; HORN, D; KELLY, J. M; CHEESEMAN, I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. **PNAS**, v. 105, n. 13, p. 5022-5027, 2008.

YAO, C; DONELSON, J. E; WILSON, M. E. The major surface protease (MSP or GP63) of *Leishmania sp.* Biosynthesis, regulation of expression, and function. **Molecular and Biochemical Parasitology**, v. 132, p. 1-16, 2003.

YAO, C. Major Surface Protease (MSP, or GP63) of Trypanosomatids, One size fits all? **Infection and Immunity**, v. 78, p. 22-31, 2009.